

Measurement and prediction of bark thickness in *Picea abies*: assessment of accuracy, precision, and sample size requirements

Stefan M. Stängle, Aaron R. Weiskittel, Carsten F. Dormann, and Franka Brüchert

Abstract: Tree and log diameters are usually measured outside bark, but inside-bark diameters are of greater economic interest and are often derived with local or regional bark thickness equations. To date, the influence of measurement method, sampling design, and sample size on bark thickness equation accuracy and precision has received limited attention. The objectives of this study were to use an extensive regional bark thickness dataset for Norway spruce (*Picea abies* (L.) Karst) in southwestern Germany to (1) quantify the accuracy and precision of bark thickness measurements with a Swedish bark gauge, (2) determine the required number of measurements to assess the within-tree variation, and (3) estimate the required sample sizes per plot and per region to develop an accurate bark thickness prediction equation. Bark gauge readings were validated with measurements derived from X-ray computed tomography (CT) and indicate that Swedish bark gauges generally overestimated bark thickness by $13.6\% \pm 28.4\%$ (mean \pm standard deviation). Results suggested having at least one measurement location every 2 m along a tree bole and at least five bark thickness measurements per each of these locations to achieve an allowable error of $<15\%$. For the study area, Monte Carlo simulations indicated that a total sample size of 50–250 trees was needed, depending on the complexity of the desired bark thickness model. Overall, this analysis indicated that there was relatively high within- and between-tree variation in bark thickness, but adequate sampling methods and sample sizes produced highly accurate bark thickness equations.

Key words: bark factor, measurement error, variance, equivalence, model performance.

Résumé : Le diamètre des arbres et des grumes est généralement mesuré sur écorce, mais le diamètre sous-écorce présente un plus grand intérêt économique et est souvent dérivé en utilisant des équations d'épaisseur d'écorce locale ou régionale. À ce jour, l'influence de la méthode de mesure, du plan d'échantillonnage et de la taille de l'échantillon sur l'exactitude et la précision des équations d'épaisseur d'écorce a reçu peu d'attention. En utilisant une importante base de données régionale d'épaisseur d'écorce pour l'épicéa commun (*Picea abies* (L.) Karst.) dans le sud-ouest de l'Allemagne, les objectifs de cette étude étaient : (1) de quantifier l'exactitude et la précision des mesures d'épaisseur d'écorce avec une jauge d'écorce suédoise, (2) de déterminer le nombre requis de mesures pour évaluer la variation intra-arbre et (3) d'estimer la taille de l'échantillon requis par parcelle et par région pour développer une équation précise de prédiction de l'épaisseur d'écorce. Les lectures de la jauge d'écorce ont été validées avec des mesures de tomographie aux rayons X informatisée et indiquent que les jauges d'écorce suédoises ont généralement surestimé l'épaisseur d'écorce de $13,6\% \pm 28,4\%$ (moyenne \pm écart type). Les résultats indiquent qu'on devrait mesurer au moins à tous les deux mètres le long d'un tronc d'arbre et prendre au moins cinq mesures d'épaisseur d'écorce à chacun de ces endroits pour obtenir une marge d'erreur $<15\%$. Pour la zone d'étude, des simulations de Monte Carlo ont indiqué que l'échantillon devrait avoir une taille totale de 50 à 250 arbres selon la complexité du modèle d'épaisseur d'écorce désiré. Dans l'ensemble, cette analyse a indiqué qu'il y avait une variation de l'épaisseur d'écorce intra-arbre et inter-arbres relativement élevée, mais des méthodes d'échantillonnage et une taille d'échantillon adéquates ont produit des équations d'épaisseur d'écorce très précises. [Traduit par la Rédaction]

Mots-clés : facteur d'écorce, erreur de mesure, variance, équivalence, performance du modèle.

Introduction

Although tree and log diameters are usually measured outside bark, wood volume calculations in forest inventories, roundwood markets, and bucking optimization algorithms generally assume inside-bark diameters. To convert outside-bark diameters to inside-bark diameters, equations for bark thickness and volume have been developed for many species (e.g., Li and Weiskittel 2011). Bark thickness can be approximated by different equations as a function of diameter outside bark, relative height in the stem, and tree variables such as diameter at breast height (dbh, 1.3 m) and total tree height (Li and Weiskittel 2011). Their application has recently become more widespread, with the shift in the commer-

cial relevance of bark from an unwanted residue to a valuable fuel and source for high-value biomaterials (Doruska et al. 2009). The accurate estimation of available bark biomass is thus becoming important to evaluate the potential of such technologies generating additional income for the forestry sector.

Both accuracy and precision are important when measuring bark thickness, as multiplicative errors accrue in volume calculations if biased bark measurements are used for predicting inside-bark diameters (Gordon 1983). The three most common techniques to measure bark thickness are the use of a Swedish bark gauge, diameter measurements with caliper or tape before and after debarking, and measurements on cross-sections. In this study, we focused on measurements with bark gauges, as they are widely

Received 1 July 2015. Accepted 5 October 2015.

S.M. Stängle and F. Brüchert. Forest Research Institute Baden-Württemberg, Wonnhaldestr. 4, 79100 Freiburg, Germany.

A.R. Weiskittel. School of Forest Resources, University of Maine, 229 Nutting Hall, Orono, ME 04469, USA.

C.F. Dormann. Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacher Str. 4, 79106 Freiburg, Germany.

Corresponding author: Stefan M. Stängle (e-mail: stefan.staengle@forst.bwl.de).

used to rapidly acquire several bark thickness observations at low cost without reducing the log value by cutting cross-sections. It has been reported that measurements with bark gauges generally overestimate bark thickness (Althen 1964; Kirschner 1976), are sensitive to the sampling season (Althen 1964), and that bias strongly depends on the operator's subjective sensation, requiring their usage to be practiced (Gray 1956; Mesavage 1969).

Variation of bark thickness has to be considered at several levels to develop a suitable bark thickness equation for a certain geographical region such as a country or district. This begins at the disc level, i.e., at each single measurement location within a tree, as the bark thickness varies around the stem (Mesavage 1969). The variation is higher for species that have rough bark and can also vary strongly within trees, usually being higher towards the base (Gordon 1983). The next spatial level to consider is each tree, as the bark to diameter ratio is not constant along the tree bole for most species (e.g., Flury 1897; Loetsch et al. 1973). For Norway spruce (*Picea abies* (L.) Karst), the bark to diameter ratio has been shown to increase with stem height (Kirschner 1976; Altherr et al. 1978). For different coniferous species, further variation in bark thickness development has been observed between stands growing under different site and growth conditions (Hoffmann 1958; Loetsch et al. 1973; Wilhelmsson et al. 2002; Laasasenaho et al. 2005; Sonmez et al. 2007) and between provenances (Persson and Downie 1992; McConnon et al. 2004; Kohnle et al. 2012). In most European countries, differences caused by growing conditions are not taken into account in the bark thickness equations that are used. However, there are countries such as Sweden where several regionalized bark thickness equations are in use (SDC Skognäringsens IT-företag 2014).

Despite its importance, the influence of measurement method, sampling design, and sample size on bark thickness equations have received limited attention in the literature. For future development of bark thickness equations, this is, however, of critical importance, as the sampling should capture the bark thickness variation in a study region at the lowest possible cost. One general rule of thumb for the calculation of a reliable bark factor, i.e., ratio of inside-bark diameter to outside-bark diameter, at breast height is to sample 20–50 trees (Husch et al. 2003). In contrast, studies for developing stem taper equations, which can either be the diameter outside or inside bark, suggest that at least 825–5000 observations in total (Kitikidou and Chatzilazarou 2008; Kitikidou 2010) or 15–80 sample trees (Kitikidou and Chatzilazarou 2008; Subedi et al. 2011) were necessary.

In this analysis, the required sample sizes at the individual disc, tree, and regional levels were estimated using extensively measured data for Norway spruce from one region in southwestern Germany. The specific objectives of this study were to (1) quantify the influence of measurement errors on the accuracy and precision of bark thickness measurements when using a Swedish bark gauge, (2) determine the necessary number of measurements (a) per location and (b) per tree to cover the within-tree variability of bark thickness, and (3) estimate the required sample size at the regional level to develop an accurate bark thickness prediction equation for sawlogs (tree boles down to a top end diameter of approximately 12 cm).

Material and methods

Data origin

Three different datasets were used to complete the above stated objectives. For objective (1), lab measurements of bark thickness on a spruce log were performed. For objective (2a), the variability of bark thickness was analyzed on stem discs that were sampled in 10 stands in the state of Baden-Württemberg, southwestern Germany, covering different site and stand conditions. For objectives (2b) and (3), bark thickness data were used from the extensive study by Altherr et al. (1978). The trees of the Altherr et al. (1978)

study were from throughout the 35 752 km² state of Baden-Württemberg in temperate forests with a broad range of site and stand conditions. A total of 133 study plots were selected throughout the state with elevations from 220 to 1100 m above sea level. Mean annual precipitation in the study plots ranged from 760 to 2000 mm, whereas annual mean air temperature varied from 5.3 to 9.4 °C. Soil types mostly covered different Cambisols but also covered Podzols, Planosols, Gleysols, and Leptosols. Site conditions such as soil depth, water holding capacity, and exposition widely varied between the study plots; however, they were not recorded in detail. Regrettably, tree age also was not recorded either, because the focus was on developing a correlation between diameter and bark thickness without considering additional factors influencing this relationship. To calculate tree heights for the Altherr et al. (1978) data, regional taper curves were used (see below).

Measurements

For objective (1), 161 predefined measurement points on the log were sampled by an experienced operator with a Swedish bark gauge (Suunto, Finland). For validation, X-ray computed tomography (CT) was used, as logs or stem discs can be scanned nondestructively and measured at a very high resolution and degree of accuracy. In particular, CT has been shown to be a technology that can be used to detect wood features such as knots and internal defects in high resolution (Wei et al. 2011). CT scans were performed with a 2008-built stationary CT scanner CTLog (MiCROTEC GmbH – slr, Italy) with a resolution of 1.1 and 5 mm for axial and longitudinal directions, respectively. For softwood species, the density contrast between bark and wood is relatively high, and therefore, automated bark detection can be performed with high accuracy (Fig. 1).

For objective (2a), the variability of bark thickness at the disc level was assessed with CT scans of 127 stem discs taken at various positions from 44 stems from different sites. From the CT image of each stem disc, one slice, which can be imagined as a virtual stem disc of 5 mm thickness, was selected. From each slice, 360 bark thickness values were extracted. By visual assessment on a screen, areas in the single slices where the bark detection algorithm was disturbed by knots or spalled bark were excluded from the analysis. Diameters of the analyzed CT slices ranged from 11.8 to 60.8 cm, and the mean bark thickness per slice varied from 4.5 to 11.2 mm (Table 1).

For the Altherr et al. (1978) study, 7712 trees from 133 stands had been felled, delimbed, and measured in the forest before any further log manipulation such as forwarding was performed. Measurement locations were at breast height and along the tree bole in 2 m increments up to a top diameter of approximately 10–15 cm, resulting in 81 975 measurement locations. Diameter and bark thickness were measured twice (approximately perpendicularly) at each location using both a caliper and a Swedish bark gauge. Double bark thickness was calculated as the sum of the two bark measurements. A detailed description of the sampling method and the data can be found in Altherr et al. (1974, 1978). As only merchantable sawlog length had been recorded, total tree height was estimated with mixed effect B-spline regression describing tree taper using the R package TapeR version 0.3.0 (Kublin and Breidenbach 2013). The taper curve equations were parameterized on newly collected diameter–height measurements from 436 Norway spruce trees ranging in dbh from 17.3 to 69.0 cm and in tree height from 14.6 to 43.0 m. The trees were measured throughout the state of Baden-Württemberg following the sampling scheme from Altherr et al. (1978) and included tree height measurements. Tree heights for the Altherr et al. (1978) material were required to calculate the relative height of each measurement location. Estimated tree heights larger than 43 m were unrealistic, and therefore, 85 trees with 1202 measurement locations were excluded from further analysis. Dimensional characteristics of the selected trees and measurement locations are summarized in Tables 2 and 3.

Fig. 1. Computed tomography (CT) image representing a stem disc of 5 mm thickness of Norway spruce. Gray levels correspond to the density, with brighter areas having a higher density. The outer border and the wood–bark border were detected automatically and are illustrated with white and black lines, respectively.

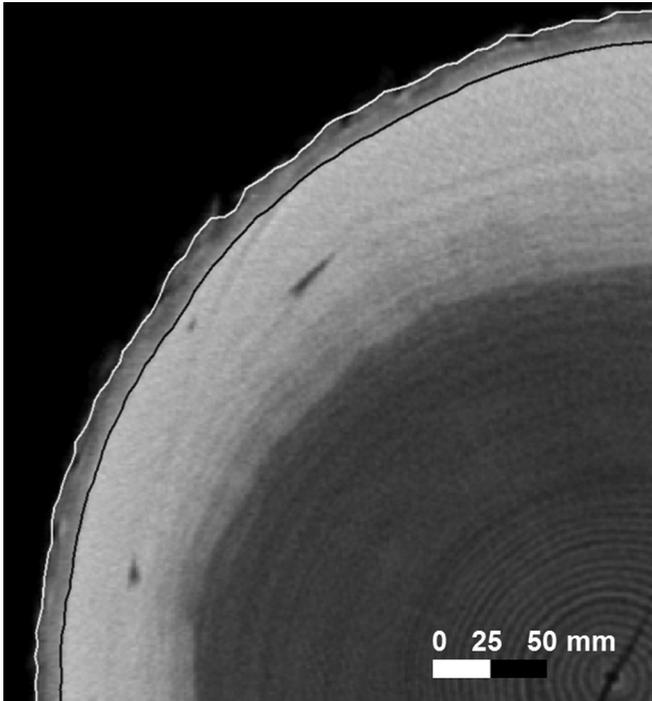


Table 1. Dimensional characteristics of the analyzed slices from X-ray computed tomography (CT) ($n = 127$).

	Mean	SD	Minimum	Maximum
Slice diameter (cm)	28.4	10.8	11.8	60.8
No. of measurements per slice	296	34.7	142	318
Mean bark thickness per slice (mm)	6.1	1.4	4.3	11.2
SD of bark thickness per slice (mm)	0.72	0.24	0.26	1.80

Note: SD, standard deviation.

Table 2. Dimensional characteristics of the sample trees ($n = 7627$) from Alther et al. (1978) including diameter at breast height (dbh), double bark thickness at breast height (DBT_{bh}), estimated total tree height (H), and number of measurement locations.

	Mean	SD	Minimum	Maximum
dbh (cm)	34.3	14.3	13.1	84.6
DBT_{bh} (mm)	19.0	6.4	6.0	48.0
H (m)	28.6	6.2	12.1	43.0
No. of measurement locations	10.6	2.8	5	14

Note: SD, standard deviation.

Table 3. Dimensional characteristics of the measurement locations ($n = 80\,773$) including diameter (d), double bark thickness (DBT), relative bark thickness (BT_{rel}), and relative height in the stem (H_{rel}).

	Mean	SD	Minimum	Maximum
d (cm)	26.8	11.1	7.9	84.6
DBT (mm)	15.7	5.6	4.0	48.0
BT_{rel} (%)	6.1	1.4	2.7	17.3
H_{rel} (%)	38.6	22.1	3.0	91.0

Note: SD, standard deviation.

Data analysis

The manual gauge readings from the sampled spruce log were compared with CT-derived bark thickness values by plotting the difference between the two measurements versus their mean with a 95% level of agreement (Bland and Altman 2003).

The bark thickness variability at one measurement location was quantified by the coefficient of variation (CV) computed on all bark thickness values of one CT slice. The number of measurements that was required to assess the mean bark thickness at one location at a predefined precision of relative allowable errors (AE) of 15% and 20% at a 95% confidence level was calculated iteratively for each slice by applying eq. 1 (Köcher et al. 1972).

$$(1) \quad n = \left(\frac{CVt}{AE} \right)^2$$

where n is the required number of measurements, $CV = SD(x)/\bar{x}$ (SD, standard deviation), and t refers to the 95% quantile of the two-tailed t distribution, with $n-1$ degrees of freedom.

The bark thickness variability along each tree bole was expressed by the variation of relative bark thickness, i.e., the proportion of the outside-bark diameter that constitutes bark, within each tree. We calculated the number of required locations for each of the 7627 trees using eq. 1, with an AE of 15% and 20% at a 95% confidence level.

For the main analysis, we compared predictions of bark thickness of different bark thickness models to assess the effect of model form and reduced sample size on the quality of equation coefficients. We selected two bark thickness models from the literature, from which eq. 2 uses the outside-bark diameter as a single explanatory variable (Dolph 1989). Equation 3 includes also the relative height, the total height, and the bark thickness at breast height as covariates (Cao and Pepper 1986). Equation 2 works well for several spruce species, and eq. 3 is suitable for several different coniferous species (see eqs. 7 and 5, respectively, in Li and Weiskittel (2011)). To account for the different structure of the models, with only eq. 3 being forced through the origin, we evaluated eq. 4 being a version of eq. 2 without intercept and eq. 5 being a version of eq. 3 including an intercept. Equation 4 was introduced by Meyer (1946) and was found to be inferior to eq. 2 for spruce trees by Li and Weiskittel (2011).

$$(2) \quad d_{ib} = \beta_0 + \beta_1 d_{ob} + \varepsilon$$

$$(3) \quad d_{ib} = d_{ob} \left(\beta_1 + \beta_2 H_{rel} + \beta_3 H_{rel}^2 + \beta_4 H + \beta_5 \frac{dbh_{ib}}{dbh_{ob}} \right) + \varepsilon$$

$$(4) \quad d_{ib} = \beta_1 d_{ob} + \varepsilon$$

$$(5) \quad d_{ib} = \beta_0 + d_{ob} \left(\beta_1 + \beta_2 H_{rel} + \beta_3 H_{rel}^2 + \beta_4 H + \beta_5 \frac{dbh_{ib}}{dbh_{ob}} \right) + \varepsilon$$

where d_{ib} is the diameter inside bark (mm), d_{ob} is the diameter outside bark (mm), H_{rel} is the relative tree height, H is the total tree height (m), dbh_{ib} is the diameter at breast height inside bark (cm), dbh_{ob} is the diameter at breast height outside bark (cm), and ε is the error term.

As eqs. 3 and 5 have three more parameters, we expect them to fit the data better than eqs. 2 and 4. The data were clustered at two levels as several observations were made per individual tree and multiple trees were grouped in plots. The random effects, however, were not considered in the tested linear models, as the aim of this study was to develop bark thickness equations that would be applicable for the whole study region. Variance components were estimated to quantify the importance of observed variability at the tree and the plot level. The ratio of between-group variance

to total variance was calculated to describe the intraclass correlation.

To compare the predictive accuracy of the different models, observations were blocked in 1000 iterations to keep independence of training and test dataset by randomly removing 33 stands as the test dataset. This corresponds to approximately 25% of the total number of stands, as suggested for block cross-validation by Burman et al. (1994) and Racine (2000). Bark thickness models were fitted on the training dataset, and diameter inside bark was predicted for the test dataset. For model comparison, we averaged the mean absolute bias from 1000 runs, which is the mean of absolute difference between observed and predicted values, and the mean root mean square error (RMSE).

For detailed analysis, we selected the much more commonly employed eqs. 2 and 3. We evaluated the performance of these two bark thickness models for different sample sizes at both the plot and regional levels by applying Monte Carlo simulations (described below) to determine the required number of trees per plot and the number of plots in the study region.

The effect of the diameter distribution of the sample trees on model performance was analyzed with a stratified sampling approach with reduced dbh ranges. To test the effect of a reduced dbh distribution of the training data on predictions for trees of the full dbh range, four different tree selection protocols were simulated, representing four dbh distributions. Trees for the training data were either drawn from all dbh classes or only from one of the three dbh classes: <25 cm, 25–40 cm, or >40 cm. Different sample sizes were simulated by reducing the size of the training dataset at two levels, i.e., the number of plots (stands) and the number of trees per plot. The tested sample sizes were 50, 20, 10, and 5 plots with 25, 10, 5, and 1 trees each. The plots and trees were drawn randomly with replacement from the training data.

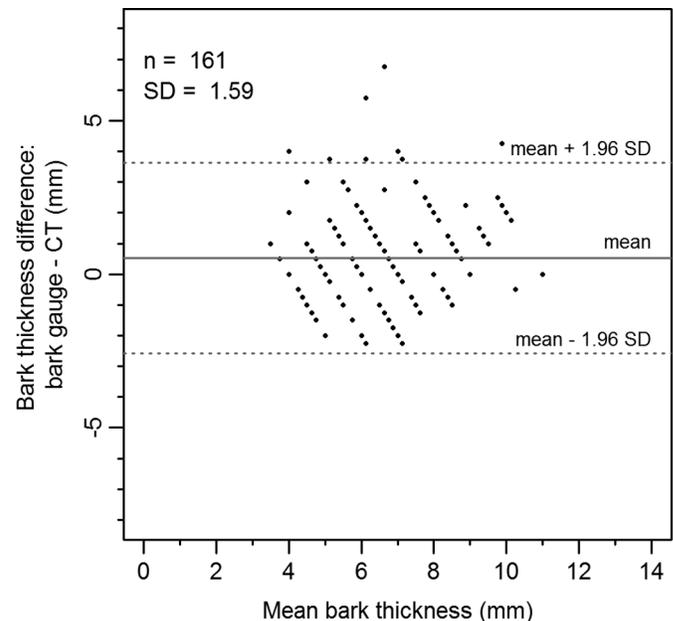
Each of these 64 combinations was used to fit both bark thickness models and to predict diameter inside bark to the test dataset 1000 times. For model evaluation, the RMSE is reported for each simulation. As the trees of the larger dbh classes are taller, the four different dbh distribution protocols showed increasing numbers of measurement locations in the training data with increasing dbh. It was assumed that coefficients of bark thickness equations obtained from the largest possible dataset yielded the best possible predictions. To test if the sample size of a simulation was sufficient to estimate equation coefficients reliably, predictions of the fitted models were compared with predictions that were made when the models had been fit on the full training data with equivalence testing. For the comparison, double bark thickness was calculated as difference between predicted diameter inside bark and measured diameter outside bark. Equivalence testing was chosen because this method allows for actually testing for similarity of data and not only for dissimilarity of data, which many traditional statistical tests do. The null hypothesis H_0 of a test of equivalence states that the means of the two compared groups are different (H_0 of dissimilarity). Therefore, the user has to predefine a region of indifference, in which one assumes equivalence (Robinson and Froese 2004). The advantage is that the statistical significance is defined by actual relevance, and it is less sensitive to sample size. In our study, the region of indifference was set to 1 mm, which defines the absolute size that the mean of the differences can reach so that H_0 is still rejected, and therefore, equivalence is assumed. The R package equivalence version 0.6.0 (Robinson 2014) was used to calculate a “two one-sided t test of paired sample equivalence” (TOST). The above described statistical analysis was performed with the software R version 3.1.1 (R Core Team 2014).

Results

Precision and accuracy of Swedish bark gauge measurements

Bark thickness readings from a gauge between 4 and 12 mm were compared with CT measurements at 161 measurement points (Fig. 2).

Fig. 2. Plot of differences between bark thickness measurements in Norway spruce assessed with a Swedish bark gauge and derived from X-ray computed tomography (CT) plotted against the mean of both measurements. The solid line represents the mean difference, and the dotted lines represent the 95% limits of agreement. SD, standard deviation.



Gauge readings overestimated the true value by 0.52 ± 1.59 mm (mean \pm standard deviation) or $13.6\% \pm 28.4\%$. With 0.13 mm, the standard error of the mean was much smaller than the average overestimation, indicating a statistically significant bias of the gauge. The spread of the differences around their mean, which describes the accuracy of gauge measurements, was similar across the entire range of tested bark thickness values. 95% of all gauge measurements can be expected to be in the range of 2.60 mm smaller and 3.64 mm larger than CT measurements.

Within-tree bark thickness variability

With increasing disc diameter, the thickness and the roughness of bark increased (Fig. 3). This led to higher absolute variability per cross-section for larger diameters. Across the ranges of tested diameters and bark thicknesses, the relative variability was similar with a CV of $11.8\% \pm 2.7\%$. For an AE of 15%, the calculated number of required bark thickness measurements per slice was 4.3 ± 1.6 , which, in practice, would mean at least five measurements at each measurement location. An AE of 20% reduced the number of required measurements to three (2.3 ± 1.5) (Fig. 4).

With increasing tree height, higher variability in relative bark thickness requires more measurement locations (Fig. 5). The calculated average number of measurement locations that was required to describe the variance of the relative bark thickness with an AE of 15% was 7.1 ± 3.7 . A 20% AE resulted in an average of 4.7 ± 2.6 measurements. The actual sampled number of locations per tree was equal or larger than the calculated required number for 90% and 99% of all tested trees for AEs of 15% and 20%, respectively (Table 4).

Model performance with different sample sizes at the plot and regional levels

Parameter estimates from the fitted bark thickness equations for all trees from all 133 study plots are listed in Table 5. High intraclass correlation was found on both the tree and plot levels, with observed correlation coefficients of 18%–37% and 17%–48%, respectively. When fitted to the training datasets of 1000 runs,

Fig. 3. Bark thickness values assessed in 127 stem discs with X-ray computed tomography (CT). SD, standard deviation.

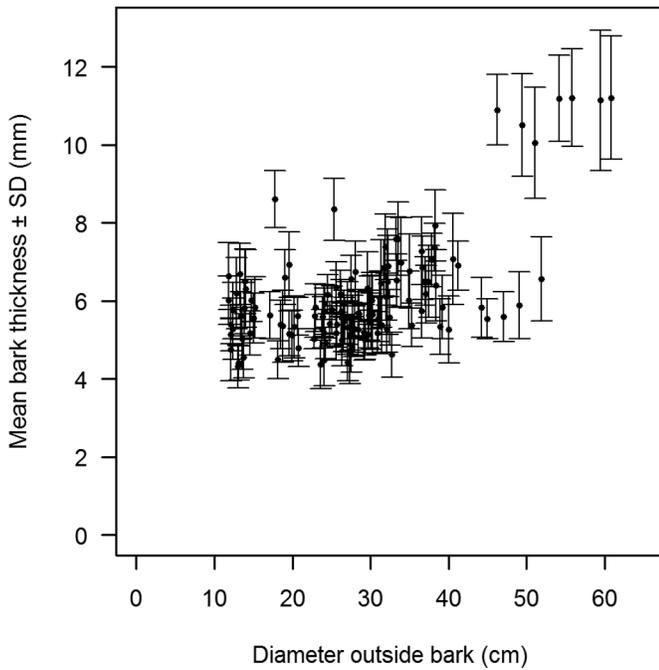
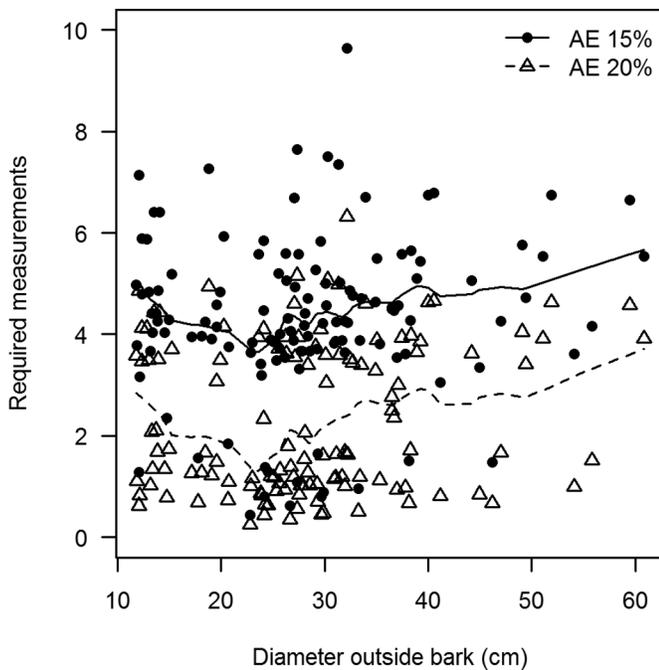


Fig. 4. Required number of bark thickness measurements per cross-section, with an allowable error (AE) of 15% (closed circles) and 20% (open triangles). Trend lines are lowest regression fits, with a solid line for an AE of 15% and a dashed line for an AE of 20%. $n = 127$ cross-sections.



bark thickness eqs. 3 and 5 with the more complex model form performed better than the simple eqs. 2 and 4, as their mean absolute biases were more than 1 mm smaller (Table 6). For the simple model form, we found that the version including intercept (eq. 2) yielded better results than the one without intercept (eq. 4). The complex model form with intercept (eq. 5) only marginally improved predictions compared with the much more commonly

Fig. 5. Density distribution of required number of measurement locations per tree for an allowable error (AE) of 15% (dark gray) and 20% (light gray) grouped by tree height classes. Average per group is shown by horizontal lines.

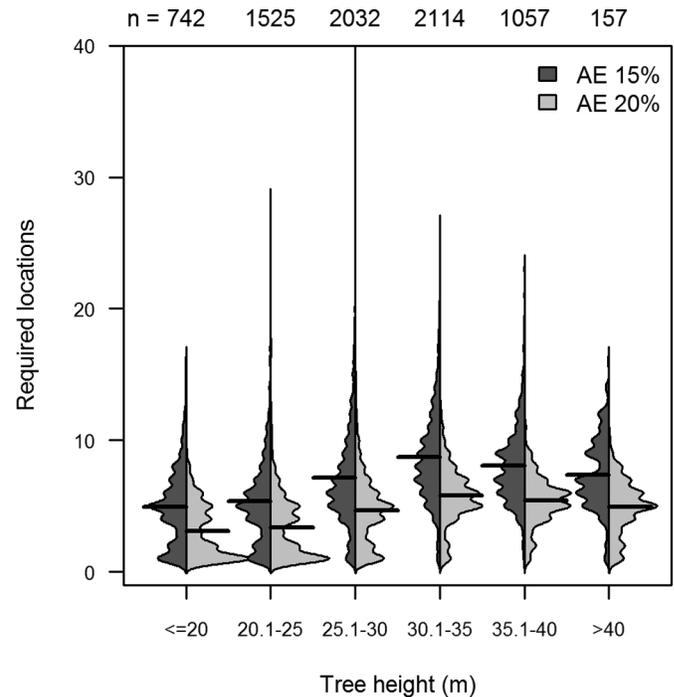


Table 4. Proportion of studied trees that were sampled at least the required number of measurement locations.

AE	Tree height (m)						Total
	<20	20–25	25–30	30–35	35–40	>40	
15%	0.77	0.88	0.89	0.91	0.97	0.99	0.90
20%	0.97	0.98	0.99	1.00	1.00	1.00	0.99

Note: Separated for different tree height groups and allowable errors (AE) of 15% and 20%.

employed eq. 3. Consequently, we selected eqs. 2 and 3 for more detailed analysis.

For almost all reduced sample size simulations, the complex eq. 3 outperformed the simpler eq. 2 in respect of the fit quality expressed by lower RMSE values (Fig. 6). The reduction of the training data to trees of certain diameter ranges had different effects on the fit quality of the two equations. A reduced fit quality can be observed for eq. 3 when only small trees (dbh < 25 cm) were used for the training data (Fig. 6b). For both equations, using only medium-sized trees (dbh between 25 and 40 cm) for model fitting resulted in almost the same RMSEs as when all trees were used (Fig. 6c). Reducing the training data to large trees (dbh > 40 cm) decreased the fit quality of eq. 2 (Fig. 6d). The mean RMSE of model fits from eq. 2 and 3 increased by less than 14% and 44%, respectively, when comparing the largest and the smallest sample sizes. A more pronounced effect can be observed for the standard deviation of the RMSEs, which increased more than tenfold (Fig. 6a).

In general, the observed effect of reducing the sample size on the equation predictions was larger for eq. 2 than for eq. 3 (Figs. 6e–6h). Predictions of eq. 2 were equivalent to the predictions of the full model in more than 95% of the runs, as long as the sample size was at least 50 plots with five trees each. For eq. 3, 10 plots with five trees each were enough to predict equivalent values (Fig. 6e). For small trees, only eq. 3 was useful to predict equivalent values but again only with large sample sizes (a minimum of 50 plots with five trees each) (Fig. 6f). When only medium

Table 5. Estimated parameters, corresponding standard errors (SE), residual standard error (σ), and tree-level and plot-level intraclass correlation coefficients (ICC_{tree} and ICC_{plot} , respectively) of the bark thickness eqs. 2–4 when fitted to the full dataset.

Parameters	Eq. 2		Eq. 3		Eq. 4		Eq. 5	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	4.57654	0.02921	—	—	—	—	1.36400	0.02882
β_1	0.95846	0.00010	0.20330	0.00231	0.94390	0.00004	0.15150	0.00253
β_2	—	—	0.02447	0.00040	—	—	0.02302	0.00040
β_3	—	—	-0.08065	0.00055	—	—	-0.08446	0.00054
β_4	—	—	-0.00021	0.00001	—	—	-0.00003	0.00001
β_5	—	—	0.79240	0.00251	—	—	0.83760	0.00265
σ	3.1720	—	1.8920	—	3.6220	—	1.8660	—
ICC_{tree} (%)	18.3	—	33.5	—	14.7	—	37.1	—
ICC_{plot} (%)	47.6	—	21.8	—	33.7	—	16.7	—

Table 6. Summary of fit statistics from the tested bark thickness models with mean absolute bias (MAB) and the root mean square error (RMSE).

Equation	Model form	RMSE (mm)	MAB (mm)
2	Simple with intercept	3.19	2.45
3	Complex without intercept	1.91	1.46
4	Simple without intercept	3.64	2.77
5	Complex with intercept	1.88	1.43

or large trees were used in the training data, both equations behaved similar as in the case when all trees were used (Figs. 6g and 6h).

Discussion

Precision and accuracy of bark thickness measurements

Bark thickness is a critical measurement to assess the economic value of timber, and it has often not been given the attention it may deserve. For example, Marshall et al. (2006) found that using an inappropriate bark thickness equation can result in a loss of value of up to 11% due to misspecifications in log bucking algorithms. However, bark thickness is rarely measured, and when it is, the degree of accuracy and precision is not reported. In this analysis, bark thickness measurements using the widely used Swedish bark gauges tended to significantly overestimate the true value, which corresponds to results of other previous studies (Althen 1964; Kirschner 1976). The low precision of the tool leads to higher variation in data acquisition and, therefore, to a larger number of required measurements. Of course, the results are species dependent as relatively accurate values can be expected for smooth bark and higher variation can be expected for rough bark. Another relevant factor is the age of the trees, because a change from smooth to rough bark can occur during the ontogeny of some species such as for many species of *Pinus*, *Fraxinus*, or *Pseudotsuga*. Additionally, the relative bark thickness at a given relative height in the stem can increase or decrease during ontogeny (Adams and Jackson 1995). Due to the limitations in accuracy and precision demonstrated in this study, bark gauges cannot be used directly to exactly determine the bark thickness at a single point. Rather, these measurements can be used to gather bark thickness values for the development of bark thickness equations, which are generally based on outside-bark diameters. The use of a correction factor to account for the general overestimation could be considered; however, differences among gauge operators and under different measuring conditions will have to be considered in further studies.

Within-tree bark thickness variability

Calculations of the required number of measurements at the disc and tree levels in this study were based on practical consid-

erations and not solely on statistical significance. For the analysis, 1 mm was selected as a region of equivalence, while the AE was set to 15% and 20%. One mm was selected, as the precision of the measurements cannot be expected to be higher, and a 15% AE is a good compromise between research and operational achievability. For different applications or in a different context, other values may be more appropriate. In this analysis, at least five measurements per cross-section were recommended to achieve an AE of 15% due to the variability of bark thickness around the stem. Other studies using Swedish bark gauges have suggested averaging two (Meyer 1946; Althen 1964; Mesavage 1969; Kirschner 1976; Wilhelmsson et al. 2002; McConnon et al. 2004; Sonmez et al. 2007; Kohnle et al. 2012) or three (Malone and Liang 2009) gauge readings per location. This might be enough for species of smooth bark but would underestimate the variability of gauge measurements per location for species having a rough bark such as Norway spruce.

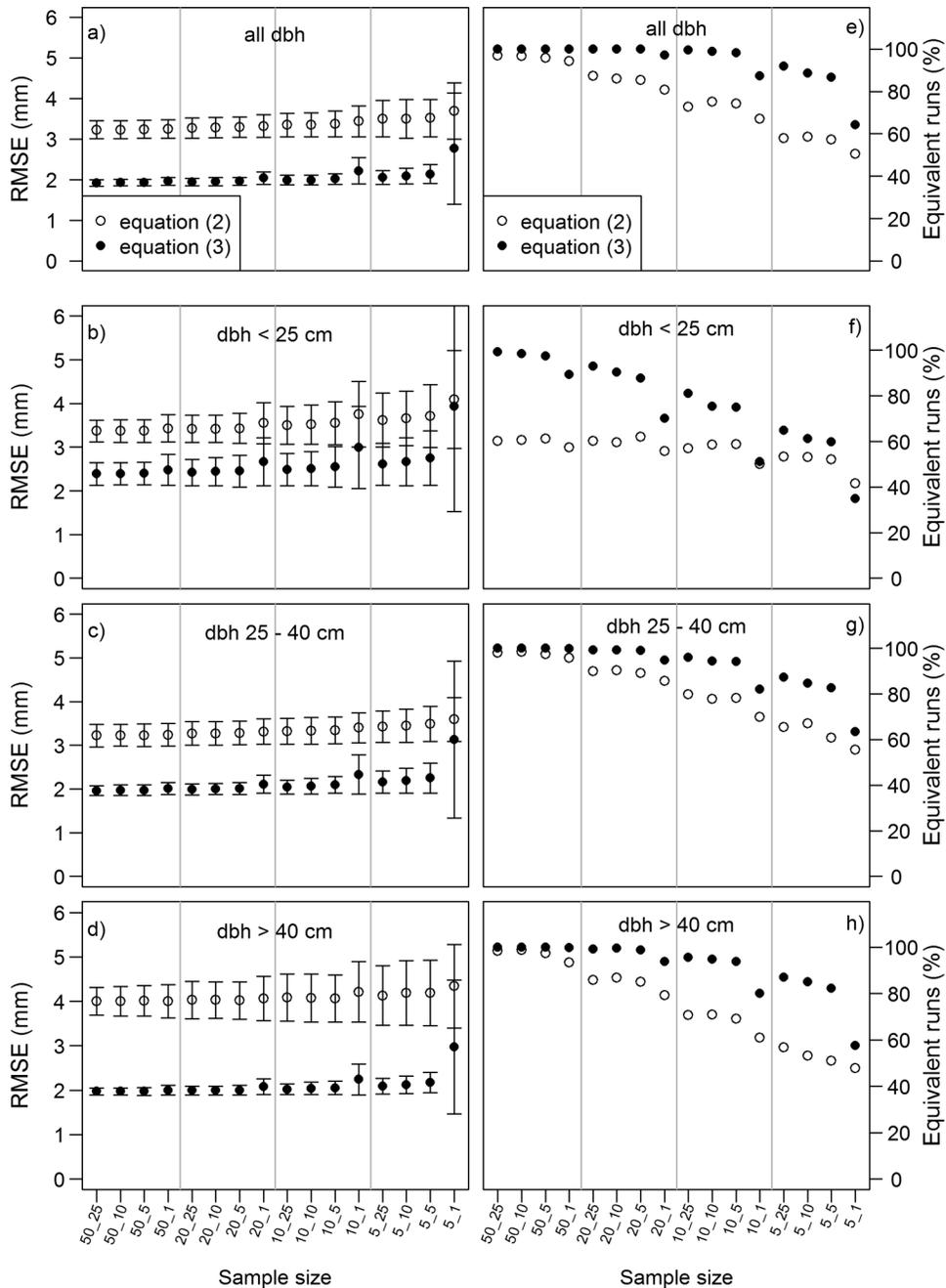
The number of measurement locations required per tree increased with dbh and tree height. Consequently, it is suggested to choose a sampling design with regular distances between measurement locations, which will lead to more measurements with increasing tree height. In contrast, a fixed number of relative positions along the tree bole, as it is suggested in other studies (e.g., Korell 1972; Feduccia and Mann 1976; Kozak and Yang 1981; Gordon 1983; Laasasenaho et al. 2005), could reduce the support for each data point and would require additional effort in the field to recompute relative positions for each single tree. The sampling design of this study with 2 m increments between the measurement locations starting from 1.3 m above ground was adequate to assess the variability within a stem in most cases and can be recommended for further studies. If, however, both bark thickness and form factor equations are developed from the same data, then it could be favorable to choose relative positions, as the supposed sampling scheme would not adequately describe the stem form, especially for young trees. In addition, the variability of bark thickness below 1.3 m can significantly increase due to changes in stem form and bark profiles. Additional research will be necessary to understand how the inclusion of observations below 1.3 m would influence samples sizes.

Sample size at plot and regional levels

Reduced sample sizes for model fitting resulted in a higher variation of RMSE in the different runs, suggesting that the relationship between diameter and bark thickness varies strongly between individual trees (Figs. 6a–6d), which was also supported by the high observed intraclass correlation.

A high variation between the plots was also shown as the amount of sampled plots strongly determined the number of iterations with predicted bark thickness values that were equivalent to predictions from the full-data model (Figs. 6e–6h). For the predictive power of models on hierarchical data, in general, the

Fig. 6. Root mean square error (RMSE) and results from the equivalence test of eqs. 2 and 3 for the 16 different sample size combinations. Shown are the mean RMSE of predictions and its standard deviation (a–d, smaller values are better) and the proportion of runs (from $n = 1000$) that resulted in equivalent ($p < 0.05$) predictions as predictions from the full training dataset (e–h) based on a similarity region of 1 mm (larger values are better). Labels on the x axis refer to the number of plots and the number of trees per plot, respectively (e.g., “50_25” means 50 plots with 25 trees each).



sample size at a higher level (in our case, plots) is more important than the sample size at a lower level (trees), especially when the intraclass correlation is high (Snijders 2005; Scherbaum and Ferreter 2008), as was found for our data. Our analysis also confirmed this above statement, as the same number of trees ($n = 50$) from three different simulations (“50 plots with one tree each”, “10 plots with five trees each”, and “five plots with ten trees each”) showed increasing RMSEs with a reduced number of plots and a greater number of trees per plot (Fig. 6a). This indicates that mixed models with a plot random effect would increase model accuracy. However, such models would likely include spatial ef-

fects and not be useful for predicting bark thickness for new diameter measurements in other stands in the study region and, therefore, would not be useful for practice. Consequently, for gains in prediction accuracy to be realized using mixed-effects models, local calibration is required, and research has indicated no universal sample size or location for calibration of mixed-effect bark thickness equations (e.g., Li and Weiskittel 2011), which complicates their implementation.

An important factor influencing the response in predictive quality of bark thickness equations on sample size was model choice. To account for the wide range of sample size requirements

of different model types, we selected a relatively simple bark thickness equation and one that was more complex. Similar to Li and Weiskittel (2011), relatively large differences between these two equations were observed, with the more complex equation generally being preferred. However, it should be noted that the simpler equation performed quite well when sample sizes were adequate (>50 trees). Both equations are preferred over a fixed ratio between inside and outside bark (e.g., Li and Weiskittel 2011).

The sample sizes assessed in this analysis depended on the data we used. Different diameter distributions, other species, or more diverse growing conditions might influence bark thickness variation in a study area and, therefore, the recommended number of sample trees. Tree heights, and therefore also relative heights of measurement locations, were estimated in this study. Predictions errors could have propagated into the bark thickness predictions of eq. 3, and thus, the suggested sample sizes could be higher than actually observed if the tree height was measured. Most published studies that sampled bark thickness of coniferous tree species had sample sizes larger than the calculated number of required sample trees found in this analysis. Typically, more than 100 trees per species (e.g., Johnson 1966; Feduccia and Mann 1976; Gordon 1983; Cao and Pepper 1986; Johnson and Wood 1987; Persson and Downie 1992; Degenhardt 1999; Wilhelmsson et al. 2002; McConnon et al. 2004; Božić et al. 2007; Li and Weiskittel 2011; Cellini et al. 2012; Wehenkel et al. 2012) and even more than 1000 trees per species (e.g., Kozak and Yang 1981; Laasasenaho et al. 2005) have been used. However, these studies did not assess how many trees would have been needed to parameterize a sufficiently accurate bark thickness model, as was done for our data.

Conclusion

This analysis found that bark thickness measurements of Norway spruce in southwestern Germany using a Swedish bark gauges consistently overestimated true values. For best results, we suggest taking at least five bark thickness measurements per location and seven locations per tree. Taller trees had a higher variation of relative bark thickness, and thus, more measurement locations were needed for them than for smaller trees. A regular distance between measurement locations is a simple way to achieve this in practice and a distance of 2 m is suggested. If a bark thickness equation is to be developed for a region, it is critical to sample trees in several plots throughout the region. For the tested simple but much less accurate bark thickness equation, we suggest a sample size of at least 50 plots with five trees each, whereas the more flexible yet more complex equation will require a sample size of at least 10 plots with five trees each. The actual sample sizes are an approximation. Ultimately, the necessary sample size will depend on the dissimilarity of the plots in the tested region and on the models chosen. This suggests that additional studies across different species and larger geographic regions are needed to fully understand how the proposed sample sizes from this analysis will be influenced.

Acknowledgements

This study was funded by the Ministry for Rural Affairs and Consumer Protection Baden-Württemberg, Germany (MLR) and was conducted at the Forest Research Institute Baden-Württemberg (FVA). The authors thank Udo Hans Sauter, Gerald Kändler, and Ulrich Kohnle for fruitful discussions and for providing access to the data from the Altherr et al. (1978) study. We also thank two anonymous reviewers for helping to improve a prior version of this manuscript.

References

Adams, D.C., and Jackson, J.F. 1995. Estimating the allometry of tree bark. *Am. Midl. Nat.* **34**(1): 99–106. doi:10.2307/2426487.

Altherr, F.W. 1964. Accuracy of the Swedish bark measuring gauge. *For. Chron.* **40**(2): 257–258. doi:10.5558/tfc40257-2.

Altherr, E., Unfried, P., Hradetzky, J., and Hradetzky, V. 1974. Statistische Rindenbeziehungen als Hilfsmittel zur Ausformung und Aufmessung unentrindeten Stammholzes: Teil I: Kiefer, Buche, Hainbuche, Esche und Roterle. Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg. Mitteilungen. 61.

Altherr, E., Unfried, P., Hradetzky, J., and Hradetzky, V. 1978. Statistische Rindenbeziehungen als Hilfsmittel zur Ausformung und Aufmessung unentrindeten Stammholzes: Teil IV: Fichte, Tanne, Douglasie und Sitka-Fichte. Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg. Mitteilungen. 90.

Bland, J.M., and Altman, D.G. 2003. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstetr. Gynecol.* **22**(1): 85–93. doi:10.1002/uog.122.

Božić, M., Čavlović, J., Vedriš, M., and Jazbec, M. 2007. Modeling bark thickness of silver fir trees (*Abies alba* Mill.). *Šumarski list.* **131**(1–2): 3–12.

Burman, P., Chow, E., and Nolan, D. 1994. A cross-validatory method for dependent data. *Biometrika.* **81**(2): 351–358. doi:10.1093/biomet/81.2.351.

Cao, Q.V., and Pepper, W.D. 1986. Predicting inside bark diameter for shortleaf, loblolly, and longleaf pines. *South. J. Appl. For.* **10**(4): 220–224.

Cellini, J.M., Galarza, M., Burns, S.L., Martinez-Pastur, G.J., and Lencinas, M.V. 2012. Equations of bark thickness and volume profiles at different heights with easy-measurement variables. *For. Syst.* **21**(1). doi:10.5424/fs/2112211-01963.

Degenhardt, A. 1999. Schaftform- und Rindenstärkenfunktionen für die Baumarten Kiefer und Roterle. *Beitr. Forstwirtschaft. u. Landsch.ökol.* **33**(1): 39–41.

Dolph, K.L. 1989. Nonlinear equations for predicting diameter inside bark at breast height for young-growth red fir in California and southern Oregon. *USDA For. Serv. Res. Note PSW-409.*

Doruska, P.F., Patterson, D., Hartley, J., Hurd, M., and Hart, T. 2009. Newer technologies and bioenergy bring focus back to bark factor equations. *J. For.* **107**(1): 38–43.

Feduccia, D.P., and Mann, W.F. 1976. Bark thickness of 17-year-old loblolly pine planted at different spacings. *USDA Forest Service, Southern Forest Experiment Station, New Orleans, Louisiana, Research Note SO-210.*

Flury, P. 1897. Einfluss der Berindung auf die Kubirung [sic] des Schaffholzes. In *Mitteilungen der Schweizerischen Centralanstalt für das forstliche Versuchswesen.* Edited by C. Bourgeois. Fäsi & Beer, Zürich. pp. 203–255.

Gordon, A. 1983. Estimating bark thickness of *Pinus radiata*. *N. Z. J. For. Sci.* **13**(3): 340–348.

Gray, H.R. 1956. The form and taper of forest-tree stems. *Imperial Forestry Institute, University of Oxford, Institute Paper 32.*

Hoffmann, J. 1958. Untersuchungen über die Rindenstärke der Fichte auf verschiedenen Standorten im südöstlichen Thüringer Wald. *Wiss. Z. TU Dresden.* **7**(2): 361–368.

Husch, B., Beers, T.W., and Kershaw, J.A. 2003. *Forest mensuration.* John Wiley & Sons, Hoboken, New Jersey.

Johnson, F. 1966. Bark factors for Douglas-fir. *USDA Forest Service, Pacific Northwest Research Station, Research Note PNW-34.*

Johnson, T.S., and Wood, G.B. 1987. Simple linear model reliably predicts bark thickness of radiata pine in the Australian capital territory. *For. Ecol. Manage.* **22**(3–4): 173–183. doi:10.1016/0378-1127(87)90103-4.

Kirschner, R. 1976. Die Rindenstärke der Hauptbaumarten Tirols. *Dissertation, University of Natural Resources and Life Science, Vienna.*

Kitikidou, K. 2010. An alternative sample size estimation for optimizing taper equations fitting. *Oper. Res. Int. J.* **10**(2): 199–208. doi:10.1007/s12351-009-0057-3.

Kitikidou, K., and Chatzilazarou, G. 2008. Estimating the sample size for fitting taper equations. *J. For. Sci.* **54**(4): 176–182.

Köcher, D., Matt, G., Oertel, C., and Schneeweiss, H. 1972. *Einführung in die Simulationstechnik.* Beuth-Verlag, Berlin/Köln/Frankfurt.

Kohnle, U., Hein, S., Sorensen, F.C., and Weiskittel, A.R. 2012. Effects of seed source origin on bark thickness of Douglas-fir (*Pseudotsuga menziesii*) growing in southwestern Germany. *Can. J. For. Res.* **42**(2): 382–399. doi:10.1139/x11-191.

Korell, U. 1972. Über Rindendicken der Fichte. *Beitr. Forstwirtschaft.* (2): 54–55.

Kozak, A., and Yang, R.C. 1981. Equations for estimating bark volume and thickness of commercial trees in British Columbia. *For. Chron.* **57**(3): 112–115. doi:10.5558/tfc57112-3.

Kublin, E., and Breidenbach, J. 2013. *TapeR: flexible tree taper curves based on semiparametric mixed models.* R package version 0.3.0.

Laasasenaho, J., Melkas, T., and Aldén, S. 2005. Modelling bark thickness of *Picea abies* with taper curves. *For. Ecol. Manage.* **206**(1–3): 35–47. doi:10.1016/j.foreco.2004.10.058.

Li, R., and Weiskittel, A.R. 2011. Estimating and predicting bark thickness for seven conifer species in the Acadian Region of North America using a mixed-effects modeling approach: comparison of model forms and subsampling strategies. *Eur. J. For. Res.* **130**(2): 219–233. doi:10.1007/s10342-010-0423-y.

Loetsch, F., Zöhrer, F., and Haller, K.E. 1973. *Forest inventory.* Vol. II. BLV Verl. Ges, München, Germany.

Malone, T., and Liang, J. 2009. A bark thickness model for white spruce in Alaska northern forests. *Int. J. For. Res.* **2009**(3): 1–5.

Marshall, H.D., Murphy, G.E., and Lachenbruch, B. 2006. Effects of bark thickness estimates on optimal log merchandising. *For. Prod. J.* **56**(11/12): 87–92.

- McConnon, H., Knowles, R.L., and Hansen, L.W. 2004. Provenance affects bark thickness in Douglas fir. *N. Z. J. For. Sci.* **34**: 77–86.
- Mesavage, C. 1969. Measuring bark thickness. *J. For.* **67**(10): 753–754.
- Meyer, H.A. 1946. Bark volume determination in trees. *J. For.* **44**(12): 1067–1070.
- Persson, B., and Downie, B. 1992. Variation in bark thickness of young *Pinus contorta* var. *latifolia* engelm. in Sweden. *Scand. J. For. Res.* **7**(1–4): 99–106. doi:10.1080/02827589209382701.
- R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Racine, J. 2000. Consistent cross-validated model-selection for dependent data: hv-block cross-validation. *J. Econometrics*, **99**(1): 39–61. doi:10.1016/S0304-4076(00)00030-0.
- Robinson, A.P. 2014. equivalence: provides tests and graphics for assessing tests of equivalence. R package version 0.6.0.
- Robinson, A.P., and Froese, R.E. 2004. Model validation using equivalence tests. *Ecol. Model.* **176**(3–4): 349–358. doi:10.1016/j.ecolmodel.2004.01.013.
- Scherbaum, C.A., and Ferrerter, J.M. 2008. Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, **12**(2): 347–367. doi:10.1177/1094428107308906.
- SDC Skognärings IT-företag. 2014. SDC's instruktioner för virkesmätning: mätning av stocks volym under bark [online]. Available from <http://www.sdc.se/default.asp?id=2833> [accessed 14 January 2015].
- Snijders, T.A.B. 2005. Power and sample size in multilevel linear models. In *Encyclopedia of Statistics in Behavioral Science*. Edited by Brian S. Everitt, and David C. Howell. John Wiley & Sons, Ltd, Chichester, UK, pp. 1570–1573.
- Sonmez, T., Keles, S., and Tilki, F. 2007. Effect of aspect, tree age and tree diameter on bark thickness of *Picea orientalis*. *Scand. J. For. Res.* **22**(3): 193–197. doi:10.1080/02827580701314716.
- Subedi, N., Sharma, M., and Parton, J. 2011. Effects of sample size and tree selection criteria on the performance of taper equations. *Scand. J. For. Res.* **26**(6): 555–567. doi:10.1080/02827581.2011.583677.
- Wehenkel, C., Cruz-Cobos, F., Carrillo, A., and Lujan-Soto, J.E. 2012. Estimating bark volumes for 16 native tree species on the Sierra Madre Occidental, Mexico. *Scand. J. For. Res.* **27**(6): 578–585. doi:10.1080/02827581.2012.661453.
- Wei, Q., Leblon, B., and La, Rocque, A. 2011. On the use of X-ray computed tomography for determining wood properties: a review. *Can. J. For. Res.* **41**(11): 2120–2140. doi:10.1139/x11-111.
- Wilhelmsson, L., Arlinger, J., Spångberg, K., Lundqvist, S.-O., Grahn, T., Hedenberg, Ö., and Olsson, I. 2002. Models for predicting wood properties in stems of *Picea abies* and *Pinus sylvestris* in Sweden. *Scand. J. For. Res.* **17**(4): 330–350. doi:10.1080/02827580260138080.