



# Comparing local calibration using random effects estimation and Bayesian calibrations: a case study with a mixed effect stem profile model

O. Bouriaud<sup>1,2</sup> · G. Stefan<sup>1</sup> · L. Saint-André<sup>3,4</sup>

Received: 31 August 2018 / Accepted: 24 May 2019 / Published online: 21 June 2019  
© INRA and Springer-Verlag France SAS, part of Springer Nature 2019

## Abstract

• **Key message** Local site-level calibration of allometric models was scrutinized. Two Bayesian calibration methods were compared to local random effects estimations. The Bayesian calibration methods proved more effective than local estimation of random effects in reducing prediction bias. The simplest literature-based calibration can be recommended. The local calibration had minor effects on stem volume estimations.

• **Context** The spatial variability of trees allometry has long prompted the necessity for local, site- or stand-level calibrations. Mixed-effect models have enabled a quantitative progress through a local calibration where data are available. More recently, Bayesian statistics brought new alternatives owing to their formal definition of the random effects from prior information.

• **Aims** To compare three local calibration methods: (i) a calibration based on the estimation of the local random effects, (ii) a Bayesian calibration where prevailing measurements are used to produce prior estimations, and (iii) a Bayesian calibration reproducing a calibration based on literature data only.

• **Methods** The three calibrations were compared using a stem taper model developed for Norway spruce in Romania. The taper model was fitted to a large dataset, then applied locally to two high-elevation sites with contrasting growing conditions.

• **Results** The local calibration of mixed-effect models resulted in small gains and high biases. The Bayesian calibrations yielded better results, mostly because the Monte Carlo Markov Chain implementation permitted to tune of all the model's parameters simultaneously. The differences in stem volume estimations were however always very small ranging from  $-5.2$  to  $3.3\%$  of the non-calibrated volume.

• **Conclusion** The Bayesian literature-like calibration performed as well as the calibration using the large dataset ( $4-97\%$  bias reduction according to the tree) and can be preferred for its ease of use.

**Keywords** Localizing · Allometry · Mixed model · Calibration · Bayesian statistics · Taper equation · Stem volume

---

**Handling Editor:** John M. Lhotka

---

**Contribution of the co-authors** OB and LSA designed the study. OB and GS did the field work and measurements. OB and LSA performed the analyses and wrote the paper. All authors read, reviewed, and approved the final manuscript.

---

✉ O. Bouriaud  
obouriaud@gmail.com

<sup>1</sup> National Forest Inventory, National Institute for Research and Development in Forestry “Marin Drăcea”, Bd Eroilor 128, 077190 Voluntari, Romania

<sup>2</sup> University Stefan cel Mare of Suceava, 720229 Suceava, Romania

<sup>3</sup> INRA, UR BEF, F54280 Champenoux, France

<sup>4</sup> CIRAD, UMR ECO&SOLS, 34000 Montpellier, France

## 1 Introduction

The estimation of tree volume remains one of the most important tasks in forestry and forest research. In spite of the investment to develop estimation methods, assessing tree volume remains challenging, mostly because it can hardly be done directly on the field. Estimating the aboveground volume on large scale is usually done by applying allometric models established either on felled trees, using remote diameter measurements on standing trees, or more recently from laser scanning (Calders et al. 2015). However, such models are basically non-generic (strongly influence/dependent on the observations used to fit them) and should be applied with care to trees outside of the population used for the model's fit. Indeed, changes in the silvicultural regime or ecological conditions

proved to be very influential on tree allometry and result in biases for both total volume and its partitioning (Forrester et al. 2017). The sampling itself may have a great influence on the model's fit, on the parameters and their confidence interval, and the residual variance (Cunia 1964). If generality is a desired property for a model, it is necessary to take all the "between stand" and the "within stand" variability sources into account for the equation building (Wirth et al. 2004). The great number of volume or biomass equation developed (e.g., Zianis et al. 2005 for temperate species, Henry et al. 2011 for tropical species, Chojnacky et al. 2014; Neumann et al. 2016) is a direct consequence of the fact that the equations have a limited geographic validity and have often been fitted on few trees. Taper equations can provide volume estimations and are based on measurement of diameters along the tree bole. Such models are therefore particularly sensitive to spatial variations in stem form and represent good indicators of the local errors that can result from applying allometric models outside of their base geographic domain. Inversely, including upper diameter measurements has been proved in many occasions to improve substantially the estimation of the volume thus making them more spatially robust (Kublin et al. 2013). Robust taper equations also require a local, site-specific or stand-specific calibration, for unbiased predictions.

The use of mixed-effect models contributed to improving the robustness of the models and reduce bias. They proved to have smaller prediction errors than fixed-effect models and were successfully utilized to do local calibration of allometric models, hence intrinsically have greater adaptability (Lappi 1986; Calama and Montero 2004; Adame et al. 2008; Cao and Wang 2011). The random parts of the models account for the hierarchical structure of the dataset whereby several trees are sampled from the same site, potentially violating the hypothesis of independence among observations—this represents the majority of samplings in the literature. The random effects also represent the geographic variation, generally a site-level or plot-level variation, and possibly nested in a larger regional or national level. Such models have been recently developed for their ability to provide predictions at different levels of the independent variables. However, the magnitude and the distribution of the random effects are very seldom reported and documented in the literature.

Predicting the random effects parameters for new levels can be made using local calibration, which requires the use of local measurements made on few trees, typically a small fraction of the number used to fit the model. Methods and limitations to the estimation of the random effects for predictions can be found in Ni and Nigh (2012). This calibration requires the estimation of the random effects parameters, which are not formally defined but for which different estimation procedures were proposed and have been successfully applied in different forestry applications (e.g., Hall and Bailey 2001; Cao and Wang 2011; Paulo et al. 2011). Yang

and Huang 2011 Another possible calibration has been recently proposed and seems to produce superior fits, particularly when the dataset is small: the Bayesian calibration, which has been recently advocated (Zapata-Cuartas et al. 2012; Zell et al. 2014). We are aware that the local estimation of random effects is generally based on the empirical Bayes approach (Vonesh and Chinchilli 1997, Meng and Huang 2009), and thus, local estimation of random effects can be considered as a Bayesian estimation. But here, we refer to Bayesian calibration only for the situation where the model parameters are all fitted based on Bayesian models, which is done here within the Monte Carlo Markov Chain framework (MCMC).

In order to test and compare the two calibration methods, the random effects parameters estimation or the Bayesian calibration, a taper model for Norway spruce (*Picea abies* L.) was fit from a detailed measurement campaign over Romania. The taper model was fit with random effects using two methods, maximum likelihood and Bayesian, then applied to trees sampled from sites not used for the fit and with different growing conditions, representing a difficult or risky extrapolation. Last, a Bayesian calibration based on fit statistics, not direct measurements, is implemented to test the potential of literature-based calibration where local data are not directly available.

## 2 Material and methods

### 2.1 The datasets

Two sets of data were used for the analysis: the first set, referred to as the fit dataset, was used to fit the models while the second set was used to perform their local calibration and validation.

#### 2.1.1 The fit data set

The 16 plots were inventoried according to the National Forest Inventory procedure (Marin et al. 2010), and 3 to 10 trees per plot were selected for destructive measurements as a constant fraction of the total number of trees in the plot. The plots consisted in two concentric circles of fixed radius:  $56 \leq \text{breast-height diameter (dbh)} \leq 285$  mm within 0–7.98 m and  $\text{dbh} \geq 285$  mm within 0–12.96 m. Within each plot, tally trees are measured both in *dbh*, total height, base of living crown, and base of dead crown. Plots were dominantly set up on clear-cut areas (building road or electric lines) so that all tree size (in diameter and height) but particularly big trees could be sampled. Thus, the selection of the trees sampled was not limited by silvicultural constraints, which would otherwise have led to felling trees from lower canopy positions. Successive diameters (outside-bark) were measured along

the stem using a caliper after felling at prescribed heights: 0.1, 0.5, 1, 1.3, 2, 2.6, and 3 m and then every 2 m to the tip.

The fit sample covered a range of *dbh* from 5.6 to 53.4 cm, and of total height from 1.3 to 41.6 m (Table 1).

### 2.1.2 The application set

Two high elevation stands were sampled in the buffer zone of forest reserves, thus differing from the stands used for the fit by their low-intensity (to no) forest management and a high elevation as compared to the fit set. The first site (47° 46' 19" N, 25° 48' 70") was located in the Rărau Giurnalău Mountains at an elevation of 1350–1450 m on a slope of 25–35°. The second site (47° 19' 02" N, 25° 15' 10") was located at a similar elevation (1400–1500 m) and slope range, about 40 km apart from the first site.

At both sites, the application set was constituted of 24 trees covering the diameter range of the stands, from 6 to 66 cm, measured after felling. Two trees were sampled per diameter class in each stand so the application sets can be randomly split in two equally sized batches of trees (12 trees) having very similar dimensions: one batch was used for the local calibration and referred to as the calibration batch, the other for describing the fit quality and quantifying the impact of the calibration, referred to as the evaluation batch.

## 2.2 The stem taper equations

Many taper equations have been published, which differ both in their form and in the number of their parameters. Flexibility seems to have been an important aspect in the development of the different equations, based on variety of functions, from

exponential to spline. Choosing the best suited taper equation could be a study on its own, given the great diversity of published equations: polynomial, variable-exponent (Kozak 1997), spline (Kublin 2003), continuous or segmented curves (Max and Burkhart 1976). Although very different in their shape and parametrisation, different models can have very comparable performance, particularly for smooth and regular stems such as that of spruce. As Arias-Rodil et al. (2015) put it, “none can be considered best.” The ultimate choice could have been based on comparing the third decimal of some statistics but the choice here was rather related to the possibility of including covariates and to the necessity to have a model addressing separately each stem portions explicitly.

Thus, the stem taper equation used in this study is a versatile equation successfully fit to a variety of species such as eucalyptus (Saint-André et al. 2002; Gomat et al. 2011) and teak (Adu-Bredu et al. 2008). Given the *dbh* and the total height of the tree (*H*), the equation relates the relative diameter *dr* ( $dr = d/dbh$ ) to the relative height *hr* ( $hr = h/H$ ) as:

$$dr = a_1 \times \left( (1 - a_2 \times hr)(1 + a_3 \times e^{-a_4 \cdot hr}) - (1 - a_2) \right) \times hr^{a_5} \quad (1)$$

This taper model has the advantage of having parameters specifically related to each of the main stem fractions: the stem taper ( $a_1$  and  $a_2$ ), the butt-swell and taper of the lower stem part ( $a_3$  and  $a_4$ ), and the stem top ( $a_5$ ).

This initial model was further developed by testing possible effects of covariables on the parameters of the model themselves (Adu-Bredu et al. 2008; Gomat et al. 2011). In our case, the model (1) was fit once on the fit dataset individually, i.e., separately for each tree, then the 126 set of parameters were

**Table 1** Comparison of the site and destructively sampled tree characteristics for the fit and evaluation datasets. Site ID; site name; elevation (m a.s.l.); stand density; age class; sampled trees; *dbh* range (cm); tree height range (m)

Site	<i>N</i> trees	Elevation (m a.s.l.)	Stand density (stem <sup>-1</sup> )	Age (years)	<i>dbh</i> range of sampled trees (cm)	Height range of sampled trees (m)	Stem volume (m <sup>3</sup> ) (min–max)
<b>Fit dataset</b>							
Demacușa	15	900–1100	760–800	85	20.8–44.9	22.5–34.5	0.364–2.476
Frasin	6	825	440	95	8.3–49.7	6.9–37.1	0.022–2.864
Crasna	17	760–810	460–560	60–90	27.9–43.1	31.5–38.2	0.471–2.419
Valea Doftanei	33	1100–1200	600–740	50–70	9.4–22.4	6.3–24.5	0.032–0.455
Padeș	27	1300–1360	40–460	60–100	5.6–37.7	4.2–25.8	0.008–1.107
Delnița	10	1000–1100		130	39.8–66.1	38.9–45.1	2.374–6.825
Valea Putnei	18			20	7.5–13.3	10.0–18.4	0.024–0.118
<b>Evaluation dataset</b>							
Rărau—batch 1	12	1350–1450	350	130	9.2–58.7	6.6–41.4	0.112–5.252
Rărau—batch 2	12	1350–1450	350	130	7.1–60.9	5.5–44.1	0.199–4.823
Călimani—batch 1	12	1400–1500	420	140	6.7–58.0	4.9–43.1	0.013–5.764
Călimani—batch 2	12	1400–1500	420	140	8.1–62.3	6.5–42.2	0.257–5.012

correlated to candidate covariables:  $h/d$ ,  $dbh$ , top height and a robustness index estimated as  $\sqrt{(\pi dbh)/H}$ . There was little evidence of correlation, except between  $a_1$  and tree height,  $ht$ . The form of the relationship was clearly non-linear, and while different forms were tested, a simple hyperbolic model seemed the most suited. This led to the more complete model (2) derived from model (1) as:

$$dr = \left( a_1 \frac{a_2}{1 + a_3} \right) \times \left( (1 - b \times hr)(1 + c \times e^{-d \cdot hr}) - (1 - b) \times hr^e + \varepsilon \right) \quad (2)$$

The model parameters for which a random site-level effect was included were the parameters  $b$ ,  $c$ , and  $d$ . It was not possible to estimate random effects for all parameters because of computational issues despite the fair amount of available data ( $N = 1896$ ). The model with random effects was in the form:

$$dr = \left( a_1 \frac{a_2}{1 + a_3} \right) \times \left( (1 - b_2 \times hr)(1 + b_3 \times e^{-b_4 \cdot hr}) - (1 - b_2) \times hr^{b_5} + \varepsilon \right) \quad (3)$$

with  $\forall k \in [2, 4]$ ,  $b_k = \beta_k + v_{k,j}$  where  $\beta_k$  is the fixed part of the parameter  $\beta_k$ , and  $v_{k,j}$  is the random, site-level component such that for each site  $j$ ,  $v_{k,j} \sim N(0, \sigma_k^2)$ , and model's residual  $\varepsilon \sim N(0, \sigma^2)$ . The variance components  $\sigma_k^2$  and  $\sigma^2$  are unknown and need to be estimated during the MCMC fit. Convergence cannot be achieved if  $b_5$  is expressed as a random-effect parameter. The final form of the model was chosen based on fit statistics for different combinations of parameters using the Akaike information criterion (AIC), the bias, the model efficiency (MEF), and a pseudo- $R^2$ . The bias, MEF, and pseudo- $R^2$  were computed as:

$$Bias = \frac{\sum (y - \hat{y})^2}{n}$$

$$MEF = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}$$

$$Pseudo-R^2 = corr(y, \hat{y})^2$$

where  $y$  is the measured relative diameter,  $\hat{y}$  the predicted value,  $\bar{y}$  the mean value over all  $n$  observations. The comparison of the fit statistics for different model forms is displayed in Table 6 (Appendix).

Autocorrelation in errors caused by the use of several measurements per tree (Meng and Huang 2009) was not explicitly corrected since most of it was absorbed by the random effects, as also concluded by Trincado G, Burkhart (2006). Another reason is that the autocorrelation, if any, cannot be supposed to be equal among observation sets and thus cannot be directly modeled.

## 2.3 Statistical methods

### 2.3.1 Mixed effect models maximum likelihood calibration

The taper model (2) was fit by maximum likelihood (ML) using the nlme library (Pinheiro et al. 2016) of R (3.2.1). The local calibration of a mixed effect model consists in estimating the random effect parameters for new individuals. According to Ni and Nigh (2012), there are two methods for estimating the random effect parameters and two for predicting the response variable, leading to four combinations, of which only two are coherent. In short, while the expected value of the random effect parameters is 0 by definition, the value used for calibration is more generally the estimation obtained at the latest iteration of the model fitting (iterative) process. Such estimation is itself based on a predictor, for instance the empirical best linear unbiased predictor (EBLUP). In this situation the construction of the prediction distribution is made based on the EBLUP expansion procedure, also known as the first-order conditional expectation (FOCE), initially proposed by Lindstrom and Bates (1990) and described and implemented by Huang et al. (2009). There are many examples of calibration following the method described by Huang et al. (2009), and it will not be described in further details here. The FOCE procedure followed to estimate the random expansion values here was based on Huang et al. (2009) and Ni and Nigh (2012). In this procedure, the EBLUPs were first estimated by iteration as presented in the R code associated to this article (Bouriaud et al. 2019), then applied to the new data.

### 2.3.2 Bayesian calibrations

The taper equation fit using the nlme package of R was subsequently used in a Bayesian Monte Carlo Markov Chain framework using Stan (Hoffman and Gelman 2011) with the library rstan (2.8.0, Stan Development Team 2016). Stan uses a Hamiltonian Monte Carlo algorithm, the No-U Turn Sampler. The default settings of the library were used for the runs presented in this study. It typically uses half of the prescribed number of iterations for the warmup, the posterior distributions being based on the successive iterations. Posterior estimates were based on post-warmup iterations only. The convergence was checked using the default functions of the rstan package using the Rhat statistics (Gelman and Hill 2007). Each post-warmup iteration provides a set of model parameters, predictions, and error estimations.

The first Bayesian calibration is based on the raw data of the fit dataset and is therefore referred to as a database Bayesian calibration (BCd). In the Bayesian framework, the taper equation was first fit to the fit dataset, which produced a first set of posterior distributions for the parameters. The equation was then fit to the application set. For this second fit, the

priors of both the parameters and the variance components were taken from the posterior distribution of the fit realized on the fit data set. Both fittings are done simultaneously within each loop of the MCMC. The distribution of the parameters was supposed to be normal with means and variance to be determined based on the posterior distributions.

The second calibration method (BCI) is the one that would be used in the situation where the fit dataset cannot be used, and the prior relies on the fitted values of the parameters only. This situation is perhaps the most typical, when the parameters of the model and their error can be obtained from the literature but the (raw) data used for the fit are not. Several examples were reported in the field (see Zell et al. 2014). In our example, the parameters are obtained from the fit on the calibration sites, hence independently from the application zone, and only the fitted value of the parameters is being used as prior information for the calibration on the application site. This calibration is referred to as the literature-based Bayesian calibration (BCI) where prior information comes from independent studies and parameter distribution is often not reported in the published papers.

### 2.3.3 Volume estimations

The bole volume was estimated by numerical integration over 100 stem portions using the Smalian formula (see, e.g., Li and Weiskittel 2010 and references herein). For the MCMC estimations, the integration was repeated for each of the post-warmup iterations of the MCMC, thus providing an estimation of the standard deviation of the volume.

### 2.3.4 Fit statistics

The fit statistics included the mean bias of model predictions (bias), the root mean square error of prediction (RMSEP), estimated as:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (Y_{i,j} - \hat{Y}_{i,j})^2}{n}}$$

and

$$bias_j = \frac{\sum_{j=1}^n Y_{i,j} - \hat{Y}_{i,j}}{n}$$

where  $n$  is the number of observations across all the trees and positions in a given dataset,  $Y_{i,j}$  is the measured stem diameter for a given position  $i$  in a tree  $j$ , and  $\hat{Y}_{i,j}$  is the corresponding estimated stem diameter. Both RMSEP and bias are computed on the dataset used for the fit and on the evaluation dataset, as a measure of the prediction capacities of the model.

The effects of the calibration on the tree bole volume was also tested based on the RMSEP and bias using the same

formulas, such that  $Y_{i,j}$  is the measured volume (based on measured diameters) of tree  $i$  in batch  $j$ , while  $\hat{Y}_{i,j}$  is the corresponding estimated volume based on predicted diameters for each of the non-calibrated or calibrated models.

## 3 Results

### 3.1 Taper model fit

The fit dataset was constituted of 126 spruce trees measured from 16 different plots across Romania with diameter ranging from 5.6 to 66 cm and a mean slenderness (ratio of total height to diameter at 1.3 m) of  $9.996 \pm 2.571$ . The taper model (2) was fit with three random plot-level terms using maximum likelihood (Table 2). The model did not present signs of bias (Figs. 1 and 2) and fitted conveniently the stem profile of the trees despite the great variability of their provenience.

Based on the same input data, the model was fit in the Bayesian MCMC framework. The prior was the estimation of the ML fit, which was meant to maximize the similarity between the ML and the Bayesian fits, since the use of different start or prior values did not prove influential in our case (Fig. 3). The metropolis acceptance probability was high (0.87) and the Rhat values very close to 1 for all the model parameters. ML and Bayesian fit resulted in very similar parameter estimates. The discrepancy in the parameters estimation between the ML and Bayesian fit was less than 4% of the estimation on average but parameter stability was higher for the ML fit (low std error/parameter value) (Table 2). The slight differences in estimated parameters had little influence on the stem profiles fitted, which were very similar for both models, with a mean absolute difference of 0.01 (Fig. 3).

### 3.2 Likelihood calibration

First, the ML estimations of the model (3) were used directly to estimate the stem profile of the trees of the application sets. The prediction errors without local calibration were important in the high-elevation dataset (Fig. 4, site 2 batch 2). The application set was subsequently split in two: one batch of 12 trees was used to estimate the random effects using the FOCE methods while the second batch of 12 trees (evaluation batch) was held for testing the model after calibration.

The estimated random effects were globally smaller than the parameters themselves (Table 3) and represented 31–92% of the standard error of the parameter prediction. The calibration resulted in a mild reduction of the RMSEP (by 6.5 to 25.3% according to the test site, estimated as percent of the non-calibrated model's RMSEP) and the bias (up to 97% reduction) but not in all situations (Fig. 4, Table 4) because some

**Table 2** Inference of the parameters using maximum likelihood (ML, left 4 columns) versus Bayesian model in Stan (fitted within MCMC using 2 chains each with 1000 post-warmup draws) for the taper model (3) fitted on the fit dataset

Parameters	ML parameter estimation				Bayesian MCMC estimation			
	Value	Standard error	t value	<i>p</i> value	Mean	Standard error of the mean	Standard deviation	Rhat
a1	1.6076	0.1313	12.2473	0.0000	1.6394	0.2037	0.0129	1.0021
a2	0.6739	0.1244	5.4161	0.0000	0.7064	0.1956	0.0125	1.0021
a3	3.0271	0.9548	3.1705	0.0015	3.3337	1.2112	0.0551	1.0008
b2	0.6304	0.0275	22.9100	0.0000	0.6342	0.0332	0.0019	0.9995
b3	0.3493	0.0369	9.4708	0.0000	0.3683	0.0424	0.0016	1.0018
b4	60.2292	6.4776	9.2981	0.0000	69.6040	9.0887	0.3565	0.9999
b5	4.4087	0.1729	25.6036	0.0000	4.4511	0.1824	0.0041	1.0000
$\epsilon_1$	0.0431				0.0431	0.0007	0.0000	0.9999

trees displayed a profile different from the others and had a strong influence on the validation statistics.

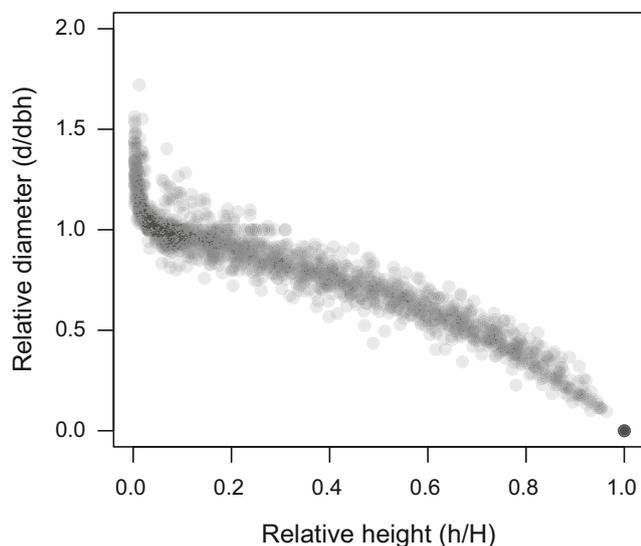
### 3.3 Bayesian calibrations

The data-based calibration (BCd) was based on the posterior estimation of the model (2) fit parameters, which were used as priors for the fit of the same model to the new test data. The new parameter values for the test dataset differed substantially from the prior values (Table 3) from 1.3 to up to 10%. After calibration, the RMSEP on the fit batch generally decreased more strongly than using ML FOCE calibration: from 4 to 65% according to the site (Table 4) and the bias from 4 to 97%. On the evaluation set, the RMSEP on the decreased by

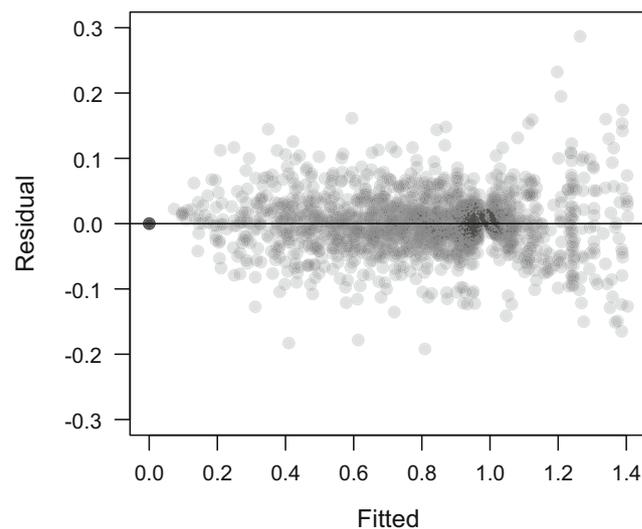
5–65% and the bias from 4 to 97% (Fig. 4). The small size-tree on site 2-batch 1 behaved differently from site 2-batch 2 and was responsible for the higher bias and smaller RMSEP decrease. However, the fit with BCd was globally much better than that of the ML.

Similarly, the literature-based Bayesian calibration BCl resulted in decreased RMSEP (4–65%) and bias (7–96%), and had values very comparable to that of the BCd (Fig. 4). Occasionally, the bias reduction was even stronger than in the BCd.

The distribution density of the diameter prediction errors (measured-predicted) showed that the absence of calibration resulted in larger errors and bias, which were strongly reduced by the local calibration (Fig. 5). The calibrated predictions (FOCE) represented a mean prediction error (over all the validation trees) of  $\sim 3.16$  mm (against 3.34 mm for non-calibrated predictions) and a maximum error of  $\sim 22.94$  mm

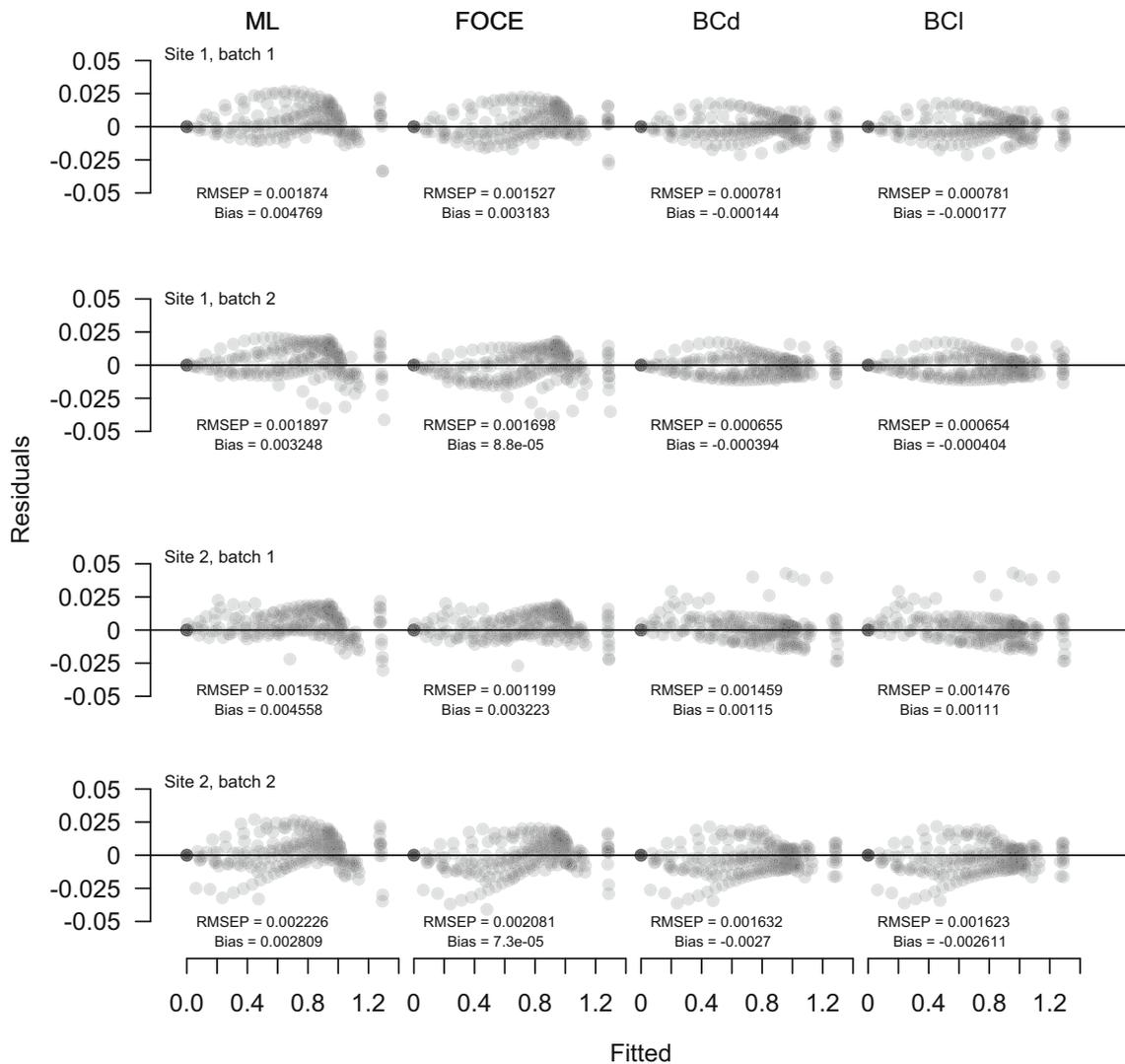
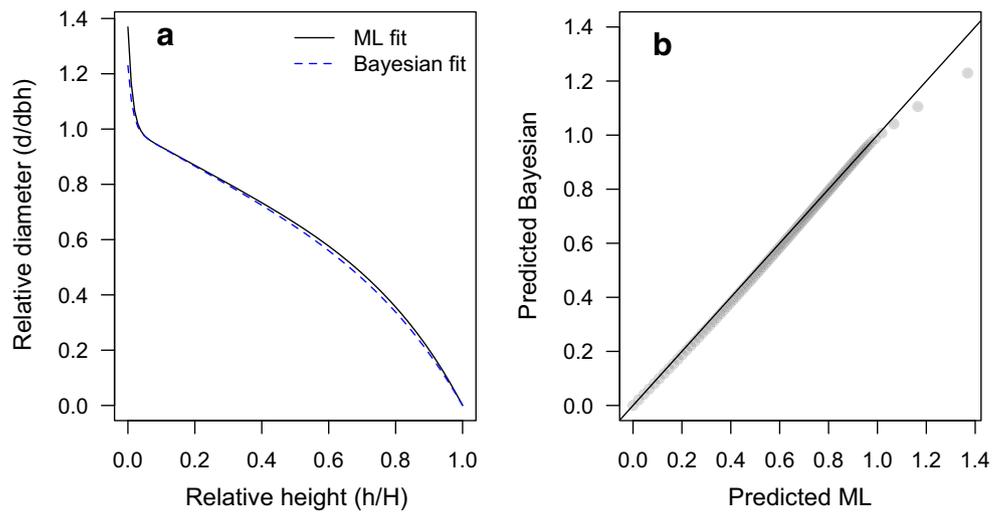


**Fig. 1** Stem profile of all trees used for the taper model fit based on the ML fit of the model (3). The relative height is the ratio between the height at any point above the ground ( $h$ ) to the top height of the tree ( $H$ ). Similarly, the relative diameter is the ratio between the diameter at any point ( $d$ ) to the breast-height diameter ( $dbh$ )



**Fig. 2** Residuals of the model (3) fitted by maximum likelihood versus model predictions

**Fig. 3** Comparison of the ML and Bayesian stem profile model fit. **a** Modeled diameter against relative height. **b** Comparison of the modeled diameters for the fit dataset



**Fig. 4** Residual graphs of the predicted relative diameter for each application site and evaluation batch. ML refers to the mixed-effect prediction not calibrated, ML FOCE to the FOCE calibration, BCd to the data-based Bayesian calibration, and BCI to the literature-based

calibration. Within each site models are calibrated on one batch (the calibration batch) of 12 trees, then applied on the other batch (the evaluation batch) and residuals are expressed as measured-predicted

**Table 3** Estimation of the random effects and model (3) parameters for each site and calibration batch. BCd refers to the Bayesian data-based calibration while BCI refers to the Bayesian literature-based calibration

Parameter	M.L. fixed	Site	Random parameter	New parameter	BCd	BCI
b2	0.6303719	Ap. 1 batch 1	-0.01015404	0.6202179	0.6297889	0.6300823
		Ap. 1 batch 2	-0.01554865	0.6148233	0.6388681	0.6394577
		Ap. 2 batch 1	-0.00846694	0.621905	0.622501	0.6223621
		Ap. 2 batch 2	-0.01487004	0.6155019	0.643453	0.6438258
b3	0.3492837	Ap. 1 batch 1	0.01355861	0.3628423	0.3611102	0.3615243
		Ap. 1 batch 2	0.01546979	0.3647535	0.3561766	0.3563752
		Ap. 2 batch 1	0.01378309	0.3630668	0.359316	0.3595053
		Ap. 2 batch 2	0.01262041	0.3619041	0.3560215	0.3566447
b4	60.229160	Ap. 1 batch 1	5.964609	66.19377	73.45418	73.7597
		Ap. 1 batch 2	5.227134	65.45629	74.76358	75.00014
		Ap. 2 batch 1	5.428521	65.65768	73.07313	73.26324
		Ap. 2 batch 2	5.724993	65.95415	76.77054	77.11607

from the non-calibrated predictions. But the Bayesian calibrations reduced even further the error magnitude with a mean error of 2.45 mm and 7% more diameters within the range of -1 to +1 mm from the measured value.

### 3.4 Consequences on the estimation of the stem volume

Stem volume is an integrated representation of the taper fit. In order to quantify the consequence of the deviations in the modeled stem profile, the bole volume resulting from the integration of the model were computed for each variants: ML without local calibration, ML with local calibration, BCd, and the BCI. The difference between local calibrated and uncalibrated predicted volumes ranged from -5.2 to 3.3% (Fig. 6), thus representing a maximum deviation of  $\sim 0.1 \text{ m}^3$  from the non-calibrated predictions at tree level. Globally, for the two test stands, local calibration resulted in larger predicted volumes (Table 5). The difference between BCd and BCI calibrations was smaller than the prediction's confidence interval and

can be deemed insignificant. Nevertheless, the calibration resulted in systematic and strong reductions of the bias of volume estimations, with the strongest reductions for the Bayesian calibrations (Table 5). The RMSEP was likewise noticeably reduced by the calibration with the exceptions of the FOCE calibration.

## 4 Discussion

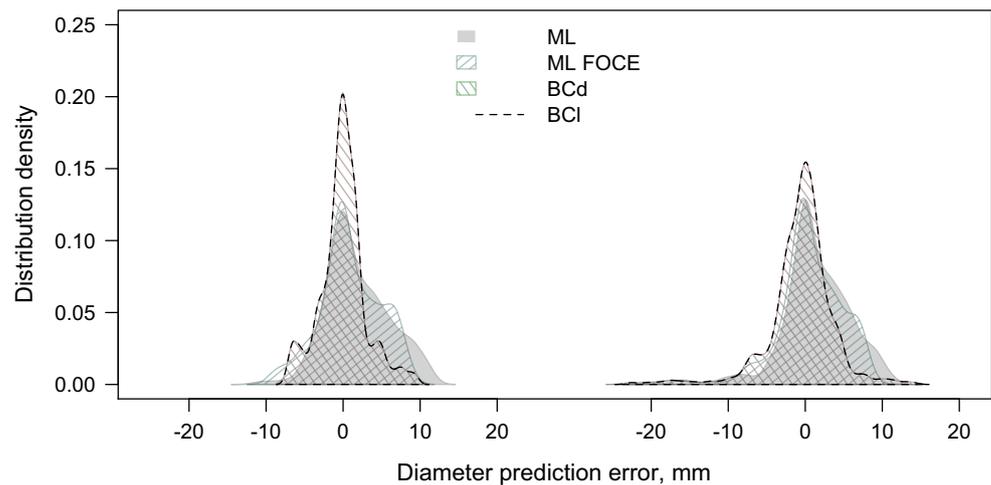
Calibrating the model improved markedly the predictions but high biases remained. The Bayesian calibrations proved superior to the ML one, particularly for their ability to reduce biases. The BCI proved as good as the more difficult calibration using raw data BCd.

These results confirm the fact that the Bayesian calibration improves substantially the prediction abilities of a model. Interestingly, the Bayesian calibration based on the sole knowledge of the parameters fit on external data (the literature-based calibration) proved as good as the more

**Table 4** RMSEP and mean bias decrease of the diameter model predictions as percent of the non-calibrated ML model. FOCE refers to the calibration based on estimating random effects using the first-order conditional expectation procedure, BCd refers to the Bayesian data-based calibration, and BCI refers to the Bayesian literature-based calibration

Calibration dataset	Evaluation dataset	Batch type	RMSEP decrease (%)			Mean bias decrease (%)		
			FOCE	BCd	BCI	FOCE	BCd	BCI
Ap. 1 batch 1	Ap. 1 batch 1	Calibration	13.20	56.15	55.98	48.69	71.93	71.04
	Ap. 1 batch 2	Evaluation	18.49	58.34	58.32	33.25	96.98	96.28
Ap. 1 batch 2	Ap. 1 batch 1	Evaluation	10.51	65.49	65.50	97.29	87.86	87.56
	Ap. 1 batch 2	Calibration	20.66	56.26	56.26	66.52	88.24	88.47
Ap. 2 batch 1	Ap. 2 batch 1	Evaluation	25.38	59.37	59.49	59.08	86.88	88.84
	Ap. 2 batch 2	Calibration	21.75	4.81	3.66	29.28	74.76	75.64
Ap. 2 batch 2	Ap. 2 batch 1	Evaluation	6.49	26.69	27.10	97.40	3.86	7.04
	Ap. 2 batch 2	Calibration	8.92	48.66	48.59	48.45	67.02	65.61

**Fig. 5** Distribution density of the prediction error of the diameter on both application sites according to the calibration method



elaborate method—but difficult to implement—which requires the raw data. Hence, the literature-based calibration seems a very efficient and desirable option because it is very simple to implement and much faster. The new sampling algorithm implemented in Stan has brought a significant improvement compared to previous MCMC algorithms, which could be slow. But even so, the fit of the data-based Bayesian calibration takes several minutes of computation, compared to several seconds only for the literature-based calibration. Analyzing aboveground biomass equations for several species, Zell et al. (2014) came to the same conclusion that a Bayesian calibration performs very well even with reduced sample size, but has the advantage over other biomass estimation procedures to provide prediction intervals.

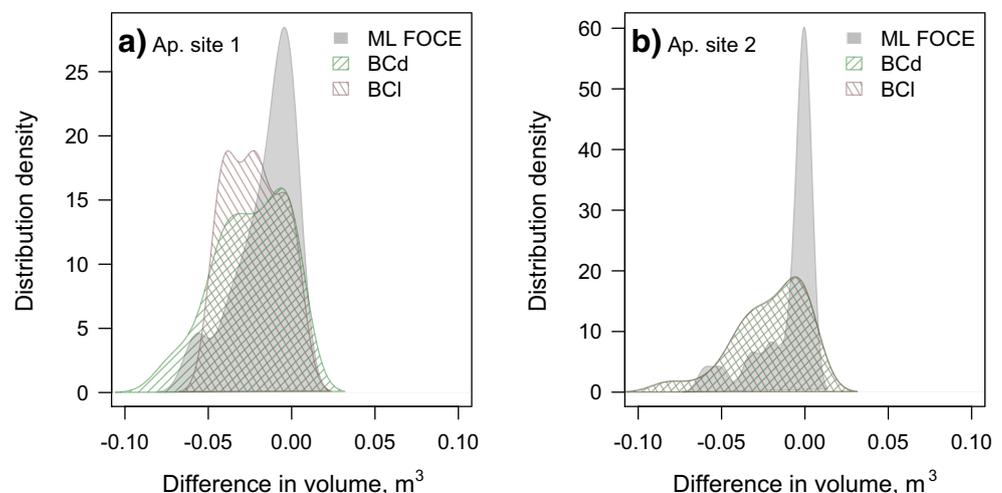
The ML calibration appeared to bring fewer improvements than the Bayesian calibration. The main reason for the discrepancy between the two families of method mostly comes from the fact that the ML calibration can only tune the parameters of the model that have a random component. In this study, three out of six parameters had a random component. Increasing the

number of random parameters may not always be possible and convergence issues can rise. For one given model, the number of parameters that can be left as mixed-effect parameters may differ according the data as, e.g., in Meng and Huang (2009). The Bayesian calibration therefore has an appreciable advantage over the calibration of mixed-effect models that all parameters are being optimized together during the calibration procedure and has no computational issues, long as the Markov chain converges.

The model fitted used diameters measured at successive heights, up to the tree top. Contrarily to other studies where the model is localized using a single additional diameter (i.e., on top of the traditional *dbh*) taken at an upper stem position (e.g., Cao 2009), the model here was fitted based on more than 5 additional measurements per tree. Despite using more measurements, biases remained and the prediction errors without calibration remained important.

The number of trees used for the calibration may be influential. Here, we used a balanced split design where 12 trees were constantly used for both the calibration and the

**Fig. 6** Distribution of the difference in estimated stem volume between the non-calibrated ML estimation and the calibrated estimations over both sites



**Table 5** Comparison of the batch-level descriptive statistics of the estimated tree volume, according to the calibration method. The mean volume and the RMSEP are expressed in liters

Dataset	Statistic	ML	FOCE	BCd	BCI
Ap. 1 batch 1	Mean volume	1551.6	1551.5	1574.0	1573.7
	RMSEP	42.2937	44.5706	29.2764	29.2764
	Bias	21.6329	0.5934	-0.9125	-0.8470
Ap. 1 batch 2	Mean volume	1885.7	1905.4	1905.4	1905.4
	RMSEP	64.3553	59.2268	0.5406	0.5406
	Bias	29.9579	19.3224	3.4273	3.3120
Ap. 2 batch 1	Mean volume	1586.6	1597.3	1613.2	1613.3
	RMSEP	39.0386	39.1232	32.6995	32.6995
	Bias	21.6443	21.7363	-0.7996	-0.5364
Ap. 2 batch 2	Mean volume	1551.6	1551.5	1574.0	1573.7
	RMSEP	53.1784	61.3128	44.2898	44.2898
	Bias	14.8974	-4.8770	-4.7998	-4.8373

evaluation, and were sampled in the same diameter size class, to avoid extrapolation pitfalls. Our study proves that a reduced number of trees are sufficient to reduce substantially the prediction error and the bias of a model. Nevertheless, it should be noted that some trees displayed divergent profiles and created difficulties both in the fit and the validation. One way to cope with such trees may be to realize successive random draws from the same sample and use the average parameter value over the draws.

The estimated random effects were always smaller than the standard error of the parameters estimation, yet contributed greatly to improve the model fit. The gain in using random effects was perhaps maximized in this study, because the stands used for the calibration and validation were purposely sampled from sites displaying very different growing conditions. This situation is however representative of protected forests where tree form differs significantly from managed and open forests and where sampling is often prohibited. In our case, gain resulting from local calibration can be seen from the parameter values obtained for the two application sites: they were substantially lower (*b*) and higher (*c* and *d*) than the values obtained for the mean population of the fit dataset (Table 4) meaning that trees of these forest reserves have a lower taper (more cylindrical) and higher butt-swell at a fixed diameter and height. A direct application of the model, without local calibration would thus result in an underestimation of the true volume.

The consequences of a non-local calibration on the estimation of the volume were in the 5% range, which can be considered as low. But because of the bias, the errors would not compensate between trees and any stand-level application would result in a severe deviation of the per-hectare volume, in our case in the sense of a sub-estimation of the stand-level or per-hectare volume.

The methodology proposed in this paper is then of major importance for forests where sampling is difficult and even

impossible (for instance in tropical rain forests). Given its parsimony, the local calibration can be performed from literature parameters and from very few felled trees (whenever possible), or from non-destructive measurements (see Picard et al. 2012), or from new technologies such as terrestrial LiDAR (Hackenberg et al. 2015). The literature-based calibration offers the possibility to reduce substantially such errors by defining an informative and constraining prior. The results can be generalized to other situations and other allometric models. The generality of the results comes from the fact that all the parameters of the allometric models can be tuned by the calibration simultaneously, and the calibration does not depend on a particular structure or model form. In our example, the model has twice as many parameters (6) as in classic volume or biomass allometric models that have only 2 or 3 parameters—plus potentially one more for the variance model in case of heteroscedasticity. This opens new possibilities to considerably improve national volume (and therefore biomass) estimations at national scales, especially for tropical countries where data availability can be very scarce for both volume and biomass (Henry et al. 2011, 2013, 2015).

## 5 Conclusions

Several conclusions may be drawn from the study. First, the local calibration of the stem profile model reduced markedly the prediction errors: 4–65% reduction in the RMSEP for all calibration methods. The Bayesian calibrations performed better than the calibration based on estimating random effects with a stronger bias reduction (4–97% according to the tree). The Bayesian calibration based on literature data and a calibration dataset performed as well as the complex calibration based on both a fit dataset and a calibration dataset. The literature-based calibration thus represents a very promising method to locally calibrate allometric models with minimal

measurements. However, regardless of the method, the local calibration of the taper model did not result in significant changes in the estimated stem volume with differences between non-calibrated and calibrated estimations ranging from  $-5.2$  to  $3.3\%$  of the non-calibrated volume.

**Acknowledgements** This study will provide bridging functions for spruces in the harmonization process of this project. The authors wish to thank the Editor John Lhotka, and anonymous reviewers for their helpful and constructive suggestions.

**Funding** This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS – UEFISCDI, project number PN-II-ID-PCE-2011-3-0781. OB, SG, and LSA are financially supported by the project H2020 DIABOLO (grant agreement no. 633464). The UR BEF, one of the research units contributing to the present study, is supported by the French National Research Agency through the Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01).

**Statement on data availability** Data and R codes are available in the Zenodo repository (Bouriaud et al. 2019) at the following address: <https://doi.org/10.5281/zenodo.2551028>

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

## Appendix

**Table 6** Fit statistics for different forms of the taper model, fitted using maximum likelihood

Parameter	Converges	AIC	Bias	$R^2$	MEF
<i>b</i>	Yes	-6029	0.002322779	0.9784704	0.9784702
<i>c</i>	Yes	-5837	0.002587035	0.9760209	0.9760208
<i>d</i>	No				
<i>e</i>	No				
<i>b,c</i>	Yes	-6269	0.001966337	0.9817743	0.9817741
<i>b,d</i>	Yes	-6195	0.00204966	0.981002	0.9810017
<i>b,e</i>	Yes	-6086	0.00221621	0.979458	0.979458
<i>c,d</i>	Yes	-5888	0.002466034	0.9771426	0.9771424
<i>c,e</i>	Yes	-6193	0.002054657	0.9809666	0.9809554
<i>b,c,d</i>	Yes	-6374	0.001816626	0.9831629	0.9831617
<i>c,d,e</i>	Yes	-6283	0.001907342	0.9823435	0.9823209

## References

- Adame P, del Río M, Canellas I (2008) A mixed nonlinear height-diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). *For Ecol Manag* 256:88–98
- Adu-Bredu S, Bi AFT, Bouillet JP, Me MK, Kyei SY, Saint-André L (2008) An explicit stem profile model for forked and un-forked teak (*Tectona grandis*) trees in West Africa. *For Ecol Manag* 255:2189–2203
- Arias-Rodil M, Diéguez-Aranda U, Rodríguez Puerta F, López-Sánchez CA, Canga Libano E, Cámara Obregón A, Castedo-Dorado F (2015) Modelling and localizing a stem taper function for *Pinus radiata* in Spain. *Can J For Res* 45:647–658
- Bouriaud O, Ștefan G, Saint-André L (2019) Supporting data and code for the article «Comparing local calibration using random effects estimation and Bayesian calibrations. A case study with a mixed effect stem profile model». V1. Zenodo. [Dataset]. <https://doi.org/10.5281/zenodo.2551028>
- Calama R, Montero G (2004) Interregional nonlinear height diameter model with random coefficients for stone pine in Spain. *Can J For Res* 34:150–163
- Calders K, Newnham G, Burt A, Murphy S, Raunonen P, Herold M, Culvenor D, Avitabile V, Disney M, Armston J, Kaasalainen M (2015) Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods Ecol Evol* 6:198–208
- Cao, Q. 2009 Calibrating a segmented taper equation with two diameter measurements. *South J Appl For* 33:58–61
- Cao Q, Wang J (2011) Calibrating fixed-and mixed-effects taper equations. *For Ecol Manag* 262:671–673
- Chojnacky DC, Heath LS, Jenkins JC (2014) Updated generalized biomass equations for north American tree species. *Forestry* 87:129–151
- Cunia T (1964) Weighted least squares method and construction of volume tables. *For Sci* 10:180–191
- Forrester DI, Benneter A, Bouriaud O, Bausch J (2017) Diversity and competition influence tree allometric relationships—developing functions for mixed-species forests. *J Ecol*, 105(3):761–774
- Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press
- Gomat HY, Deleporte P, Moukini R, Mialounguila G, Ognouabi N, Saya AR, Vigneron P, Saint-André L (2011) What factors influence the stem taper of Eucalyptus: growth, environmental conditions, or genetics? *Ann For Sci* 68:109–120
- Hackenberg J, Wassenberg M, Spiecker H, Sun D (2015) Non destructive method for biomass prediction combining TLS derived tree volume and wood density. *Forests* 6:1274–1300
- Hall DB, Bailey RL (2001) Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *For Sci* 47:311–321
- Henry M, Picard N, Trotta C, Manlay RJ, Valentini R, Bernoux M, Saint-André L (2011) Estimating tree biomass of sub-Saharan African forests: a review of available allometric equations. *Silva Fenn* 45:477–569
- Henry M, Bombelli A, Trotta C, Alessandrini A, Birigazzi L, Sola G, Vieilledent G, Santenoise P, Longuetaud F, Valentini R, Picard N, Saint-André L (2013) GlobAllomeTree: international platform for tree allometric equations to support volume, biomass and carbon assessment. *IForest* 6:326–330
- Henry M, Rejou-Mechain M, Jara MC, Wayson C, Piotta D, Westfall J, Fuentes JM, Guier FA, Lombis HC, Lopez EC, Lara RC, Rojas KC, Pasquel JD, Montoya AD, Vega JF, Galo AJ, Lopez OR, Marklund LG, Milla F, Cahidez JDN, Malavassi EO, Perez J, Zea CR, Garcia L, Pons RR, Sanquetta C, Scott C, Zapata-Cuartas M, Saint-André L (2015) An overview of existing and promising technologies for national forest monitoring. *Ann For Sci* 72:779–788
- Hoffman MD, Gelman A (2011) The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *arXiv preprint arXiv:1111.4246*, 2011
- Huang S, Meng SX, Yang Y (2009) Estimating a nonlinear mixed volume-age model with and without taking into account serially-correlated errors: differences and implications. *Mod Appl Sci* 3:3–20

- Kozak A (1997) Effects of multicollinearity and autocorrelation on the variable-exponent taper functions. *Can J For Res* 27:619–629
- Kublin E (2003) Einheitliche Beschreibung der Schaftform—Methoden und Programme—BDATPro. *Forstw Cbl* 122(3):183–200. <https://doi.org/10.1046/j.1439-0337.2003.00183.x>
- Kublin E, Breidenbach J, Kändler G (2013) A flexible stem taper and volume prediction method based on mixed-effects B-spline regression. *Eur J For Res* 132:983–997
- Lappi J (1986) Mixed linear models for analyzing and predicting stem form variation of Scots pine. The Finnish Forest Research Institute
- Li R, Weiskittel AR (2010) Comparison of model forms for estimating stem taper and volume in the primary conifer species of the North American Acadian Region. *Ann For Sci* 67:1–16
- Lindstrom MJ, Bates DM (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics* 46:673–687
- Marin G, Bouriaud O, Dumitru M, Nitu D (2010) Development of Romania's National Forest Inventory. In: Tomppo E, Gschwanter T, Lawrence M (eds) *National Forest Inventories: pathways for common reporting*. Springer
- Max TA, Burkhart HE (1976) Segmented polynomial regression applied to taper equations. *For Sci* 22:283–289
- Meng SX, Huang S (2009) Improved calibration of nonlinear mixed-effects models demonstrated on a height growth function. *For Sci* 55:238–248
- Ni C, Nigh GD (2012) An analysis and comparison of predictors of random parameters demonstrated on planted loblolly pine diameter growth prediction. *Forestry* 85:271–280
- Neumann M, Moreno A, Mues V, Härkönen S, Mura M, Bouriaud O, Lang M, Achten WM, Thivolle-Cazat A, Bronisz K, Merganič J. 2016 Comparison of carbon estimation methods for European forests. *For Ecol Manage* 361(1):397–420
- Paulo JA, Tomé J, Tomé M (2011) Nonlinear fixed and random generalized height–diameter models for Portuguese cork oak stands. *Ann For Sci* 68:295–309
- Picard N, Saint-André L, Henry M (2012) Manuel de construction d'équations allométriques pour l'estimation du volume et la biomasse des arbres : de la mesure de terrain à la prédiction = Manual for building tree volume and biomass allometric equations: from field measurements to prediction = Manual de construcción de ecuaciones alométricas para estimar el volumen y la biomasa de los árboles: del trabajo de campo a la ecuaciones alométricas para estimar el volumen y la biomasa. Montpelier: CIRAD, FAO, 219 p. <http://www.fao.org/docrep/018/i3058f/i3058f.pdf>. Accessed May 2019
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R development core team (2016). nlme: linear and nonlinear mixed effects models. R package version 3.1–120
- Saint-André L, Bouillet J-P, Deleporte P, Mabiala A, Ognouabi N, Baillières H, Nouvellon Y, Moukini R (2002) Integrative modelling approach to assess the sustainability of the eucalyptus plantations in Congo. In Proc. 4th Workshop on 'Connection between Forest Resources and Wood Quality: Modelling Approaches and Simulation Softwares'. Harrison Hot Springs Resort, British Columbia, Canada, September 8–15, 2002, pp 611–621
- Stan Development Team (2016) RStan: the R interface to Stan, version 2.5. Available: <http://mc-stan.org/rstan.html>. Accessed Nov 2016
- Trincado G, Burkhart HE (2006) A generalized approach for modeling and localizing stem profile curves. *For Sci* 52:670–682
- Vonesh E, Chinchilli VM. 1997. *Linear and Non-Linear Models for the Analysis of Repeated Measurements* Marcel Decker. Inc, New York, NY.
- Wirth C, Schumacher J, Schulze E-D (2004) Generic biomass functions for Norway spruce in Central Europe—a meta-analysis approach toward prediction and uncertainty estimation. *Tree Physiol* 24: 121–139
- Yang Y, Huang S (2011) Comparison of different methods for fitting nonlinear mixed forest models and for making predictions. *Can J For Res* 41:1671–1686
- Zapata-Cuartas M, Sierra CA, Alleman L (2012) Probability distribution of allometric coefficients and Bayesian estimation of aboveground tree biomass. *For Ecol Manage* 277:173–179
- Zell J, Bösch B, Kändler G (2014) Estimating above-ground biomass of trees: comparing Bayesian calibration with regression technique. *Eur J For Res* 133:649–660
- Zianis D, Muukkonen P, Mäkipää R, Mencuccini M (2005) Biomass and stem volume equations for tree species in Europe. *Silv Fenn Monographs* 4:63

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.