Combining a predicted diameter distribution with an estimate based on a small sample of diameters

Lauri Mehtätalo, Carlos Comas, Timo Pukkala, and Marc Palahí

Abstract: The diameter distribution of a forest stand is of great interest in many situations, including forest management planning and the related prediction of growth and yield. The estimation of the diameter distribution may be based on, for example, a measured sample of diameters or the application of previously estimated parameter prediction models (PPMs), which relate the parameters of an assumed distribution function to some stand characteristics. We propose combining these two information sources. The approach is adopted from the mixed-effects modelling theory. The PPMs are treated as mixed-effects models, the residuals being stand effects. These stand effects are predicted using a small sample of tree diameters with the best linear predictor. A study conducted with a Spanish pine data set showed that in a situation where the predictors of the PPM include errrors, the prediction can be improved even by using a sample plot of as few as five sample trees. Vice versa, a distribution based on a sample plot of 3–15 sample trees can be significantly improved by utilizing existing PPMs. An additional simulation study was conducted to further investigate how the violation of different underlying assumptions of the method affects the performance.

Résumé : La distribution des diamètres d'un peuplement forestier est d'un grand intérêt dans de nombreuses situations, y compris pour la planification de l'aménagement forestier et la prédiction de la croissance et du rendement. L'estimation de la distribution des diamètres peut être fondée, par exemple, sur la mesure d'un échantillon des diamètres ou sur l'application de modèles de prédiction de paramètres (MPP) estimés antérieurement et qui relient les paramètres d'une fonction de distribution donnée à certaines caractéristiques du peuplement. Nous proposons de combiner ces deux sources d'information en adoptant l'approche de la modélisation à effets mixtes. Les modèles MPP sont traités comme des modèles à effets mixtes, les résidus étant considérés comme les effets du peuplement. Ces effets du peuplement sont prédits à l'aide d'un petit échantillon de diamètres d'arbre et avec le meilleur prédicteur linéaire. Une étude basée sur un ensemble de données de pin espagnol a montré que dans une situation où les prédicteurs du modèle MPP comportent des erreurs, la prédiction peut être améliorée même en utilisant une placette d'échantillonnage contenant aussi peu que cinq arbres échantillons. Vice versa, une distribution basée sur une placette contenant de 3 à 15 arbres échantillons peut être significativement améliorée en utilisant les modèles MPP existants. Une étude de simulation additionnelle a été effectuée pour étudier plus en profondeur comment la violation des différentes hypothèses sous-jacentes de la méthode affecte les performances.

[Traduit par la Rédaction]

Introduction

In Spain, stand development is predicted using either individual-tree models (Palahí et al. 2004; Trasobares and Pukkala 2004; Trasobares et al. 2004) or, if they are not available for the main tree species, then stand-level models (Álvarez González et al. 1999; Diéguez-Aranda et al. 2005, 2006; Castedo Dorado et al. 2007). There are also models that predict the temporal change of diameter distribution (Del Rio 1998). Some models also require the diameter distribution as input data.

When stand models are used to simulate stand development, which is the prevailing practice, for instance, in Galicia (northwest of Spain), diameter distribution models are often used to convert stand-level information into tree-level information. Trees or diameter classes drawn from the predicted distribution are used to simulate thinning treatments and to calculate the removal of various timber assortments. In Catalonia (northeast of Spain), where individual-tree models are commonly used, simulation uses as input data either diameter classes or individual trees measured in inventory plots. Parameter prediction models have been developed to predict the diameter distribution from stand characteristics (Palahí et al. 2006) in situations where the inventory data consist only of stand-level variables. The predicted distribution may be calibrated so that it agrees with all of the measured stand variables (Kangas and Maltamo 2000; Mehtätalo 2004; Palahí et al. 2006). However, the prediction may still be too inaccurate, and improving the prediction by using aditional diameter measurements could resolve the problem.

The most common practice in forest inventory is to use small field plots in which the diameters of all trees larger

Received 27 August 2010. Accepted 8 January 2011. Published at www.nrcresearchpress.com/cjfr on 8 April 2011.

L. Mehtätalo and T. Pukkala. Department of Forest Sciences, University of Eastern Finland, P.O. Box 111, 80110 Joensuu, Finland.
C. Comas. Department of Mathematics, Universitat Jaume I, Campus Riu Sec, E-12071 Castellón, Spain.
M. Palahí. EFIMED-Mediterranean regional Office of the European Forest Institute, Castella 33, Barcelona, Spain.

Corresponding author: L. Mehtätalo (e-mail: lauri.mehtatalo@uef.fi).

doi:10.1139/X11-008

than a certain minimum size are individually measured. In stand inventory, several plots may be measured within the stand. The trees of the inventory plots are then pooled to get a single empirical distribution for the calculations. In the Spanish National Forest Inventory (NFI), which uses concentric plots, the sampled distribution is used as such to calculate results or make projections.

The relevant question is whether the measured inventory data are used in the best possible way. The variable-radius sample plots measured in a stand inventory or the NFI, which have a radius of 25 m for the largest trees, are small compared with the total stand area (often tens of hectares). Therefore, the measured size distributions contain a considerable amount of sampling error and may not be sufficiently accurate for estimating the size distribution of the stand. One possibility to improve this estimate would be to use the properties of the whole stand from which the sample was drawn. Such information would be provided, for example, by stand characteristics that have been assessed for the whole stand area, not only for the sample plot area.

The available data determine how the diameter distribution is estimated for computations. If the sample of the trees measured for diameter in a stand is sufficiently large, then the measured sample can be used as such. If, on the other hand, the measured sample is relatively small, then the sample distribution can be smoothed (Droessler and Burk 1989; Mehtätalo 2004) or a theoretical distribution can be fitted to it (Bailey and Dell 1973). The smoothing can be done, for example, using a kernel smoothing method (Härdle 1990).

A very common situation, however, is that no diameter sample has been measured in the field. Instead, some stand characteristics, such as stand age, site fertility, basal area, and mean diameter, have been assessed. In this case, the alternatives for diameter distribution prediction are the parameter prediction model (PPM) method and the parameter recovery model (PRM) method. In PPM, the prediction utilizes an estimated regression relationship between the measured stand characteristics and the parameters of an assumed diameter distribution function. This regression is estimated from data including measured diameter distributions from a representative sample of stands. The PRM method is a very similar approach, but the models are constructed for moments or percentiles instead (Hyink and Moser 1983).

Many methods, including PPM, PRM, and maximum likelihood (ML) fit, need an assumption about a theoretical distribution. Of the several utilized distribution functions, the Weibull function is probably the most commonly used one. It is the only widely used two-parameter alternative, even though it also has been used in the three-parameter form for the prediction of diameter distributions. Other convenient and widely used alternatives are beta, Johnson's S_B, and the logit–logistic distribution (Wang and Rennolls 2005), all of which have four parameters. The common practice in forest inventory is to measure only trees larger than a fixed minimum diameter. With such data, a truncated version of an assumed distribution should be used (Palahí et al. 2007).

The PPM method for predicting diameter distributions is based on the estimation of general rules about the shape of the diameter distribution in stands of different characteristics. On the other hand, the approaches based on a sample are based on the sample information on the target stand only. Because the information sources of these two methods are different, it might be expected that combining these two approaches would lead to predictions that are more accurate than either of the two methods alone. Such a method could be used either in imroving predictions based on PPM method by using a small sample of diameters or, alternatively, to improve the estimates based on a sample of diameters by using measured stand characteristics from the whole stand area.

The aim of this study is to estimate the underlying diameter distribution that generated the tree diameters by utilizing different kinds of information. The methods evaluated here can be classified as follows: (*i*) those that are based on a measured diameter sample alone, (*ii*) those based on measured stand characteristics alone, and (*iii*) those combining both a sample of diameters and measured stand characteristics. The main interest is to utilize a small sample of diameters in a situation where stand characteristics have also been measured. In particular, we develop approaches for combining the PPM method with a sample of diameters and compare their accuracy with that of other existing methods that could be used with such data.

Material

Modelling data

The modelling data consisted of the field plots of the Second NFI in Catalonia. The plots were placed at 1 km intervals in the x and y directions. Therefore, the plots cover all stand types and conditions, reflecting the variation in stand characteristics in Catalonia. For this study, the plots with at least 20 Scots pine (*Pinus sylvestris* L.), Austrian pine (*Pinus nigra* Arnold), or allepo pine (*Pinus halepensis* P. Mill.) trees were selected. For more information on the data, see Palahí et al. (2006).

The NFI uses circular plots of radius 25 m. However, only trees with a diameter above 42.5 cm are measured on the whole plot area. For smaller trees, the sample plot radius depends on the tree diameter so that trees with a diameter at breast height of between 22.5 and 42.5 cm are sampled using a radius of 15 m, trees with a diameter at breast height of between 12.5 and 22.5 cm using a radius of 10 m, and trees with a diameter at breast height of between 7.5 and 12.5 cm using a radius of 5 m. Trees with a diameter below 7.5 cm are not measured at all. This design implies that the observations of a given plot constitute a weighted sample from a distribution that is left-truncated at diameter t = 7.5. The weights depend on tree diameter as follows:

$$w(d) = \begin{cases} 5^2/25^2 & 7.5 < d \leq 12.5 \\ 10^2/25^2 & 12.5 < d \leq 22.5 \\ 15^2/25^2 & 22.5 < d \leq 42.5 \\ 1 & d \geq 42.5 \end{cases}$$

assuming that the underlying spatial distribution of trees is homogeneous.

Assuming that $F_t(d|\boldsymbol{\theta})$ is the underlying diameter distribution of those trees of a stand that have a diameter at breast height above 7.5 cm (e.g., the left-truncated Weibull distribution, see below) and $f_t(d|\boldsymbol{\theta})$ is the corresponding density, the sampling design yields the weighted density

Table 1. Some characteristics of the modelling and test data sets.

	Modelling data (n = 712)				Test data (n = 26)			
	Minimum	Mean	Maximum	SD	Minimum	Mean	Maximum	SD
No. of trees	21.0	30.8	77.0	9.28	34	81	126.0	26.7
Alt, 100 m	0	9.34	19	4.19	3.94	10.0	16.35	4.61
DGM (cm)	9.65	20.8	53.9	6.4	10.8	23.3	39.7	7.3
N (1/ha)	170	1110	4230	563	331	1093	2037	486
G (m^2/ha)	13.8	24.8	59.4	8.1	9.3	31.3	73.2	16.1
$\widehat{\alpha_{\mathrm{ML}}}$	0.54	3.0	8.6	1.2	0.5	2.7	4.2	1.15
$\widehat{\beta_{\mathrm{ML}}}$	1.0	17.6	49.5	6.6	1	17.5	37.1	9.4

Note: Alt, altitude above sea level; DGM, median of the diameter distribution weighted by basal area; N, number of stems; G, basal area. $\widehat{\alpha_{\mathrm{ML}}}$ and $\widehat{\beta_{\mathrm{ML}}}$ are the ML estimates for the shape and scale parameters of the truncated Weibull distribution.

$$g(d) = \frac{w(d)f_t(d|\boldsymbol{\theta})}{\int_t^{\infty} w(u)f_t(u|\boldsymbol{\theta})\mathrm{d}u}$$

for the sample. Constant t is the treshold diameter of the truncation and vector $\boldsymbol{\theta}$ includes the parameters of the underlying diameter distribution of the stand. Under the applied sampling design, the density of the observed data becomes

$$[1] \qquad g(d|\boldsymbol{\theta}) = \begin{cases} \dfrac{5^2}{25^2 I} f_t(d|\boldsymbol{\theta}) & 7.5 < d \leq 12.5 \\ \dfrac{10^2}{25^2 I} f_t(d|\boldsymbol{\theta}) & 12.5 < d \leq 22.5 \\ \dfrac{15^2}{25^2 I} f_t(d|\boldsymbol{\theta}) & 22.5 < d \leq 42.5 \\ \dfrac{1}{I} f_t(d|\boldsymbol{\theta}) & d \geq 42.5 \end{cases}$$

where

$$\begin{aligned} I &= \int_t^{\infty} w(u)f_t(u|\boldsymbol{\theta})\mathrm{d}u \\ &= 1 - F_t(42.5|\boldsymbol{\theta}) + \frac{15^2}{25^2}[F_t(42.5|\boldsymbol{\theta}) - F_t(22.5|\boldsymbol{\theta})] \\ &\quad + \frac{10^2}{25^2}[F_t(22.5|\boldsymbol{\theta}) - F_t(12.5|\boldsymbol{\theta})] \\ &\quad + \frac{5^2}{25^2}[F_t(12.5|\boldsymbol{\theta}) - F_t(7.5|\boldsymbol{\theta})] \end{aligned}$$

We assumed the underlying diameter distribution of the trees with diameter above 7.5 cm on plot k to be of the form of the truncated Weibull distribution, which has the cumulative distribution function

$$F_t(d|\alpha_k, \beta_k) = 1 - \exp\left[\left(\frac{t}{\beta_k}\right)^{\alpha_k} - \left(\frac{d}{\beta_k}\right)^{\alpha_k}\right]$$

and density

$$[2] \qquad f_t(d|\alpha_k, \beta_k) = -\frac{\alpha_k}{\beta_k}\left(\frac{d}{\beta_k}\right)^{\alpha_k - 1} \exp\left[\left(\frac{t}{\beta_k}\right)^{\alpha_k} - \left(\frac{d}{\beta_k}\right)^{\alpha_k}\right] \qquad (t \leq d \leq \infty,\ t, \alpha_k, \beta_k > 0)$$

where d is tree diameter and α_k and β_k are the parameters that specify the distribution for plot k. In the recent study of Palahí et al. (2007), the truncated Weibull function was found to be the best alternative from among several candidates in the data set of this study. In particular, the truncated distribution is favoured over the well-known three-parameter version of the Weibull distribution because a fixed measurement limit for the diameter implies censored data that are realistically modelled with a left-truncated distribution.

The truncated Weibull distribution was fitted for each plot by maximizing the log likelihood $\ell(\boldsymbol{\theta}) = \sum \ln g(d|\boldsymbol{\theta})$, where g is as specified in eq. 1, f_t is as given in eq. 2, and $\boldsymbol{\theta}_k = (\alpha_k, \beta_k)$, to get the ML estimates $\widehat{\boldsymbol{\theta}_k}$ for plot k (see "Using the diameter sample for estimation" section below). The estimation was carried out using the R function mle in the package stats4 (R Development Core Team 2010). The summary of the modelling data and the obtained estimates are given in Table 1.

Assuming that the forest stands are homogeneous, the estimated parameters for the plot are unbiased (we believe that asymptotic unbiasedness of ML-estimates holds with the applied sample size $n > 20$) also for the stand. However, they include a small amount of estimation errors due to the use of a sample in estimation. Nevertheless, they provide the best available estimates for the Weibull parameters for the stand and are therefore assumed to represent the diameter distribution of the whole stand.

Test data

The test data were originally collected to test growth and other models developed using the NFI data. The plots were subjectively placed in *P. sylvestris*, *P. nigra*, and *P. halepensis* forests. The aim was to cover the whole range of variation in stand age, density, and structure. The possible stand conditions were classified into a few classes of stand age (tree size), stand density, and stand structure and a few sample plots were measured from every class.

The test plots had a fixed radius, which varies between 10 and 25 m according to the stand density (the mean radius was 16.2 m). The plot size was selected so as to have approximately 100 trees in each plot. Since the utilized NFI plots contained, on average, only 30 trees, it can be said that the test plots are clearly "larger" than the NFI plots. In particular, we make an assumption that the plots are so large that the trees can be used as such to represent the population of the trees in the stand.

The truncated Weibull distribution was fitted to the plots of the test data set using the ML (see "Using the diameter sample for estimation" section). This estimate was later taken as the true underlying diameter distribution of the stand. The interesting variables of the test data set are summarized in Table 1.

Methods

Tree diameter of tree i in stand k, D_{ki}, is assumed to be a random variable with a specified continuous distribution. We assume that the diameter of this particular tree i was generated by an underlying stand-specific model. All trees of the stand are assumed to have been generated by the same model, i.e., the diameters are identically distributed. Thus, even if we were to measure all of the trees of the stand, we would only observe a finite number of realizations from this continuous model. To simplify notations, we drop the index i from D_{ki} and use D_k instead.

The next subsections present the prediction methods of this study, which are classified as follows: (*i*) those that are based on a measured diameter sample alone (empirical, kernel, and ML fit), (*ii*) those based on measured stand characteristics alone (PPM), and (*iii*) those combining both a sample of diameters and measured stand characteristics (best linear predictor (BLP) and simple mean approach (MEAN)).

Using the diameter sample for estimation

ML fit

Assuming the diameters d_{ki}, $i = 1, ..., n_k$, being observed from stand k using equal sampling probability (cf. eq. 1), the method of ML searches such an estimate for the parameter vector $\boldsymbol{\theta}_k = (\alpha_k, \beta_k)'$ that maximizes the log-likelihood

$$\ell(\boldsymbol{\theta}_k) = \sum_{i=1}^{n_k} \ln f(d_{ki}|\boldsymbol{\theta}_k)$$

With a large sample, the estimates $\widehat{\boldsymbol{\theta}}_k$ have the variance–covariance matrix

$$\mathrm{var}(\widehat{\boldsymbol{\theta}}_k) \approx [-\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_k)]^{-1}$$

where $\boldsymbol{H}$ is the Hessian matrix, which includes the second derivatives of the log-likelihood with respect to the elements of $\boldsymbol{\theta}$ (Casella and Berger2002). The ML estimate based on sample s from stand k is noted later on by $\widehat{\boldsymbol{\theta}_{ks,\mathrm{ML}}}$.

If the sample size n is sufficiently large, the obtained estimates are unbiased and have a bivariate normal distribution with the variance–covariance matrix given above. The bivariate normality implies that all of the marginal distributions of the ML estimates $\widehat{\alpha_{k,\mathrm{ML}}}$ and $\widehat{\beta_{k,\mathrm{ML}}}$ are also normal and the correlation between them is linear. However, what constitutes "sufficiently large" depends on the situation. In this study, we will use very small samples, and thus the asymptotic properties do not necessarily hold.

Other methods

The simplest way to utilize a sample distribution is to use it as such to present the diameter distribution of the stand. However, especially with a small sample size, this may not be a good strategy due to the large sampling errors associated with small samples. Thus, methods can be used that smooth the sampled distribution. The ML method of the previous subsection can be seen as one smoothing method. Another, more flexible method is kernel smoothing (Droessler and Burk 1989; Härdle 1990; Uuttera and Maltamo 1995; Mehtätalo 2004).

The idea of kernel smoothing is to plot each observed diameter on the x-axis, replace these points with a density function called "kernel" (e.g., Gaussian density), sum up these densities, and finally rescale to unity by dividing them by the sample size (Härdle 1990):

$$f_h(d) = \frac{1}{n}\sum_{i=1}^{n} k_h(d - d_{ki})$$

The cumulative distribution function corresponding to the kernel density results from replacing the densities with the corresponding cumulative distribution function

$$F_h(d) = \frac{1}{n}\sum_{i=1}^{n} K_h(d - d_{ki})$$

where k_h and K_h are the density and distribution functions of the utilized kernel with the smoothing parameter h and d_{ki}, $i = 1, ..., n$, are the sampled tree diameters from plot k. The resulting estimates of the density and distribution function are based on the real data and are smoother than the histogram of the sample but more flexible than a fitted parametric density function (e.g., the Weibull function), although far more local.

The utilized kernel function is usually a symmetric distribution function with a mean of zero, such as the Gaussian kernel, which was used here. The degree of smoothing is determined by the smoothing parameter, e.g., the standard deviation of the Gaussian kernel, which is set subjectively. In selecting the bandwidth, we utilized Silverman's rule of thumb (Silverman 1986, p. 48) $h_{\mathrm{s}} = 0.9\min(\widehat{\sigma}, \widehat{R}/1.34)n^{1/5}$, where $\widehat{\sigma}$ is the estimated standard deviation and $\widehat{R}$ is the interquartile range of the sample. We also evaluated the kernel method using the values $0.5h_{\mathrm{s}}$ and $2h_{\mathrm{s}}$ for the bandwidth.

Because of the applied diameter limit of 7.5 cm, the kernel-smoothed distribution wasalso truncated at the minimum diameter t. Thus, the final truncated kernel cumulative distribution function of the data was computed by taking only the part above the limit t and rescaling it as

$$F_{ht}(d) = \frac{1}{1 - F_h(t)}[F_h(d) - F_h(t)]$$

The truncated kernel densities were computed correspondingly.

Using the stand characteristics for prediction

In the PPM method, the the parameter specifying the truncated Weibull diameter distribution $\boldsymbol{\theta}_k = (\alpha_k, \beta_k)'$ was assumed to be a random vector that takes a unique value for each stand. Furthermore, it is assumed that the value for stand k depends on stand characteristics $\boldsymbol{x}_k$ according to the multivariate model

$$[3] \qquad \boldsymbol{\theta}_k = E(\boldsymbol{\theta}|\boldsymbol{x}_k) + \boldsymbol{e}_k$$

where $\boldsymbol{x}_i$ includes the values of the predictors for stand k (e.g., basal area G, basal area median diameter DGM, and stand age T) and $\boldsymbol{e}_k$ includes the stand level residuals for the parameters of stand k. This model assumes that there is an underlying relationship that expresses the dependence of the expected value of the parameters α and β on some stand characteristics $\boldsymbol{x}_k$. The residuals of the PPM, $\boldsymbol{e}_k$, express how much the realized values of the stand-specific estimates $\widehat{\boldsymbol{\theta}_k}$ deviate from this conditional expectation. They can also be interpreted as stand effects of a random coefficient model. The variance–covariance matrix of stand effects, var($\boldsymbol{e}_k$), specifies how much the stand effects vary around the conditional expectations. The notation $E(\boldsymbol{\theta}|\boldsymbol{x}_k)$ emphasizes that we assume the fixed part of the model to be true, i.e., that it satisfactorily expresses the conditional expectation of the parameters for the given value of vector $\boldsymbol{x}_k$.

In practice, we only have the ML estimates of the Weibull parameters available. That is, the response of the model is $\widehat{\boldsymbol{\theta}_k}$, not $\boldsymbol{\theta}_k$. This does not cause any bias problems to the estimation of the fixed part of the model because the ML estimates are unbiased. However, it means that the residual variances from these models overestimate the true between-stand variation. This overestimation depends on the amount of estimation errors in the ML estimates. Formally, this effect can be seen by rewriting model eq. 3 in terms of the ML estimates $\widehat{\boldsymbol{\theta}_k}$ and the estimation errors ϵ_k as

$$\widehat{\boldsymbol{\theta}_k} + \epsilon_k = E(\boldsymbol{\theta}|\boldsymbol{x}_k) + \epsilon_k$$

Moving the estimation errors to the right-hand side gives

$$[4] \qquad \widehat{\boldsymbol{\theta}_k} = E(\boldsymbol{\theta}|\boldsymbol{x}_k) + \boldsymbol{e}_k - \epsilon_k = E(\boldsymbol{\theta}|\boldsymbol{x}_k) + \boldsymbol{e}_k^*$$

This shows that when the PPM is fitted to the ML estimates of the parameters, the residual includes two components: the true stand effect and the estimation error of the response (i.e., that of the the ML estimate). If these two components are independent, then the residual variance is just the sum of the corresponding variances (because var(X – Y) = var(X) + var (Y) – 2 cov[X,Y]). This needs to be taken into account if the residual variances are used in prediction but can be ignored when only the fixed part of the PPM is used.

The modelling data of 712 NFI plots were used for fitting prediction models for the ML estimates of the shape and scale parameters of the truncated Weibull distribution. With the truncated Weibull distribution, we came to the following models for shape and scale parameters, respectively:

$$[5] \qquad \widehat{\ln(\alpha_k)} = E[\ln(\alpha)|\boldsymbol{x}_k] + e_{\alpha,k}$$

$$[6] \qquad \widehat{\beta_k} = E(\beta|\boldsymbol{x}_k) + e_{\beta,k}$$

where the residuals may be correlated, i.e., cov($e_{\alpha,k}$, $e_{\beta,k}$) $\neq 0$. These transformations were selected to obtain symmetric distributions for the residuals.

Two different models were fitted: one using altitude, G, the number of stems (N), and basal area DGM as predictors (PPM1) and one that uses all of these except N as predictors (PPM2). The performance of these two models is very different because the three stock characteristics together (G, N, and DGM) include such information on the shape that cannot be obtained using only two of them (Siipilehto 1999). That is, if we use only stand density and median diameter, the median diameter would specify the location of the distribution and the stand density would scale the distribution to the measured density. This would be possible with a distribution of any shape. However, if we additionally use G, it would restrict also the shape of the distribution.

The models were first fitted separately to find the best model form. The final models were then fitted using the seemingly unrelated regression, which slightly improves the efficiency of the estimates. However, the main motivation for using the seemingly unrelated regression was the resulting estimate of the covariance between error terms, cov(e_α, e_β), which would be needed later (see "The BLP approach" section below).

The PPM prediction for a new stand k with known x_k is computed using the estimated expression for the conditional expectation

$$[7] \qquad \widehat{\boldsymbol{\theta}_{k,\mathrm{PPM}}} = E(\widehat{\boldsymbol{\theta}|\boldsymbol{x}_k})$$

Because we do not have any information on the stand effects ($e_{\alpha,k}$ and $e_{\beta,k}$), they are given their expected value, i.e., 0.

Combining PPM predictions and ML fits

The BLP approach

In addition to $\boldsymbol{x}_k$, we may also have other information on the diameter distribution of stand k, such as a small sample of the tree diameters. Here, we propose to combine these two information sources. We adopt from the mixed-effects models (e.g., Pinheiro and Bates 2000; McCulloch and Searle 2001) the idea of predicting the random effects of a given class using observed values of the response in that class. This approach has been found to be very useful in many forest inventory applications, including taper curves, site index curves and height–diameter curves (Lappi 1986, 1997; Lappi and Bailey 1988). In our application, PPMs take the place of the mixed-effects model and the ML estimates based on a sample of diameters provide the observations of the response of these models. The estimation error, which is estimated separately for each ML fit as the asymptotic variance–covariance matrix of these estimates, specifies the accuracy of these observations. This provides a difference to the BLP used in the mixed-effects model.

To develop the BLP approach, assume that vector $\boldsymbol{\theta}_k$ can be unbiasedly estimated with a known accuracy, var(ϵ_{ks}). Such an estimate can be obtained by fitting the assumed distribution function to a random sample of diameters from the stand. The large-sample properties of the ML estimator support these assumptions. However, with very small sample size, the asymptotic properties of the ML estimator may not be met and the assumptions on the observations may not be valid.

The estimate based on sample s from stand k consists of three components: the expectation conditional on stand characteristics (i.e., the PPM prediction), the stand effect related to this particular stand, and the estimation error related to this particular sample. Thus, it can be written as

$$\widehat{\boldsymbol{\theta}_{ks,\mathrm{ML}}} = \boldsymbol{\theta}_k + \epsilon_{ks} = E(\boldsymbol{\theta}|\boldsymbol{x}_k) + \boldsymbol{e}_k + \epsilon_{ks}$$

It is realistic to assume that the estimation error is not correlated with the stand effect. Then, the covariance between the estimate and the unobserved stand effect is

$$\mathrm{cov}(\boldsymbol{e}_k + \epsilon_{ks}, \boldsymbol{e}_k) = \mathrm{var}(\boldsymbol{e}_k)$$

The first- and second-order properties of vectors $\boldsymbol{\theta}_k$ and $\widehat{\boldsymbol{\theta}_{ks,\mathrm{ML}}}$ can be summarized as

$$\begin{bmatrix}\boldsymbol{\theta}_k \\ \widehat{\boldsymbol{\theta}_{ks,\mathrm{ML}}}\end{bmatrix} \sim \left(\begin{bmatrix}E(\boldsymbol{\theta}|\boldsymbol{x}_k) \\ E(\boldsymbol{\theta}|\boldsymbol{x}_k)\end{bmatrix}, \begin{bmatrix}\boldsymbol{V}_1 & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}' & \boldsymbol{V}_2\end{bmatrix}\right)$$

where $\boldsymbol{V}_1 = \mathrm{var}(\boldsymbol{e}_k)$, $\boldsymbol{V}_{12} = \mathrm{var}(\boldsymbol{e}_k)$, and $\boldsymbol{V}_2 = \mathrm{var}(\boldsymbol{e}_k) + \mathrm{var}(\epsilon_{ks})$.

The BLP of $\boldsymbol{\theta}_k$ is (McCulloch and Searle 2001; Lappi et al. 2006)

$$\mathrm{BLP}(\boldsymbol{\theta}_k) = E(\boldsymbol{\theta}_k|\boldsymbol{x}_k) + \boldsymbol{V}_{12}\boldsymbol{V}_2^{-1}[\widehat{\boldsymbol{\theta}_{ks,\mathrm{ML}}} - E(\boldsymbol{\theta}_k|\boldsymbol{x}_k)]$$

This prediction combines the information of the sample data with that of the PPM in a way that is optimal, assuming that the the model is correct, the assumptions on the measurements are met, and the utilized variances are known (Robinson 1991). Optimality means that from among all linear unbiased predictors, the BLP has the smallest variance. If the estimation errors and stand effects are normally distributed, then the BLP is also the best predictor, i.e., it is unbiased and has the lowest variance from among all possible predictors. In computations, the matrices are replaced with their estimates and $E(\boldsymbol{\theta}|\boldsymbol{x}_k)$ is replaced with its estimate $\widehat{\boldsymbol{\theta}_{k,\mathrm{PPM}}}$. The resulting estimator is noted as $\widehat{\boldsymbol{\theta}_{k,\mathrm{BLP}}}$.

Fitting of the PPMs yielded only an estimate for $\mathrm{var}(\boldsymbol{e}_k^*)$ (eq. 4), not for $\mathrm{var}(\boldsymbol{e}_k)$ (eq. 4). Noting that the stand effects and estimation errors are independent, we estimated $\mathrm{var}(\boldsymbol{e}_k)$ by subtracting the average estimation error from the estimated residual variance–covariance matrix as

[8] $$\mathrm{var}(\boldsymbol{e}_k) = \mathrm{var}(\boldsymbol{e}_k^*) - \mathrm{var}(\epsilon_k)$$

The resulting estimate of $\mathrm{var}(\boldsymbol{e}_k)$ was used to predict the stand effects using the above-described BLP approach. The average estimation error in eq. 8 was based on the approximate asymptotic variance–covariance matrices of the ML fits to the modeling data. We first took the medians of the estimation error variances over the plots of the modelling data, which were used as the variances in the estimated matrix $\mathrm{var}(\epsilon_k)$. To ensure a positive definite matrix, the median correlation over the plot-specific estimated correlations was used to estimate the required covariance. Median variances and correlations were used instead of means to decrease the effect of some very high estimates.

We assumed here that the coefficients of the PPMs are fixed, even though they are actually random. This strategy was taken to avoid extra complexity of the methodology after noticing that the effect of this randomness on the prediction error variance would have been only a few percentage points.

Measurement errors of predictors in the BLP approach

In most practical cases, the measurements of stand characteristics include errors (e.g., Kangas et al. 2004). When these erroneous stand characteristics are used as predictors in the PPMs, they lead to lower than expected accuracy. In this subsection, we present an approach to take these errors into account.

Assume that the PPM (eq. 3) is of the linear form

$$\boldsymbol{\theta}_k = \boldsymbol{X}_k\boldsymbol{B} + \boldsymbol{e}_k$$

where $\boldsymbol{X}(n \times p)$ is the assumed model matrix, $\boldsymbol{B}(p \times q)$ is the matrix of regression coefficients, n is the number of observations, q is the length of $\boldsymbol{\theta}$ (here, $q = 2$), and p is the total number of coefficients in the q regression equations. The same model is used for prediction when the measurements of $\boldsymbol{X}$ include errors. In this case, the true, unknown parameter can be written as

$$\boldsymbol{\theta}_k = \widetilde{\boldsymbol{X}}_k\boldsymbol{B} + \widetilde{\boldsymbol{e}}_k$$

where the design matrix $\widetilde{\boldsymbol{X}}_k$ includes the predictors with measurement errors and $\widetilde{\boldsymbol{e}}_k$ is stand effect of this model. Setting the two models for $\boldsymbol{\theta}_k$ equal and rearranging terms gives an expression for the random stand effect of the latter model:

$$\widetilde{\boldsymbol{e}}_k = (\boldsymbol{X}_k - \widetilde{\boldsymbol{X}}_k)\boldsymbol{B} + \boldsymbol{e}_k$$

If the measurements of $\boldsymbol{X}$ are unbiased, the stand effects have a mean of 0. The variance–covariance matrix of the stand effect can be written using the known variance–covariance matrices of the measurement error and the stand effect of the original PPMs as

[9] $$\mathrm{var}(\widetilde{\boldsymbol{e}}_k) = \boldsymbol{B}'\mathrm{var}(\widetilde{\boldsymbol{X}}_k - \boldsymbol{X}_k)\boldsymbol{B} + \mathrm{var}(\boldsymbol{e}_k)$$

Thus, the measurement errors increase the between-stand variation. In the BLP equation, this can be taken into account by replacing $\mathrm{var}(\boldsymbol{e}_k)$ in the definition of matrices $\boldsymbol{V}_1$, $\boldsymbol{V}_{12}$, and $\boldsymbol{V}_2$ by $\mathrm{var}(\widetilde{\boldsymbol{e}}_k) = \boldsymbol{B}'\mathrm{var}(\widetilde{\boldsymbol{X}}_k - \boldsymbol{X}_k)\boldsymbol{B} + \mathrm{var}(\boldsymbol{e}_k)$.

A simple mean approach

As a simple ad hoc alternative to the BLP, we also consider an estimator that is obtained simply by averaging the ML estimators and the PPM prediction as

$$\widehat{\boldsymbol{\theta}_{k,\mathrm{MEAN}}} = \frac{1}{2}(\widehat{\boldsymbol{\theta}_{k,\mathrm{PPM}}} + \widehat{\boldsymbol{\theta}_{k,\mathrm{ML}}})$$

Green and Clutter (2000) combined alternative estimators of diameter class proportions by taking a mean of the estimates inversely proportional to variances. However, the covariances between the class-specific estimates were not taken into account. Our simple mean approach does not take into account the variances and covariances. On the other hand, the BLP of the previous subsection takes them both into account.

Evaluating the alternative methods with empirical data

We predicted the diameter distribution for the 26 evaluation plots using each of the methods presented earlier. To evaluate the effect of sample size on the accuracy of prediction, we used samples of size $n = 3, 4, ..., 20$ from the plots and used them for prediction. In the basic simulation (SIM1), n trees closest to the plot centre were taken as the sample.

The simulations were conducted separately with the two estimated models PPM1 and PPM2. The known true values for the predictors of the models were used in prediction. In addition, a third simulation was conducted, PPM2E, where normally distributed, independent, zero-mean measurement errors were added to the predictors DGM and G of the models PPM2. The variances of the measurement errors were

Table 2. Coefficients of models PPM1 (see Table 1 for abbreviations).

	Estimate	SE
Model for ln(α), R^2 = 0.28, residual SE = 0.36		
Intercept	2.137	0.213
100 × N × DGM/G	–0.126	0.008
Alt	–0.0824	0.0122
ln(Alt + 1)	0.485	0.104
ln(G)	–0.117	0.0493
Model for ln(β), R^2 = 0.85, residual SE = 2.58		
Intercept	81.146	4.275
ln(G)	10.972	0.561
ln(N)	–11.357	0.493
DGM	0.746	0.0411
ln(DGM)	–12.639	1.052
Alt	–0.640	0.0876
ln(Alt + 1)	3.716	0.746

Table 3. Coefficients of models PPM2 (see Table 1 for abbreviations).

	Estimate	SE
Model for ln(α), R^2 = 0.08, residual SE = 0.41		
Intercept	–1.745	0.572
DGM	–0.0330	0.0120
ln(DGM)	1.171	0.264
Alt	–0.0521	0.0144
ln(Alt + 1)	0.304	0.121
ln(G)	–0.0830	0.0252
Model for ln(β), R^2 = 0.76, residual SE = 3.26		
Intercept	–16.666	4.575
DGM	0.691	0.0963
ln(DGM)	6.235	2.122
Alt	–0.389	0.116
ln(Alt + 1)	2.193	0.976

10% of the true value for G and 7% of the true value for DGM. Compared with the errors reported in Finnish studies (e.g., Kangas et al. 2004), these error rates are rather low. Because our model included G and DGM as predictors both in the natural and logarithmic scales, computing the matrix $\mathrm{var}(\widetilde{\boldsymbol{X}}_k - \boldsymbol{X}_k)$ in eq. 9 required some additional approximations. The variance of ln(X) was approximated with ln(var (X) + X^2) – 2 ln X (see Mehtätalo and Kangas 2005, eq. A2). In addition, as $\ln X$ is rather linear for a narrow range of X, it was assumed that $\mathrm{corr}(X, \ln X) = 1$.

The different methods for predicting and smoothing the diameter distribution were compared using the so-called error index (Reynolds et al. 1988), which measures the goodness of fit. The index for stand k is the sum of absolute differences in true and predicted diameter class frequencies:

$$ei1_k = \sum \| f - \widehat{f} \|$$

where f and $\widehat{f}$ are the true and predicted diameter class frequencies. The summation is taken over all diameter classes with nonzero predicted or true frequency. We utilized the class width of 2.5 cm, starting from the truncation limit of 7.5 cm. The value of the index ranges from 0 (identical distributions) to 2 (nonoverlapping distributions). To make comparisons between methods, we averaged the index over the 26 evaluation plots.

Results

Estimated PPM models

The parameter estimates for the models PPM1 and PPM2 are reported in Tables 2 and 3. The estimated variance–covariance matrices of stand effects for PPM1 were

$$\widehat{\mathrm{var}(\boldsymbol{e}_k)} = \mathrm{var}\begin{pmatrix} e_{\alpha,k} \\ e_{\beta,k} \end{pmatrix} = \begin{pmatrix} 0.0792 & 0.451 \\ 0.451 & 3.183 \end{pmatrix}$$

and for PPM2 were

$$\widehat{\mathrm{var}(\boldsymbol{e}_k)} = \mathrm{var}\begin{pmatrix} e_{\alpha,k} \\ e_{\beta,k} \end{pmatrix} = \begin{pmatrix} 0.115 & 0.837 \\ 0.837 & 7.195 \end{pmatrix}$$

These estimates were obtained by subtracting the estimated variance–covariance matrix of estimation errors

$$[10] \qquad \widehat{\mathrm{var}(\epsilon_k)} = \begin{pmatrix} 0.0489 & 0.345 \\ 0.345 & 2.450 \end{pmatrix}$$

from the residual variance covariance matrix of the PPMs. The between-model correlations of stand effects computed from these matrices are 0.899 for PPM1 and 0.919 for PPM2. As expected, dropping N from among the predictors increased the estimated variances of stand effects considerably: in the model for ln(α) from 0.28^2 to 0.34^2 and in the model for β from 1.78^2 to 2.68^2.

Figure 1 shows the residual plots of PPM1 on the predicted values as well as the normal QQ plots. For the log shape model, the residual variance has a slight decreasing trend on the predicted values. The normality is met quite well, even though there are slightly heavy tails. For the scale model, the residual variance appears rather constant. However, the distribution is left skewed and quite different from the normal distribution. The plots for model PPM2 were quite similar, but the decreasing residual variance for the model of ln(α) was not as pronounced as with PPM1.

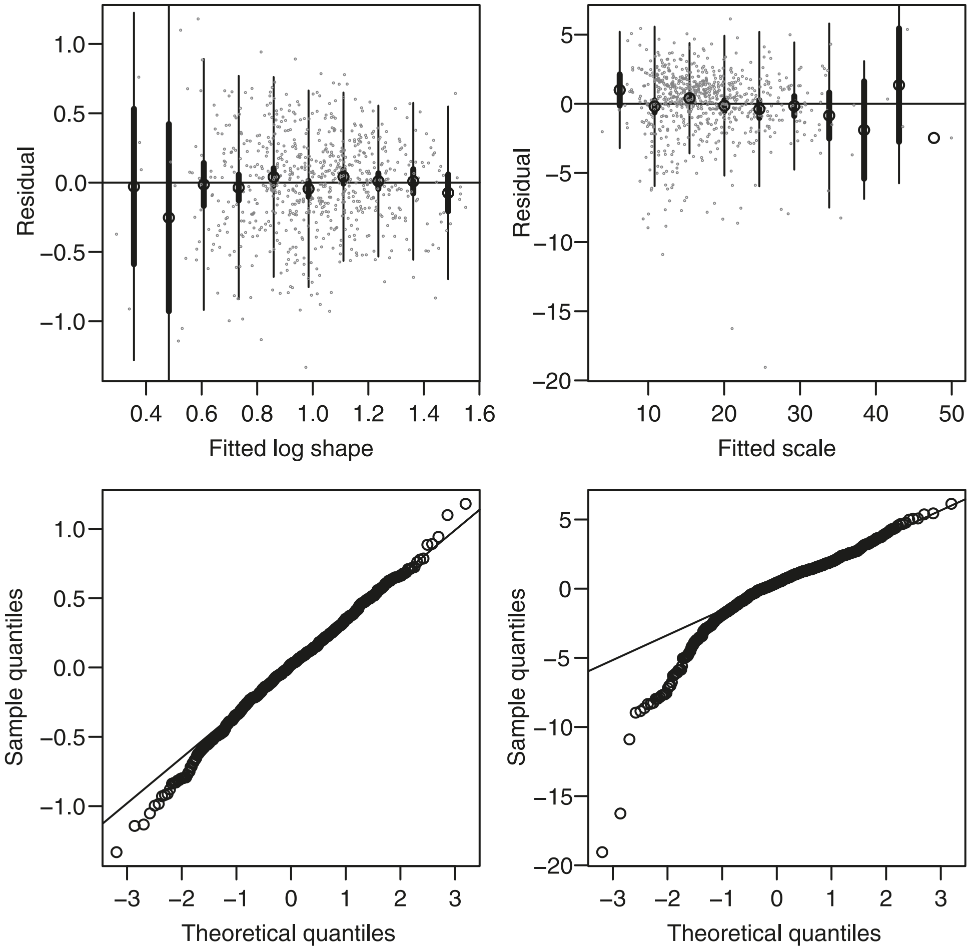
Evaluation of the prediction methods

Figure 2 shows the error index of different methods as a function of sample size. The lowest lines in the plots show the error index of the ML fit to the complete data, which was regarded as the true underlying distribution of the stand (TRUE). This is the share of the index that is related to the sampling errors and possible poor fit of the truncated Weibull distribution.

The methods that use measured stand characteristics only are PPM1, PPM2, and PPM2E. The mean error index for PPM1, with N as an aditional predictor, is 0.37 (Fig. 2, top panel). This value is remarkably lower than the mean error index for the model PPM2 (0.44, middle panel), which did not have N as a predictor. The value further increases to 0.46 when the predictors of model PPM2 include measurement errors (Fig. 2, bottom panel).

Of the methods that use the sample alone (empirical, ML fit, and kernel), the kernel method with a bandwith of $2h_{\mathrm{s}}$ performs the best for sample sizes below 14. For higher sam-

Fig. 1. Residual plots of PPM1. The left panel shows the residuals for the model of ln(α)and the right panel those for ln(β). The vertical lines in the upper plots illustrate the residual standard deviation (mean ± 2 SD) in 10 classes of equal width. The horizontal lines show the confidence interval for the mean (mean ± 2 SE of the mean).



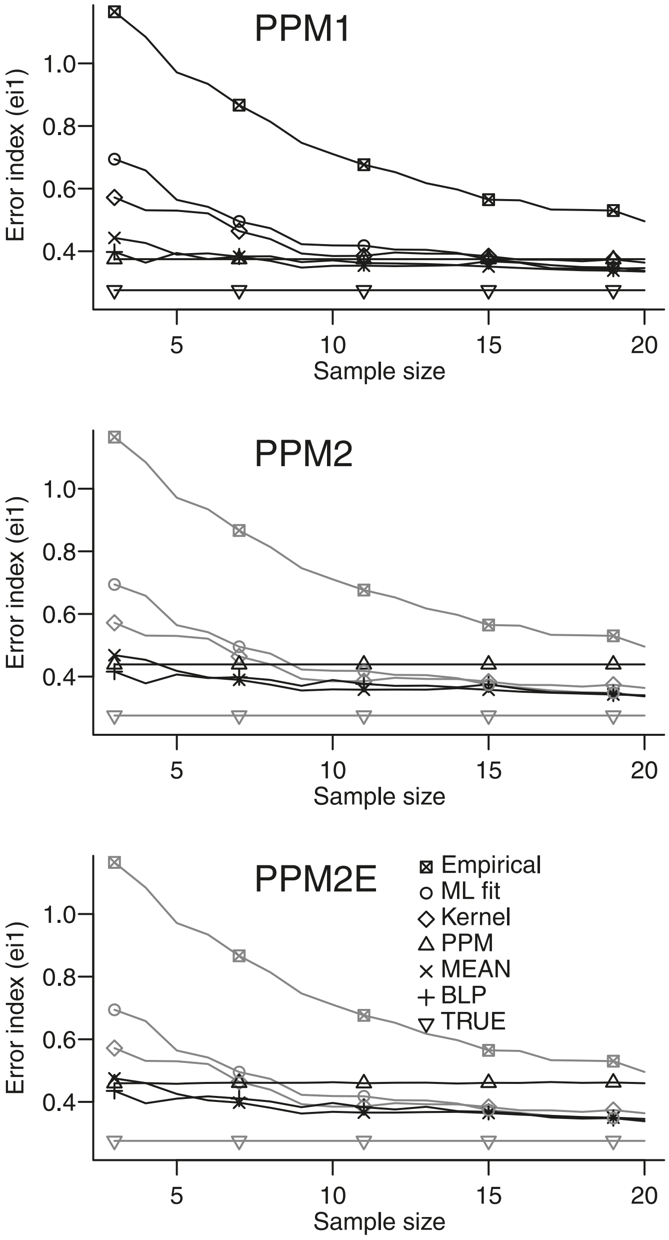
ple sizes, the ML fit performs equally well. The error index of the ML fit was always much lower than that of the raw empirical distribution. As the sample size increases from 3 to 20, the mean error index of the ML fit decreases from 0.67 to 0.33. The ML fit line crosses the PPM line at sample sizes 15 (PPM1) and 9 (PPM2 and PPM2E), which indicates that measuring stand characteristics with the accuracy assumed in each of the methods PPM1, PPM2, and PPM2E leads to as accurate diameter distributions as using a sample of this size alone.

For all sample sizes, the BLP and MEAN methods, which utilize both measured stand characteristics and sample diameters, perform generally better than the methods using the sample alone. They also perform better than the PPM method, except for the smallest sample sizes. Especially with models PPM2 and PPM2E, using the BLP method with only three sample trees improves the fit compared with the PPM method alone. With PPM1, measuring sample tree diameters improves the fit only slightly with rather large sample sizes (eight or more). The reason for this behaviaour is that the already quite accurate PPM1 method needs more sample trees to be further improved than the less accurate PPM2 method. BLP is better than MEAN for the smallest sample sizes, whereas MEAN is otherwise better. However, the lines for BLP and MEAN (see Fig. 2) are so fluctuating and close to each other that no final conclusions on the differences in the performance of these two methods can be made on the basis of these plots.

Discussion

The analyses of this study suggested that when the predictors of the PPM1s are measured accurately, the PPM method alone would lead to such accurate predictions of diameter distribution that it could be further improved only by using quite large diameter samples ($n > 8$). However, accurate measurements of the predictors are not expected, especially for the number of stems. For example, Kangas et al. (2004) found errors of up to 80.6% in the measured values of the number of stems in Finland. Errors in basal area and mean diameter were also very high (31.8% and 19.6%). Thus, the accuracy of the PPM method in practice may be much lower than simulations PPM1 and PPM2 would indicate. We believe that our simulations with PPM2E may be the closest to what can be expected in practice. If this is the case, our results indicate that PPM predictions can be improved with measurements of diameters from very small samples. Vice versa, estimates based solely on samples of any size between 3 and 20 trees can be further improved by utilizing informa-

Fig. 2. Mean error index (version 1) of different methods with different numbers of sample trees using PPM1, PPM2, and PPM2E (with measurement errors. To facilitate comparisons, the gray lines in the bottom two graphs repeat from the top graph the lines based on the sample alone. The line labeled as "TRUE" shows the error index of the ML fit to the complete data.



tion based on the PPMs. However, when the sample size exceeds 15 trees, the additional gain achieved by using PPMs may be marginal.

The measurement errors of the predictors in PPM2E were assumed to be independent of the estimation errors of the distribution that is used in prediction of the stand effects. This assumption means that the improvements in the predictions may not be very strong if both the measured sample of diameters and the predictors of the PPMs are based on the same sample plot. If the model is to be used in such situations, additional analysis of the performance is recommended.

The three pine species used in this study (*P. sylvestris*, *P. nigra*, and *P. halepensis*) were pooled in the analyses. Although using species-specific PPMs might have improved the results of this method slightly, fitting models separately for each tree species would imply smaller data sets and less precise estimates, which would have had the opposite effect.

In the evaluation of the methods, not all the trees of the stand were known. The total number of trees from each plot was, on average, 80 trees, and the maximum sample size that we utilized (n = 20) used about one fourth of them. However, the measured data were used as if it were the whole tree population of the stand. Furthermore, all of the trees were taken from the same plot. If they had been taken from different plots of the same stand, one could have expected the values for the error index to have been higher. However, this hardly had changed the order of the methods with respect to how well they perform.

The number of evaluation plots, 26 plots, was not very high. However, the trends observed with these data are very systematic and logical, so we do not have any reason to question the results. A larger evaluation data set would have made it possible to recognize even smaller differences between the methods, e.g., the difference between the MEAN and BLP methods. However, the practical meaning of such a small difference might be unimportant.

Combining measured diameter data with other sources of information has also been studied previously. For example, Green and Clutter (2000) compared a so-called precision-weighted composite estimator and a pseudo-Bayesian estimator for diameter class frequencies. If diameter class frequencies are seen as the probabilities of the multinomial distribution, this work can be considered to tackle the same problem by applying the continuous Weibull distribution. Mehtätalo (2005) proposed an approach where the percentile-based diameter distribution was localized using measured sample order statistics. The predicted percentiles were interpreted as parameters of an underlying percentile-based diameter distribution. The measured sample order statistics were interpreted as measured percentiles, i.e., measurements of the parameters that specify the diameter distribution. This paper applied the same idea to the Weibull diameter distribution.

The BLP prediction can be expected to perform well when the assumed Weibull model is correct, observed tree diameters are independent and identically distributed, and stand effects and utilized ML estimates have a multinormal distribution. Appendix A presents additional simulations where the effects of these properties were explored through simulations. It was shown that if the model is correct, samples are independent, and stand effects are normal, BLP is clearly the best among the evaluated methods, except for the smallest sample sizes in the most accurate PPMs. With smallest sample sizes, the reason for BLP not being the best is most likely the high bias of the ML estimates.

When BLP is used for prediction of stand effects, one should be aware of how well the underlying assumptions are met. In this study, the poorer than expected performance of BLP was explained by that the assumptions about the distribution of stand effects were not met. When applying the proposed method in new data sets, it is important to model such transformations of the parameters that have homogeneous, symmetric residuals with linear cross-model correlation. The same transformations needs to be used also in the ML fitting to the small sample so that the approximate asymptotic standard errors would be estimated for the same transformation. In this study, this meant that the truncated Weibull distribution was parameterized in computations using scale and logarithmic shape instead of the conventional shape and scale parameters as such.

Usually modellers are concerned about the properties of residuals and stand effects to make the tests on the fixed coefficients reliable. However, if localization with BLP is the intended use of the model, then the modelling of the random part also has a stronger impact on the predictions. This should motivate modellers to pay more attention to the behaviour of the residuals and stand effects of the fitted models. For example, Lappi (2006) decided upon an approach where the variance–covariance structures of the taper models were modelled using smoothing splines. The main motivation for such a complicated variance–covariance structure was not the estimation of the fixed part of the model but the use of the model in prediction using BLP.

References

Álvarez González, J., Rodríguez Soalleiro, R., and Vega Alonso, G. 1999. Elaboración de un modelo de crecimiento dinámico para rodales regulares de *Pinus pinaster* ait en galicia. Investig. Agrarias: Sist. Recursos For. **8**(2): 319–334.

Bailey, R., and Dell, T. 1973. Quantifying diameter distributions with the Weibull function. For. Sci. **19**: 97–104.

Casella, G., and Berger, R.L. 2002. Statistical inference. 2nd ed. Duxbury, Pacific Grove, Calif.

Castedo Dorado, F., Diéguez-Aranda, U., and Álvarez González, J. G. 2007. A growth model for *Pinus radiata* D. Don stands in north-western Spain. Ann. For. Sci. **64**(4): 453–465. doi:10.1051/forest:2007023.

Del Rio, M. 1998. Régimen de claras y modelo de producción para *Pinus sylvestris* L. en los sistemas central e ibérico. Technical report. Ph.D. thesis, Centro Investigación Forestal, Instituto Nacional de Investigaciones Agrarias, Madrid, Spain.

Diéguez-Aranda, U., Castedo Dorado, F., and Álvarez González, J. G. 2005. Funciones de crecimiento en área basimétrica para masas de *Pinus sylvestris* L. procedentes de repoblación en Galicia. Investig. Agrarias: Sist. Recursos For. **14**(2): 253–266.

Diéguez-Aranda, U., Castedo Dorado, F., Álvarez González, J.G., and Rojo Alboreca, A. 2006. Dynamic growth model for scots pine (*Pinus sylvestris* L.) plantations in Galicia (north-western Spain). Ecol. Model. **191**(2): 225–242. doi:10.1016/j.ecolmodel.2005.04.026.

Droessler, T.D., and Burk, T.E. 1989. A test of nonparametric smoothing of diameter distributions. Scand. J. For. Res. **4**(1): 407–415. doi:10.1080/02827588909382577.

Green, E.J., and Clutter, M. 2000. Using auxiliary information to estimate stand tables. Can. J. For. Res. **30**(6): 865–872. doi:10.1139/cjfr-30-6-865.

Härdle, W. 1990. Smoothing techniques with implementation in S. Springer, New York.

Hyink, D.M., and Moser, J.W.J. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. For. Sci. **29**(1): 85–95.

Kangas, A., and Maltamo, M. 2000. Calibrating predicted diameter distribution with additional information. For. Sci. **46**(3): 390–396.

Kangas, A., Heikkinen, E., and Maltamo, M. 2004. Accuracy of partially visually assessed stand characteristics: a case study of Finnish inventory by compartments. Can. J. For. Res. **34**(4): 916–930. doi:10.1139/x03-266.

Lappi, J. 1986. Mixed linear models for analyzing and predicting stem form variation of scots pine. Communicationes Instituti Forestalis Fenniae 134. Finnish Forest Research Institute, Suonenjoki, Finland.

Lappi, J. 1997. A longitudinal analysis of *h*/*d* curves. For. Sci. **43**: 555–570.

Lappi, J. 2006. A multivariate, nonparametric stem-curve prediction method. Can. J. For. Res. **36**(4): 1017–1027. doi:10.1139/X05-305.

Lappi, J., and Bailey, R. 1988. A height prediction model with random stand and tree parameters:an alternative to traditional site index methods. For. Sci. **34**: 907–927.

Lappi, J., Mehtätalo, L., and Korhonen, K.T. 2006. Generalizing sample tree information. *In* Forest inventory — methodology and application. *Edited by* A. Kangas and M. Maltamo. Springer New York. pp. 85–106.

McCulloch, C.E., and Searle, S.R. 2001. Generalized, linear, and mixed models. Wiley-Interscience, New York.

Mehtätalo, L. 2004. An algorithm for ensuring compatibility between estimated percentiles of diameter distribution and measured stand variables. For. Sci. **50**(1): 20–32.

Mehtätalo, L. 2005. Localizing a predicted diameter distribution using sample information. For. Sci. **51**(4): 292–302.

Mehtätalo, L., and Kangas, A. 2005. An approach to optimizing field data collection in an inventory by compartments. Can. J. For. Res. **35**(1): 100–112. doi:10.1139/x04-139.

Palahí, M., Pukkala, T., Miina, J., and Montero, G. 2004. Individual-tree growth and mortality models for Scots pine (*Pinus sylvestris* L.) in north-east Spain. Ann. For. Sci. **60**(1): 1–10. doi:10.1051/forest:2002068.

Palahí, M., Pukkala, T., and Trasobares, A. 2006. Modelling the diameter distribution of *Pinus sylvestris*, *Pinus nigra* and *Pinus halepensis* forest stands in Catalonia using the truncated weibull function. Forestry, **79**(5): 553–562. doi:10.1093/forestry/cpl037.

Palahí, M., Pukkala, T., and Trasobares, A. 2007. Comparison of beta, Johnson's SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). Eur. J. For. Res. **126**(4): 563–571.

Pinheiro, J.C., and Bates, D.M. 2000. Mixed-effects models in S and Splus. Springer-Verlag, New York.

R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from http://www.R-project.org.

Reynolds, M.R.J., Burk, T.E., and Huang, W.C. 1988. Goodness-of-fit tests and model selection procedures for diameter distribution models. For. Sci. **34**(2): 373–399.

Robinson, G.K. 1991. That BLUP is a good thing: the estimation of random effects. Stat. Sci. **6**(1): 15–32. doi:10.1214/ss/1177011926.

Siipilehto, J. 1999. Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fenn. **33**: 281–301.

Silverman, B.W. 1986. Density estimation. Champman and Hall, London, U.K.

Trasobares, A., and Pukkala, T. 2004. Using past growth to improve individual-tree diameter growth models for uneven-aged mixtures of *Pinus sylvestris* L. and *Pinus nigra* Arn. in Catalonia, north-east Spain. Ann. For. Sci. **61**(5): 409–417. doi:10.1051/forest:2004034.

Trasobares, A., Pukkala, T., and Miina, J. 2004. Growth and yield model for uneven-aged mixtures of *Pinus sylvestris* L. and *Pinus* *nigra* Arn. in Catalonia, north-east Spain. Ann. For. Sci. **61**(1): 9–24. doi:10.1051/forest:2003080.

Uuttera, J., and Maltamo, M. 1995. Impact of regeneration method on stand structure prior to first thinning. Comparative study North Karelia, Finland vs. Republic of Karelia, Russian Federation. Silva Fenn. **29**: 267–285.

Wang, M., and Rennolls, K. 2005. Tree diameter distribution modelling: introducing the logit–logistic distribution. Can. J. For. Res. **35**(6): 1305–1313. doi:10.1139/x05-057.

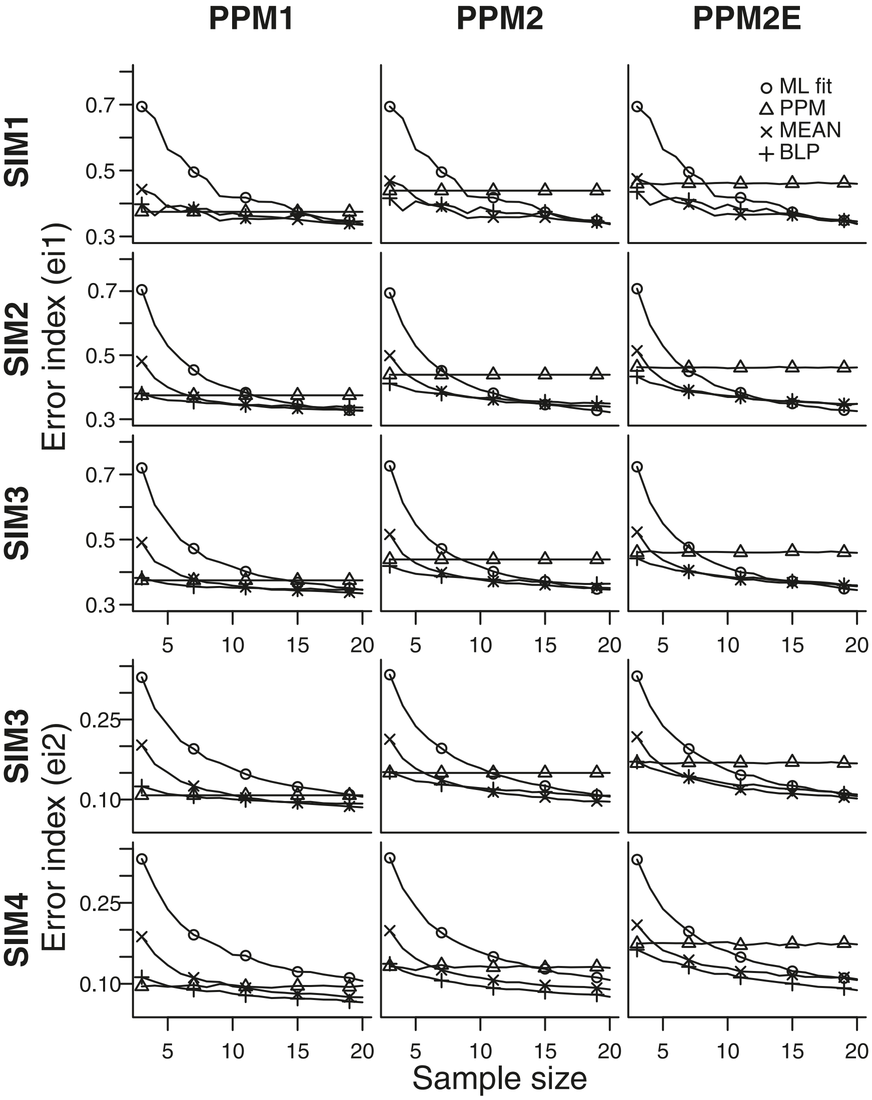
Appendix A. The effect of underlying assumptions on BLP

The underlying assumptions of the BLP method may not be very well met in simulation SIM1 because (*i*) the sample obtained from a concentric plot may not be independent due to spatial autocorrelation of tree size within a stand, (*ii*) the true diameter distribution of the stand may be too different from the assumed Weibull distribution, (*iii*) the assumptions on the stand effects of the PPMs may not be valid, especially they may not be identically distributed with constant variance and their joint distribution may not be bivariate normal, and (*iv*) the ML estimates of the parameters from a small sample may not meet the bivariate normality criteria. The folowing subsection explores the effects of reasons *i*, *ii*, and *iii* and the last subsection explores the effect of reason *iv*.

Spatial autocorrelation model shape, and normality of estimates

The effects of spatial autocorrelation, model shape, and normality of model residuals were explored through additional simulations. In the first of these, SIM2, the sample trees were selected randomly from among the sample trees of the stand instead of selecting them from the centre of the plot. This ensured that the utilized diameters were sampled independently from the measured trees, i.e., it removed (at

Fig. A1. Mean error index of different methods with different numbers of sample trees in simulations 1–4 using PPM1, PPM2, and PPM2E (with measurement errors). Version 1 of the error index was used in SIM1, SIM2, and SIM3 and version 2 in SIM 3 and SIM4; both versions of the index are shown for SIM3 to enable comparisons with both SIM4 and SIM2.



least partially) the effect of spatial autocorrelation of tree size. In SIM3, the trees were not sampled from the true measured trees of the plot, but a sample of the desired size was simulated from the assumed true truncated Weibull distribution, which was based on the ML fit to the complete data. Compared with SIM2, SIM3 removed the effect of having a true diameter distribution in the stand that was not of the Weibull form. Finally, in SIM4, the values of the true Weibull parameters of the stand were also simulated by adding multivariate normal errors to the PPM predictors of the parameters using the estimated variance–covariance matrix of stand effects (var(e_k)) and a mean of 0. Then, a sample of tree diameters was generated in a similar manner as in SIM3. Compared with SIM3, SIM4 removed the effect of having nonnormal stand effects of the PPMs (cf. Fig. 1).

In SIM2, SIM3, and SIM4, sampling could be replicated several times per plot. A total of 100 replicates was found to sufficiently remove the effect of sampling errors from the results. Random numbers were generated using the functions mvrnorm and runif of the R packages base and MASS (R Development Core Team 2010). The random numbers from a truncated Weibull distribution were generated by the probability integral method (Casella and Berger 2002, p. 55).

An alternative version 2 of the error index can be used in simulation studies where the underlying pdf of tree diameters is known (Mehtätalo et al. 2008). This was the case in SIM3 and SIM4. For comparing the estimated and true truncated Weibull distributions, it is defined as

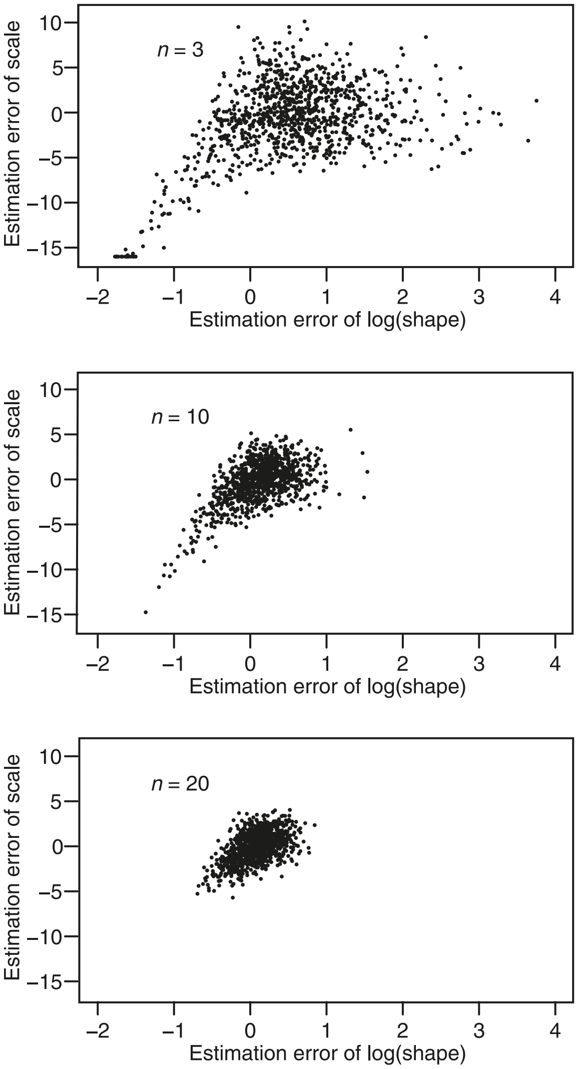
$$ei2_k = \int_{-t}^{\infty} \| f(d|\widehat{\boldsymbol{\theta}_{k,1}}) - f(d|\boldsymbol{\theta}_{k,\mathrm{TRUE}}) \| \,\mathrm{d}y$$

where $f(d|\widehat{\boldsymbol{\theta}_{k,1}})$ is the density based on the estimates and $f(d|\boldsymbol{\theta}_{k,\mathrm{TRUE}})$ is the density based on the true values of the parameters. This index ($ei2_k$) obtains lower values than the previous version ($ei1_k$) because it does not include the classification and sampling errors. For SIM3, both versions of the index are reported to allow comparison with both SIM2 and SIM4.

The lowest row of panels in Fig. A1 show that the BLP method leads to the lowest mean error index in SIM4 (except for PPM1 and PPM2 with the smallest sample sizes). This result is expected based on the properties of BLP. In SIM3, where the true parameter values were used instead of the simulated ones, the order of the ML fit, MEAN, and BLP methods remains quite similar to that in SIM4, but the difference between BLP and MEAN vanishes with large sample sizes and the difference between the PPM method and BLP decreases. This indicates that the assumptions on the homogeneity and normality of the stand effects are quite a critical issue in the performance of the BLP method.

In the upper three rows of panels in Fig. A1, SIM2 does not show any clear difference from SIM3. This indicates that the possible discrepancy from the Weibull assumption did not have any visible effect on the ranking of the methods. However, SIM1 differs from SIM2 slightly. In SIM2, PPM crosses the ML fit with the sample sizes 13 (PPM1), 8 (PPM2), and 7 (PPM2E), which are slightly lower values than the values where PPM crosses the ML fit in SIM1 (15, 9, and 9). This indicates that measurements taken from a small plot do not give as much information as a similar number of measurements spread over a larger area. The reason for this is the spatial autocorrelation of tree size within the stand. Mehtätalo et al. (2006) reported similar results with quantile tree measurements.

Fig. A2. Realized errors of the ML estimates of ln(α) and ln(β) in simulations of 1000 samples of each of the sample sizes 3, 10, and 20.



Normality of the ML estimates

To explore the properties of the ML estimates from small samples (assumption *iv* in the list above), a small simulation was conducted. In that simulation, repeated samples of sizes 3, 10, and 20 trees were drawn from a truncated Weibull distribution and the truncated Weibull distribution was fitted to these samples using ML. The bivariate normality of the obtained parameter estimates was explored graphically.

The plots in Fig. A2 show examples of the joint distribution of the estimation errors of ln(α) and ln(β) in a stand where $\alpha = 3$ and $\beta = 17$. With a sample size of 3, the estimates are biased and the marginal distributions are quite different from the normal distribution, with heavy left tails. In addition, the correlation between the estimation errors is rather nonlinear. Similar trends can be seen also with a sample size of 10. With a sample size of 20, the estimation errors seem to be quite close to normal, even though a detectable bias still occurs in the estimate of the logarithmic shape.

References

Casella, G., and Berger, R.L. 2002. Statistical inference. 2nd ed. Duxbury, Pacific Grove, Calif.

Mehtätalo, L., Maltamo, M., and Kangas, A. 2006. The use of quantile trees in predicting the diameter distribution of a stand. Silva Fenn. **40**(3): 501–516.

Mehtätalo, L., Gregoire, T.G., and Burkhart, H.E. 2008. Comparing strategies for modelling tree diameter percentiles from remeasured plots. Environmetrics, **19**(5): 529–548. doi:10.1002/env.896.

R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from http://www.R-project.org.