

Multidimensional scaling of first-return airborne laser echoes for prediction and model-assisted estimation of a distribution of tree stem diameters

Steen Magnussen¹ · Jean-Pierre Renaud²

Received: 8 March 2016 / Accepted: 19 August 2016 / Published online: 2 September 2016
© INRA and Springer-Verlag France 2016

Abstract

• **Key message** We demonstrate how multidimensional scaling can be used to combine forest inventory field data and airborne laser scanner data to obtain both predictions and model-assisted estimation of a tree stem diameter distribution.

• **Context** The size distribution of forest trees is important both for management planning and analysis purposes. Yet field samples are rarely large enough to assuage a desired accuracy of a direct estimation in all areas of interest. Improvements in spatial coverage and accuracy are possible with a census—or a very large sample of one or more cost-effective auxiliary variables that can inform one about the tree size distribution.

• **Aims** The objective of this study is to demonstrate how a relative frequency distribution of canopy heights from airborne laser scanner data can be used to improve direct estimates of a tree size distribution.

• **Methods** Multidimensional scaling is used to link a relative frequency distribution of canopy heights to an observed plot-level distribution of tree size.

• **Results** A multivariate linear model can be used for both predictions and model-assisted estimation of a tree stem diameter distribution.

• **Conclusion** Multidimensional scaling can provide a multivariate linear link between two relative frequency distributions and is therefore ideally suited for both stand-level predictions and design-based inference of tree size distributions.

Keywords Forest inventory · Stratified random sampling · Relative frequency distribution · LiDAR · Goodness-of-fit · Bootstrap confidence intervals

Handling Editor: Barry Alan Gardiner

Contribution of the co-author

Jean-Pierre provided data for the study, participated in discussions about methodology, presentation, and interpretation of results. J-P reviewed and commented on two earlier drafts.

✉ Steen Magnussen
steen.magnussen@canada.ca

Jean-Pierre Renaud
jean-pierre.renaud-02@onf.fr

¹ Natural Resources Canada, Canadian Forest Service, Victoria, British Columbia, Canada

² Département RDI, Office National des Forêts, Fort-de-France, France

1 Introduction

The size distribution of trees in a stand, a stratum, or a forest is an important summary statistic of the resource. It is used by forest management for planning purposes (Clutter et al. 1983, ch. 9; Kangas et al. 2015, ch. 4) and by analysts to assess historic trends and current and future resource values (del Río et al. 2016; Hyytiäinen and Haight 2012; Valencia et al. 2016). The stem diameter of a tree at a reference height of 1.3 m (DBH) is a key attribute correlated with the height, volume, biomass, and utilization (Bailey 1980; Clutter et al. 1983, ch. 4.2.2). Through a design-based field sampling and measurements of DBH in fixed area plots, a forest inventory provides an unbiased estimate of the distribution of DBH in the sampled population and in areas of interest provided they contain at least one sample plot (Fuller 2009, p. 69 and p. 304).

In national forest inventories with thousands of plots, a sample-based estimate of a national or even a regional DBH distribution may be quite accurate (Gasparini et al. 2013), but for most enterprise level inventories—with a few hundred plots distributed across several strata—the accuracy of direct estimate of a DBH distribution in an area of interest (e.g. a stratum or a stand) may not be satisfactory for planning or analysis (Korn and Graubard 1998; Magnussen et al. 2013; Sison and Glaz 1995).

Airborne laser scanner (ALS) data and photogrammetric point clouds (PPC), in combination with optical remotely sensed data, offer options for extending tree size and species-related information to areas not covered by a field plot (Brosofske et al. 2014; Magnussen et al. 2013; Wallerman and Holmgren 2007). Through modelling, nearest neighbour imputations, or single tree detection (STD), predictions of the desired tree size distribution can be made for an area of interest (Maltamo and Gobakken 2014; Vauhkonen et al. 2014a).

There are currently a large number of approaches and methods that can produce satisfactory results (Brosofske et al. 2014; Hyypä et al. 2008). Evidently, the so-called area-based approach (ABA)—which uses extracted plot-level multivariate remotely sensed explanatory variables—is easier to implement than STD. The preferred approach is context specific and depends on available expertise, available data, and inventory objectives. Implementation issues and modus operandi are important factors for choosing the combination of auxiliary data (ALS, PPC, optical data) and method (ABA, STD) for extracting a tree size distribution.

At the level of a forest management unit, the interest is in the diameter distribution within a stand, but for planning and analysis of larger areas, the interest may shift to a diameter distribution within a stratum or an entire population. An estimator that combines stand-level (model based) predictions and design-unbiased estimates for larger areas (e.g. a stratum) would therefore be attractive. Current estimators are all model based with a focus on stand-level predictions; their application to larger areas will generate estimates that are model based, i.e. with properties that depends (strongly) on the correctness of the model (Gregoire 1998; Magnussen 2015).

In this study, we demonstrate that multidimensional scaling (MDS) provides an estimator that can be used both for stand-level predictions and for model-assisted inference of the diameter distribution of trees in a larger area of interest (e.g. a stratum). In our demonstration, we apply a MDS for the linear transformation of a discrete probability distribution function of first-return ALS data of canopy height to a discrete probability distribution of tree stem diameters. Examples are provided for ALS first returns and inventory field data from the four sites in France covering a diversity of forest stands (Bouvier et al. 2015; Véga et al. 2016).

2 Material and methods

2.1 Field data

Field data from three sites in Aillon, Bure, and the Vosges were collected as part of the Foresee research project (<http://foresee.fcba.fr>). Data from St. Gobain were collected by the Office National des Forêts (ONF). Field measurements were performed during the winter period (see Table 1) from 255 fixed area plots located in Aillon (49), Bure (28), St. Gobain (133), and the Vosges (45). All but one plot had a radius of 15 m with an area of 706.9 m². The exception was a 100 m × 100 m plot in Aillon. Species and diameter at a reference height of 1.3 m (DBH) above ground was, as a rule, recorded with a tape for all live trees with a DBH ≥ 17.5 cm. Note, this lower diameter threshold is a standard in many ONF inventories (Duplat and Perrotte 1981, p. 84). We did not have data on the number and DBH of smaller trees. It is assumed that the forests in the four sites have been stratified by leading species (dominant by basal area) and that the plots constitute a simple random sample (without replacement) within each post stratum. The strata included in this study have as leading species: beech (*Fagus sylvatica* L.), hornbeam (*Carpinus betulus* L.), fir (*Abies alba* Mill.), linden (*Tilia* spp. L.), maple (*Acer pseudoplatanus*, L. (Falk)), and oak (*Quercus* spp. L.). A summary of the field data is in Table 2.

The DBH values in a plot were distributed to $k = 15$ equal-width (6 cm) intervals (see ALS data for details on k). Interval midpoints were 20, 26, ..., 104 cm. A few trees with a DBH greater than 107 cm were assigned to the 15th bin.

2.2 ALS data

ALS acquisition details are given in Table 1. ALS flights were done in the same year as the field measurements, except for the Vosges, where the latter were performed one growing season after the ALS acquisition. Scanners (Riegl or Optech) were mounted on fixed-wing aircrafts flying at 500 to 1500 m above the ground (Table 1). The pulse frequency ranged from 71 to 100 kHz with a range of scan angle from ±13° to ±30°. The mean density of echoes ranged from 2.6 to 18 m⁻². A Terrasolid software (www.terrasolid.com) was used for echo classification to ground and non-ground. The mean pulse densities of ground echoes ranged from 0.5 to 4.1 m⁻².

ALS first-return canopy heights (X_{hi}) in the h th stratum and the i th plot were binned to k height classes and converted to a discrete relative frequency distribution. The number of bins was determined using the suggestion by Freedman and Diaconis (1981). With a mean of 8491 first returns per plot, an inter-quantile range of 31 m, and a range of 43 m, we took $k = 15$. The first bin was fixed to the interval from 0 to 1.3 m, while the remaining bins divided the interval from 1.3 to 43 m in 14 equal-width bins.

Table 1 Site areas, period of field measurements, and acquisition details for the ALS data

Site	Area (km ²)	Field measurements	ALS flight period	LiDAR type	Flight altitude (m)	Mean echo density m ⁻²	Scanning angle (degrees)	Band swath (m)
Aillon	25.5	03–04/2011	08/2011	Riegl LMS-Q560	550	16.5	±30°	621
Bure	40	02/2010	03/2010	Riegl LMS-Q560	600	18.7	±20°	417
St. Gobain	104	01–03/2015	03/2015	Riegl LMS-680i	550	18	±30°	615
The Vosges	1362	02–04/2012	04/2011	Optech ALTM3100	1500	4.6	±13°	673

2.3 Sampling design

For the purpose of our demonstration of MDS in a context of both stand-level predictions and model-assisted estimators and inference, we have assumed a stratified simple random sampling without replacement in the ten strata across the four sites (Table 2). Stratification was nominally based on dominant tree species (by a basal area). Sample sizes (n_h) in each stratum ($h = 1, \dots, H = 11$) are in Table 2. Strata sizes N_h are unknown but fixed arbitrarily in this study for the sake of demonstration (see below). Sample units were fixed area circular plots with a horizontal radius of 15 m. A single plot in Aillon measured 100 m × 100 m.

2.4 Estimators

It is assumed that the discretized relative frequency distribution of DBH (\mathbf{Y}) and heights of first-return ALS echoes (\mathbf{X}) are associated through a MDS via an ortho-normal transformation matrix \mathbf{B} (Mardia et al. 1979, ch. 14). For the i th plot in the h th stratum, we have the following model:

$$\mathbf{Y}_{hi} = \mathbf{B}_h \mathbf{X}_{hi} + \mathbf{e}_{hi}, \quad h = 1, \dots, H, i = 1, \dots, N_h \quad (1)$$

where \mathbf{e}_{hi} is a residual term and N_h is the size of stratum h expressed in the number of sample plots that nominally would

fill the stratum. For a finite area stratum with an area of A_h and sample plots with an area of a , N_h becomes the integer value of $A_h \times a^{-1}$. Note, we do not claim that the model in Eq. 1 is a true model; it is merely a working model formulated prior to observing the data and therefore external to the sample. There is no search for a best model form nor a selection of explanatory variables that would violate the requirements to a working model (Särndal et al. 1992, ch. 6.7).

In an ortho-normal transformation, the dimension of \mathbf{Y}_{hi} , \mathbf{X}_{hi} , and \mathbf{e}_{hi} are the same (k). As argued in section 2.2, k is 15 throughout. Consequently, the dimension of \mathbf{B}_h is 15×15 . The multivariate linear model in Eq. 1 ensures that a unique ortho-normal \mathbf{B}_h exists. To see this, pre-multiply Eq. 1 with the inverse, viz. transpose of \mathbf{B}_h and use that $\mathbf{B}_h' \mathbf{B}_h = \mathbf{I}_h$ (Searle 1982, p.320). An estimate of \mathbf{B}_h was obtained by constrained weighted multivariate least-squares techniques. The imposed constraints ensure that the Euclidean length of each column and row vector in \mathbf{B}_h is 1.0. The weights were the inverse of the sample inclusion probability (cf. section on sampling design). For the i th sample unit (plot), we have the following:

$$\mathbf{Y}_{hi} = \hat{\mathbf{B}}_h \mathbf{X}_{hi} + \hat{\mathbf{e}}_{hi} = \hat{\mathbf{Y}}_{hi} + \hat{\mathbf{e}}_{hi} \quad (2)$$

where $\hat{\mathbf{Y}}_{hi}$ is an estimate of the expected relative frequency distribution of DBH in the i th plot within stratum h . A 95 %

Table 2 Summaries of circular field plots with a radius of 15 m (0.07 ha)

Site	Plots	Dom. Species	Number of species (DBH ≥ 17.5 cm)			Stems with DBH ≥ 17.5 cm			DBH ≥ 17.5 cm		
			Min	Median	Max	Min	Median	Max	Min	Mean	Max
Aillon	12	Beech	2	3	5	7	22	52	17.6	35.2	97.2
	37	Fir	2	3	5	10	20	66	17.0	38.0	94.5
Bure (OPE)	15	Beech	1	3	5	15	23	35	17.2	30.9	83.1
	13	Hornbeam	3	5	7	11	30	45	17.2	35.4	108.9
St. Gobain	15	Beech	1	3	5	8	11	21	17.0	37.5	87.0
	62	Hornbeam	2	3	6	7	14	26	18.0	33.8	104.0
	15	Linden	2	4	6	9	16	23	18.0	32.4	102.0
	14	Maple	2	4	7	7	13	22	18.0	34.5	88.0
The Vosges	27	Oak	1	3	6	7	13	23	18.0	36.6	122.0
	23	Beech	1	2	7	9	25	69	17.2	28.9	104.7
	22	Fir	2	3	5	16	28	72	17.2	34.0	109.5

confidence interval for \hat{Y}_{hi} was obtained from the 0.025 and 0.975 quantile in a distribution of 120 parametric bootstrap estimates of $\hat{Y}_{hi}^b = \hat{B}_h^b X_{hi}$, $b = 1, \dots, 120$ where \hat{B}_h^b is the b th bootstrap estimate of B_h . Resampling with replacement was at the level of plots.

The estimate of B_h can be used directly for a prediction of the diameter distribution in a stand (st) for which only the vector of auxiliary variables X_{st} is known. The prediction would be $\hat{B}_h X_{st}$ and by necessity model based (synthetic) (Chambers and Clark 2012, p. 162). Note that a prediction is ‘automatically’ scaled to the stand level without a need to subdivide the stand into virtual plots and repeat predictions for each.

The mean of \hat{Y}_{hi} taken across the plots in a stratum (\bar{Y}_{hi}) is a model-based estimator of the stratum mean of Y_{hi} , i.e. \bar{Y}_{hi} . It is a model-assisted estimator only if the average residual vector \hat{e}_h is a vector of zeros. A 95 % confidence interval for \bar{Y}_{hi} was derived from the 120 bootstrap estimates of B_h and the n_h observations of X_{hi} .

With a census of X_{hi} (\bar{X}_h) in stratum h , a model-assisted estimator of the stratum relative frequency distribution of DBH in this stratum becomes (Särndal et al. 1992, ch. 7.8) the following:

$$\hat{Y}_h = \hat{B}_h \bar{X}_h + \hat{e}_h \tag{3}$$

with expectations computed as weighted averages, whereby weights are the inverse of the sample inclusion probabilities $\pi_{hi}, i = 1, \dots, n_h$. We shall assume a stratified simple random sampling without replacement. An approximation to the variance–covariance matrix ($\hat{\Sigma}$) of \hat{Y}_h is (Binder 1983) the following:

$$\hat{\Sigma}(\hat{Y}_h) = N_h^{-2} \sum_{i=1}^{n_h} \sum_{j \in s} \frac{\pi_{hij}^{-1} - \pi_{hi} \pi_{hj}}{\pi_{hij}} \hat{e}_{hi} \otimes \hat{e}_{hj} \tag{4}$$

where $e_{hi} \otimes e_{hj}$ is the Kronecker product of the two residual vectors (Searle 1982, p. 265) yielding a 15×15 matrix of cross-products.

In this study, we did not have census vectors \bar{X}_h of relative class frequencies of ALS first-return canopy heights. For the sake of demonstration, they were simulated. First, we fixed strata sizes to $N_h = 10^4 n_h$ and then simulated $N_h - n_h$ randomly scaled and perturbed versions of $X_{hi} i = 1, \dots, n_h$. The $N_h - n_h$ vectors were first selected with-replacement and unequal probability from the n_h actual observations. The probability of selection was assigned at random from a uniform distribution on the interval 0.8 to 1.0. The scaling of a selected vector was done with a random draw from a uniform distribution on the interval 0.7 to 1.3. Finally, a random perturbation was added via a random draw from a uniform distribution on the interval -0.08 to 0.08 . All simulated vectors were constrained to satisfy a sum-to-one of the k relative frequencies.

In case of a post-stratification to dominant species, the estimator of a post stratum mean (cf. Eq. 3) and variance (cf. Eq. 4) requires a modification as outlined in, for example, Särndal (2011) in order to capture the random allocation of samples to post strata.

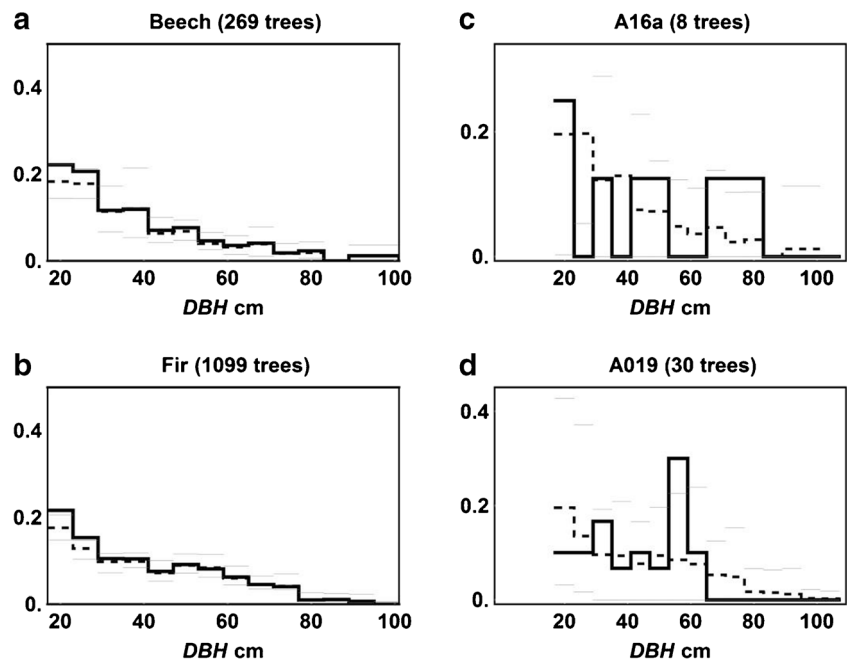
To gauge the fit between a stratum mean of n_h fitted DBH distributions \hat{Y}_h and the observed mean \bar{Y}_h , we obtained 120 with-replacement bootstrap replications of \hat{Y}_h and repeatedly tested the hypothesis $H_0 : \bar{Y}_h^b = \bar{Y}_h$, $b = 1, \dots, 120$ where the superscript denotes an estimate in the b th bootstrap sample (Bickel and Krieger 1989). An Anderson–Darling (AD) test (Anderson and Darling 1952) provided 120 test statistics, and Holm’s sequential rejection test procedure at the 1–0.05 level (Holm 1979) was used to gauge the overall level of significance. Each AD test was carried out with the observed (tree-level) distribution of DBH values and matching numbers of predicted values of DBH drawn at random from the inverses to the fitted cumulative distribution functions of DBH. This test procedure was intended to bestow robustness on the AD test which is sensitive to the presence of low frequency DBH classes (abound in this study). The AD tests were complemented by the less powerful Kolmogorov–Smirnov (KS) test to gauge goodness of fit in high frequency DBH bins around the center of a distribution (Conover 1980, p. 344). The equality of the mean DBH in the fitted and observed relative frequency distributions was tested with a t test imbedded in the above bootstrap test procedure.

The combination of 15 DBH bins and a relative low number of recorded stems in the studied field plots (Table 2) effectively rendered the statistical power too low (< 0.6) for a meaningful goodness-of-fit testing at the plot level.

3 Results

Fitted strata means of the relative frequency distribution of DBH (\hat{Y}_{hi}) matched the observed means to within a mean absolute difference (across 15 bins) that varied from 0.003 (Bure, hornbeam) to 0.016 (the Vosges, beech) with a mean of 0.007. The maximum absolute difference (in a single bin) varied from 0.006 (St. Gobain, Maple) to 0.08 (the Vosges, beech) with an average of 0.024. We associate the relatively larger errors in St. Gobain with its relatively low number of plot trees and a corresponding relatively poorly determined distribution of DBH in a single plot. Figures 1, 2, 3, and 4 provide a visual impression of the achieved stratum-level results. The figures also show a comparison between the observed and expected distribution of $DBH \geq 17.5$ cm in a randomly drawn plot. Plot-level results are clearly not convincing. The relatively poor fit at the plot level illustrates that a relative frequency

Fig. 1 Observed (*full line*) and expected (*dashed line*) relative frequency distributions of DBH ≥ 17.5 cm in Aillon. Leading species stratum means are in subplots **a** and **b**. Plot-level results are in subplots **c** and **d**. Bootstrap 95 % confidence intervals are in *grey*. See text below Eq. 2 for details



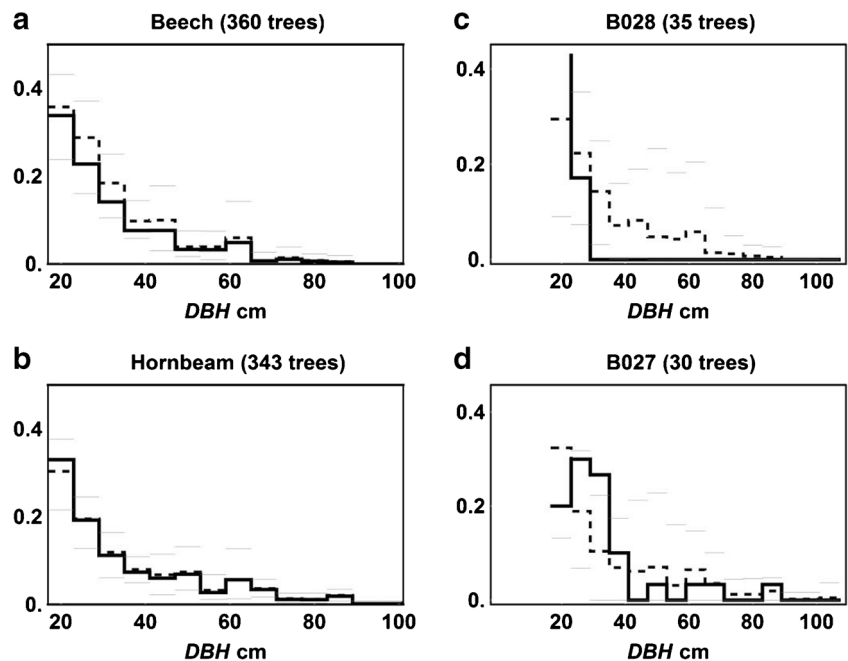
distribution of first-return canopy heights—at the given spatial resolution of a field plot—contains little information about an underlying distribution of DBH ≥ 17.5 cm.

We noted that the number of trees in a plot was more important than the number of field plots for obtaining a good stratum-level fit because in each plot, several DBH classes had a tree count of zero. Results from the AD and KS goodness-of-fit tests reflected, to a larger degree, the sample size of trees than practical relevant issues related to the fitting procedure. We ascribe this to the fact that MDS is linear and asymptotically unbiased. We rejected all hypotheses of

equality between the observed and fitted distributions for fir in Aillon and beech in the Vosges. A high rejection rate of 0.6 was registered for hornbeam in St. Gobain. In the remaining eight cases, we accepted the null hypothesis. A rejection of the equality of the observed and fitted distribution triggered, in most cases, also a rejection of the *t* test of equal mean DBHs. However, the differences were never greater than 2.7 cm and deemed unimportant for practical applications.

A stratum-level census of the relative frequency distributions of canopy height in ALS first returns (simulated) was converted to a stratum-level model-assisted estimate of the

Fig. 2 Observed (*full line*) and expected (*dashed line*) relative frequency distributions of DBH ≥ 17.5 cm in Bure (OPE). Leading species stratum means are in subplots **a** and **b**. Plot-level results are in subplots **c** and **d**. Bootstrap 95 % confidence intervals are in *grey*. See text below Eq. 2 for details



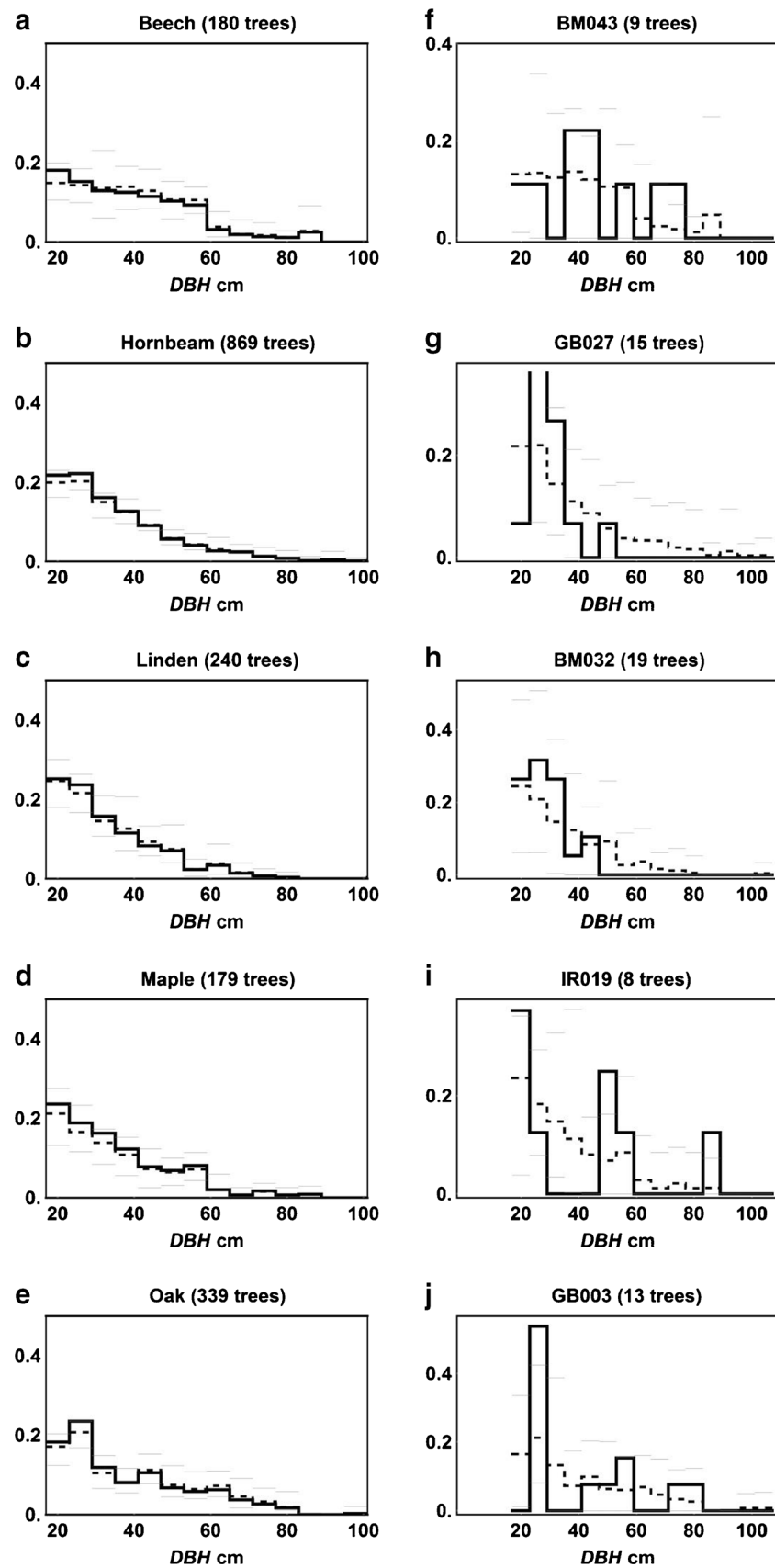
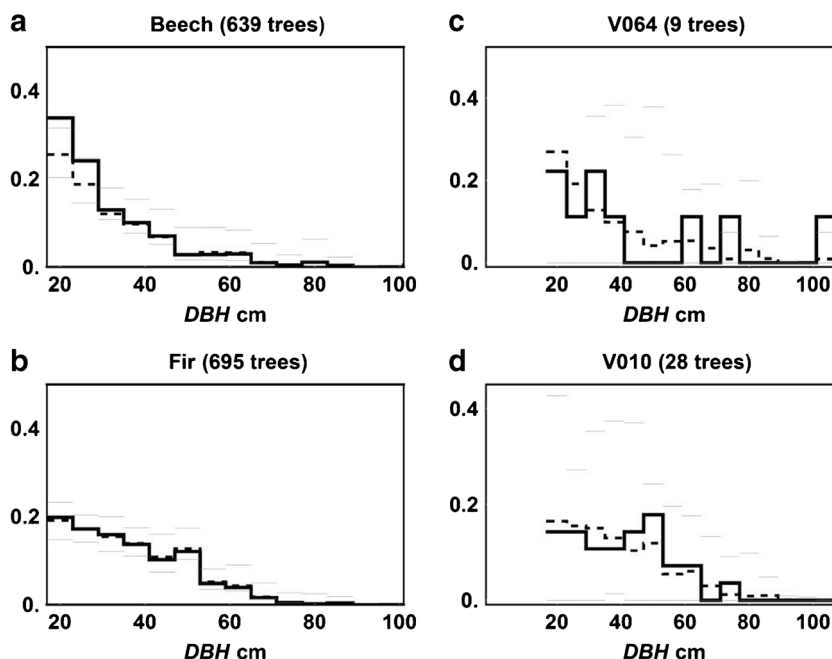


Fig. 3 Observed (full line) and expected (dashed line) relative frequency distributions of DBH ≥ 17.5 cm in St. Gobain. Leading species stratum means are in subplots a to e. Plot-level results are in subplots f to j. Bootstrap 95 % confidence intervals are in grey. See text below Eq. 2 for details

Fig. 4 Observed (*full line*) and expected (*dashed line*) relative frequency distributions of DBH ≥ 17.5 cm in the Vosges. Leading species stratum means are in subplots **a** and **b**. Plot-level results are in subplots **c** and **d**. Bootstrap 95 % confidence intervals are in *grey*. See text below Eq. 2 for details



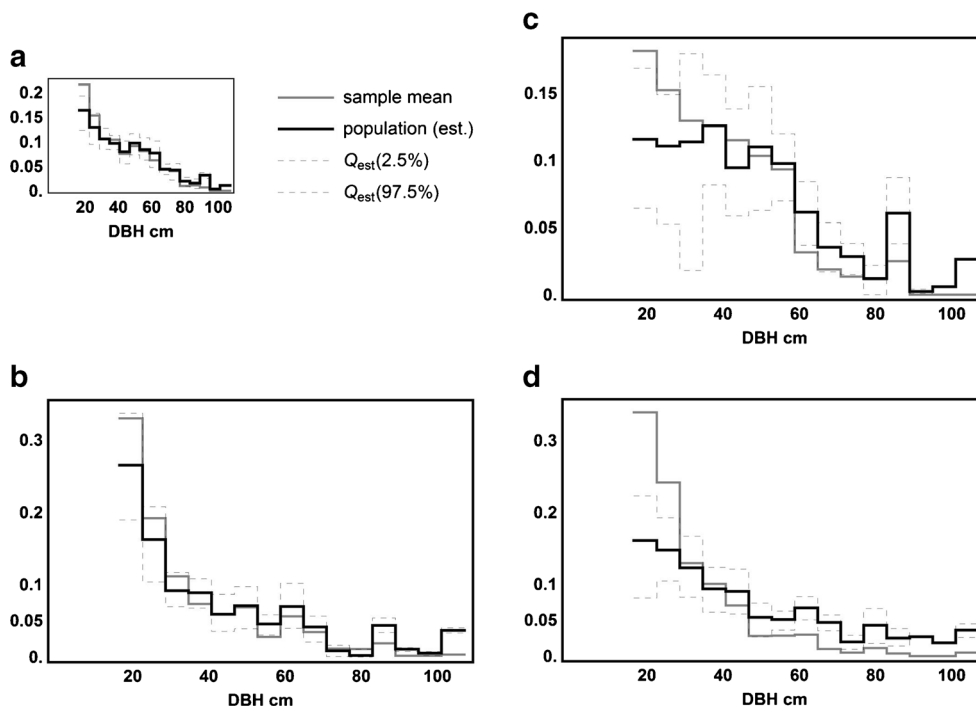
relative frequency distribution of DBH ≥ 17.5 cm via Eq. 3. Bootstrap-based 95 % confidence intervals were computed from 120 bootstrap estimates of the mean residual vector in Eq. 3. The mean absolute error in a DBH class estimate is in the order of 0.04, and the width of 95 % confidence intervals varied from 0.001 for the largest DBH class to 0.21 for the class with a DBH between 17.5 and 23 cm. With a prevailing expectation of an inverse *J*-shaped distributions of DBH ≥ 17.5 cm in our study, the lower frequencies of large trees predicate the trend for the relative errors to increase

rapidly as DBH increases. For the three largest DBH classes with relative frequencies below 5 %, the relative errors were typically in excess of 30 %. The four randomly selected site specific examples of stratum-level results are shown in Fig. 5.

4 Discussion

The proposed MDS of first-return ALS canopy heights for the purpose of predicting a tree size distribution in a stand

Fig. 5 Examples of estimated (*full line*) stratum relative frequency distributions of DBH ≥ 17.5 cm in the following: **a** Aillon and fir, **b** Bure (OPE) and hornbeam, **c** St. Gobain and beech, and **d** the Vosges and beech. Sample means are indicated (*dashed line*). Bootstrap 95 % confidence intervals are in *grey*. See text below Eq. 2 for details



appears—when comparing an observed distribution to an expected distribution in a plot—to be as effective as tried alternatives such as the following: fitting a parametric distribution (Magnussen et al. 2013; Maltamo et al. 2007; Thomas et al. 2008), decile fitting (Bollandsås et al. 2013; Gobakken and Næsset 2005), k -nearest neighbour imputations (Lindberg et al. 2013), or modelling of tree size based on STD methods (Kankare et al. 2015; Vauhkonen et al. 2014b). For all methods, results at the plot level are typically poor, partly due to a relatively low number of observations per plot and a large among-plot variation. Yet an advantage of MDS is the ease of scaling; predictions can be made for any area directly on the basis of a census of the relative frequencies of canopy heights. In this regard, MDS shares properties of the cumulant-based model proposed by Magnussen et al. (2013). Most traditional ALS auxiliary data metrics (Næsset 1997; Næsset 2004) are scale-dependent quantiles (Magnussen and Boudewyn 1998) and require that \mathbf{X} is obtained with the spatial support of a field plot providing an observation of \mathbf{Y} . Even for relative small areas, this can add a non-trivial time-consuming computational effort. Nowhere is this more pronounced than with a STD approach (Ene et al. 2012; Gupta et al. 2010; Heurich 2008).

The proposed MDS method also works with sample plots of different sizes. This is important because sample plots near the edge may have a portion of their area outside of the sampling frame (Gregoire and Valentine 2008, p. 59; Mandallaz 2008, p. 223). This is also why we did not exclude the single 100×100 m plot in Aillon.

Forest planning and analysis may require a conversion of a relative frequency distribution of DBH to an actual size distribution (Bollandsås et al. 2013; Maltamo et al. 2004; Maltamo et al. 2012; Xu et al. 2014). This conversion, however, is beyond the scope of our study.

We recognize that prediction of a DBH distribution in stand is likely to be the most important application for our proposed MDS (Magnussen 1986; Maltamo et al. 2004; Maltamo et al. 2006). However, inference about a DBH distribution in a larger area, like a stratum, a population, or any area of interest represented by a sufficient probability sample is also an important area of targeted application (Chen 2004; Lehtonen and Veijanen 2009). Our results, and results from cited studies, suggest that as few as 12 plots with ten or more trees provide an estimate of a size distribution in an area of interest suitable for planning and analysis. The attraction of MDS in these applications rests with the ortho-normal transformation matrix \mathbf{B} ensuring a linear transformation of \mathbf{X} to \mathbf{Y} and permitting a model-assisted inference with any probability sampling design (Särndal et al. 1992, ch. 6). If a census of \mathbf{X} is not available, but instead a large sample estimate of \mathbf{X} , a model-assisted approach to inference is still possible with a few simple modifications (Mandallaz et al. 2013;

Ståhl et al. 2016). Cited alternatives to MDS may claim a design-based empirical difference estimator (Baffetta et al. 2009; Magnussen 2013) disregarding a critical requirement for predictions to be generated independently of the observed sample (Särndal et al. 1992, ch. 6.3).

For the purpose of the prediction or an estimation of a DBH distribution, stratifying by leading species is intuitive as it can be assumed that the distribution of first-return canopy heights to a large degree reflects the crown architecture of the leading species (Heurich 2008). In stands with no clear dominance of a single species, it must be expected that the relationship between a relative frequency distribution of first returns will be weaker than in our study.

We quantified uncertainty in an expected relative frequency distribution by bootstrap quantiles (Bickel and Krieger 1989) as they are easy to obtain and communicate. The alternatives would be Agresti–Coull or Clopper–Pearson type intervals (Dean and Pagano 2015). While the absolute errors in our study may seem reasonable for stratum-level inference, it remains clear that the number of trees measured needs to be increased significantly if the maximum tolerable relative error on an estimated relative frequency of trees in an important DBH class is around 20 % (King and Madansky 2013).

Our demonstration of the MDS was limited to data with a lower DBH threshold of 17.5 cm. Since first-return canopy heights arrive from an unknown number of trees and from different depths of the canopy, one can intuitively surmise that they have a greater information content on DBH in trees with an exposed tree crown (Magnussen et al. 1999) than for trees not visible from above. From that follows that estimation errors of a truncated distribution would be higher when a portion of the first returns arrives from trees with a DBH below the threshold. We could not pursue these aspects with our data.

The relatively high threshold of 17.5 cm meant that the average of the expected distributions of DBH within a leading species stratum in our study resembled an inverse J -shaped distribution (Moser 1976) with a near exponential decline in relative class frequencies between the first and last class. However, in a single plot, the distribution of DBH was often irregular. As with any regression model, if the information in \mathbf{X} from a single plot is only weakly correlated with \mathbf{Y} , the prediction of \mathbf{Y} for the plot will be close to the predicted average of \mathbf{Y} in the sample (Draper and Smith 2014, ch. 3).

In our exposé, it was assumed that the forest in the four sites was stratified by leading species (by basal area). The information necessary for a design-based stratification may not be available until after the field sampling is complete (Andersen et al. 2011; Saborowski et al. 2010; Tomppo et al. 2008; von Lüpke et al. 2012). In this case, a model-assisted inference is still possible, but the additional variance generated from classification errors and random post-strata sample sizes must be taken into account (Cochran 1977, ch. 5A.2 and 5A.9; Magnussen et al. 2015; Tipton et al. 2013).

It is evident that the proposed MDS and estimators would apply equally to distributions of other quantitative continuous forest inventory variables (e.g. tree height, basal area, stem volume, and biomass).

5 Conclusion

A multidimensional scaling provides a multivariate linear link between two relative frequency distributions and is therefore ideally suited for both a design-based inference of tree size distributions within a stratum and stand-level predictions.

Acknowledgments We thank the Savoie General Council (CG73) for the supply of LiDAR data from Aillon.

Two referees and the editor prompted significant improvements to an earlier draft of our submission.

Compliance with ethical standards

Funding Part of this work has received financial support from the French National Research Agency (ANR) (Foresee project: ANR-10-BIOE-08-07) as well as support under the program ‘Investissements d’avenir’ (ANR-11-LabX-002-01, laboratoire d’Excellence ARBRE).

References

- Andersen H-E, Strunk J, Temesgen H, Atwood D, Winterberger K (2011) Using multilevel remote sensing and ground data to estimate forest biomass resources in remote regions: a case study in the boreal forests of interior Alaska. *Can J Rem Sens* 37:1–16
- Anderson TW, Darling DA (1952) Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann Math Stat* 23: 193–212
- Baffetta F, Fattorini L, Franceschii S, Corona P (2009) Design-based approach to the kNN technique for coupling field and remotely sensed data in forest surveys. *Remote Sens Environ* 113:463–475
- Bailey RL (1980) Individual tree growth derived from diameter distribution models. *For Sci* 26:626–632
- Bickel PJ, Krieger AM (1989) Confidence bands for a distribution function using the bootstrap. *J Am Stat Assoc* 84:95–100
- Binder DA (1983) On the variances of asymptotically normal estimators from complex surveys. *Int Stat Rev* 51:279–292. doi:10.2307/1402588
- Bollandsås OM, Maltamo M, Gobakken T, Næsset E (2013) Comparing parametric and non-parametric modelling of diameter distributions on independent data using airborne laser scanning in a boreal conifer forest. *Forestry* 86:493–501
- Bouvier M, Durrieu S, Fournier RA, Renaud J-P (2015) Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. *Remote Sens Environ* 156: 322–334
- Brosofske KD, Froese RE, Falkowski MJ, Banskota A (2014) A review of methods for mapping and prediction of inventory attributes for operational forest management. *For Sci* 60:733–756. doi:10.5849/forsci.12-134
- Chambers RL, Clark RG (2012) An introduction to model-based survey sampling with applications. Oxford statistical science series, vol 37. Oxford University Press, New York, p. pp 265
- Chen WJ (2004) Tree size distribution functions of four boreal forest types for biomass mapping. *For Sci* 50:436–449
- Clutter JL, Fortson JC, Pienaar LV, Brister GH, Bailey RL (1983) Timber management: a quantitative approach. Wiley, New York, p. pp 333
- Cochran WG (1977) Sampling techniques. Wiley, New York, p. pp 380
- Conover WJ (1980) Practical nonparametric statistics. Wiley, New York, p. 511
- Dean N, Pagano M (2015) Evaluating confidence interval methods for binomial proportions in clustered surveys. *J Surv Statist Meth* 3: 484–503. doi:10.1093/jssam/smv024
- del Río M et al. (2016) Characterization of the structure, dynamics, and productivity of mixed-species stands: review and perspectives. *Eur J For Res* 135:23–49
- Draper NR, Smith H (2014) Applied regression analysis, 3 edn. Wiley, New York
- Duplat P, Perrotte G (1981) Inventory and estimation of the growth of forest stands. Office National des Forêts, Paris, p. 432
- Ene L, Næsset E, Gobakken T (2012) Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. *Int J Remote Sens* 33:5171–5193. doi:10.1080/01431161.2012.657363
- Freedman D, Diaconis P (1981) On the histogram as a density estimator: L 2 theory. *Probab Theory Relat Fields* 57:453–476
- Fuller WA (2009) Sampling statistics. Wiley, New York, p. 454
- Gasparini P, Di Cosmo L, Cenni E, Pompei E, Ferretti M (2013) Towards the harmonization between National Forest Inventory and Forest Condition Monitoring. Consistency of plot allocation and effect of tree selection methods on sample statistics in Italy. *Environ Monit Assess*:6155–6171
- Gobakken T, Næsset E (2005) Weibull and percentile models for lidar-based estimation of basal area distribution. *Scand J For Res* 20: 490–502
- Gregoire TG (1998) Design-based and model-based inference in survey sampling: appreciating the difference. *Can J For Res* 28:1429–1447
- Gregoire TG, Valentine HT (2008) Sampling strategies for natural resources and the environment. Chapman & Hall/CRC, Boca Raton, p. 465
- Gupta S, Weinacker H, Koch B (2010) Comparative analysis of clustering-based approaches for 3-D single tree detection using airborne fullwave lidar data. *Rem Sens* 2:968–989
- Heurich M (2008) Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *For Ecol Manag* 255:2416–2433
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70
- Hyypä J, Hyypä H, Leckie D, Gougeon F, Yu X, Maltamo M (2008) Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *Int J Remote Sens* 29:1339–1366
- Hyttiäinen K, Haight RG (2012) Optimizing continuous forest cover management. In: Pukkala T, Gadow KV (eds) Continuous forest cover management. Springer, Dordrecht, pp. p 195–p 228. doi:10.1007/978-94-007-2202-6
- Kangas A, Kurttila M, Hujala T, Eyvindson K, Kangas J (2015) Forest Management Planning. In: Decision Support for Forest Management. Springer, p 11–p 21.
- Kankare V, Liang X, Vastaranta M, Yu X, Holopainen M, Hyypä J (2015) Diameter distribution estimation with laser scanning based multisource single tree inventory. *ISPRS J Photogr Rem Sens* 108: 161–171
- King B, Madansky A (2013) On sampling design issues when dealing with zeros. *J Surv Statist Meth* 1:144–170. doi:10.1093/jssam/smt006

- Korn EL, Graubard BI (1998) Confidence intervals for proportions with small expected number of positive counts estimated from survey data. *Surv Meth* 24:193–201
- Lehtonen R, Veijanen A (2009) Design-based methods of estimation for domains and small areas. In: Rao CR (ed) *Handbook of statistics*, vol Volume 29, Part B. Elsevier, p 219–p 249. doi:10.1016/S0169-7161(09)00231-4
- Lindberg E, Holmgren J, Olofsson K, Wallerman J, Olsson H (2013) Estimation of tree lists from airborne laser scanning using tree model clustering and k-MSN imputation. *Rem Sens* 5:1932–1955
- Magnussen S (1986) Diameter distributions in *Picea abies* described by the Weibull model. *Scand J For Res* 1:493–502
- Magnussen S (2013) An assessment of three variance estimators for the k-nearest neighbour technique. *Silv Fenn* 47:1–19
- Magnussen S (2015) Arguments for a model based inference? *Forestry* (Oxford) 88:317–325. doi:10.1093/forestry/cpv002
- Magnussen S, Boudewyn P (1998) Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Can J For Res* 28:1016–1031
- Magnussen S, Andersen H-E, Mundhenk P (2015) A second look at endogenous poststratification. *For Sci* 61:624–634
- Magnussen S, Eggermont P, LaRiccia VN (1999) Recovering tree heights from airborne laser scanner data. *For Sci* 45:407–422
- Magnussen S, Næsset E, Gobakken T (2013) Prediction of tree-size distributions and inventory variables from cumulants of canopy height distributions. *Forestry: an international. J For Sci* 86:585–593. doi:10.1093/forestry/cpt022
- Maltamo M, Gobakken T (2014) Predicting tree diameter distributions. In: *Forestry applications of airborne laser scanning*. Springer, p 177–191.
- Maltamo M, Eerikainen K, Pitkanen J, Hyyppä J, Vehmas M (2004) Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sens Environ* 90:319–330
- Maltamo M, Malinen J, Packalén P, Suvanto A, Kangas J (2006) Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Can J For Res* 36:426–436
- Maltamo M, Mehtätalo L, Vauhkonen J, Packalén P (2012) Predicting and calibrating tree attributes by means of airborne laser scanning and field measurements. *Can J For Res* 42:1896–1907
- Maltamo M, Suvanto A, Packalén P (2007) Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. *For Ecol Manag* 247:26–34
- Mandallaz D (2008) *Sampling techniques for forest inventories*. Chapman and Hall, Boca Raton, p. 251
- Mandallaz D, Breschan J, Hill A (2013) New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based Monte Carlo approach with applications to small-area estimation. *Can J For Res* 43:1023–1031. doi:10.1139/cjfr-2013-0181
- Mardia KV, Kent JT, Bibby JM (1979) *Multivariate analysis*. Academic Press, p. 524
- Moser JJ (1976) Specification of density for the inverse J-shaped diameter distribution. *For Sci* 22:177–180
- Næsset E (1997) Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sens Environ* 61:246–253. doi:10.1016/S0034-4257(97)00041-2
- Næsset E (2004) Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scand J For Res* 19:164–179
- Saborowski J, Marx A, Nagel J, Böckmann T (2010) Double sampling for stratification in periodic inventories—infinite population approach. *For Ecol Manag* 260:1886–1895. doi:10.1016/j.foreco.2010.08.035
- Särndal C-E (2011) Combined inference in survey sampling. *Pak J Stat* 27:359–370
- Särndal CE, Swensson B, Wretman J (1992) *Model assisted survey sampling*. Springer Series in Statistics. Springer-Verlag, New York, p. 694
- Searle SR (1982) *Matrix algebra useful for statistics*. Wiley, New York, p. 438
- Sison CP, Glaz J (1995) Simultaneous confidence intervals and sample size determination for multinomial proportions. *J Am Stat Assoc* 90:366–369
- Ståhl G et al. (2016) Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *For. Ecosyst* 3:5
- Thomas V, Oliver RD, Lim K, Woods M (2008) LiDAR and Weibull modeling of diameter and basal area. *For Chron* 84:866–875
- Tipton J, Opsomer J, Moisen G (2013) Properties of endogenous post-stratified estimation using remote sensing data. *Remote Sens Environ* 139:130–137
- Tomppo E, Olsson H, Stahl G, Nilsson M, Hagner O, Katila M (2008) Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens Environ* 112:1982–1999
- Valencia V, Naeem S, García-Barrios L, West P, Sterling EJ (2016) Conservation of tree species of late succession and conservation concern in coffee agroforestry systems. *Agric Ecosyst Environ* 219:32–41
- Vauhkonen J, Maltamo M, McRoberts R, Næsset E (2014a) Introduction to forestry applications of airborne laser scanning. In: Maltamo M, Næsset E, Vauhkonen J (eds) *Forestry applications of airborne laser scanning, Managing Forest Ecosystems*, vol 27. Springer, Dordrecht, NL, pp. 1–18
- Vauhkonen J, Packalén P, Malinen J, Pitkanen J, Maltamo M (2014b) Airborne laser scanning-based decision support for wood procurement planning. *Scand J For Res* 29:132–143
- Véga C, Renaud J-P, Durrieu S, Bouvier M (2016) On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters. *Remote Sens Environ* 175:32–42
- von Lüpke N, Hansen J, Saborowski J (2012) A three-phase sampling procedure for continuous forest inventory with partial re-measurement and updating of terrestrial sample plots. *Eur J For Res* 131:1979–1990. doi:10.1007/s10342-012-0648-z
- Wallerman J, Holmgren J (2007) Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data. *Remote Sens Environ* 110:501–508
- Xu Q, Hou Z, Maltamo M, Tokola T (2014) Calibration of area based diameter distribution with individual tree based diameter estimates using airborne laser scanning. *ISPRS J Photogr Rem Sens* 93:65–75