DISCUSSION

Open Access

CrossMark

Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation

Göran Ståhl[1], Svetlana Saarela[1*] iD, Sebastian Schnell[1], Sören Holm[1], Johannes Breidenbach[2], Sean P. Healey[3], Paul L. Patterson[3], Steen Magnussen[4], Erik Næsset[5], Ronald E. McRoberts[3] and Timothy G. Gregoire[6]

Abstract

This paper focuses on the use of models for increasing the precision of estimators in large-area forest surveys. It is motivated by the increasing availability of remotely sensed data, which facilitates the development of models predicting the variables of interest in forest surveys. We present, review and compare three different estimation frameworks where models play a core role: model-assisted, model-based, and hybrid estimation. The first two are well known, whereas the third has only recently been introduced in forest surveys. Hybrid inference mixes design-based and model-based inference, since it relies on a probability sample of auxiliary data and a model predicting the target variable from the auxiliary data..We review studies on large-area forest surveys based on model-assisted, model-based, and hybrid estimation, and discuss advantages and disadvantages of the approaches. We conclude that no general recommendations can be made about whether model-assisted, model-based, or hybrid estimation should be preferred. The choice depends on the objective of the survey and the possibilities to acquire appropriate field and remotely sensed data. We also conclude that modelling approaches can only be successfully applied for estimating target variables such as growing stock volume or biomass, which are adequately related to commonly available remotely sensed data, and thus purely field based surveys remain important for several important forest parameters.

Keywords: Design-based inference, Model-assisted estimation, Model-based inference, Hybrid inference, National forest inventory, Remote sensing, Sampling

Introduction

Use of models in large-area surveys of forests is attracting increased interest. The reason is the improved availability of auxiliary data from various remote sensing platforms. Aerial photographs (e.g., Næsset 2002a, Bohlin et al. 2012) and optical satellite data (e.g., Reese et al. 2002) have been available and used operationally for many decades, while data from profiling (e.g., Nelson *et al.* 1984, Nelson et al. 1988) and scanning lasers (e.g., Næsset 1997) and radars (Solberg et al. 2010) have become available for practical applications more recently. Some of the new types of remotely sensed data, such as data from laser scanners, have already become widely applied in forest inventories (e.g., Næsset 2002b). A common application involves the development of models that are applied wall-to-wall over an area of interest (e.g., Næsset 2004), often for providing data for forest management. However, this type of data is increasingly applied also in connection with large-area forest surveys, such as national-level forest inventories (Tomppo et al. 2010, Asner et al. 2012).

Applications of models in large-area forest surveys often use the model-assisted estimation framework (Särndal et al. 1992) where a model is used to support the estimation following probability sampling within the context of design-based inference (Gregoire 1998). Importantly, an inadequately specified model will not make the estimators biased in this case, but only affect the variance of the estimators. Examples of large-area forest inventory applications include Andersen et al. (2011) who applied the technique in Alaska, Gregoire et al. (2011) and Gobakken et al. (2012), who applied it in Hedmark County, Norway, and Saarela et al. (2015a) who used it in Kuortane, Finland.

* Correspondence: svetlana.saarela@slu.se
[1]Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden
Full list of author information is available at the end of the article

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 2 of 11

Some applications of models in large-area forest surveys involve model-based inference (Gregoire 1998), which to a larger extent than model-assisted estimation relies on model assumptions. In this case an inadequately specified model might make the estimators both biased and imprecise. On the other hand, with accurate models this mode of inference can be very efficient (e.g., Magnussen 2015). Examples of applications in forest inventory include McRoberts (2006, 2010), who used model-based inference for estimating forest area based on Landsat data in northern Minnesota, U.S.A., Ståhl et al. (2011) who used it for estimating biomass in Hedmark, Norway, using laser data, and Healey et al. (2012) who applied the technique in California, U.S.A., using data from the space-borne Geoscience Laser Altimeter System (GLAS).

Non-parametric modelling, applying methods such as the *k*-Nearest Neighbours (kNN) technique (Tomppo and Katila 1991, Tomppo et al. 2008), has a long tradition in forest inventories. These techniques typically have been applied for providing small-area estimates through combining field sample plots and various sources of remotely sensed data. However, the kNN technique has also been used in connection with model-assisted estimation (e.g., Baffetta et al. 2009, 2011, Magnussen and Tomppo 2015) and model-based inference (e.g., McRoberts et al. 2007).

The objective of this paper was to present, review and discuss how models are applied in the case of model-assisted and model-based estimation in large-area forest surveys, and to discuss advantages and disadvantages of the two estimation frameworks in this context. We also present, review and discuss a newly introduced estimation framework where probability sampling is applied for the selection of auxiliary data, upon which model-based inference is applied in a second phase. This framework in denoted hybrid inference, after Corona et al. (2014).

We restrict the study to large-area estimation. This is the case of national forest inventories and greenhouse gas inventories under the United Nations Framework Convention on Climate Change (e.g., Tomppo et al. 2010). Importantly, in this case there is no need to make assumptions about residual error terms linked to individual population elements, which is a core issue in model-based small-area estimation (e.g., Breidenbach and Astrup 2012, Breidenbach et al. 2015). The reason is that the residual error terms will have almost no influence on the results, as will be demonstrated below. However, we do not specify how large a "large area" must be, but use the term as a general concept.

Below, we present the basics of model-assisted, model-based, and hybrid inference (chapter 2). Subsequently we present a brief review of the application of these methods in forest surveys (chapter 3), and, finally, we

discuss advantages and disadvantages of the different approaches and make conclusions (chapters 4 and 5).

Basics of model-assisted, model-based and hybrid estimation

In this chapter we summarize some basic concepts related to the use of models in large-area forest surveys. We restrict the scope to cases where models are applied for improving estimators (or predictors) once sample or wall-to-wall data have been collected. However, models may also be used in the design phase for improving the sample selection (e.g., Fattorini et al. 2009, Grafström et al. 2014), but such cases are not covered in this article.

Design-based inference
This paper requires a basic understanding of the concepts design-based and model-based inference (e.g., Cassel et al. 1977, Särndal 1978, Gregoire 1998, McRoberts 2010).

Design-based inference typically assumes a finite population of elements to which one or more fixed target quantities are linked. The objective normally is to estimate some fixed population parameter, such as the total or the mean of these quantities (e.g., Gregoire and Valentine 2008). In order to estimate the fixed but unknown parameters a probability sample is selected from the population according to some appropriate sampling design, which assigns positive inclusion probabilities to each element. Mathematical formulas (estimators) are used for estimating the parameters based on the sample data. The estimates are random variables due to the random selection of samples, i.e., the estimators produce different values depending on which population elements are included in the sample.

The Horvitz-Thompson estimator can be applied to any probability sampling design with inclusion probabilities known at least for the sampled units (e.g., Särndal et al. 1992). Using this estimator, a population total, τ, is estimated as

$$\hat{\tau} = \sum_{i \in s} \frac{y_i}{\pi_i} \tag{1}$$

Here, y_i is the variable of interest for the *i:th* sampled element, π_i is the inclusion probability, and s is the sample.

The precision of an estimator is usually expressed through its variance, which is a fixed quantity given the population, the design, and the estimator. The variance usually can be estimated through a variance estimator, and confidence intervals can be computed as a means to provide decision makers with the range of values wherein the true population parameter is located with a defined probability.

In case of the Horvitz-Thompson estimator, a general formula for the variance is

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 3 of 11

$$var(\hat{\tau}) = \sum_{i \in U} \sum_{j \in U} (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \qquad (2)$$

In addition to the previously introduced notation, π_{ij} is the joint probability of inclusion for unit i and j. The step from the variance to a variance estimator and a confidence interval normally is straightforward (e.g., Gregoire and Valentine 2008).

Some key features of design-based inference are:

- The values that are linked to the population elements are fixed
- The population parameters about which we wish to infer information are also fixed
- Our estimators of the parameters are random because a probability sample is selected according to some sampling design, such as simple random sampling
- The probability of obtaining different samples can be deduced from the design and used for inference

The foundations of design-based inference were laid out by Neyman (1934) and it is the standard mode of inference in most statistical surveys, including sample-based national forest inventories (Tomppo et al. 2010) that are carried out in a large number of countries.

Design-based inference through model-assisted estimation

Models can be used to improve estimators under the design-based framework. An important category of such estimators are known as model-assisted estimators (Särndal et al. 1992). The general form of such estimators, for estimating a population total, is

$$\hat{\tau}_{ma} = \sum_{i \in U} \hat{y}_i + \sum_{i \in s} \frac{(y_i - \hat{y}_i)}{\pi_i} \qquad (3)$$

where the first part of the estimator is a sum of model estimates of each element in the population; the second term is a Horvitz-Thompson estimator of the total of the deviations between observed values and values estimated by the model; the subscript '*ma*' is used to point out that the estimator is model-assisted. Thus, the model-assisted estimator can be seen as composed of a first crude estimator which is refined through a correction term that makes it asymptotically unbiased when the model is external (in which case Eq. 3 is often referred to as a difference estimator), and approximately unbiased when the model is internal (in which case Eq. 3 is often referred to as a generalised regression estimator). In case the model is external the variance is

$$var(\hat{\tau}_{ma}) = \sum_{i \in U} \sum_{j \in U} (\pi_{ij} - \pi_i \pi_j) \frac{e_i}{\pi_i} \frac{e_j}{\pi_j} \qquad (4)$$

This is almost the same expression as the variance in Eq. (2), but the y_i- terms have been replaced by $e_i = y_i - \hat{y}_i$. If an accurate model is used the latter terms should be much smaller than the former, and thus the variance of the model-assisted estimator should be much smaller than the variance of the ordinary Horvitz-Thompson estimator, although this is not immediately clear when comparing Eq. 2 and Eq. 4.

Model-based inference

In contrast to design-based inference (including model-assisted estimators), a basic assumption underlying model-based inference is that the values that are linked to the elements in the population are realizations of random variables. As a consequence, target survey quantities such as population totals and means are also random variables. Thus, due to the different points of view underlying design-based and model-based inference some caution must be exercised when comparing results from the two inferential frameworks. For example, with model-based inference the random population total (or mean) may be predicted or (as in this study) the expected value of the population total may be estimated. For large population the difference between these two quantities, in relative terms, typically is minor although for small populations the relative difference may be substantial. However, just like design-based inference, model-based inference in many cases is a useful and straightforward approach for quantifying target features of a population (e.g., Chambers and Clark 2012). In forest inventories, examples of such cases are surveys of remote areas with poor road infrastructure and small-area estimation for forest management. In both cases the field sample sizes typically are small or acquired through non-probability sampling whereas remotely sensed data are available wall-to-wall.

A basic assumption of model-based inference is that the random values of the population elements follow some specific model, e.g., a model based on auxiliary data derived from remote sensing. Thus, in the standard case, auxiliary data are available for all population elements. A simple and fairly general example is the linear model, i.e., (in matrix form)

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \qquad (5)$$

where $\boldsymbol{Y}$ is an $N \times 1$ matrix of the target variable, $\mathbf{X}$ an $N \times p$ matrix of auxiliary data, $\boldsymbol{\beta}$ is a $p \times 1$ matrix of model parameters, and an $N \times 1$ matrix of random variables that follow some joint probability distribution; N is the population size; in a forest survey it might be the number of grid cells which tessellate the study area.

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 4 of 11

Our objective typically is to predict a random population quantity, e.g., the mean or the total, following the selection of a sample for estimating the model parameters. Regardless of how the sample is selected, the observations are realizations of random variables due to the model assumptions. Once the model parameters are estimated, we can use the estimated model, $\hat{Y} = X\hat{\beta}$, for predicting the population quantities of interest based on the auxiliary data; in standard cases these are assumed available for all population elements. Introducing $\mathbf{1}$ as an $N \times 1$ vector of "1"-entries, the random population total $\tau^* = \mathbf{1}'Y = \mathbf{1}'X\beta + \mathbf{1}'\varepsilon$ may be predicted as

$$\hat{\tau}^* = \mathbf{1}'\hat{Y} \qquad (6)$$

Note the distinction in nomenclature between *estimating* a fixed but unknown value (a population parameter) and *predicting* a random variable (e.g., Särndal 1978, Gregoire 1998). Note also that some authors (Chambers and Clark 2012) present the model-based predictor as a sum of two terms: the sum of the values of the sampled elements and the sum of the predictions for the non-sampled elements. The difference between such a predictor and Eq. (6) would, however, be very small in case a small sample is selected from a large population.

Turning to the mean square error of the predictor in Eq. (6) we need to acknowledge that uncertainty is introduced both by the estimation of the model parameters and by the random residual terms linked to each population element. Since the residuals may often be spatially auto-correlated estimating the mean square error of the Eq. (6) predictor may be very complicated.

However, an important feature of large-area surveys is that the relative difference between τ^* and E (τ^*) typically is very small (e.g., Chambers and Clark 2012, p. 16). The relative difference is $\mathbf{1}'\varepsilon/(\mathbf{1}'X\beta + \mathbf{1}'\varepsilon)$, which intuitively can be seen to tend to zero as N tends to infinity, since in the cases we focus on the $X_i\beta$ -terms are almost always positive and typically much larger (in absolute value) than the residual terms, which may be either negative or positive. Thus, instead of predicting τ^*, in large-area estimation we can estimate E (τ^*), which simplifies the model-based inference. The estimator will be identical to Eq. (6), i.e., $E(\hat{\tau}^*) = \mathbf{1}'\hat{Y}$, but it is now an estimator rather than a predictor. The variance (due to the model) of this estimator is simpler to derive, since it does not involve any residual terms; thus uncertainty in this case is introduced only through the model parameter estimation.

The variance of the estimator of E (τ^*) is

$$var\left(E[\hat{\tau}^*]\right) = \mathbf{1}'Xcov\left(\hat{\beta}\right)X'\mathbf{1} \qquad (7)$$

The matrix $cov\left(\hat{\beta}\right)$ is the variance-covariance matrix of the model parameter estimates. A variance estimator is obtained by inserting the estimated covariance matrix in Eq. (7).

Thus, some key features of model-based inference are:

- The values linked to population elements are random variables
- Since the individual values are random variables so is the population total or mean that we wish to predict
- A model for the relationship between the target variable and one or more auxiliary variable(s) can adequately conform to the trend in Y.
- Auxiliary data are commonly available for all population elements
- After having selected a sample – that need not be random – for estimating the model parameters, we apply the fitted model for predicting the target population quantity or estimating the expected value of this quantity.

Hybrid inference: a special case of model-based inference
Auxiliary data may not be available prior to a forest survey and they may be very expensive to collect for all units in a population, as required for standard application of model-based inference. In such cases a probability sample of auxiliary data can be acquired, based on which the population total or mean of the auxiliary variable is estimated following design-based inference. A model can still be specified and applied regarding the relationship between the study variable and the auxiliary variables, and thus model-based inference can be applied once the auxiliary variable totals (or means) have been estimated through design-based inference.

Thus, design-based principles are applied in a first phase and model-based principles in a second phase. This approach was termed *hybrid* inference by Corona et al. (2014) and in the present paper we follow that terminology. In a previous study by Mandallaz (2013) it was called *pseudo-synthetic* estimation. In a study by Ståhl et al. (2011) it was simply called model-based inference, although later denoted *model-dependent* estimation by Gobakken et al. (2012). However, the term model-dependent estimation appears to have been first proposed by Hansen et al. (1978, 1983) to include all sampling strategies that depend on the correctness of a model; according to Hansen et al. (1978) "a model-dependent design consists of a sampling plan and estimators for which either the plan or the estimators, or both, are chosen because they have desirable properties

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 5 of 11

under an assumed model, and for which the validity of inferences about the population depends on the degree to which the population conforms to the assumed model." Thus, standard model-based inference as well as hybrid inference, and other approaches, belong to Hansen's model-dependent category.

In the case of hybrid inference, expected values and variances are derived by considering both the design through which auxiliary data were collected and the model used for predicting values of population elements based on the auxiliary data. Thus, assuming we use a linear model, a general estimator of $E(\tau^*)$ is given as

$$E(\hat{\tau^*}) = \sum_{i \in s} \frac{X_i \hat{\beta}}{\pi_i} = \boldsymbol{\pi}' X \hat{\boldsymbol{\beta}} \qquad (8)$$

where s is the sample of auxiliary data, π_i is the probability of including population element i into the auxiliary data sample, $\boldsymbol{\pi}$ is an n-length column vector of $(1/\pi_i)$ – values, and X is an $n \times p$ matrix of sampled auxiliary data. The model parameters are estimated from a sample that is assumed to be independent from the sample of auxiliary data.

In deriving the variance of the estimator in Eq. (8), note that the part $\boldsymbol{\pi}' X$ of the estimator is a $1 \times p$ matrix of design-unbiased estimators of population totals of auxiliary data, which we denote $\hat{\tau}_x$. This matrix is multiplied by the matrix of estimated model parameters, i.e., the result is a sum of estimated population totals of auxiliary variables times the corresponding model parameter estimate, such as $\hat{\tau}_{Xj} \cdot \hat{\beta}_j$. In each term the two components are independent, but the estimators of the auxiliary variable totals as well as the estimators of the parameters are typically correlated. Thus, the variance (due to the sample and the model) is

$$var\left(E[\hat{\tau^*}]\right) = var\left(\hat{\boldsymbol{\tau}}_x \hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}' cov(\hat{\boldsymbol{\tau}}_x) \boldsymbol{\beta}$$
$$+ \ \boldsymbol{\tau}_x cov\left(\hat{\boldsymbol{\beta}}\right) \boldsymbol{\tau}_x' + \ Tr\left(cov(\hat{\boldsymbol{\tau}}_x) cov\left(\hat{\boldsymbol{\beta}}\right)\right) \qquad (9)$$

where $cov(\hat{\boldsymbol{\tau}}_x)$ is the covariance matrix of the estimators of the auxiliary variable totals and $cov\left(\hat{\boldsymbol{\beta}}\right)$ is the covariance matrix of the model parameter estimators. The *Tr*-operator is the trace, i.e., the sum of the diagonal entries in the matrix. The diagonal entries in $cov(\hat{\boldsymbol{\tau}}_x)$ are of the kind presented in Eq. (2). The off-diagonal entries are computed in a similar fashion (Särndal et al. 1992). The covariance matrix of the model parameter estimators normally, under ordinary least squares regression assumptions, is derived as $\sigma^2(X'X)^{-1}$ where σ^2 is the residual variance, given the regression model. In case of heteroskedastic residual variance, alternative estimators can be applied (e.g., Saarela et al. 2015b). We do not offer a proof of Eq. (9), but readers familiar with the

variance of a product of two independent random variables (i.e., $var(WZ) = E(W)^2 var(Z) + E(Z)^2 var(W) + var(W)var(Z)$) can identify the similarity with Eq. (9).

Although it seems likely that hybrid type estimators have been applied outside forest inventories, we have not yet found any description of them in non-forest publications.

In Fig. 1 an overview of the "positions" of standard design-based estimation (without using models), model-assisted estimation, hybrid estimation, and model-based estimation is shown with regard to how much these estimation techniques rely on (i) the correctness of the model and (ii) the use of probability sampling.

A brief review of the use of models in large-area forest surveys

From the methods section it is clear that models can be used in several ways for improving the estimation of target quantities in large-area forests surveys. Our review is separated into the following cases:

- Use of models in the context of design-based inference through model-assisted estimation
- Use of models in the context of model-based inference through model-based estimation
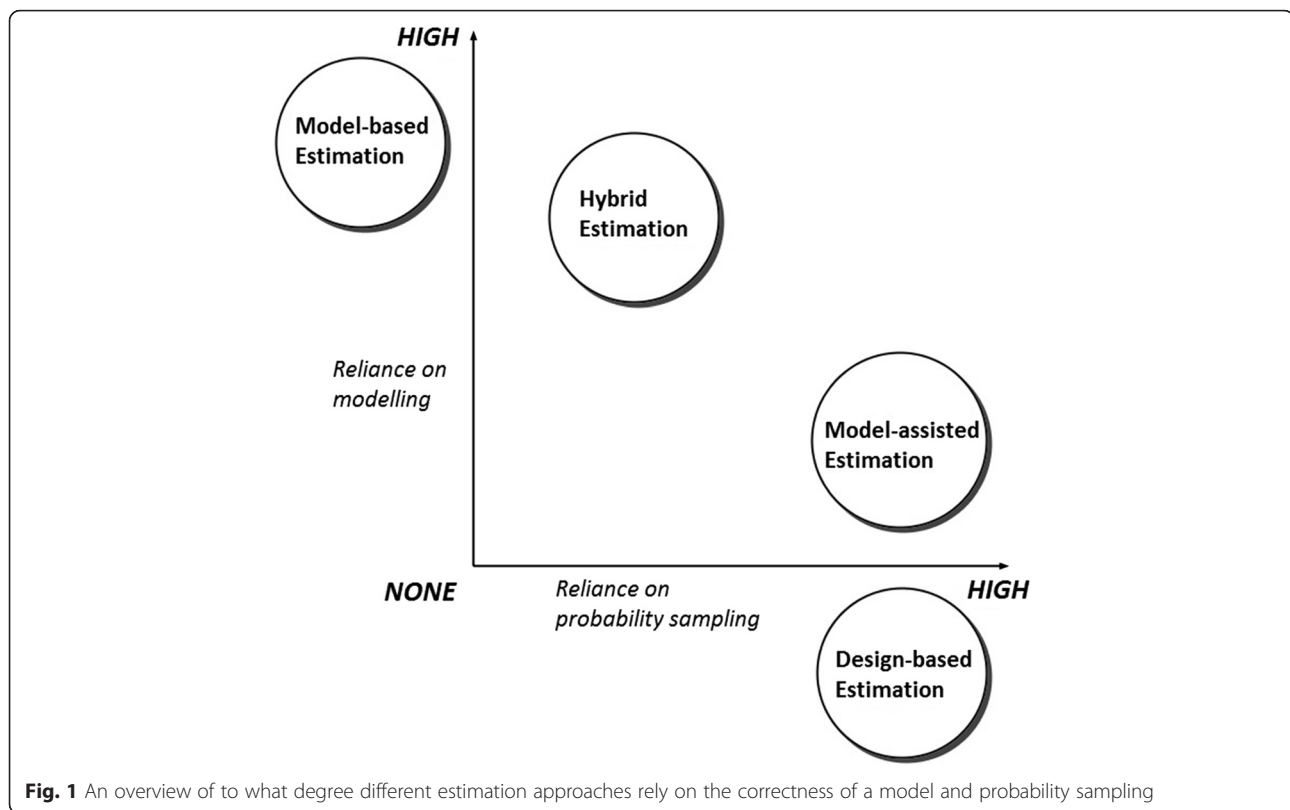- Use of models in the context of hybrid inference

Model-assisted estimation in large-area forest surveys

Formal model-assisted estimators appear to be fairly recently introduced to large-area forest surveys, although standard regression estimators (i.e., a simple kind of model-assisted estimators) have been applied in forest surveys for a long time. An important example of the latter kind is the Swiss national forest inventory (Köhl and Brassel 2001) where air photo interpretation has been combined with field surveys for a long time and the Italian national forest inventory, where a three-phase sampling approach is applied (Fattorini et al. 2006).

An early model-assisted study was conducted by Breidt et al. (2005), who used spline models in estimating population totals in a simulation study linked to surveys of forest health. Model-assisted estimation was found to perform well in the context of a two-phase survey with multiple auxiliary variables.

Opsomer et al. (2007) used model-assisted estimation in a two-phase systematic sampling design, applying generalized additive models linking ground measurements with auxiliary information from remote sensing. The study was an extension of the study by Breidt and Opsomer (2000), where univariate models and a single-phase sampling strategy were applied.

In Boudreau et al. (2008), model-assisted estimation was used for estimating biomass in Quebec, Canada, based on data from a laser profiler, GLAS satellite data,

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 6 of 11

Fig. 1 An overview of to what degree different estimation approaches rely on the correctness of a model and probability sampling

and land cover maps based on data from Landsat-7 ETM+. The study demonstrated that GLAS data could improve large-scale monitoring of aboveground biomass at large spatial scales; however, the presented estimators were not denoted "model-assisted". Nelson et al. (2009) built upon the study by Boudreau et al. (2008) and introduced some new, partly model-based, estimation techniques. Andersen et al. (2009) presented a study based on model-assisted estimation where the biomass of western Kenai, Alaska, was estimated based on samples of field and laser scanner data.

In Gregoire et al. (2011) model-assisted estimation was used for estimating aboveground biomass in Hedmark County, Norway, using sample data from laser profilers and scanners. The study triggered the start of a series of studies where the model-assisted theory, developed by Särndal et al. (1992), was applied for large-scale forest surveys based on samples of laser scanner data. Næsset et al. (2011) applied and compared two sources of auxiliary information, laser scanner data and interferometric synthetic aperture radar data for model-assisted estimation of biomass over a large boreal forest area in the Aurskog-Høland municipality in Norway and quantified to what extent the two types of auxiliary data improved the estimated precision. Gobakken et al. (2012) compared the performance of model-assisted estimation with model-based prediction of aboveground biomass in

Hedmark County, Norway using data from airborne laser scanning as auxiliary data. The two approaches were found to yield similar results. Nelson et al. (2012) conducted a similar study over the same area using data from a profiling rather than scanning airborne laser, while Næsset et al. (2013b) evaluated the precision of the two-stage model-assisted estimation conducted by Gobakken et al. (2012). The authors noted the sensitivity of variance estimators to unequal sample strip length and systematically selected strips. The latter issue was further pursued by Ene et al. (2012), who showed that the variance was often severely overestimated when estimators assuming simple random sampling were applied in this context. Similar results were reported by Magnussen et al. (2014).

Strunk et al. (2012a, 2012b) investigated different aspects of model-assisted estimation. For example, the authors found that the laser pulse density had almost no effect on the precision of model-assisted estimators of core parameters, such as basal area, volume, and biomass.

Saarela et al. (2015a) proposed to use probability-proportional-to-size sampling of laser scanning strips in a two-phase model-assisted sampling study where the total growing stock volume was estimated in a boreal forest area in Kuortane, Finland. It was also found that full cover of Landsat auxiliary information improved the

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 7 of 11

precision of estimators compared to using only sampled LiDAR strip data.

Massey et al. (2014) evaluated the performance of the model-assisted estimation technique in connection with the Swiss national forest inventory. The authors also addressed several methodological issues and, among other things, evaluated the performance of non-parametric methods in connection with model-assisted estimation and the close connection between difference estimators and regression estimators.

As some of the first laser scanning campaigns carried out for inventory purposes at the turn of the millennium have been repeated in recent years, change estimation assisted by laser data have become an important research area. Bollandsås et al. (2013), Næsset et al. (2013a, 2015), Skowronski et al. (2014), McRoberts et al. (2015), and Magnussen et al. (2015) analysed different approaches to modelling of change in biomass, such as separate modelling of biomass at each point in time and then estimate the difference, direct modelling of change with different predictor variables, such as the variables at each time point or their differences, and longitudinal models. These modelling techniques have been combined with different design-based and model-based estimators to produce change estimates and confidence intervals. Sannier et al. (2014) investigated change estimation based on a series of maps, which provided the auxiliary data for model-assisted difference estimation. A comprehensive review and discussion of change estimation can be found in McRoberts et al. (2014, 2015). Melville et al. (2015) evaluated three model-based and three design-based methods for assessing the number of stems using airborne laser scanning data. The authors reported that among the design-based estimators, the most precise estimates were achieved through stratification.

Stephens et al. (2012) applied double sampling regression estimators in the design-based framework for estimating carbon stocks in New Zealand forests using laser data as auxiliary information.

Chirici et al. (2016) compared the performance of two types of airborne LiDAR-based metrics in estimating total aboveground biomass through model-assisted estimators. The study area was located in Molise Region in central Italy. Corona et al. (2015) dealt with the use of map data as auxiliary information in a similar context.

Model-based and hybrid inference in large-area forest surveys

McRoberts (2006, 2010) applied model-based inference for estimating forest area using Landsat data as auxiliary information and field plots data. The studies were performed in northern Minnesota, U.S.A. In the studies the expected value of the total forest area was estimated, as

a means to reduce the complexity of the variance estimators.

A large number of studies have applied model-based prediction for mapping forest attributes across large areas using remotely sensed auxiliary information. Baccini et al. (2008) used moderate resolution imaging spectro-radiometer (MODIS) and GLAS for mapping aboveground biomass across tropical Africa. Armston et al. (2009) used Landsat-5 TM and Landsat-7 ETM+ sensors for prediction foliage projective cover across a large area in Queensland, Australia. Asner et al. (2010) applied model-based prediction for mapping the aboveground carbon stocks using satellite imaging, airborne LiDAR and field plots over 4.3 million ha of Peruvian Amazon. Helmer et al. (2010) used time series from 24 Landsat TM/ETM+ and Advance Land Imager (ALI) scenes for mapping forest attributes on the island of Eleuthera. These are only examples of a very large number of studies where wall-to-wall remotely sensed data have been applied for mapping and monitoring forest resources. However, a majority of these studies do not apply a formal model-based inferential framework. For example, in case the uncertainty of estimators is addressed, usually the strict model-based inference approach [Eq. (7)] is not applied but instead some other, often ad-hoc, method that does not correctly reflect the uncertainty of the estimator or predictor involved.

Saarela et al. (2015b) evaluated the effects of model form and sample size on the precision of model-based estimators in the study area Kuortane, Finland, and identified minor to moderate differences in results when different model forms were applied. In a simulation study, Magnussen (2015) demonstrated the usefulness of model-based inference for forest surveys and argued that this approach has several advantages over traditional design-based sampling. McRoberts et al. (2014a,b) assessed the effects of uncertainty in model predictions of individual tree volume model predictions on large-area volume estimates in the survey framework of hybrid inference.

As previously mentioned, Corona et al. (2014) proposed to use the term hybrid inference for the case where a probability sample of auxiliary data may be selected, on which model-based inference is applied; the study by Corona et al. mainly dealt with small-area estimation issues. Ståhl et al. (2011), Gobakken et al. (2012), Nelson et al. (2012) and Magnussen et al. (2014) used hybrid inference for estimating the forest resources in Hedmark county, Norway, based on combinations of laser scanner data, laser profiler data, and field data. In the study by Magnussen et al. two populations were simulated using the data. Healey et al. (2012) applied the technique in California, using GLAS data. In a study of boreal forests in Canada, Margolis et al. (2015) likewise

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 8 of 11

used GLAS data, in combination with airborne laser data, to estimate aboveground biomass.

Geographical mismatches between remotely sensed data and field measurements may considerably affect the precision of estimators in large-area surveys. The effects of such errors in model-based and model-assisted estimation were evaluated by Saarela et al. (2016).
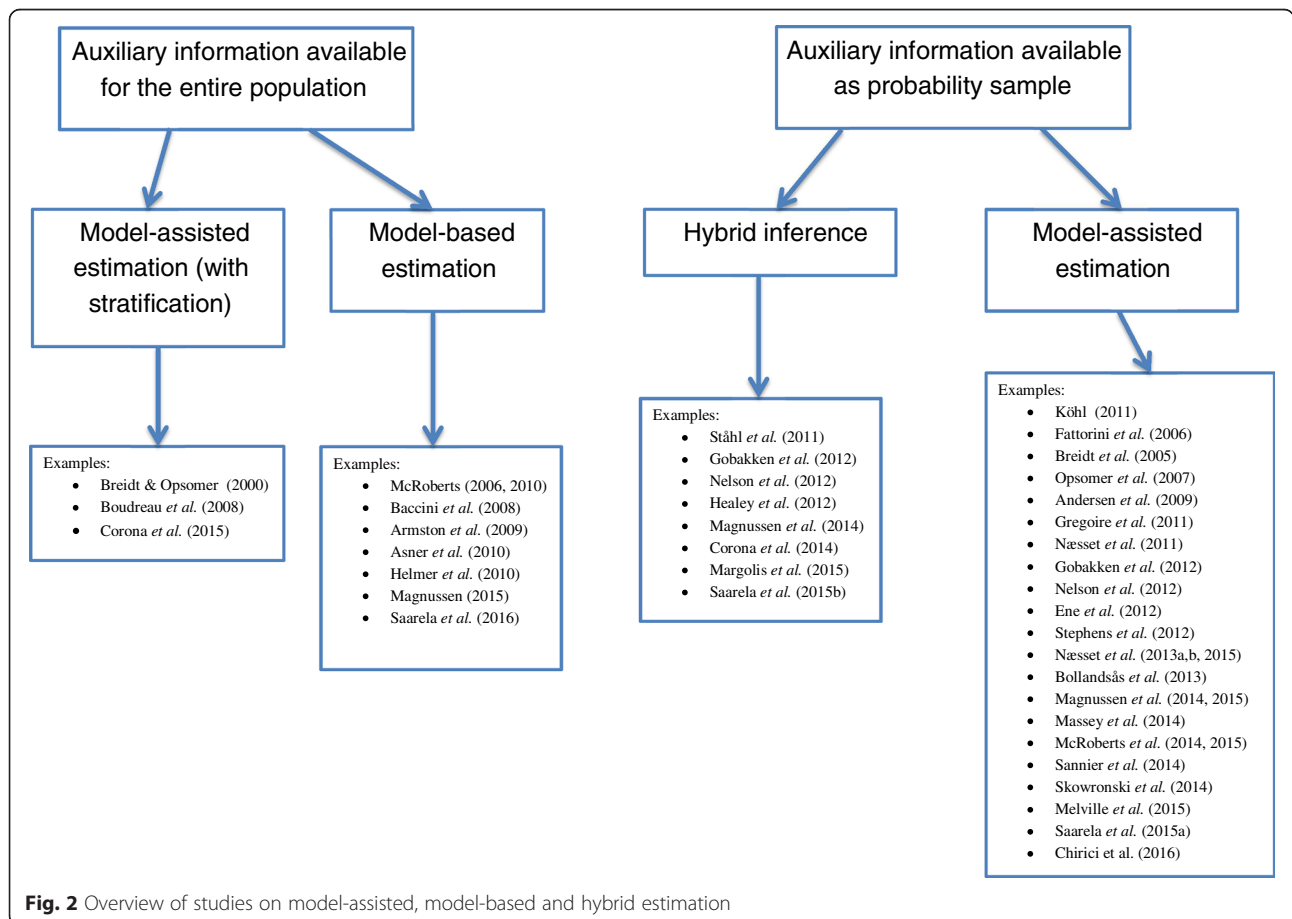
The findings from the brief literature review are summarized in Fig. 2.

Discussion

The review revealed that use of models in large-scale forest inventories is widespread, although statistically strict applications of model-assisted estimators, model-based inference, or hybrid inference are rather limited. While the model-assisted estimation framework is attracting large interest, model-based inference and hybrid inference are not applied as much. A large number of studies apply approaches that could be classified as model-based inference, although they do not pursue any strict uncertainty analyses. In this context there is room for substantial improvement regarding how mean square errors or variances are estimated.

An advantage of model-assisted estimation, as compared to model-based and hybrid inference, is that the unbiasedness of estimators of totals and means do not rely on the correctness of the model; the model is only applied for enhancing a design-based estimator (Särndal et al. 1992). Whereas there is a theoretical chance that a model-assisted estimator is worse (in terms of variance) than a strictly design-based estimator if the model is extremely poor, a well specified model might substantially increase the precision of the model-assisted estimator compared to the strictly design-based estimator. This was shown by, e.g., Ene et al. (2012) and Saarela et al. (2015a).

If well specified models are available model-based inference is definitely a competitive alternative to design-based inference through model-assisted estimation (McRoberts et al. 2014a, b, Magnussen 2015). It has advantages since it does not rely on a probability sample from the target area. Such samples may sometimes not be feasible due to poor infrastructure conditions, restricted access to private land, or the presence of areas that are for some reason dangerous to visit in the field. Further, in case a probability sample has been selected, based upon which models are developed and applied,

Fig. 2 Overview of studies on model-assisted, model-based and hybrid estimation

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 9 of 11

model-based inference and model-assisted estimation usually lead to similar total estimates. In case the condition $\sum_{i\in s}^{n} \frac{(y_i - \hat{y}_i)}{\pi_i} = 0$ holds the estimated values will be identical. However, Saarela et al. (2016) showed that the model-based variance estimators are less prone to problems with geolocation mismatches between field plots and remotely sensed auxiliary data.

Hybrid inference is a straightforward approach in cases where auxiliary data are not available wall-to-wall and such data are expensive to acquire. In such cases a sample of auxiliary data can be selected, upon which the auxiliary variable totals and means can be estimated and used together with model predictions that link the auxiliary variables with the target variable. The approach so far appears to have been applied only in a limited number of forest inventories, although implicitly it has been used for a long time in forest inventories where models (such as volume, biomass and growth models) have been applied based on data from forest plots (Ståhl et al. 2014).

Overall, the use of models relies on auxiliary data that are correlated with or otherwise related with the target variable. Considering the variables normally included in national forest inventories (Tomppo et al. 2010) it is likely that a large number of variables would be very difficult to model in terms of remotely sensed data. This might be the case for forest floor vegetation, soil properties, and several types of forest damage. Modelling approaches linked to such variables would probably not improve the precision of estimators. Thus, a large number of variables, such as site index, forest floor vegetation, soil type, etc., are likely to require probability field samples.

Conclusions

We conclude by noting that all three approaches studied: model-assisted estimation, model-based inference, and hybrid inference, have advantages and disadvantages when applied in large-area forest surveys. A main advantage of model-assisted estimation is that unbiasedness of estimators does not rely on the suitability of the model, but the model only helps to improve the precision of an estimator known to be (approximately) unbiased. Model-based and hybrid inference rely on the suitability of the model, but may have several advantages under conditions where access to field plots is difficult or expensive. All three approaches rely on the possibility to develop accurate models, which is possible for several important forest variables (such as biomass), but not for all variables that are included in a normal national forest inventory.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden. [2]Norwegian Institute for Bioeconomy Research, Ås, Norway. [3]USDA Forest Service, Washington, D.C., USA. [4]Canadian Forest Service, Pacific Forestry Centre, British Columbia, Canada. [5]Norwegian University of Life Sciences, Ås, Norway. [6]School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA.

References
Andersen HE, Barrett T, Winterberger K, Strunk J, Temesgen H (2009) Estimating forest biomass on the western lowlands of the Kenai Peninsula of Alaska using airborne lidar and field plot data in a model-assisted sampling design. In: Proceedings of the IUFRO Division 4 Conference: "Extending Forest Inventory and Monitoring over Space and Time"., pp 19–22

Andersen HE, Strunk J, Temesgen H (2011) Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the Upper Tanana Valley of Interior Alaska. West J Appl Forestry 26:157–164

Armston JD, Denham RJ, Danaher TJ, Scarth PF, Moffiet TN (2009) Prediction and validation of foliage projective cover from Landsat-5 TM and Landsat-7 ETM+ imagery. J Appl Remote Sensing 3:33540–33540, http://dx.doi.org/10.1117/1.3216031

Asner GP, Powell GV, Mascaro J, Knapp DE, Clark JK, Jacobson J, Hughes RF (2010) High-resolution forest carbon stocks and emissions in the Amazon. Proc Natl Acad Sci 107:16738–16742, http://dx.doi.org/10.1073/pnas.1004875107

Asner GP, Mascaro J, Muller-Landau HC, Vieilledent G, Vaudry R, Rasamoelina M, Hall S, van Breugel M (2012) A universal airborne LiDAR approach for tropical forest carbon mapping. Oecologia 168:1147–1160, http://dx.doi.org/10.1007/s00442-011-2165-z

Baccini A, Laporte N, Goetz SJ, Sun M, Dong H (2008) A first map of tropical Africa's above-ground biomass derived from satellite imagery. Environ Res Lett 3:9

Baffetta F, Fattorini L, Franceschi S, Corona P (2009) Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys. Remote Sensing Environ 113(3):463–475, http://dx.doi.org/10.1016/j.rse.2008.06.014

Baffetta F, Corona P, Fattorini L (2011) Design-based diagnostics for k-NN estimators of forest resources. Can J Forest Res 41:59–72

Bohlin J, Wallerman J, Fransson JE (2012) Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. Scand J Forest Res 27:692–699, http://dx.doi.org/10.1080/02827581.2012.686625

Bollandsås OM, Gregoire TG, Næsset E, Øyen B-H (2013) Detection of biomass change in a Norwegian mountain forest area using small footprint airborne laser scanner data. Stat Methods Appl 22:113–129, http://dx.doi.org/10.1007/s10260-012-0220-5

Boudreau J, Nelson RF, Margolis HA, Beaudoin A, Guindon L, Kimes DS (2008) Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing Environ 112:3876–3890, http://dx.doi.org/10.1016/j.rse.2008.06.003

Breidenbach J, Astrup R (2012) Small area estimation of forest attributes in the Norwegian National Forest Inventory. Eur J Forest Res 131:1255–1267, http://dx.doi.org/10.1007/s10342-012-0596-7

Breidenbach J, McRoberts RE, Astrup R (2015) Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sensing Environ (in press). http://dx.doi.org/10.1016/j.rse.2015.07.026

Breidt FJ, Opsomer JD (2000) Local polynomial regression estimators in survey sampling. Ann Stat 2000:1026–1053

Breidt FJ, Claeskens G, Opsomer JD (2005) Model-assisted estimation for complex surveys using penalised splines. Biometrika 92:831–846, http://dx.doi.org/10.1093/biomet/92.4.831

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 10 of 11

Cassel CM, Särndal CE, Wretman JH (1977) Foundations of inference in survey sampling. Wiley, New York

Chambers R, Clark R (2012) An introduction to model-based survey sampling with applications. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198566625.001.0001

Chirici G, McRoberts RE, Fattorini L, Mura M, Marchetti M (2016) Comparing echo-based and canopy height model-based metrics for enhancing estimation of forest aboveground biomass in a model-assisted framework. Remote Sensing Environ 174:1–9, http://dx.doi.org/10.1016/j.rse.2015.11.010

Corona P, Fattorini L, Franceschi S, Scrinzi G, Torresan C (2014) Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. Can J Forest Res 44:1303–1311, http://dx.doi.org/10.1139/cjfr-2014-0203

Corona P, Fattorini L, Pagliarella MC (2015) Sampling strategies for estimating forest cover from remote sensing-based two-stage inventories. Forest Ecosystems 2(1):1–12, http://dx.doi.org/10.1186/s40663-015-0042-7

Ene LT, Næsset E, Gobakken T, Gregoire TG, Ståhl G, Nelson R (2012) Assessing the accuracy of regional LiDAR-based biomass estimation using a simulation approach. Remote Sensing Environ 123:579–592, http://dx.doi.org/10.1016/j.rse.2012.04.017

Fattorini L, Marcheselli M, Pisani C (2006) A three-phase sampling strategy for large-scale multisource forest inventories. J Agric Biol Environ Stat 11(3):296–316, http://dx.doi.org/10.1198/108571106X130548

Fattorini L, Franceschi S, Pisani C (2009) A two-phase sampling strategy for large-scale forest carbon budgets. J Stat Plann Inference 139(3):1045–1055, http://dx.doi.org/10.1016/j.jspi.2008.06.014

Gobakken T, Næsset E, Nelson R, Bollandsås OM, Gregoire TG, Ståhl G, Holm S, Ørka HO, Astrup R (2012) Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. Remote Sensing Environ 123:443–456, http://dx.doi.org/10.1016/j.rse.2012.01.025

Grafström A, Saarela S, Ene LT (2014) Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. Can J Forest Res 44:1156–1164, http://dx.doi.org/10.1139/cjfr-2014-0202

Gregoire TG (1998) Design-based and model-based inference in survey sampling: appreciating the difference. Can J Forest Res 28:1429–1447, http://dx.doi.org/10.1139/x98-166

Gregoire TG, Valentine HT (2008) Sampling strategies for natural resources and the environment. CRC Press, Taylor & Francis Group, Boca Raton

Gregoire TG, Ståhl G, Næsset E, Gobakken T, Nelson R, Holm S (2011) Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway This article is one of a selection of papers from Extending Forest Inventory and Monitoring over Space and Time. Can J Forest Res 41:83–95, http://dx.doi.org/10.1139/X10-195

Hansen MH, Madow WG, Tepping BJ (1978) On inference and estimation from sample surveys. In: Proceedings of the Survey Research Methods Section,, pp 82–107

Hansen MH, Madow WG, Tepping BJ (1983) An evaluation of model-dependent and probability-sampling inferences in sample surveys. J Am Stat Assoc 78:776–793, http://dx.doi.org/10.1080/01621459.1983.10477018

Healey SP, Patterson PL, Saatchi S, Lefsky MA, Lister AJ, Freeman EA (2012) A sample design for globally consistent biomass estimation using lidar data from the Geoscience Laser Altimeter System (GLAS). Carbon Balance Manage 7:1–9, http://dx.doi.org/10.1186/1750-0680-7-10

Helmer EH, Ruzycki TS, Wunderle JM, Vogesser S, Ruefenacht B, Kwit C, Ewert DN (2010) Mapping tropical dry forest height, foliage height profiles and disturbance type and age with a time series of cloud-cleared Landsat and ALI image mosaics to characterize avian habitat. Remote Sensing Environ 114:2457–2473, http://dx.doi.org/10.1016/j.rse.2010.05.021

Köhl M, Brassel P (2001) Zur Auswirkung der Hangneigungskorrektur auf Schätzwerte im Schweizerischen Landesforstinventar (LFI) [Investigation of the effect of the slope correction method as applied in the Swiss National Forest Inventory of estimates.]. Schweizerische Zeitschrift fur Forstwesen 152(6):215–225, http://dx.doi.org/10.3188/szf.2001.0215

Magnussen S (2015) Arguments for a model-dependent inference? Forestry 88(3):317–325, http://dx.doi.org/10.1093/forestry/cpv002

Magnussen S, Tomppo E (2015) Model-calibrated k-nearest neighbor estimators. Scandinavian J Forest Res 1–11. http://dx.doi.org/10.1080/02827581.2015.1073348

Magnussen S, Næsset E, Gobakken T (2014) An estimator of variance for two-stage ratio regression estimators. Forest Sci 60(4):663–676, http://dx.doi.org/10.5849/forsci.12-163

Magnussen S, Næsset E, Gobakken T (2015) LiDAR-supported estimation of change in forest biomass with time-invariant regression models. Can J Forest Res 45(999):1514–1523, http://dx.doi.org/10.1139/cjfr-2015-0084

Mandallaz D (2013) Design-based properties of some small-area estimators in forest inventory with two-phase sampling. Can J Forest Res 43:441–449, http://dx.doi.org/10.1139/cjfr-2012-0381

Margolis HA, Nelson RF, Montesano PM, Beaudoin A, Sun G, Andersen HE, Wulder M (2015) Combining satellite lidar, airborne lidar and ground plots to estimate the amount and distribution of aboveground biomass in the Boreal forest of North America. Can J Forest Res 45(7):838–855, http://dx.doi.org/10.1139/cjfr-2015-0006

Massey A, Mandallaz D, Lanz A (2014) Integrating remote sensing and past inventory data under the new annual design of the Swiss National Forest Inventory using three-phase design-based regression estimation. Can J Forest Res 44:1177–1186, http://dx.doi.org/10.1139/cjfr-2014-0152

McRoberts RE (2006) A model-based approach to estimating forest area. Remote Sensing Environ 103:56–66, http://dx.doi.org/10.1016/j.rse.2006.03.005

McRoberts RE (2010) Probability-and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. Remote Sensing Environ 114:1017–1025, http://dx.doi.org/10.1016/j.rse.2009.12.013

McRoberts RE, Tomppo EO, Finley AO, Heikkinen J (2007) Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sensing Environ 111:466–480

McRoberts RE, Bollandsås OM, Næsset E (2014) Modeling and estimating change. In: Maltamo M, Næsset E, Vauhkonen J. (eds) Forestry Applications of Airborne Laser Scanning. Concepts and Case Studies. Springer, pp. 293–314. http://dx.doi.org/10.1007/978-94-017-8663-8_15

McRoberts RE, Næsset E, Gobakken T, Bollandsås OM (2015) Indirect and direct estimation of forest biomass change using forest inventory and airborne laser scanning data. Remote Sensing Environ 164:36–42, http://dx.doi.org/10.1016/j.rse.2015.02.018

Melville GJ, Welsh AH, Stone C (2015) Improving the efficiency and precision of tree counts in pine plantations using airborne LiDAR data and flexible-radius plots: model-based and design-based approaches. J Agric Biol Environ Stat 20(2):229–257, http://dx.doi.org/10.1007/s13253-015-0205-6

Næsset E (1997) Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing Environ 61:246–253, http://dx.doi.org/10.1016/S0034-4257(97)00041-2

Næsset E (2002a) Determination of mean tree height of forest stands by means of digital photogrammetry. Scand J Forest Res 17: 446–459. http://dx.doi.org/10.1080/028275802320435469

Næsset E (2002b) Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing Environ 80: 88–99. http://dx.doi.org/10.1016/S0034-4257(01)00290-5

Næsset E (2004) Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. Scand J Forest Res 19:554–557, http://dx.doi.org/10.1080/02827580410019544

Næsset E, Gobakken T, Solberg S, Gregoire TG, Nelson R, Ståhl G, Weydahl D (2011) Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. Remote Sensing Environ 115:3599–3614, http://dx.doi.org/10.1016/j.rse.2011.08.021

Næsset E, Bollandsås OM, Gobakken T, Gregoire TG, Ståhl G (2013a) Model-assisted estimation of change in forest biomass over an 11year period in a sample survey supported by airborne LiDAR: A case study with post-stratification to provide "activity data". Remote Sensing Environ 128: 299–314. http://dx.doi.org/10.1016/j.rse.2012.10.008

Næsset E, Gobakken T, Bollandsås OM, Gregoire TG, Nelson R, Ståhl G (2013b) Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway. Remote Sensing Environ 130: 108–120. http://dx.doi.org/10.1016/j.rse.2012.11.010

Næsset E, Bollandsås OM, Gobakken T, Solberg S, McRoberts RE (2015) The effects of field plot size on model-assisted estimation of aboveground biomass change using multitemporal interferometric SAR and airborne laser scanning data. Remote Sensing Environ 168:252–264, http://dx.doi.org/10.1016/j.rse.2015.07.002

Nelson R, Krabill W, Maclean G (1984) Determining forest canopy characteris-tics using airborne laser data. Remote Sensing Environ 15:201–212, http://dx.doi.org/10.1016/0034-4257(84)90031-2

Nelson R, Krabill W, Tonelli J (1988) Estimating forest biomass and volume using airborne laser data. Remote Sensing Environ 24:247–267, http://dx.doi.org/10.1016/0034-4257(88)90028-4

Ståhl *et al. Forest Ecosystems* (2016) 3:5

Page 11 of 11

Nelson R, Boudreau J, Gregoire TG, Margolis H, Næsset E, Gobakken T, Ståhl G (2009) Estimating Quebec provincial forest resources using ICESat/GLAS. Can J Forest Res 39:862–881, http://dx.doi.org/10.1139/X09-002

Nelson R, Gobakken T, Næsset E, Gregoire TG, Ståhl G, Holm S, Flewelling J (2012) Lidar sampling - using an airborne profiler to estimate forest biomass in Hedmark County, Norway. Remote Sensing Environ 123:563–578, http://dx.doi.org/10.1016/j.rse.2011.10.036

Neyman J (1934) On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. J R Stat Soc 97:558–606, http://dx.doi.org/10.2307/2342192

Opsomer JD, Breidt FJ, Moisen GG, Kauermann G (2007) Model-assisted estimation of forest resources with generalized additive models. J Am Stat Assoc 102:400–409, http://dx.doi.org/10.1198/016214506000001491

Reese H, Nilsson M, Sandström P, Olsson H (2002) Applications using estimates of forest parameters derived from satellite and forest inventory data. Comput Electron Agric 37:37–55, http://dx.doi.org/10.1016/S0168-1699(02)00118-7

Saarela S, Grafström A, Ståhl G, Kangas A, Holopainen M, Tuominen S, Nordkvist K, Hyyppä, J (2015a) Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information. Remote Sensing Environ 158: 431–440. http://dx.doi.org/10.1016/j.rse.2014.11.020

Saarela S, Schnell S, Grafström A, Tuominen S, Nordkvist K, Hyyppä J, Kangas A, Ståhl G (2015b) Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume in Kuortane, Finland. Can J Forest Re 45:1524–1534. http://dx.doi.org/10.1139/cjfr-2015-0077

Saarela S, Schnell S, Tuominen S, Balazs A, Hyyppä J, Grafström A, Ståhl G (2016) Effects of positional errors in model-assisted and model-based estimation of growing stock volume. Remote Sensing Environ 172:101–108, http://dx.doi.org/10.1016/j.rse.2015.11.002

Sannier C, McRoberts RE, Fichet LV, Makaga EMK (2014) Using the regression estimator with Landsat data to estimate proportion forest cover and net proportion deforestation in Gabon. Remote Sensing Environ 151:138–148, http://dx.doi.org/10.1016/j.rse.2013.09.015

Särndal CE (1978) Design-based and model-based inference in survey sampling [with discussion and reply]. Scand J Stat 5(1):27–52

Särndal CE, Swensson B, Wretman J (1992) Model Assisted Survey Sampling. Springer. http://dx.doi.org/10.1007/978-1-4612-4378-6

Skowronski NS, Clark KL, Gallagher M, Birdsey RA, Hom JL (2014) Airborne laser scanner-assisted estimation of aboveground biomass change in a temperate oak-pine forest. Remote Sensing Environ 151:166–174, http://dx.doi.org/10.1016/j.rse.2013.12.015

Solberg S, Astrup R, Bollandsås OM, Næsset E, Weydahl DJ (2010) Deriving forest monitoring variables from X-band InSAR SRTM height. Can J Remote Sensing 36:68–79, http://dx.doi.org/10.5589/m10-025

Ståhl G, Holm S, Gregoire TG, Gobakken T, Næsset E, Nelson R (2011) Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. Can J Forest Res 41:96–107, http://dx.doi.org/10.1139/X10-161

Ståhl G, Heikkinen J, Petersson H, Repola J, Holm S (2014) Sample-based estimation of greenhouse gas emissions from forests – A new approach to account for both sampling and model errors. Forest Sci 60:3–13, http://dx.doi.org/10.5849/forsci.13-005

Stephens PR, Kimberley MO, Beets PN, Paul TS, Searles N, Bell A, Brack C, Broadley J (2012) Airborne scanning LiDAR in a double sampling forest carbon inventory. Remote Sensing Environ 117:348–357, http://dx.doi.org/10.1016/j.rse.2011.10.009

Strunk JL, Reutebuch SE, Andersen HE, Gould PJ, McGaughey RJ (2012a) Model-assisted forest yield estimation with light detection and ranging. West J Appl Forestry 27: 53–59. http://dx.doi.org/10.5849/wjaf.10-043

Strunk J, Temesgen H, Andersen HE, Flewelling JP, Madsen L (2012b) Effects of lidar pulse density and sample size on a model-assisted approach to estimate forest inventory variables. Can J Remote Sensing 38: 644–654. http://dx.doi.org/10.5589/m12-052

Tomppo E. Katila M (1991) Satellite image-based national forest inventory of Finland for publication in the IGARSS'91 digest. In: Geoscience and Remote Sensing Symposium, 1991. IGARSS'91. Remote Sensing: Global Monitoring for Earth Management., International (Vol. 3, pp. 1141–1144). http://dx.doi.org/10.1109/igarss.1991.579272

Tomppo E, Olsson H, Ståhl G, Nilsson M, Hagner O, Katila M (2008) Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sensing Environ 112(5):1982–1999

Tomppo E, Gschwantner T, Lawrence M, McRoberts RE, Gabler K, Schadauer K, Vidal C, Lanz A, Ståhl G, Cienciala E (2010) National forest inventories. Pathways for Common Reporting. Springer, 541–553. http://dx.doi.org/10.1007/978-90-481-3233-1