Linear Models for Airborne-Laser-Scanning-Based Operational Forest Inventory With Small Field Sample Size and Highly Correlated LiDAR Data

Virpi Junttila, Tuomo Kauranne, Andrew O. Finley, and John B. Bradford

***Abstract*—Modern operational forest inventory often uses remotely sensed data that cover the whole inventory area to produce spatially explicit estimates of forest properties through statistical models. The data obtained by airborne light detection and ranging (LiDAR) correlate well with many forest inventory variables, such as the tree height, the timber volume, and the biomass. To construct an accurate model over thousands of hectares, LiDAR data must be supplemented with several hundred field sample measurements of forest inventory variables. This can be costly and time consuming. Different LiDAR-data-based and spatial-data-based sampling designs can reduce the number of field sample plots needed. However, problems arising from the features of the LiDAR data, such as a large number of predictors compared with the sample size (overfitting) or a strong correlation among predictors (multicollinearity), may decrease the accuracy and precision of the estimates and predictions. To overcome these problems, a Bayesian linear model with the singular value decomposition of predictors, combined with regularization, is proposed. The model performance in predicting different forest inventory variables is verified in ten inventory areas from two continents, where the number of field sample plots is reduced using different sampling designs. The results show that, with an appropriate field plot selection strategy and the proposed linear model, the total relative error of the predicted forest inventory variables is only 5%–15% larger using 50 field sample plots than the error of a linear model estimated with several hundred field sample plots when we sum up the error due to both the model noise variance and the model's lack of fit.**

***Index Terms*—Bayesian linear model, model-based forest inventory, regularization, sampling design, singular value decomposition (SVD).**

Manuscript received January 30, 2013; revised June 12, 2013, November 5, 2013, March 4, 2014, October 20, 2014, February 20, 2015, and March 16, 2015; accepted April 12, 2015. Date of publication May 14, 2015; date of current version June 10, 2015. The work of A. O. Finley was supported in part by the U.S. Department of Agriculture (USDA) Forest Service through the Forest Inventory and Analysis National Program and the Forest Health Technology Enterprise Team, by the National Science Foundation under Grants EF-1137309, EF-1253225, and DMS-1106609, and by NASA Carbon Monitoring System grants. The work of V. Junttila was supported in part by the USDA/NASA under Grant 10-JV-11242307-037 and in part by NASA Carbon Monitoring System grants. The work of J. B. Bradford was supported in part by the USDA/NASA Grant 10-JV-11242307-037 and in part by the U.S. Geological Survey Climate and Land Use and Ecosystems Mission Areas. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

V. Junttila is with the Department of Applied Mathematics, Lappeenranta University of Technology, 53851 Lappeenranta, Finland (e-mail: virpi.junttila@lut.fi).

T. Kauranne is with the Department of Applied Mathematics, Lappeenranta University of Technology, 53851 Lappeenranta, Finland, and also with Arbonaut Oy Ltd., 80130 Joensuu, Finland (e-mail: tuomo.kauranne@lut.fi).

A. O. Finley is with the Department of Forestry, the Department of Geography, and the Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824 USA (e-mail: finleya@msu.edu).

J. B. Bradford is with the Southwest Biological Science Center, U.S. Geological Survey, Flagstaff, AZ 86001 USA (e-mail: jbradford@usgs.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2015.2425916

I. Introduction

THE prediction of forest inventory variables over large domains is used in many applications, including forest management and policy decisions, environmental research, and environmental monitoring. Predictions with an associated uncertainty can be estimated by the statistical extension of field-measured inventory variables or by mathematical models based on field measurements combined with remotely sensed data. The most common sources of remotely sensed data are aerial and satellite images, and most recently, light detection and ranging (LiDAR) data. Many studies, some of which are discussed in the following, have detailed models to improve the prediction and mapping of important economic and ecological forest inventory variables.

The methods based on the use of LiDAR data have recently gained popularity, [1], [2], to the extent that some countries, e.g., Finland [3], have chosen it as the only method to be used in medium-scale operational forest inventory. Here, we consider the medium scale as inventory initiatives at a stand level over no more than a million hectares. In national-scale inventory over several millions of hectares with a high spatial resolution, satellite-image-based methods are still the best choice because of the low cost and high temporal frequency of satellite imagery. The main disadvantage of satellite images is lower precision and the lack of fit of models to the data because of issues such as saturation than what is possible using LiDAR. Another disadvantage is that satellite images provide little information about below-canopy attributes. In this paper, our focus is on medium-scale forest inventory.

A variety of modeling approaches to forest inventory prediction can be found in literature, including nonparametric approaches such as the k-nearest neighbor (k-NN) and random forest, linear regression, and geostatistical approaches (e.g., see [4]–[9] and the references therein). Although these remote-sensing-based methods are cost effective and accurate, they require a representative field sample for estimating parameters.

A linear regression model using LiDAR predictors is known to accurately predict forest inventory variables such as the median tree height, the basal area, the volume, and the above-ground biomass (AGB) (e.g., see [5], [6], [10], and [11]). A linear regression model is an extrapolation method, which

must be used with caution, i.e., there is a risk of negative predictions or absurdly large positive predictions, particularly when the predictor values in the population are beyond the ranges observed in the sample (e.g., see [12]). The remotely sensed measurements must be accompanied by field measurements to estimate the model parameters. The number of field sample plots needed for the accurate predictions of the forest inventory variables is generally several hundreds for an area between 10 000 and 100 000 ha (e.g., see [3]). When producing estimates based on models for several thousand hectares, the field work becomes expensive. In particular, on difficult terrain, such as tropical forests, the cost of field measurements can be prohibitively high. Thus, minimizing the number of field sample plots needed to attain the desired prediction precision of forest inventory variables is important to forest inventory tasks.

New inventory sites generally lack prior information of the forest inventory variable values, and precise prior assumptions about the model behavior cannot be expected before field measurements are performed. In practice, prior to field sampling, an investigator defines the criteria to select the sample plots. Then, the field measurements of forest variable(s) of interest, which are referred to as the responses, are carried out in these plots. The parameters of a suitable prediction model are then estimated using a combination of the responses and features derived from LiDAR data, which is referred to as the predictors of the model. These responses and predictors from the same field sample plots are referred to as the training set of the model. The resulting model with estimated parameters is used to predict the forest variables in the rest of the inventory area.

The effect of a reduced number of field sample plots on the forest variable prediction accuracy has been explored in several recent model-based forest inventory studies. Junttila *et al.* [13] considered the effect of LiDAR predictors and the varying intensities of inventory plots for predicting a host of forest variables using ordinary least squares (OLS) and sparse Bayesian regression models. Hawbaker *et al.* [14] assessed two sampling designs for informing a regression model used for predicting the biomass and other forest variables. Here, field data were drawn from a simple random sample and a stratified sample that was informed using the mean and standard deviation of LiDAR canopy height estimates. In a similar study that used the k-NN method to predict forest variables, Maltamo *et al.* [15] considered several plot selection strategies, including random selection, random selection within prestratification according to the forest type, and the selection of plots based on the properties of the LiDAR data given as prior information. In recent studies, Junttila *et al.* [16] have verified the effect of different classical designs based on LiDAR predictors and the spatial information of candidate plots, and Gobakken *et al.* [17] have tested the use of laser data as auxiliary information in the selection of field plot locations. The conclusions were that the use of laser-scanning-data-based information in the field plot selection procedure helps improve the reliability of predictions. In addition, in the survey sampling community, the idea of using auxiliary variables, i.e., LiDAR data and the spatial location in forest inventory, in sample selection has been widely discussed (e.g., see [18] and the references therein).

The aim of our study is to maximize the accuracy and precision of forest inventory variable predictions when only a small number of field plots are available for the model parameter estimation procedure. Our study relies on the sampling designs and results described previously [16], and our aim is to improve these results independent of the sampling design. Our focus is in the mathematical properties of the regression model and the training set data, which may result in the erroneous estimates of model parameters and may therefore induce a lack of fit to the corresponding predictions. When reducing the number of plots by algorithmic means, it is possible that the resulting set of field plots does not constitute a valid probability sample of the forest. By this we mean that it is generally not possible to determine the selection probabilities of plots, and it may happen that some parts of the population obtain probability zero. However, the selection probabilities of plots are not used in model parameter estimation, which is the only direct use of plots in model-based inference (e.g., see [12]).

Further sources of estimation errors, which are defined here as both the lack of fit and poor precision, include the possible correlation of predictors (multicollinearity), the potential for a training set that does not represent the characteristics of the predicted forest inventory variable values, and a small number of training set plots compared with the number of predictors. For this reason, we have chosen to validate the estimates on sites where a relatively large sample of plots is available.

The LiDAR measurement distributions within a plot area are used to predict plot-level variables using multiple statistical estimates, which serve as predictors in the regression model. Predictors generally contain different combinations of characteristics estimated from LiDAR pulse returns, such as mean values, standard deviations, percentiles, the minimum and maximum values of height measurements, the percentage of vegetation returns of given heights, the percentiles of measurement intensities, etc. (e.g., see [13] and [19]–[22]). These predictors are usually highly correlated. In regression models, such a situation is called multicollinearity. The severity of multicollinearity in regression analysis can be quantified, e.g., using a variance inflation factor (VIF), which provides an indicator that measures how much the variance of an estimated regression parameter is increased because of collinearity. If a predictor is a (nearly) linear combination of other predictors, the VIF value is high, and severe multicollinearity occurs. Strong multicollinearity can cause numerical instability, i.e., singularities in parameter estimation equations and less precise estimates. One way to resolve this problem is to reduce the number of collinear predictors until there is only one remaining out of the set. Other methods are regularization, which is also known as ridge regression [23], where the singular or close-to-singular parameter covariance matrix is modified with additive small diagonal elements, and the reformulation of an independent data matrix to a form only containing diagonal and orthogonal matrices using principal component regression (see [24] and [25]).

In practice, the field sample plots selected to comprise the training set only represent a fraction of the forest area. Thus, the predictions of the forest inventory variables in the rest of the area, which are based on a linear candidate model estimated with the field sample plot measurements and the remotely

sensed data, are likely to contain inaccuracy. Compared with the optimal model-based solution, i.e., a model constructed with the information of the whole forest, a possible source of inaccuracy is an inefficiently chosen set of field sample plots. With a small training set size, the variances of the estimated model parameter distributions become larger (more uncertainty in the parameter estimates), and using a random set of field sample plots, the training set data may show completely spurious correlations between some responses and predictors. This leads to highly variable estimates of the model parameters, as shown in [16].

Moreover, the number of LiDAR predictors can be large compared with the number of field sample plots, particularly in cases where the number of field sample plots is minimized. This can lead to overfitting, which produces exceedingly accurate estimates for the training set data but poorly generalizes to the prediction in the unobserved set. To overcome this problem, different approaches to choose fewer predictors are stepwise regression, [25]–[27], cross validation [28], or ridge regression performed with a Bayesian regularization approach [13]. The aim of these methods is to reduce the number of predictors used by giving zero values to the model parameters related to the redundant predictors, i.e., to give a sparse parameter vector. In operational use, the predictor selection algorithm needs to be adjustable, reliable with different sorts of data, and computationally fast.

In this paper, a modified linear regression model is used to enable the use of a minimal amount of input data (sample plots) such that the model prediction accuracy will not suffer from the truncation of an independent data set. A combination of singular value decomposition (SVD), which is closely related to the principal component analysis, with Bayesian regularization is employed.

The proposed linear regression model used in the prediction of forest inventory variables is based on a sparse Bayesian model (see [13] and [29]) and a regularized linear model with orthogonalized predictors (see [30] and [31]) with some modifications. In these studies, the regularization of the linear model is performed in a Bayesian formulation, where the model precision is automatically compared with the model complexity such that the model parameters are estimated by simple optimization based on ideas first introduced in [32]. In the latter studies, the orthogonal decomposition of the predictors is used together with regularization. In our study, the orthogonal decomposition is substituted with the SVD. In the presence of highly correlated predictors, the SVD condenses the most important signals into a few orthogonal and normally distributed vectors and corresponding singular values. The prediction accuracy of the proposed model is verified using ten different forest inventory areas, which are located in the USA and in Finland, and three different sampling designs.

The remainder of this paper is organized as follows. Section II describes the linear model modifications beginning from the method to overcome problems arising from high correlation among predictors (see Section II-A). To overcome problems arising from a small amount of data compared with the number of predictors, this method is combined with a regularized Bayesian linear model in Section II-B. The model verification procedure is described in Section III, where we first define several approaches to the thinning of the field sample plot grid over the inventory area in Section III-A; then, we introduce the forest data in Section III-B, and we discuss the error validation in Section III-C. The results are shown and analyzed in Section IV, and finally, in Section V, we summarize the advantages of the proposed method.

II. Methods

The forest area of interest can be conceptually partitioned into a grid of equally sized units. We shall use the terms "plot" and "grid cell" to refer to the population unit of a forest. A plot refers to a field sample plot, whereas a grid cell refers to the spatial unit of the estimates calculated with a model. Plots and grid cells are assumed to have a roughly similar size but not shape because plots are often circular but circles do not tesselate the plane. In practice, n field sample plot locations are selected from a set of N_{cand} candidate locations distributed over the whole area of interest, and the field plot measurements are then used to estimate the model parameters. The resulting model is used to predict the forest variables over the forested area either in a set of N_{pred} plots of interest or in N_{pred} grid cells covering the whole area [26]. In this paper, the model is designed to be used with either of these population units, but we use the term plot here because the validation data are measured in circular plot units, and for consistency and legibility reasons, this terminology is retained throughout this paper.

In this paper, a linear model is used to predict the forest variables in each of the N_{pred} plots of interest. The auxiliary data are assumed to be known for all the plots of interest and for all the n field sample plots, i.e., for $N = N_{\mathrm{pred}} + n$ plots. In each plot, i.e., $i,\ i = 1, 2, \ldots, N$, remotely sensed LiDAR data are measured, and M separate and plot-level predictors $\boldsymbol{z}_i$ are derived from these data. These predictors (also called covariates), which serve as the predictors in the regression model, are stored together with an intercept in an $N \times (M + 1)$ matrix $\boldsymbol{Z}$. In addition to the remotely sensed data, the central coordinates of the plots are known, i.e., $\boldsymbol{s}_i = (s_{xi}, s_{yi})$. The geographical coordinates and the predictors drawn from the remotely sensed data form the auxiliary data of the study site.

The field measurements of the forest inventory variables of interest are performed in the set of n selected field sample plots. The indexes of these n plots are stored to an $n \times 1$ index vector $\boldsymbol{t}$. After a set of K forest inventory variable values is measured in these plots, it is stored in the columns of an $n \times K$ response matrix $\boldsymbol{Y_t}$. These data combined with the corresponding auxiliary data are used as the training set of the forest inventory variable prediction model, i.e., all the parameters of the model are estimated using the training set data.

In this paper, each forest inventory variable, i.e., an $n \times 1$ response vector, is predicted with a separate model. For reasons of clarity, vector $\boldsymbol{y_t}$ represents any individual forest inventory variable field measurement vector (a column of matrix $\boldsymbol{Y_t}$). Each forest inventory variable prediction model is assumed to be a parametric and linear regression model as follows:

$$\boldsymbol{y_t} = \boldsymbol{Z_t w} + \boldsymbol{\epsilon}. \tag{1}$$

Here, $\boldsymbol{w}$ is an $(M + 1) \times 1$ vector of the parameters related to the regression model predictors and the intercept (also known

as regression parameters). In this paper, the $n \times 1$ error vector $\boldsymbol{\epsilon}$ is assumed to be normally distributed with zero mean and variance σ^2.

Prior to field sample measurements, there are no data to estimate model parameters $\{\boldsymbol{w}, \sigma^2\}$. In a new inventory site, the field sample plots are selected from the set of candidate plots in the whole inventory area. With a suitable selection method, this set is assumed to represent well the variability of the forest characteristics of the whole inventory area. These methods are discussed later in Section III-A.

A. SVD of Predictors

A classical approach to overcome the problem of multicollinearity is the SVD. Here, the centered and scaled LiDAR predictor matrix of size $N \times M$ is decomposed in the form $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$, where $\boldsymbol{U}$ is an $N \times N$ matrix consisting of the eigenvectors of matrix $\boldsymbol{Z}\boldsymbol{Z}^T$, and $\boldsymbol{V}$ is an $M \times M$ matrix consisting of the eigenvectors of matrix $\boldsymbol{Z}^T\boldsymbol{Z}$. The columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are called left and right singular vectors, respectively. $\boldsymbol{S}$ is an $N \times M$ diagonal matrix, where the upper $M \times M$ block is a diagonal matrix consisting of the singular values, i.e., λ_m, which are given in the order $\lambda_1 > \lambda_2 > \cdots > \lambda_M$, and the rest of the components are zeros. Singular values are square roots of the eigenvalues of $\boldsymbol{Z}^T\boldsymbol{Z}$ and $\boldsymbol{Z}\boldsymbol{Z}^T$. The corresponding singular vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ are given in the same order. These vectors are orthogonal, i.e., $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}_N$ and $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}_M$. Here, $\boldsymbol{I}_N$ refers to an $N \times N$ identity matrix.

A large singular value means that the corresponding singular vector captures a large amount of the variation in the original data $\boldsymbol{Z}$. Thus, only using the q first columns of matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ and the corresponding q largest singular values, the original data matrix can be approximated by $\boldsymbol{Z} \simeq \boldsymbol{Z}_q = \boldsymbol{U}_q\boldsymbol{S}_q\boldsymbol{V}_q^T$. The value of q can be estimated as the largest value for which

$$\sum_{m=1}^{q-1} \lambda_m / \sum_{m=1}^{M} \lambda_m \leq P \tag{2}$$

holds with a given explanation ratio, e.g., 90% ($P = 0.9$). This new data matrix, i.e., $\boldsymbol{Z}_q$, explains $P \times 100\%$ of the variability of the original matrix only using q orthogonal predictors. Thus, the problem of multicollinearity can be overcome with the SVD.

If the intercept, which is a column with constants, is added to the decomposed predictor matrix $\boldsymbol{Z}_q \Leftarrow (\boldsymbol{1}, \boldsymbol{Z}_q)$, the decomposition matrices become

$$\begin{aligned} &\boldsymbol{U}_q \Leftarrow (\boldsymbol{1}, \boldsymbol{U}_q) \qquad \operatorname{diag}(\boldsymbol{S}_q) \Leftarrow (1\ \lambda_1\ \lambda_2\ \cdots\ \lambda_q)^T \\ &\boldsymbol{V}_q \Leftarrow \begin{pmatrix} 1 & \boldsymbol{0}_{1\times q} \\ \boldsymbol{0}_{q\times 1} & \tilde{\boldsymbol{V}}_q \end{pmatrix}. \end{aligned} \tag{3}$$

In real-world cases, the amount of available measurements of the response, i.e., n, may be much smaller than the total number of plots of interest, i.e., $n \ll N$, and only n sample observations are used in the training set of the model. An important feature of auxiliary data $\boldsymbol{Z}$ is that it is given for each plot of interest. Thus, instead of only basing the SVD on the set of predictor values given for field sample plots, the singular values and vectors are based on the predictor values given for all the plots of interest. With this approach, we can maintain more information and avoid a situation where incorrect features of the auxiliary data may become dominating, and the prediction model would lean on misleading assumptions of the signal given in $\boldsymbol{Z}$.

After the SVD, matrix $\boldsymbol{Z}$ is divided into two submatrices. Thus, the original data included in the training set of the model consist of the rows $\boldsymbol{t}$ of predictor matrix $\boldsymbol{Z}$, i.e., $\boldsymbol{Z}_{\boldsymbol{t}}$, corresponding to the rows of response vector $\boldsymbol{y}_{\boldsymbol{t}}$ (observations). The model is then used to predict the $N - n$ unknown response values in the rest of the plots, i.e., verification set $\boldsymbol{v}$, using the corresponding independent matrix values, i.e., $\boldsymbol{Z}_{\boldsymbol{v}}$.

After the SVD is first performed for the full $N \times M$ matrix, the training set and verification set singular value matrices become $\boldsymbol{Z}_{q,\boldsymbol{t}} = \boldsymbol{U}_{q,\boldsymbol{t}}\boldsymbol{S}_q\boldsymbol{V}_q$ and $\boldsymbol{Z}_{q,\boldsymbol{v}} = \boldsymbol{U}_{q,\boldsymbol{v}}\boldsymbol{S}_q\boldsymbol{V}_q$, respectively. Matrices $\boldsymbol{U}_{q,\boldsymbol{t}}$ and $\boldsymbol{U}_{q,\boldsymbol{v}}$ consist of the corresponding rows of the original orthogonal matrix $\boldsymbol{U}_q$. Note that the orthogonality rule no longer holds for distinct matrices $\boldsymbol{U}_{q,\boldsymbol{t}}$ and $\boldsymbol{U}_{q,\boldsymbol{v}}$.

In the regression model of the training set, $\boldsymbol{y}_{\boldsymbol{t}} = \boldsymbol{Z}_{q,\boldsymbol{t}}\tilde{\boldsymbol{w}} + \boldsymbol{\epsilon}$, where $\boldsymbol{Z}_{q,\boldsymbol{t}}$ is an $n \times (q+1)$ matrix, and $\tilde{\boldsymbol{w}}$ is a $(q+1) \times 1$ regression parameter vector; the formulation becomes $\boldsymbol{y}_{\boldsymbol{t}} = \boldsymbol{U}_{q,\boldsymbol{t}}\boldsymbol{S}_q\boldsymbol{V}_q^T\tilde{\boldsymbol{w}} + \boldsymbol{\epsilon}$. Since the right-hand-side singular vector, i.e., $\boldsymbol{V}_q$, is equal for every subset of data input and there is no need to verify the exact regression parameter values, $\tilde{\boldsymbol{w}}$, which is the model based on the training set, can be stated as

$$\boldsymbol{y}_{\boldsymbol{t}} = \boldsymbol{X}_{\boldsymbol{t}}\boldsymbol{w} + \boldsymbol{\epsilon} \tag{4}$$

where $\boldsymbol{X}_{\boldsymbol{t}} = \boldsymbol{U}_{q,\boldsymbol{t}}\boldsymbol{S}_q$, and $\boldsymbol{w} = \boldsymbol{V}_q^T\tilde{\boldsymbol{w}}$.

Similar to the assumption in (1), we assume that deviation $\boldsymbol{y}_{\boldsymbol{t}} - \boldsymbol{X}_{\boldsymbol{t}}\boldsymbol{w}$ is normally distributed with variance σ^2, and we write

$$N(\boldsymbol{y}_{\boldsymbol{t}}|\boldsymbol{X}_{\boldsymbol{t}}\boldsymbol{w}, \sigma^2\boldsymbol{I}_n) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}_{\boldsymbol{t}} - \boldsymbol{X}_{\boldsymbol{t}}\boldsymbol{w})^T(\boldsymbol{y}_{\boldsymbol{t}} - \boldsymbol{X}_{\boldsymbol{t}}\boldsymbol{w})}. \tag{5}$$

The solution of this regression model is called the truncated SVD (tSVD) method, where the truncation refers to the $M - q$ singular vectors dismissed in the model because of the corresponding small singular values. For the discussion of the truncation and different approaches to choose the level of truncation, see [33]. In this paper, we call this method OLS_q. This method is closely related to the principal components analysis method, principal components regression (PCR), and partial least squares regression (PLSR), which have been used in forest inventory studies with a small amount of data (e.g., see [20] and [34]). However, the PCR and the PLSR only utilize the data of the training set to the eigenvalue-type analysis. This may result in missing some important information about the predictor characteristics, particularly in the case of a small training set size. If we use all the q orthogonal predictors, i.e., predictor $\boldsymbol{X}_{\boldsymbol{t}}$, in the PCR or the PLSR, the model parameters are equal to that of OLS_q.

B. Effect of Small Training Set

A reduced number of field sample plots affect the reliability of the resulting predictions by offering less data for the value estimation of parameter vector $\boldsymbol{w}$. With different field sample

plot selection strategies, different subsets of sample plots are selected, and the estimation reliability varies depending on its quality. If the level of truncation defines the number of predictors used, i.e., q, such that it is close to the number of field sample plots n, the accuracy can become very poor due to model overfitting, which is independent of the quality of the selected subset.

A common method to solve this problem is the SVD, and we only use those new predictors that have a singular value that is large enough. However, in a complicated linear model with unknown relationships between responses and predictors, it is difficult to determine which predictors can be ignored such that no important data are left out from the model. Thus, in this case, the SVD alone is not a sufficient solution for the problem of overfitting.

Instead, regularization can be used. Regularization gives an additional penalty to the model complexity, forcing the associated regression model parameter values toward zero (shrinkage). In a Bayesian formulation, regularization can be formulated by giving a normal prior distribution for regression parameters $\boldsymbol{w}$ such that the prior mean is zero, and some variance α^{-1} is allowed. With a small variance ($\alpha \to \infty$), the corresponding regression parameter value becomes close to zero, and with a large variance ($\alpha \to 0$), nonzero regression parameter values are allowed.

In this paper, the prior distribution associated with the intercept, i.e., $\boldsymbol{w}_0$, has variance α_0^{-1}, and the prior distributions associated with the orthogonal predictors, i.e., $\boldsymbol{w}_i$, have a common variance $\alpha_i^{-1} = \alpha^{-1}$, $i = 1, \ldots, q$. This approach is similar to the uniformly rectangular OLS (UROLS) algorithm given in a Bayesian formulation, e.g., in [31]. Our method is similar to it with small modifications, i.e., the SVD is performed for all auxiliary data (the orthogonality no longer holds for the training set), and the model formulation is written according to the formulation given in [29].

The regression problem with regularization becomes

$$N(\boldsymbol{y_t}|\boldsymbol{X_t w}, \sigma^2 \boldsymbol{I}_n) N(\boldsymbol{w}|\boldsymbol{0}, \boldsymbol{A}^{-1}) \tag{6}$$

where covariance matrix $\boldsymbol{A}$ is an $(q+1) \times (q+1)$ diagonal matrix with diagonal elements α_i, $i = 0, 1, 2, \ldots, q$.

Following the formulation of Tipping [29], the unknown parameters $\{\boldsymbol{w}, \sigma^2, \alpha_0, \alpha\}$ can be estimated by an iterative procedure. The mean $\boldsymbol{\mu_w}$ and covariance $\boldsymbol{\Sigma_w}$ of parameter values $\boldsymbol{w}_m$ can be analytically estimated by the normal posterior as follows:

$$N(\boldsymbol{y_t}|\boldsymbol{X_t w}, \sigma^2 \boldsymbol{I}_n) N(\boldsymbol{w}|\boldsymbol{0}, \boldsymbol{A}) = N(\boldsymbol{w}|\boldsymbol{\mu_w}, \boldsymbol{\Sigma_w}) N(\boldsymbol{y_t}|\boldsymbol{0}, \boldsymbol{C}) \tag{7}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1} &= \sigma^{-2} \boldsymbol{X}_{\boldsymbol{t}}^T \boldsymbol{X_t} + \boldsymbol{A} \\ &= \boldsymbol{S}_q \underbrace{\left(\sigma^{-2} \boldsymbol{U}_{q,\boldsymbol{t}}^T \boldsymbol{U}_{q,\boldsymbol{t}} + \boldsymbol{S}_q^{-1} \boldsymbol{A} \boldsymbol{S}_q^{-1}\right)}_{=\boldsymbol{\Sigma}_U^{-1}} \boldsymbol{S}_q \end{aligned} \tag{8}$$

$$\boldsymbol{\mu_w} = \sigma^{-2} \boldsymbol{\Sigma_w} \boldsymbol{X}_{\boldsymbol{t}}^T \boldsymbol{y_t} = \boldsymbol{S}_q^{-1} \underbrace{\sigma^{-2} \boldsymbol{\Sigma}_U \boldsymbol{U}_{q,\boldsymbol{t}}^T \boldsymbol{y_t}}_{=\boldsymbol{\mu}_U} \tag{9}$$

$$\boldsymbol{C} = \sigma^{-2} \left(\boldsymbol{I}_n + \boldsymbol{U}_{q,\boldsymbol{t}} \boldsymbol{S}_q^2 \boldsymbol{A}^{-1} \boldsymbol{U}_{q,\boldsymbol{t}}^T\right). \tag{10}$$

In fact, model residual $\boldsymbol{y_t} - \boldsymbol{X_t \mu_w}$ is equal to $\boldsymbol{y_t} - \boldsymbol{U}_{q,\boldsymbol{t}} \boldsymbol{\mu}_U$, which corresponds to the use of predictors $\boldsymbol{U}_{q,\boldsymbol{t}}$ instead of $\boldsymbol{U}_{q,\boldsymbol{t}} \boldsymbol{S}_q$ such that each of regularization matrix $\boldsymbol{A}$'s components α_m is substituted with $\alpha_m \lambda_m^{-2}$.

Using the same formulation as Tipping, parameters $\{\sigma^2, \alpha_0, \alpha\}$ can be estimated using a type-II maximum likelihood approach, i.e., the maximization of the evidence $N(\boldsymbol{y_t}|\boldsymbol{0}, \boldsymbol{C})$ (e.g., see [29] and [32] for more details). Predictors $\boldsymbol{X_t}$ are replaced by decomposition $\boldsymbol{U}_{q,\boldsymbol{t}} \boldsymbol{S}_q$, and the update models for the parameters become

$$(\hat{\alpha}_0)^{\text{new}} = \frac{\gamma_0}{\boldsymbol{\mu}_{U,0}^2} \tag{11}$$

$$(\hat{\alpha})^{\text{new}} = \frac{\sum_{i=1}^{q} \gamma_i}{\sum_{i=1}^{q} \boldsymbol{\mu}_{U,i}^2 \boldsymbol{S}_{q,ii}^{-2}} \tag{12}$$

$$(\hat{\sigma}^{-2})^{\text{new}} = \frac{N - \sum_{i=0}^{q} \gamma_i}{(\boldsymbol{y_t} - \boldsymbol{U}_{q,\boldsymbol{t}} \boldsymbol{\mu}_U)_{\boldsymbol{t}}^{T(\boldsymbol{y}} - \boldsymbol{U}_{q,\boldsymbol{t}} \boldsymbol{\mu}_U)} \tag{13}$$

with $\gamma_i = 1 - \hat{\alpha}_i \lambda_i^{-2} \boldsymbol{\Sigma}_{U,ii}$ ($\lambda_0 = 1$). Iterating (11)–(13) and updating $\boldsymbol{\Sigma}_U$ and $\boldsymbol{\mu}_U$ until convergence is achieved, optimal estimates σ_*^2, α_{0*}, and α_* can be found.

The prediction of the rest of the plots, i.e., $\boldsymbol{v}_j$, $j = 1, \ldots, N-n$, is given as the following distribution:

$$p\left(\boldsymbol{y}_{\boldsymbol{v}_j}|\boldsymbol{y_t}, \alpha_*, \sigma_*^2\right) = \int N\left(\boldsymbol{y}_{\boldsymbol{v}_j}|\boldsymbol{X}_{\boldsymbol{v}_j} \boldsymbol{w}, \sigma_*^2\right) N(\boldsymbol{w}|\boldsymbol{\mu_w}, \boldsymbol{\Sigma_w}) d\boldsymbol{w} \tag{14}$$

where $\boldsymbol{\mu_w} = \boldsymbol{S}_q^{-1} \boldsymbol{\mu}_U$, and $\boldsymbol{\Sigma_w} = \boldsymbol{S}_q^{-1} \boldsymbol{\Sigma}_U \boldsymbol{S}_q^{-1}$. Since the right-hand-side distributions are normal, the prediction of a new plot also follows a normal distribution, i.e.,

$$p\left(\boldsymbol{y}_{\boldsymbol{v}_j}|\boldsymbol{y_t}, \alpha_*, \sigma_*^2\right) = N\left(\boldsymbol{y}_{\boldsymbol{v}_j}|\hat{\boldsymbol{y}}_{\boldsymbol{v}_j}, \hat{\sigma}_{\boldsymbol{v}_j}^2\right) \tag{15}$$

with

$$\hat{\boldsymbol{y}}_{\boldsymbol{v}_j} = \boldsymbol{U}_{q,\boldsymbol{v}_j} \boldsymbol{\mu}_U \tag{16}$$

$$\hat{\sigma}_{\boldsymbol{v}_j}^2 = \sigma_*^2 + \boldsymbol{U}_{q,\boldsymbol{v}_j} \boldsymbol{\Sigma}_U \boldsymbol{U}_{q,\boldsymbol{v}_j}^T. \tag{17}$$

Thus, the proposed model is basically a linear ridge regression model, where the predictors of the training set consist of the rows of left singular vector $\boldsymbol{U}_q$ that represent the field sample plots and the diagonal regularization matrix, i.e., $\boldsymbol{A}\boldsymbol{S}^{-2}$ that defines which predictors are allowed to have nonzero parameter values. Regularization value $\alpha \lambda_m^{-2}$ is smallest for the orthogonal predictors that explain well the original data variation (with a large singular value λ_m), and it is largest for those that do not explain much of the data variation (with a small singular value λ_m). The common parameter α retains the given proportions between different orthogonal predictors throughout the model. With a small training data set, the new model, i.e., UROLS$_q$, does not treat each candidate predictor uniformly. Instead, the regression parameters associated with different predictors are searched in a manner that favors the parameters associated with the orthogonal predictors that explain well the original data.

III. Verification Procedure

In this paper, the estimation algorithm discussed in Section II is used to verify the effect of thinning the field sample plot

grid in the inventory areas. The number of field sample plots is reduced, and the training set is chosen from all the N_{cand} candidate plots available in the area. To produce a realistic set of field sample plots, three different approaches to field sample plot selection are tested, i.e., a selection criterion based on the systematic distribution of the plots in the geographic distance space, a selection criterion based on the systematic distribution of the plots in the orthogonal predictor space, and a selection criterion based on the minimization of the error variance of linear model predictions, respectively. The measurements of the forest inventory variables and the q SVD components of LiDAR predictors in the selected field sample plots serve as the training set of linear models. In this paper, we verify the prediction accuracy of the linear model with tSVD predictors (OLS_q) and that of the linear model with tSVD predictors and uniform regularization (UROLS_q). The effect of truncation according to a given explanation rate P of the original data in the prediction accuracy (the root MSE (RMSE) and the bias) is verified with all available data for both models. In addition, the effect of thinning field measurements on the prediction accuracy is described.

A. Field Sample Plot Selection

The selection criterion of the field sample plots is given with a mathematical function called the utility function. It is a function of the index vector $\boldsymbol{t}$ of the candidate subset of n plots chosen from a total of N_{cand} plots, i.e., $u = u(\boldsymbol{t})$. The shape of the utility function depends on the goal of the selection procedure, such that the goal is achieved by maximizing the utility function.

In the field sample plot selection step of the inventory procedure, there is no prior knowledge of which predictors are important in the model of the given forest inventory variable(s) since no field measurements are available. Thus, all the predictors are treated with equal weight, i.e., all orthogonal candidate predictors are scaled identically.

1) Designs Based on Maximal Spread of Data: The maximal spread of the field sample plots in the geographic distance space or the predictor space is described in this paper using the MaxiMin approach (e.g., see [35] and [36] for details). It is a space-filling design covering the given feature space as widely as possible.

In the MaxiMin spread design, the feature space can be the geometrical space or the space defined by the predictors. The aim of the design is to include sample plots from the full range of values of the given space. To achieve this goal, the distance between plots is maximized in the plot selection, i.e., selecting plots that are as far as possible from all the other selected plots. Thus, the MaxiMin design utility function is the minimum distance (in the given coordinates) among the selected plots, and the aim is to maximize this distance.

The distance between plots i and j can be defined as the squared Euclidean distance as follows:

$$d(i,j)^2 = \sum_{l=1}^{L} (\boldsymbol{Z}_{i,k} - \boldsymbol{Z}_{j,k})^2 \tag{18}$$

where L is the number of distinct predictors in the given feature vector $\boldsymbol{Z}$. $\boldsymbol{Z}$ can be either LiDAR predictors or spatial variables. For the spatial case, L is the number of spatial coordinates (generally two or three); for the predictor space case, it is the number of predictors used. The design utility function for the maximization problem becomes

$$u_{\mathrm{MaxiMin},z}(\boldsymbol{t}) = \min d(i,j)^2, \quad i,j \in \boldsymbol{t} \tag{19}$$

where i and j are two distinct plots that are both included in the index vector $\boldsymbol{t}$ of the given candidate selection consisting of n plots.

Generally, if the design is only based on the geographic distance information, the maximization of this utility function results in an approximately regular grid over the study area. Thus, the MaxiMin design with spatial coordinates $\boldsymbol{Z} = \boldsymbol{S}$, i.e., $u_{\mathrm{MaxiMin},s}(\boldsymbol{t})$, can be used to set the plot selection that covers the spatial area as widely as possible. Alternatively, one could consider distributing sample plot locations to cover the space defined by the q orthogonal predictors, i.e., $\boldsymbol{U}_q$, and use the MaxiMin design based on predictor matrix $\boldsymbol{Z} = \boldsymbol{U}_q$, i.e., $u_{\mathrm{MaxiMin},u}(\boldsymbol{t})$.

2) Design Based on Minimization of Prediction Uncertainty: The minimization of the linear model prediction uncertainty is based on the given variance estimator, i.e., (17). The estimator contains two components, where the first, i.e., σ_*^2, describes the model residual variance, and the second describes the error caused by the uncertainty in the parameter estimates. The first component is unknown at the phase of the field sample plot selection and is ignored at this step. In the second component, there are parameters σ_*^2 and $\boldsymbol{A}$ included in Σ_U. Since there is no prior knowledge of the importance of the predictors in the linear model, the prior distribution of the regression model parameters is assumed to be noninformative. Thus, the initial assumption of variance of the prior values α_i^{-1} is set large, i.e., $\alpha_i \to 0$, and component $\boldsymbol{A}\boldsymbol{S}_q^{-2}$ is negligible. Now, residual variance σ_*^2 serves as a constant scaling factor, which can be set to 1 in this formula. The utility function is thus

$$u_{\mathrm{LIN}}(\boldsymbol{t}) = -\sum_{j=1}^{N_{\mathrm{cand}}-n} \left(\boldsymbol{U}_{q,\boldsymbol{v}_j} \left(\boldsymbol{U}_{q,\boldsymbol{t}}^T \boldsymbol{U}_{q,\boldsymbol{t}} \right)^{-1} \boldsymbol{U}_{q,\boldsymbol{v}_j}^T \right). \tag{20}$$

The utility function is maximized; thus, we use the negative values of the minimized variance model.

The linear-model-based utility function is a summation of all the prediction error variance values in the given verification plots. The maximization of the given utility function minimizes the average error variance.

3) Utility Function Optimization: The maximization of utility function $u(\boldsymbol{t})$ as a function of n plot selection indexes $\boldsymbol{t}$ is a computationally hard combinatorial problem. Depending on the set of plots, utility function $u(\boldsymbol{t})$ has different values. With a large number of candidate plots, i.e., N_{cand}, finding the optimal subset of n plots would take a long time. In this paper, the maximization problem is solved with simulated annealing (e.g., see [37] and [38] for more details). The method finds a local optimum, but there is no guarantee that the solution is the global optimum of the problem. However, the solutions are

nearly optimal, and the repetition of the optimization procedure can be used to check that the solution is appropriate.

The basic idea in simulated annealing is to start with an initial design, i.e., $\boldsymbol{t}_{\text{init}}$, of n randomly chosen plots and then generate a new candidate design with an equal amount of plots, i.e., $\boldsymbol{t}_{\text{new}}$, and a rule to accept the new design. At each step, a new candidate design $\boldsymbol{t}_{\text{new}}$ is formed by deleting one randomly chosen plot from the initial design and replacing it by a plot that is randomly chosen from the set on the $N_{\text{cand}} - n$ plots that are not included in the initial design. The acceptance rule for the new candidate design depends on the utility function values of the initial design and the new candidate design. In the maximization problem, if utility function value $u(\boldsymbol{t}_{\text{new}})$ is larger than $u(\boldsymbol{t}_{\text{init}})$, the new candidate design is accepted. A new design with a smaller utility function value is only accepted with probability $\exp(-(u(\boldsymbol{t}_{\text{init}}) - u(\boldsymbol{t}_{\text{new}}))/T)$, where T is a user-specified positive constant, which defines how much the utility function is allowed to alter. This constant is made to decrease in the course of maximization. More details of its definition can be found in the given references. The value of T is reduced by some factor, i.e., 0.9 in this paper for every 100 iterations. If the new design is accepted, it becomes the new initial design, i.e., $\boldsymbol{t}_{\text{init}} = \boldsymbol{t}_{\text{new}}$, and the search continues from it. If the new candidate design is not accepted, a new random swap of plots from the original initial design is tested.

Simulated annealing requires a lot of plot swap attempts, and in this paper, the upper limit is set to 30 000; to control that, the true optimum is found, and it may take several independent solutions to verify it.

B. Material

A range of ten different test cases is used to verify the performance of the given method in remote-sensing-data-based forest inventory. The test cases are located in the USA and Finland, and the model performance is verified with distinct approaches in the two continents due to differences in data sizes.

1) USA: Field Measurements: The American data used in this paper are located in three different areas in the USA. The Fraser Experimental Forest (FEF) is located in central Colorado near the town of Fraser. The Marcell Experimental Forest (MEF) is located in northern Minnesota. The Niwot Long-Term Ecological Research Site (NIWOT) is located near the town of Nederland, Colorado.

The tree species at FEF primarily consist of *Abies lasiocarpa* and *Picea engelmannii* at higher elevations and *Pinus contorta* at lower elevations. The climate at FEF is characterized by cold and relatively long winters, with mean annual temperature and precipitation of 0 °C and 737 mm, respectively. The tree species at MEF include mixed upland forests and peat lands. Upland forests are generally dominated by *Populus tremuloides* and *grandidentata*, but they contain substantial components of *Betula papyrifera, Pinus resinosa, Pinus strobus*, and *Pinus banksiana*. Lowland tree species include *Larix laricina, Picea mariana, Fraxinus nigra*, and *Thuja occidentalis*. The climate at MEF is subhumid continental, with a mean annual precipitation of 785 mm, a mean annual temperature of 3 °C, and air temperature extremes of −46 °C and 38 °C. The tree species in inventory site NIWOT primarily include a mix of *Abies lasiocarpa, Picea engelmannii*, and *Pinus contorta*, with minor components of *Pinus flexilis* and *Populus tremuloides*. The mean annual temperature and precipitation are 4 °C and 800 mm, respectively.

The field data at each site were collected using methods similar to the forest inventory and analysis style plot design [39]. Each plot location consists of four subplots with a radius of 8–10 m (depending on the site), with one subplot located in the center and the three others located 35 m away from the plot center at 0°, 120°, and 240°, respectively. Additional plots only consisting of one subplot were also used in this analysis. FEF, MEF, and NIWOT resulted in a total of 60, 99, and 62 subplots, respectively. Within each subplot, individual tree diameters at breast height and the height for both live and dead trees were measured and used in species-specific allometric models to estimate the AGB (stem, branch, and foliage biomass). Additional details about the field measurements and allometric equations used for biomass estimation are available in [40]. The individual tree AGB estimates were totaled for each subplot and converted to the AGB in megagrams per hectare. The AGB measurements were then square root transformed to better approximate a normal distribution before model fitting.

2) USA: LiDAR Data: The field measurement data for the three test sites are accompanied with LiDAR measurements covering the test areas. The pulse densities of the LiDAR measurements are unknown. The height aboveground measurements for the first-return pulses were calculated by subtracting the point elevations from a digital terrain model constructed from the LiDAR data. The coverage area of the LiDAR data was then parsed into 20 m × 20 m grid cells (chosen to approximate the area of a subplot), and return density profiles were constructed for each cell by summing the number of returns over the cell into 1-m height bins. The LiDAR echoes that occurred in the 0–1-m height class were removed from the data set to limit the influence of ground returns. To correct for the variation in the number of returns over each grid cell due to the flight path overlap and the laser scan angle, each profile was normalized by dividing each bin count value by the maximum bin count value. This process was repeated to obtain the LiDAR density profiles for the first returns over the subplot locations where the AGB was measured.

The original candidate LiDAR predictors $\boldsymbol{z}_i$ used in the regression model in this paper were the 10th, 20th, 25th, 30th, 40th, 50th, 60th, 70th, 75th, 80th, 90th, 95th, and 100th height percentiles of the first-return LiDAR pulses of each plot i. Thus, the predictors over the whole area are stored in an $N \times 13$ matrix $\boldsymbol{Z}$. The predictors are centered and normalized to a zero-mean unit variance normal distribution. The spatial coordinates of the center of each plot, i.e., s_{xi} and s_{yi}, are also known.

3) Finland: Field Measurements: In this paper, seven separate inventory sites in different parts of Finland were estimated. They are located at Matalansalo, Juuka, Loppi-Janakkala, Pello, Lieksa, Kuhmo, and Karttula. The tree species of the different sites contain Scots pine (*Pinus sylvestris*), Norway spruce (*Picea Abies*), and hardwoods, which mostly comprises birch (*Betula pendula* or *Betula pubescent*). The percentile part of

pine in the different inventory sites was 53.2%, 67.2%, 45.3%, 38.6%, 61.1%, 64.8%, and 29.1%, respectively. For spruce, the percentage are 34.5%, 21.7%, 41.5%, 34.1%, 24.1%, 24.0%, and 45.0%, respectively, and for hardwoods, they are 12.3%, 11.0%, 13.2%, 27.3%, 14.8%, 11.2%, and 25.9%, respectively. The annual heat sums (degree days) are 1150, 1000, 1250, 850, 1050, 900, and 1100, respectively.

The field sample plots on the inventory sites were selected with a number of different sampling strategies. Some sites were equipped with a regular sample plot grid, and some others were equipped with regularly sampled plot clusters; others were equipped with randomly selected plots. These differences were ignored in the estimation process as they are likely to vary in operational use as well. In the given inventory sites, in the order given previously, a total of 472, 511, 441, 553, 483, 470, and 538 field sample plots, respectively, are measured (the data set size N of the different inventory sites). Field sample plots were always circular plots with a 9-m radius. They were positioned at first with a handheld Global Positioning System, and the exact position of the plot center was later calculated during measurement and differentially corrected offline. This method generally ensured a position error of less than 1 m, which has been deemed adequate to align the LiDAR data and the field plots so that their areas overlap to a degree exceeding 90%. In each inventory site, five different forest inventory variables were measured, i.e., the median tree diameter, the median tree height, the stem number, the basal area, and the volume.

4) Finland: LiDAR Data: The test data for the test sites also consist of LiDAR measurements from the area. The LiDAR scanning of the different areas was conducted from 2004 to 2008. Three different types of scanners were used, i.e., the Optech ALTM 3600, the Leica ALS-50, and the Leica ALS-60. The flying height varied between 700 and 2000 m, and the scanner pulse frequency varied between 58 900 and 125 100 Hz. The LiDAR data were clipped to the plot extent before extracting the LiDAR predictors from it.

The set of candidate predictors derived from the LiDAR measurements for each sample plot used in the estimation of each forest inventory variable is similar to the set that we used in [41]. It consists of the percentile points and cumulative percentile parts of the first and last pulse heights of nonground hits (a height above 2 m), the percentile intensities of the first and last pulse intensities of nonground hits, the mean of the first pulse heights over 5 m, the standard deviation of the first pulse height, and the number of measurements less than 2 m of the first and last pulse heights divided by the total number of the same measurements of each plot. These 38 candidate predictors for each plot of the site were stored in the $N \times 38$ matrices.

In this paper, the sample plots of the test areas are used as the candidate plots for the field plot selection and as the plots of interest in the forest variable prediction. Thus, $N_{\text{cand}} = N$ for the test cases.

C. Error Validation

To achieve optimal prediction results, i.e., the baseline (BL), of the given methods OLS_q and UROLS_q, we use all the available data, the leave-one-out (LOO) cross-validation procedure for the Finnish data, and the leave-two-out (LTO) cross-validation procedure for the smaller sized American data.

In LOO, each plot i, $i = 1, \ldots, N$, at the time is left out as the verification plot, and the other $N-1$ plots are used as the training set of the study site. The prediction for each plot is performed separately, but the error is calculated using the predictions $\tilde{\boldsymbol{y}}_i$ of all the plots $i = 1, 2, \ldots, N$ as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\tilde{\boldsymbol{y}}_i - \boldsymbol{y}_i)^2}{N}} \qquad D = \frac{\sum_{i=1}^{N} (\tilde{\boldsymbol{y}}_i - \boldsymbol{y}_i)}{N} \tag{21}$$

where RMSE is the RMSE of the predictions, and D is the mean deviation. The total mean error, i.e., the MSE, is $\text{MSE} = \sqrt{\text{RMSE}^2 + D^2}$. The relative errors are

$$\text{RMSE}\% = \frac{\text{RMSE}}{\overline{\boldsymbol{y}}} \cdot 100\% \qquad D\% = \frac{D}{\overline{\boldsymbol{y}}} \cdot 100\% \tag{22}$$

where $\overline{\boldsymbol{y}} = (\sum_{i=1}^{N} \boldsymbol{y}_i)/(N)$ is the mean of the true values of the forest inventory variables.

The size of the American test data is relatively small, and the LOO cross-validation procedure may give unstable results. Thus, we use the LTO procedure to verify the prediction results. It is similar to LOO, except that, instead of one, two plots at a time are left as the validation plots, and the rest of the $N-2$ plots serve as the training set of the study site. That is, we go through all the possible pairs (in random order), leading to $N \times (N-1)/2$ repetitions, e.g., 1770 in the case of FEF. The errors are then calculated for the resulting $N \times (N-1)/2$ predictions $\tilde{\boldsymbol{y}}_i$ comparing with the corresponding set of forest inventory variable values.

When the number of sample plots is diminished, $n < N$ field sample plots are selected from the N candidates, and the model parameters are estimated using the data of those plots as the training set. Similar to the procedure discussed previously, we use the LOO and LTO cross-validation procedures with a small modification: after selecting the verification plot(s) of each repetition, the training set of n plots is selected from the rest, i.e., the $N-1$ (LOO) and $N-2$ (LTO) plots that serve as the candidate plot set of the study site. After the repetitions, the errors are calculated as given earlier.

In the case of LTO, the validation procedure is very time consuming because it includes over a thousand repetitions. However, after some hundreds of repetitions, the results generally converge to a fixed value. Thus, we can stop the verification procedure when the estimates are such that the change in relative errors (RMSE% and D%) has not changed more than 0.1 units for 50 repetitions and such that the corresponding optimal estimation values verified earlier for a full data set (the BL) have also converged such that the relative errors are at most 0.1 units from the BL error values. The errors are then calculated for the resulting N_{iter} predictions $\tilde{\boldsymbol{y}}_i$ comparing with the corresponding set of forest inventory values.

In the Finnish data, there are multiple forest inventory variables that are predicted with the same set of field sample plots and the same model formulation just by adjusting the model parameters to fit the given data. In this case, $K = 5$ forest inventory variables are predicted. In the Finnish operational

forest inventory, all these variables are required. Thus, an overall estimate of the accuracy of the model is given by averaging over the MSEs that are normalized according to the forest inventory variables' optimal LOO solution MSE, i.e., BL $\mathrm{MSE}_{\mathrm{BL}}$, as follows:

$$\mathrm{MSE}_{\mathrm{rel,tot}} = \frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{MSE}_k}{\mathrm{MSE}_{\mathrm{BL},k}} \tag{23}$$

and for one specific forest inventory variable, i.e., the volume, the relative estimate is

$$\mathrm{MSE}_{\mathrm{rel,vol}} = \frac{\mathrm{MSE}_{\mathrm{vol}}}{\mathrm{MSE}_{\mathrm{BL,volume}}}. \tag{24}$$

In this paper, all the forest inventory variables were considered equally important. If $\mathrm{MSE}_{\mathrm{rel}}$ (tot or vol) is larger than 1, the average error is $(\mathrm{MSE}_{\mathrm{rel}} - 1) \times 100\%$ larger than the error of the optimal LOO solutions, i.e., the relative additive error due to the diminished number of field sample plots.

Different designs are labeled as LIN, $\mathrm{MM}_x x$, and MM_s, corresponding to the approaches where the optimal design is selected with the aim to minimize the prediction error variance, i.e., utility function (20), and the aim to fill the predictor and the spatial space with a nearly regular grid, i.e., utility function (19), with $\boldsymbol{Z} = \boldsymbol{X}$ and $\boldsymbol{Z} = \boldsymbol{S}$. The prediction accuracy of the two modeling methods is verified using the field sample plots selected with each of these designs.

IV. Results and Discussion

The accuracy of model-based estimation in forest inventory basically depends on two aspects, i.e., how well the predictors explain the variability of the response (the level of the signal of the response in the remote sensing data) and how well the model parameters are estimated. With a small training set, i.e., a sparse set of field sample plots (30–100) in the forest inventory, both aspects affect the resulting predictions.

Before result evaluation, the convergence of the LTO and LOO methods was checked. In each validation case, the results converged within some hundreds of LOO/LTO repetitions. Despite this, in case of the Finnish data, all the N repetitions in the LOO method were performed. In the case of LTO, the repetitions were only performed until the convergence was achieved, as discussed previously. The reason for this can be seen in Fig. 1, where an example of the relative error value convergence in the LTO cross-validation procedure is shown.

A. Effect of Truncation Level

Using the orthogonal predictors derived from the original predictors with the SVD, the number of predictors can be reduced. However, how does one decide on the optimal number of orthogonal predictors to retain? The explanation rate, i.e., P, in (2) only defines how well the variability of the given set of original and correlated predictors is explained if the number of orthogonal predictors is reduced. However, some of the variability that explains the responses may be excluded if too many orthogonal predictors are left out from the model. On the other hand, if the number of predictors is large compared with the number of field plots, the model parameter estimation becomes uncertain, producing less reliable predictions.

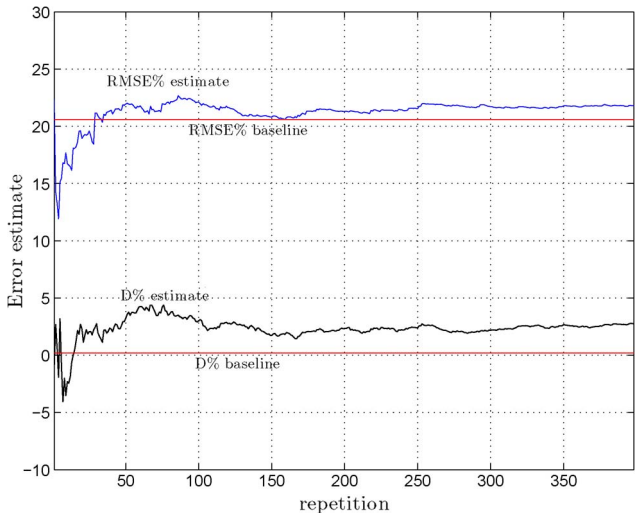

Fig. 1. Inventory area FEF: RMSE% and $D\%$ values of the UROLS_q method as functions of the LTO repetitions (design MM_s, $n = 18$, and $P = 0.95$).

This effect is verified using the LOO/LTO estimates with full data and varying the rate of variability explanation P (see Tables I and II). The RMSE% and $D\%$ results are given for the two linear models, i.e., OLS_q and UROLS_q. None of the results have large deviation ($D\%$), but the RMSE% results vary depending on the model and the number of used orthogonal predictors. With a small number of predictors (a low explanation rate P), the RMSE% of the predictions increases since the given predictors do not explain the response variability as well as the model with more orthogonal predictors. However, if the number of predictors becomes too large (a high P), the model parameter estimation may suffer due to overfitting, and the RMSE% of the predictions again increases. This can be seen particularly in the results given in Table I since the number of input data, i.e., N, is small in the American data. In the Finnish data set, the effect of overfitting is not as clear.

The problems concerning the large number of used predictors (with a relatively small number of field plots) can be overcome using the UROLS_q model, which gives similar estimates independent of the rate between the number of predictors and the input data size.

Why does the UROLS_q method give more precise results? This effect is illustrated in Fig. 2, where the singular values of the orthogonal predictors and the absolute values of the parameters for an LTO solution of the FEF square-rooted AGB derived with $P = 0.95$ and $q = 9$ predictors are shown. The orthogonal predictors that explain the largest percentage of variability in the original predictors also have the largest absolute model parameter values in the model.

In UROLS_q, the orthogonal predictors with large singular values are favored, and the predictors with small singular values are not allowed to become as important in the model. Thus, in the case of potential overfitting, the effective number of predictors is less than the real number of predictors, and overfitting is avoided.

TABLE I
LTO RESULTS FROM STUDY SITES FEF, MEF, AND NIWOT: RMSE% (D%) RESULTS WITH DIFFERENT PORTIONS OF USED PREDICTORS. P IS THE EXPLANATION RATIO USED, AND q IS THE CORRESPONDING NUMBER OF ORTHOGONAL PREDICTORS USED

	FEF ($N = 60$)			MEF ($N = 99$)			NIWOT ($N = 62$)		
P	q	OLS$_q$	UROLS$_q$	q	OLS$_q$	UROLS$_q$	q	OLS$_q$	UROLS$_q$
1.00	13	23.9 (0.5)	20.6 (0.2)	13	31.8 (0.2)	29.8 (-0.4)	13	19.6 (0.2)	17.6 (0.1)
0.99	12	23.3 (0.6)	20.6 (0.2)	12	30.9 (0.0)	29.8 (-0.3)	13	19.6 (0.2)	17.6 (0.1)
0.95	9	22.1 (0.5)	20.6 (0.2)	10	30.3 (-0.3)	29.8 (-0.2)	11	19.0 (0.3)	17.6 (0.1)
0.90	6	21.3 (0.4)	20.6 (0.2)	7	30.0 (-0.2)	29.7 (-0.3)	9	18.3 (0.3)	17.6 (0.1)
0.75	3	20.6 (0.1)	20.5 (0.2)	4	29.9 (-0.2)	29.9 (-0.2)	4	17.8 (0.1)	17.5 (0.1)
0.50	1	20.9 (0.2)	20.9 (0.1)	2	30.2 (-0.1)	30.3 (-0.1)	2	17.5 (0.1)	17.5 (0.1)

TABLE II
MATALANSALO ($N = 472$) LOO AND RMSE% (D%) RESULTS WITH DIFFERENT PORTIONS OF USED PREDICTORS. P IS THE EXPLANATION RATIO USED, AND q IS THE CORRESPONDING NUMBER OF ORTHOGONAL PREDICTORS USED

		Median tree diameter		Median tree height		Stem number		Basal area		Volume	
P	q	OLS$_q$	UROLS$_q$	OLS$_q$	UROLS$_q$	OLS$_q$	UROLS$_q$	OLS$_q$	UROLS$_q$	OLS$_q$	UROLS$_q$
1.00	39	14.0 (0.0)	13.8 (0.0)	8.9 (0.0)	8.7 (0.0)	27.5 (-0.1)	27.1 (-0.1)	16.7 (0.0)	16.0 (0.0)	20.8 (0.0)	19.9 (0.0)
0.99	32	13.9 (0.0)	13.8 (0.0)	8.8 (0.0)	8.7 (0.0)	27.3 (-0.1)	27.1 (-0.1)	16.5 (0.0)	16.0 (0.0)	20.5 (0.0)	19.9 (0.0)
0.95	22	13.7 (0.0)	13.8 (0.0)	8.8 (0.0)	8.7 (0.0)	27.2 (-0.1)	27.2 (-0.1)	16.0 (0.0)	16.0 (0.0)	19.8 (0.0)	19.9 (0.0)
0.90	16	14.3 (0.0)	14.2 (0.0)	8.8 (0.0)	8.8 (0.0)	27.6 (-0.1)	27.6 (-0.1)	16.1 (0.0)	16.1 (0.0)	20.2 (0.0)	20.1 (0.0)
0.75	8	14.3 (0.0)	14.3 (0.0)	8.8 (0.0)	8.8 (0.0)	28.2 (0.0)	28.2 (0.0)	16.3 (0.0)	16.3 (0.0)	20.2 (0.0)	20.1 (0.0)
0.50	3	16.2 (0.0)	16.2 (0.0)	9.4 (0.0)	9.4 (0.0)	30.7 (0.0)	30.7 (0.0)	16.7 (0.0)	16.7 (0.0)	21.5 (0.0)	21.5 (0.0)

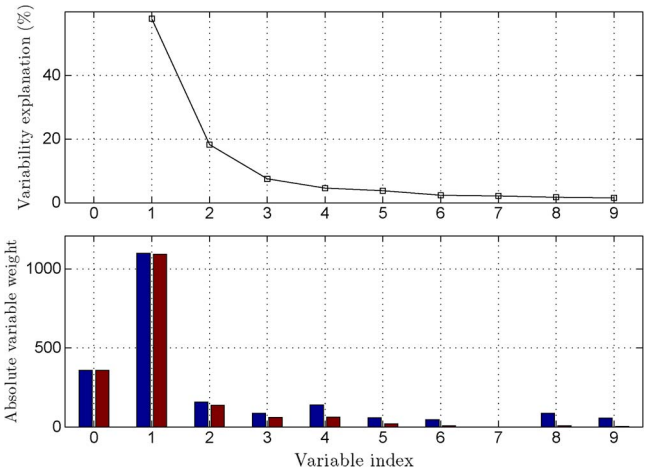

Fig. 2. FEF case, with $P = 0.95$. (Top) Percentage of the variability explanation of the orthogonal predictors. (Bottom) Absolute (left bar) OLS$_q$ and (right bar) UROLS$_q$ parameters $\boldsymbol{\mu}_U$ calculated with full data.

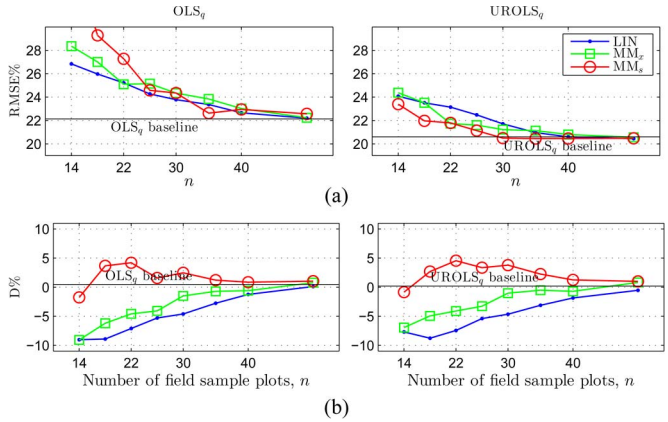

Fig. 3. Inventory area FEF: RMSE% and D% of the LOO estimates as functions of field sample plot number n and designs LIN, MM$_x$, and MM$_s$. Explanation rate $P = 95$%, and $q = 9$ orthogonal predictors were used. (a) RMSE%. (b) D%.

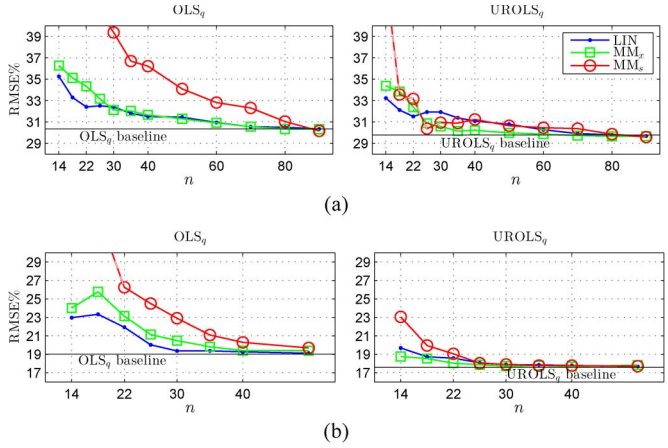

Fig. 4. RMSE% for inventory areas MEF and NIWOT. Explanation rate $P = 95$% was used for the predictors. (a) MEF. (b) NIWOT.

B. *Effect of Reduced Number of Field Sample Plots*

In this paper, the training set only consists of the selected n field sample plots. The field sample plots are selected using the LIN, MM$_x$, and MM$_s$ designs. The effect of the different designs with different values of n and the effect of regularization for the prediction precision of different forest inventory variables are shown in Figs. 3–5 (for inventory sites FEF, MEF and NIWOT, and Matalansalo, respectively).

The effects of the given designs on the prediction reliability are similar to those shown already in [16]. The difference between the performance of models OLS$_q$ and UROLS$_q$ can be seen in the RMSE% results, whereas the systematic deviation trend remains the same using both models (see Fig. 3). Using the OLS$_q$ model, the LIN design is very prone to systematic deviation but gives the smallest RMSE%, whereas the results of the MM$_s$ design are more accurate but have intolerably high RMSE% with the small sample sizes. Using the proposed model UROLS$_q$, the accuracy is approximately the same as using the OLS$_q$ model, but the RMSE% values become smaller. In fact, the RMSE% values of designs MM$_x$ and MM$_s$ are very close to that obtained with the LIN design. The same trend of

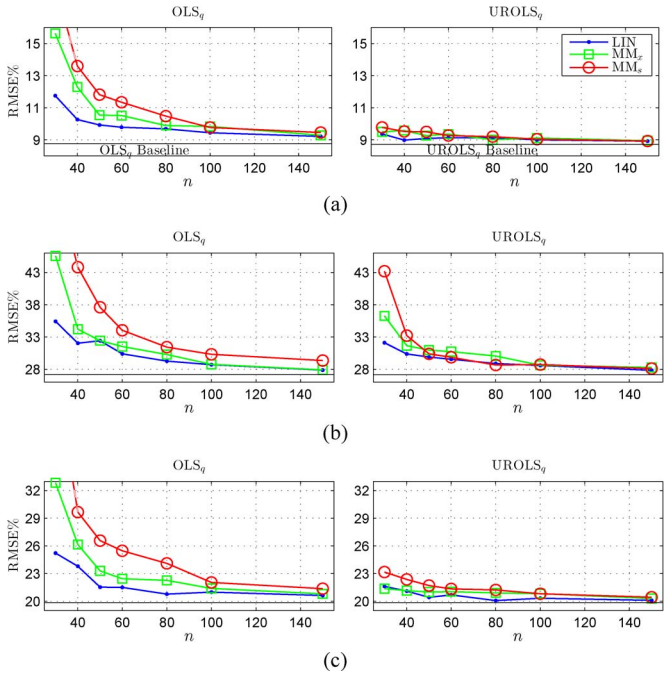

Fig. 5. Prediction RMSE% of three forest inventory variables for inventory area Matalansalo. Explanation rate $P = 95\%$ was used for the predictors. (a) Median tree height. (b) Stem number. (c) Volume.

systematic deviation of different designs can be seen in the tests of all the inventory variables in all the inventory sites. Both models OLS$_q$ and UROLS$_q$ possess approximately the same level of accuracy. Furthermore, we focus on the visualization of the RMSE% in the results of other test cases.

Independent of the design used, the predictions made with a small sample size and a relatively large number of predictors are more precise when estimated with the UROLS$_q$ model than with OLS$_q$. The problem of overfitting is overcome with the regularization that emphasizes the orthogonal predictors that most explain the original data. Thus, even with a few dozen field sample plots, the prediction precision is close to that estimated with several hundred field sample plots. The differences between the designs vanish using the proposed model. Using the UROLS$_q$ model, clear differences can be only seen with the smallest sample sizes, i.e., 14 in the American test cases and 30 in the Finnish test cases. Thus, we can also use designs that are less efficient in terms of linear model fitting but not prone to systematic deviation without losing substantial precision.

In Fig. 6, the total MSE% values compared with the BL LOO solution (MSE$_{\mathrm{rel,tot}}$) and similar relative measures for forest inventory variable volume-specific MSE% values (MSE$_{\mathrm{rel,vol}}$) are provided for two Finnish inventory sites as functions of the number of field sample plots n. The results of inventory site Matalansalo are similar to the other Finnish inventory sites, and inventory site Pello gives the worst relative MSE values. Altogether, the UROLS$_q$ method gives precise predictions even with small sample sizes. The total error among the seven Finnish inventory sites is 10%–50% worse than the optimal solution, even with only 30 field sample plots. With 50 sample plots, the total error is only about 5%–15% more, which is independent of the design used.

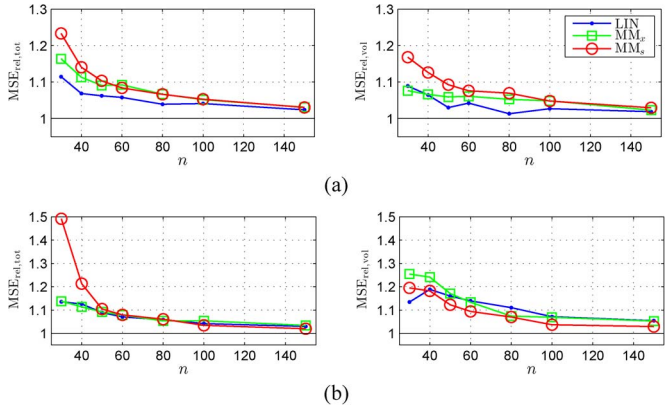

Fig. 6. MSE$_{\mathrm{rel}}$ results of the UROLS$_q$ method for inventory sites Matalansalo and Pello in Finland. Predictor explanation rate $P = 0.95$ was used. (a) Matalansalo. (b) Pello.

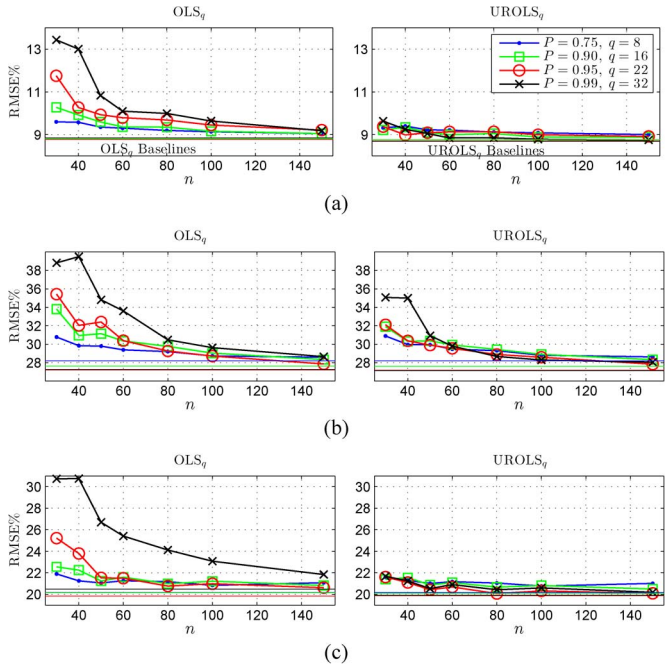

Fig. 7. RMSE% of Matalansalo forest inventory variables median tree height, stem number, and volume for predictions estimated with the OLS$_q$ and UROLS$_q$ methods using a different number of predictors. Note that, if $q \geq n$, only $q = n - 5$ most important orthogonal predictors are used in the field sample plot selection and prediction model. (a) Median tree height. (b) Stem number. (c) Volume.

However, how does the choice of the explanation rate, i.e., P, of the used orthogonal predictor set affect the solutions when the number of field sample plots is diminished? The RMSE% results of the forest inventory variables measured in inventory area Matalansalo in Finland, resulting from sampling and modeling the data with the LIN design using different number of orthogonal predictors and being chosen with different values of explanation rate P, are shown in Fig. 7. This figure shows that, with a small number of predictors and a sampling design based on the minimization of the linear model's prediction error variance, the OLS$_q$ and UROLS$_q$ errors are equally independent of the number of field sample plots n. However, with a larger number of orthogonalized predictors and a small number of

field sample plots, the regularized regression model produces substantially smaller RMSE% values. As long as $n \gtrsim q + 10$ holds, the predictions are nearly optimal even with a predictor explanation rate of 99%. With this high explanation rate, almost all the forest inventory variable information included in the predictors is utilized.

C. Model-Based or Design-Based Inference With Reduced Number of Plots

The plot selection criteria presented in this paper do not guarantee that the plot sample constitutes a probability sample of the forest nor do they automatically produce variance estimates. Indeed, the purpose of this paper has been to demonstrate that it is possible to obtain reasonably accurate and precise estimates with LiDAR-model-based estimation methods with a rather modest number of sample plots. Constructing appropriate sampling designs is a topic for a separate study. This paper should not be based on a preexisting comprehensive sample but rather on a bootstrapping approach where the sample is incrementally collected based on a chosen sampling criterion.

This paper uses model-based statistical inference as its basis, as elaborated in [12] and [42]. If the model used possesses a systematic error with respect to the total population statistics, then the total population statistics summed up from such model-based census estimates will also contain that same systematic error, unlike the inference on design-based estimates (e.g., see [12] and [43]). The regression methods employed in this paper are so constructed that they will not have a systematic error with respect to the sample plot data used for model parameter estimation. Since we have validated these model-based forest inventory variable predictions against a large sample available on all test sites, the model lack of fit can be judged if that large sample has been collected probabilistically. This is the case with some of the test sites, such as Matalansalo and Juuka, but not on most of the sites. Moreover, in any case, none of the methods proposed here retain the probability sample property when the number of plots is reduced.

The plot selection criteria discussed here could be used to estimate weight parameter functions on which to base a weighted one-level or two-level sample, starting from a design-based virtual sample. Such a sample could possess the desirable features of being a probabilistic sample of the forest and contains a variance estimate (e.g., see [12], [43], and [44]), but this will call for more investigation.

Notwithstanding such weighted sampling, it is advisable to expend some of the savings in the field work cost obtained by the criteria presented here into collecting an independent validation sample that is not used in model parameter estimation. Such a sample can be used to assess if the chosen training set has resulted in a sufficiently accurate and precise model since comparing estimates from the model with the independent sample will create an error distribution that should have zero mean and a sufficiently small standard deviation. If the model fails this test, the independent sample can be added to the training set and the model building and validation sample plot collection processes iterated until a sufficiently precise and accurate model has been obtained.

V. Summary

A small training set size paired with an almost equal number of predictors in linear models generally causes an additional uncertainty in predictions. Problems arising from the use of a large number of possibly correlated predictors with respect to the number of data points may be overcome by only using the most important orthogonal predictors derived from the original data with the SVD. However, discarding part of the data variability may cause loss of important information, and the prediction accuracy may suffer. Using a regularized model combined with orthogonal SVD predictors reduces this effect. The regularized model proposed here prefers those orthogonal predictors in the model that have large singular values, i.e., it explains most of the variability in the original predictor set but allows the use of less important predictors, depending on the model precision and the data size.

Using regularized linear regression based on SVD predictors improves the reliability of remote-sensed-data-based forest inventory models and thus enables accurate assessments to be generated from LiDAR data with only a few dozen field sample plots. To further improve the model, it should be also extended to cover the spatial correlation of the residuals.

Acknowledgment

The authors would like to thank M. Maltamo and P. Packalén of the University of East Finland and V. Leppänen, H. Holm, J. Peuhkurinen, and M. Gunia of Arbonaut Ltd., for providing the Finnish data and features of this paper.

References

[1] L. Sängstuvall, I. Gillgren, O. Lindgren, E. H. Iversen, and T. Brethvad, "Forest inventory using LiDAR at Bergvik Skog AB," in *Proc. SilviLaser*, Vancouver, BC, Canada, Sep. 16–19, 2012, p. 93.

[2] G. P. Asner *et al.*, "High-fidelity national carbon mapping for resource management and REDD+," *Carbon Balance Manage.*, vol. 8, no. 7, p. 7, Jul. 2013.

[3] M. Maltamo *et al.*, "Airborne laser scanning based stand level management inventory in Finland," in *Proc. 11th Int. Conf. LiDAR Appl. Assessing Forest Ecosyst. SilviLaser*, Hobart, TAS, Australia, Oct. 16–20, 2011, pp. 1–10.

[4] R. E. McRoberts, "Estimating forest attribute parameters for small areas using nearest neighbors techniques," *Forest Ecol. Manage.*, vol. 272, pp. 3–12, May 2012.

[5] E. Næsset, "Determination of mean tree height of forest stands using airborne laser scanning data," *ISPR J. Photogramm. Remote Sens.*, vol. 52, no. 2, pp. 49–56, Apr. 1997.

[6] S. Magnussen, E. Næsset, and T. Gobakken, "Reliability of LiDAR derived predictors of forest inventory attributes: A case study with Norway spruce," *Remote Sens. Environ.*, vol. 114, no. 4, pp. 700–712, Apr. 2010.

[7] N. A. C. Cressie, *Statistics for Spatial Data*. New York, NY, USA: Wiley, 1993.

[8] J. P. Chilés and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*. New York, NY, USA: Wiley, 1999.

[9] A. O. Finley, S. Banerjee, and R. E. McRoberts, "Hierarchical spatial models for predicting tree species assemblages across large domains," *Ann. Appl. Stat.*, vol. 3, no. 3, pp. 1052–1079, Sep. 2009.

[10] J. E. Means *et al.*, "Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the Western Cascades of Oregon," *Remote Sens. Environ.*, vol. 67, no. 3, pp. 298–308, Mar. 1999.

[11] J. L. Rooker Jensen, K. S. Humes, T. Conner, C. J. Williams, and J. DeGroot, "Estimation of biophysical characteristics for highly variable mixed-conifer stands using small-footprint LiDAR," *Can. J. Forest Res.*, vol. 36, no. 5, pp. 1129–1138, May 2006.

[12] R. E. McRoberts, E. Næsset, and T. Gobakken, "Inference for lidar-assisted estimation of forest growing stock volume," *Remote Sens. Environ.*, vol. 128, pp. 268–275, Jan. 2013.
[13] V. Junttila, M. Maltamo, and T. Kauranne, "Sparse Bayesian estimation of forest stand characteristics from airborne laser scanning," *Forest Sci.*, vol. 54, no. 5, pp. 543–552, Oct. 2008.
[14] T. J. Hawbaker *et al.*, "Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design," *J. Geophys. Res.*, vol. 114, no. G2, Jun. 2009, Art. ID. G00E04.
[15] M. Maltamo, O. M. Bollandsås, E. Næsset, T. Gobakken, and P. Packalén, "Different plot selection strategies for field training data in ALS-assisted forest inventory," *Forestry*, vol. 84, no. 1, pp. 23–31, Jan. 2011.
[16] V. Junttila, A. O. Finley, J. B. Bradford, and T. Kauranne, "Strategies for minimizing sample size for use in airborne LiDAR-based forest inventory," *Forest Ecol. Manage.*, vol. 292, pp. 75–85, Mar. 2013.
[17] T. Gobakken, L. Korhonen, and E. Næsset, "Laser-assisted selection of field plots for an area-based forest inventory," *Silva Fennica*, vol. 47, no. 5, p. 943, 2013.
[18] J. C. Deville and Y. Tillé, "Efficient balanced sampling: The cube method," *Biometrika*, vol. 91, no. 4, pp. 893–912, Dec. 2004.
[19] J. E. Means *et al.*, "Predicting forest stand characteristics with airborne scanning lidar," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 11, pp. 1367–1371, Nov. 2000.
[20] E. Næsset, O. M. Bollandsås, and T. Gobakken, "Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data," *Remote Sens. Environ.*, vol. 94, no. 4, pp. 541–553, Feb. 2005.
[21] J. A. van Aardt, R. H. Wynne, and R. G. Oderwald, "Forest volume and biomass estimation using small-footprint lidar-distributional parameters on a per-segment basis," *Forest Sci.*, vol. 52, no. 6, pp. 636–649, Dec. 2006.
[22] M. J. Falkowski, J. S. Evans, S. Martinuzzi, P. E. Gessler, and A. T. Hudak, "Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA," *Remote Sens. Environ.*, vol. 113, no. 5, pp. 946–956, May 2009.
[23] A. E. Hoerl and R. W. Kennard, "Ridge regression: Applications to nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 69–82, Feb. 1970.
[24] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. San Diego, CA, USA: Academic, 1980.
[25] Y. Li, H. E. Andersen, and R. McGaughey, "A comparison of statistical methods for estimating forest biomass from Light Detection and Ranging (LiDAR)," *Western J. Appl. Forestry*, vol. 23, no. 4, pp. 223–231, 2008.
[26] E. Næsset, "Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 88–99, Apr. 2002.
[27] P. Treitz *et al.*, "LiDAR sampling density for forest resource inventories in Ontario, Canada," *Remote Sens.*, vol. 4, no. 4, pp. 830–848, Mar. 2012.
[28] V. LeMay and H. Temesgen, "Comparison of nearest neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables," *Forest Sci.*, vol. 51, no. 2, pp. 109–119, 2005.
[29] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
[30] S. Chen, E. S. Chng, and K. Alkadhimi, "Regularized orthogonal least squares algorithm for constructing radial basis function networks," *Int. J. Control*, vol. 64, no. 5, pp. 829–837, Jul. 1996.
[31] S. Chen, "Local regularization assisted orthogonal least squares regression," *Neurocomputing*, vol. 69, no. 4–6, pp. 559–585, Jan. 2006.
[32] D. J. C. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–447, May 1992.
[33] P. Xu, "Truncated SVD methods for discrete linear ill-posed problems," *Geophys. J. Int.*, vol. 135, no. 2, pp. 505–514, Nov. 1998.
[34] T. Cai, C. Ju, and X. Yang, "Comparison of ridge regression and partial least squares regression for estimating above-ground-biomass with Landsat images and terrain data in Mu Us Sandy Land, China," *Arid Land Res. Manage.*, vol. 23, no. 3, pp. 248–261, Jul. 2009.
[35] M. D. Morris and T. J. Mitchell, "Exploratory designs for computational experiments," *J. Stat. Planning Inference*, vol. 43, no. 3, pp. 381–402, Feb. 1995.
[36] M. Trosset, "Approximate maximin distance designs," in *Proc. Section Phys. Eng. Sci.*, 1999, pp. 223–227.
[37] J. Zhou, "D-optimal minimax regression designs on discrete design space," *J. Stat. Planning Inference*, vol. 138, no. 12, pp. 4081–4092, Dec. 2008.
[38] Z. Fang and D. Wiens, "Integer-valued minimax robust designs for estimation in heteroscedastic, approximately linear models," *J. Amer. Stat. Assoc.*, vol. 95, no. 451, pp. 807–818, Sep. 2000.
[39] W. Bechtold and P. Patterson, The enhanced forest inventory and analysis program—National sampling design and estimation procedures, Gen. Tech. Rep. SRS-80, USDA Forest Serv., Southern Res. Station 85, Asheville, NC, USA, 2005.
[40] J. B. Bradford *et al.*, "Carbon pools and fluxes in small temperate forest landscapes: Variability and implications for sampling design," *Forest Ecol. Manage.*, vol. 259, no. 7, pp. 1245–1254, Mar. 2010.
[41] V. Junttila, T. Kauranne, and V. Leppänen, "Estimation of forest stand parameters from LiDAR using calibrated plot databases," *Forest Sci.*, vol. 56, no. 3, pp. 257–270, Jun. 2010.
[42] H. E. Andersen, S. Reutebuch, R. J. McGaughey, M. d'Oliveira, and M. Keller, "Monitoring selective logging in western Amazonia with repeat lidar flights," *Remote Sens. Environ.*, vol. 151, pp. 157–165, Aug. 2014.
[43] H. E. Andersen, J. Strunk, H. Temesgen, D. Atwood, and K. Winterberger, "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. Remote Sens.*, vol. 37, no. 6, pp. 1–16, 2011.
[44] A. Kangas and M. Maltamo, Eds. *Forest Inventory—Methodology and Applications*. Berlin, Germany: Springer-Verlag, 2006.

Virpi Junttila received the D.Sc. degree in applied mathematics from the Lappeenranta University of Technology, Lappeenranta, Finland.

She is currently working as a Postdoctoral Researcher with the Department of Applied Mathematics, Lappeenranta University of Technology. Her research interests include developing automated model-based forest inventory methods with the optimal use of small-sample-size laser scanning data.

Tuomo Kauranne received the Ph.D. degree in applied mathematics from Lappeenranta University of Technology, Lappeenranta, Finland.

He is an Associate Professor of applied mathematics with the Department of Applied Mathematics, Lappeenranta University of Technology, Lappeenranta, Finland. His research interests include Bayesian statistical methods in forest inventory, data assimilation in atmospheric models, and the computational models of commodity markets. He is the President of Arbonaut Oy Ltd., Joensuu, Finland, which is a forest information company that has conducted airborne-laser-scanning-based forest resource assessment campaigns on over 5 000 000 ha on six continents.

Andrew O. Finley received the Ph.D. degree from the University of Minnesota, Minneapolis, MN, USA.

He is an Associate Professor with Michigan State University, East Lansing, MI, USA, where he has a joint appointment in the Department of Forestry and the Department of Geography and is an Adjunct Professor with the Department of Statistics and Probability. His research interests include developing methodologies for monitoring and modeling environmental processes, Bayesian statistics, spatial statistics, and statistical computing.

John B. Bradford received the Ph.D. degree from Colorado State University, Fort Collins, CO, USA.

He is a Research Ecologist with the Southwest Biological Research Center, U.S. Geological Survey, Flagstaff, AZ, USA. His research interests include the consequences of changing climatic conditions and land-use practices on terrestrial ecosystem structures and functions, focusing on carbon cycling, ecohydrology, and plant species distributions.