Forestry Advance Access published October 15, 2014

Forestry An International Journal of Forest Research

Institute of Chartered Foresters

Forestry 2014; **0**, 1–12, doi:10.1093/forestry/cpu036

Evaluation of the spatial linear model, random forest and gradient nearest-neighbour methods for imputing potential productivity and biomass of the Pacific Northwest forests

Hailemariam Temesgen[1]* and Jay M. Ver Hoef[2]

[1]*Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR, USA*
[2]*National Marine Mammal Laboratory, NOAA-NMFS Alaska Fisheries Science Center, 7600 Sand Point Way NE, Bldg 4, Seattle, WA 98115-6349, USA*

*Corresponding author. Tel: +1 5417378548; Fax: +1 5417374613; E-mail: hailemariam.temesgen@oregonstate.edu.

Received 25 April 2014

Increasingly, forest management and conservation plans require spatially explicit information within a management or conservation unit. Forest biomass and potential productivity are critical variables for forest planning and assessment in the Pacific Northwest. Their values are often estimated from ground-measured sample data. For unsampled locations, forest analysts and planners lack forest productivity and biomass values, so values must be predicted. Using simulated data and forest inventory and analysis data collected in Oregon and Washington, we examined the performance of the spatial linear model (SLM), random forest (RF) and gradient nearest neighbour (GNN) for mapping and estimating biomass and potential productivity of Pacific Northwest forests. Simulations of artificial populations and subsamplings of forest biomass and productivity data showed that the SLM had smaller empirical root-mean-squared prediction errors (RMSPE) for a wide variety of data types, with generally less bias and better interval coverage than RF and GNN. These patterns held for both point predictions and for population averages, with the SLM reducing RMSPE by 30.0 and 52.6 per cent over two GNN methods in predicting point estimates for forest biomass and potential productivity.

Introduction

To manage forest resources in perpetuity, forest analysts, forest managers and policy-makers need to plan for the future using high-quality, spatially explicit, up-to-date inventory estimates. Forest biomass and potential productivity are two variables that are critical for land transactions, locating timber resources and new processing facilities, writing silvicultural prescriptions, drafting conservation plans and linking to climate change and carbon accounting (Latta *et al.*, 2010). Yet, due to the prohibitive cost of collecting detailed information over an extensive land base, most forests lack productivity and biomass values for unsampled locations.

Forest productivity is measured in a wide variety of ways. In the western US, forest productivity is assessed by the forest inventory and analysis (FIA) program (Czaplewski, 1999; Roesch and Reams, 1999) whereby site trees are selected to produce a site index for forested field plots. The site index is then used in combination with normal yield tables to determine potential productivity, i.e. the maximum potential cubic metre volume per hectare per year (potential mean annual increment; PMAI) that would be produced over the long term at a given site for a normally stocked stand (Hanson *et al.*, 2002). In this study, PMAI was used as response variable to represent productivity, indicating the average annual productivity of wood volume ($m^3 \cdot ha^{-1} \cdot year^{-1}$) that would be realized over time.

Forest biomass is important attribute for quantifying the roles of forests as carbon source or sink and for sustainable forest management. The emergence of biomass as a critical variable in assessing sequestration of atmospheric carbon and in providing critical information to forest resource management and policy decision-making has focussed attention on its estimation and prediction for non-sampled sites. Due to variation in moisture contents, dry forest biomass estimates (DRYBIOT) are the basis for forest carbon inventories and most international negotiations. In this study, DRYBIOT was used as response variable, and represents the total above-ground oven-dry biomass of live trees>2.5 cm in diameter.

Different parametric and non-parametric methods have been proposed for imputing PMAI and biomass for unsampled locations by linking measured ground variables and available auxiliary variables. Recently, nearest-neighbour (NN) methods, such as gradient nearest-neighbour (GNN, Ohmann and Gregory, 2002), *k*-nearest neighbour, (*k*-NN, McRoberts *et al.*, 2002) and random forests (RF, Breiman, 2001; Eskelson *et al.*, 2009b), have been developed and have gained widespread use in imputing (augmenting) data for point (mapping) and total (block) predictions.

The reasons for the wide use of the NN methods in forestry include:

(1) Simultaneous prediction of multiple variables: many variables of interest can be estimated at once for each location.
(2) The availability of low-cost and wall-to-wall climate and remotely sensed auxiliary variables that are related to variables of interest (*Y* variable set) (LeMay *et al*., 2008).
(3) The maintenance of logical relationships among *Y* variables and the provision of estimates that are within the bounds of biological reality (Moeur and Stage, 1995). Because a match is derived from the sampled locations, the estimate will, necessarily, exist in the population. This is particularly true if a single neighbour is used in the imputation whereby the covariance structure of multiple attributes of forests at a target location is maintained. This property benefits many applications in forest management and planning, which often require information about many variables simultaneously for all or most sites.
(4) No assumption on the distributional characteristics for the auxiliary variables or for the variables of interest (LeMay and Temesgen, 2005): because NN methods are distribution-free, they mitigate the difficulties encountered with alternative parametric and semiparametric multivariate modelling approaches (Koistinen *et al*., 2008) due to locally varying relationships between *X* and *Y* in response to variation in species, age, forest structure, soil and climate (Temesgen *et al*., 2003;McRoberts *et al*., 2007).

Despite their wide use, NN methods are neither unbiased nor consistent (LeMay and Temesgen, 2005). While in certain forest inventory applications with comprehensive field data, bias has generally not been an issue (Packalen and Maltamo, 2007), recent studies that compared NN methods using light detection and ranging reported bias that ranged from −8.78 to 0.25 per cent and 0.16 to 16.16 per cent when using most similar and RF imputation, respectively (Breidenbach *et al*., 2010; Gagliasso *et al*., 2014). In addition, NN methods neither extrapolate nor interpolate well for conditions with limited samples.

An alternative to the NN approach to mapping and estimating total biomass and productivity is to use a spatial linear model (SLM), which includes kriging and universal kriging. This approach was initially developed for a similar goal: predicting geographic values or totals for mining resources. Kriging has also been used to preserve spatial and attribute correlation of FIA data (Moeur and Hershey, 1999) and to account for spatial dependence in forest inventory and monitoring (Lappi, 2001). Other applications of kriging in forestry include mapping of forest resources (Gunnarsson *et al*., 1998), analysis of insect pests (Aleong *et al*., 1991) and predicting forest biomass and productivity (Ver Hoef and Temesgen, 2013). Numerous other examples and applications can be found in Cressie (1990, 1993).

Ver Hoef and Temesgen (2013) compared the suitability and performance of *k*-NN and SLM theoretically and empirically, and reported that SLM is a better option for point and total prediction. Despite the growing research on GNN and RF and their wide use for mapping and estimating totals for the Pacific Northwest forests, detailed analyses that compare the performance, efficiency and suitability of RF, GNN and SLM for predicting (or mapping) biomass and forest productivity at point and block-level are lacking. The overall goal of this article is to make similar comparisons of SLM to GNN and RF that were missing from Ver Hoef and Temesgen (2013). For potential uses and applications of *k*-NN and SLM in forestry, see McRoberts *et al*. (2002, 2007) and Ver Hoef and Temesgen (2013), respectively. Below, we provide a brief overview of GNN, RF and SLM.

Gradient nearest neighbour

GNN methods are also devised for prediction (mapping), both for single sites and for block averages. In general, the use of GNN methods in forestry is based on the notion that there are two datasets: one that includes a complete set of variables, and one that does not. The first dataset is called the reference (or training) set and the latter is called the target set. The missing variables (or data in the target set) are referred to as the *Y*-variable dataset, and the variables that are held in common across both datasets are referred to as the *X*-variables. For each observation in the target set, one or more nearest neighbours are found in the reference set, based on the magnitudes of differences in the *X*-variables. If the magnitude of differences between a target observation and a given reference observation, summed across all *X*-variables, is found to be a minimum, when compared with differences in relation to each of the remaining alternative observations in the reference set, then that particular reference observation is identified as the NN. This process may be extended from the NN to the next NN, and so on, up to the *k*-NNs, as defined by users of this technique. Once the NNs have been identified, the missing *Y*-variables can then be estimated using some function of the associated *k*-NNs (e.g. mean, weighted mean or mode), and those values are then assigned to the target observation as the best estimate of the missing data.

Unlike other NN methods, GNN uses canonical correspondence analysis (CCA) to assign weights to ancillary variables in selecting NNs. Ohmann and Gregory (2002) asserted that multivariate ordination space used to weight the ancillary variables provides a better overall view of ecological, biophysical and biodiversity gradients over the land base. While this assertion might be true for some ecological variables (e.g. biodiversity, presence or absence species), GNN was not the most precise in predicting commonly used forest inventory variables compared with other NN methods (e.g. Hudak *et al*., 2008; Eskelson *et al*., 2009b; Gagliasso *et al*., 2014). For detailed history on GNN, see Ohmann and Gregory (2002).

Random forest

Non-parametric methods such as the classification and regression tree (CART) are used to model or predict multisource data with nonlinear response variables. Bagging trees and RF have the best performances among CART methods (Prasad *et al*., 2006). RF has been used in various applications, including predicting soil type (Lemercier *et al*., 2012), soil organic carbon content (e.g. Kheir *et al*., 2010), basal area and trees per hectare (Hudak *et al*., 2008), above-ground biomass (Houghton *et al*., 2007) and individual tree attributes (Yu *et al*., 2011).

The RF approach has been reported to be an efficient prediction approach, especially when the number of predictor variables is very large (Svetnik *et al*., 2003) and when interactions and correlations among variables are complex and large in number (Strobl *et al*., 2008). Prasad *et al*. (2006) and Liaw and Wiener (2002) assert

that RF is efficient for high-dimensional data, and they list its advantages as follows: (1) it is non-parametric and does not require the specification of a functional form (e.g. a general linear model). Unlike multiple linear regression or maximum likelihood methods, no single dominant data structure (e.g. normality) is assumed or required; (2) pre-selection of variables is not needed (a robust stepwise selection method is used) and the variables can be a mixture of continuous and categorical variables; (3) the same variable can be reused in different parts of a tree because context dependency is automatically recognized; and (4) these methods are robust to the effects of outliers and missing data.

The RF approach is not without its limitations. In an application for classification purposes, the RF approach proved to be biased towards the most frequent classes. To avoid bias, Chen *et al.* (2004) suggested the use of balanced RFs or weighted RFs. Strobl *et al.* (2008) found bias when covariates were highly correlated. In addition, like other NN methods, RF cannot extrapolate beyond the training data and may not interpret well for conditions with few samples (McInerney and Nieuwenhuis, 2009). Unlike SLM, RF approach cannot provide probabilistic prediction accuracy.

Spatial linear model

The spatial variation of PMAI and DRYBIOT values is complex but is not generally unstructured. Due to parent materials and microsite similarities, the values of PMAI and DRYBIOT are spatially dependent at some scale, with this dependence referred to as spatial autocorrelation. This structure is assumed to be random, which can be described by a semivariogram that summarizes the spatial variation.

A semivariogram expresses the degree of similarity between two observations separated by a given distance (the lag). An empirical semivariogram can be computed from sampled data using the following expression:

$$\gamma(h) = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} \{z(x_i) - z(x_i + h)\}^2$$

where $\gamma(h)$ is the estimated semivariance at a separation distance, or lag h, and $z(x_i)$ and $z(x_i + h)$ are the observed values at x_i and $x_i + h$, of which there are $M(h)$ pairs (Cressie 1993). In most cases, the semivariance increases as the distance separating pairs of points (i.e. the lag) increases, indicating that points close together tend to have more similar values than those that are far apart.

The SLM is an extension of geostatistical methods, combining regression with classical geostatistics, which were developed for prediction (also called universal kriging), both for single sites or block averages. For a history, see Cressie (1990). In comparison to GNN and RF, SLM predictions were designed to minimize the RMSPE. The SLM has optimality properties with regard to normally distributed data and/or linear predictors, as reviewed in Ver Hoef and Temesgen, (2013). However, the SLM is generally robust to misspecification of the covariance model (Stein, 1988 and Putter and Young, 2001), and because SLM predictors are linear (as weighted averages of data), they are fairly robust to non-normal data due to a spatially correlated version of the central limit, (e.g. Bolthausen, 1982), allowing for inference based on a standard normal distribution (e.g. for prediction intervals). There are many extensions of the SLM when predictions cannot be assumed approximately normally distributed (Diggle *et al.*, 1998). Although geostatistical methods classically estimated the covariance model by binning data into distance classes and using a least-squares fit (Cressie, 1985) to a semivariogram, as described above; modern estimation of the SLM uses restricted maximum likelihood (REML) estimation (Patterson and Thompson, 1971; Harville, 1977). REML not only allows for less biased estimation of regression effects and covariance parameters, but also removes the arbitrary nature of binning.

Like any method, SLM has some disadvantages. These tend to be related to the problem of estimating the spatial covariance function. The data are used twice, called empirical best linear unbiased prediction (Zimmerman and Cressie, 1992); once to estimate the covariance parameters and secondly used for best linear unbiased prediction (Cressie, 1993). Theory suggests that SLM should be optimal, but much of that theory is based on an assumed spatial stochastic model, which does not take into account the estimation of the covariance parameters. Also, the theory does not indicate how much better SLM might be; GNN and RF have fewer assumptions. Perhaps GNN and/or RF are easier to implement and faster to compute than SLM, and a small loss in efficiency is compensated by ease of use and computational speed (Finley and McRoberts, 2008).

Objectives

While GNN and RF are widely used for mapping and estimating totals for western forests, detailed analyses that compare their performance, efficiency and suitability to SLM and other geostatistical methods for predicting DRYBIOT and PMAI at the point and block level are lacking. The overall goal of this article is to compare the performances of the SLM, RF and GNN approaches for predicting both point and total DRYBIOT and PMAI of Pacific Northwest forests through simulations and resampling of real forestry data. Building on Ver Hoef and Temesgen (2013), this manuscript extends that work by (1) simulating lognormal variables; (2) comparing SLM with two RF and two gradient NN methods; (3) using a larger dataset collected in Washington and Oregon; and (4) discussing both theoretical and applied issues.

Methods

Simulation of artificial data

The data simulation procedures are given in detail in Ver Hoef and Temesgen (2013). We give a brief summary here. We created spatially patterned and cross-correlated X-variables. All datasets were repeatedly simulated on a 20×20 regular grid evenly spaced between -1 and 1 on both coordinate axes. Eight covariates, X_1–X_8, were simulated on the grid. Each X_i was simulated from an autoregressive recursion where X_i depended on X_{i-1}, and each of these was, in turn, spatially autocorrelated (the autoregression creates the cross-correlation). We simulated the data so there was significant cross-correlation between X_1–X_4, and there was significant cross-correlation between X_5–X_8, but the set X_1–X_4 was independent of the set X_5–X_8. We then simulated the response variable, Y, from the SLM with covariates X_1–X_8, where dependence on the Xs varied from strongly positive to zero to strongly negative. The group X_1–X_4 had less autocorrelation and smaller variances than did group X_5–X_8. The response variable was related to both groups through regression coefficients, but the coefficients were zero for X_4 and X_5. The response Y had a small independent component and strongly spatially autocorrelated errors, with a range parameter of 3, for

data with x- and y-coordinates ranging from −1 to 1. The exact values used in the simulation are given in Ver Hoef and Temesgen (2013). The spatial autocorrelation model for simulation was a spherical model, although all estimations used an exponential autocorrelation model, so there was model mis-specification for autocorrelation.

We simulated three types of response variables. In all three types, the data were initially simulated from a normal (Gaussian) distribution, and this was the first type of response variable. In the second type, we exponentiated the response variable and used it as the mean when drawing Poisson random variables. In the third type, we exponentiated the response variable, creating autocorrelated lognormal variables.

In summary, several model mis-specifications were made: inclusion of covariates with no effect (X_4 and X_5), exclusion of covariates with real effects (X_3 and X_6), and, in the case of SLM, generating data from a spherical autocovariance model, but fitting the data with an exponential autocovariance model. One of the so-called advantages to the NN methods is that they are non-parametric, and so can be used on a wide variety of data. Ver Hoef and Temesgen (2013) outlined why the SLM is 'robust', although it is not completely non-parametric, which was summarized here in the Introduction. Part of the demonstration of robustness is to simulate non-normal data, simulate data under a different model than is being used for analysis, etc. Thus, aim to demonstrate that the SLM, too, can be used on a wide variety of data.

For each of the three simulation methods listed above, 2000 datasets were simulated (with 400 simulated values per dataset). For each simulation, 100 locations were sampled randomly, and from each sample, the other 300 were predicted, along with the overall total for each prediction method.

Forest productivity and biomass data

Our study uses FIA data collected in Oregon and Washington. The FIA databases are part of the national inventory of forests for the US (Roesch and Reams 1999; Czaplewski 1999). A tessellation of hexagons, each ∼2400 ha in size, is superimposed across the nation, with one field plot randomly located within each hexagon. Approximately the same number of plots is measured each year. Each field plot is composed of four subplots, with each subplot composed of three nested, fixed-radius areas used to sample trees of different sizes (Figure 1). Forested areas that are distinguished by structure, management history, or forest type are identified as unique condition-classes.

For our study area there were 3356 forested FIA plots measured between 2001 and 2006. Biomass (DRYBIOT), maximum PMAI, elevation and primary species identifier (i.e. Douglas-fir (*Pseudotsuga menziesii*), ponderosa pine (*Pinus ponderosa*) or western hemlock (WH; *Tsuga heterophylla*)) were obtained from the FIA annual database. For climate data, we used monthly temperature and precipitation normal data for the period 1971–2000 that was produced by the Parameter-elevation Regressions on Independent Slopes Model (PRISM). The model provided an 800-m grid that produced differences between measured plot elevation and overlaid PRISM grid elevation up to 350 mm in the mountainous areas of Oregon and Washington. Following the data compilation, the Pearson product-moment correlation coefficients were used to describe relationships between PMAI and climate variables (Table 1).

NN imputation methods are donor-based methods. Variables of interest are those forest attributes that are only measured on a subset of plots (e.g. PMAI and DRYBIOT). Auxiliary variables are attributes that are measured on all plots. In this study, the 3356 sample points (n) were randomly divided into reference and target points. Reference points refer to sampled points that had both DRYBIOT and PMAI, while target points refer to unsampled points that had only climate, topographic and map attributes. Reference points formed the pool of potential similar neighbours that could be selected to impute forest biomass and PMAI and related attributes onto target points. They were used to develop a similarity function in selecting a neighbour point in the NN analysis. The target points were used to represent unsampled points (missing DRYBIOT and PMAI data). They were used to validate the accuracy of the NN approach by comparing the observed to the predicted DRYBIOT and PMAI values. The predicted values were obtained by substituting the DRYBIOT and PMAI of the most similar reference point.

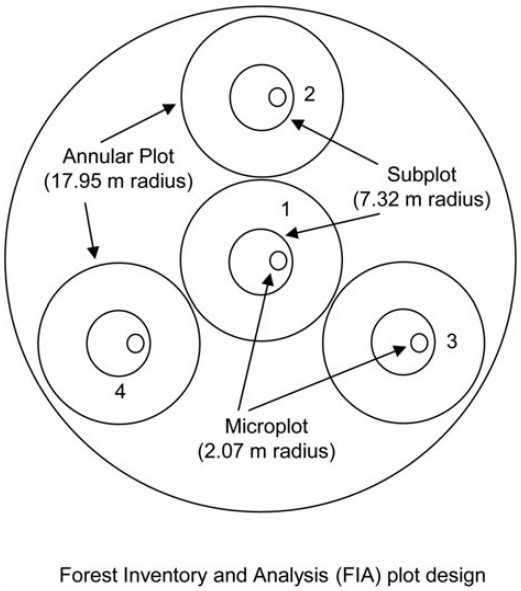

Figure 1 Each field plot is composed of four subplots, with each subplot composed of three fixed-radius areas used to sample trees of different sizes.

For SLM, we used 3356 known FIA values and resampled to predict PMAI and DRYBIOT. These were subsampled in two different experiments. For each resampling experiment listed below, the datasets were subsampled 500 times. For each subsample, a sample of 672 values was chosen randomly without replacement, and from each sample the remaining 2684 locations were predicted, along with an estimate of the total for all 3356 values. The covariates used were temperature, precipitation, Climate Moisture Index (CMI), an indicator variable for shade tolerance based on WH trees and elevation. The 500 resamplings were performed for PMAI and DRYBIOT.

Prediction methods

Seven different prediction methods were examined; these consisted of five different NN methods, multiple regression (a special case of an SLM that assumes the random errors are independent), and an SLM:

- GNN1: GNN that uses Mahalanobis distance with $k = 1$.Unlike the k-NN methods, GNN uses a CCA.
- GNN5: GNN that uses Mahalanobis distance with $k = 5$.
- RF1: RF that uses one neighbour.
- RF5: RF that uses five neighbours.
- BestNN: k-NN that uses both Mahalanobis distance and weighted Mahalanobis distance and tries $k \sim 1,2,\ldots,30$, and then chooses the distance matrix and k with the smallest cross-validation RMSPE from the observed data.
- SLM: a SLM that uses the same covariates as all NN methods as main effects only, with an exponential spatial autocovariance model estimated by REML, and using prediction and variance equations, as described in Ver Hoef and Temesgen (2013).

Table 1 Pearson correlation coefficients for the FIA plot data ($n = 3356$ plots)

Attribute	Temperature (°C)	Precipitation (cm)	CMI	Elevation (m)	Dry biomass (kg)	PMAI ($m^3 \cdot ha^{-1} \cdot year^{-1}$)	Latitude (degrees)
Precipitation (cm)	0.393						
CMI	0.229	0.837					
Elevation (m)	−0.789	−0.512	−0.467				
Dry biomass (kg)	0.179	0.372	0.276	−0.181			
PMAI (m^3 ha^{-1} $year^{-1}$)	0.555	0.629	0.6	−0.685	0.304		
Latitude (degrees)	−0.678	−0.626	−0.461	0.638	−0.298	−0.598	
Longitude (degrees)	−0.255	0.107	0.291	−0.213	−0.003	0.123	0.308

- LM: multiple regression, like SLM, but assuming all random errors are independent.

The five NN methods differ primarily in terms of the way they handle the distance calculations, the number of NNs (k) or the weighting functions they use, and the assumptions they make.

Nearest neighbor imputation

The GNN and RF methods were conducted using the yaImpute package in R language (http://www.r-project.org) (Crookston and Finley, 2008). The similarity between reference and target plots is defined using a weighted Euclidean distance for best most similar neighbour (MSN) and GNN:

$$D_{ij}{}^2 = (X_i - X_j)\, W (X_i - X_j)' \quad (1)$$

where W is the weight matrix, X_i is a vector of standardized values of the ancillary variables for the ith target plot; and X_j is a vector of standardized values of ancillary variables for the jth reference plot. The ancillary variables for both target and reference plots were standardized using the mean and variance of the ancillary variables of the reference plots. For bestNN, the weight used is $W = \Gamma \Lambda^2 \Gamma'$, where Γ is the matrix of standardized canonical coefficients for the ancillary variables and Λ^2 is the diagonal matrix of squared canonical correlations between ancillary attributes and variables of interest (Moeur and Stage 1995).

In GNN, distances are calculated in multivariate ordination space. Weights are assigned using a projected ordination score on the ancillary variables and eigenvalues on the axes and ancillary variables. In addition to weighting, the prediction procedure involves the following four steps (Ohmann and Gregory 2002).

- Using CCA, conduct direct gradient analysis and develop a model to quantify relations between response and auxiliary variables.
- For each unsampled location, predict scores for the first eight axes by applying coefficients from the model developed in (1) using their auxiliary variables, and identify nearest reference point based on Euclidean and predicted scores, weighted by their eigenvalues.
- Impute the response variables of the selected NN sample to the unsampled locations.

A binary CART (Breiman, 1984) approach was used to develop classification rules for estimating DRYBIOT and PMAI and to characterize relationships between potential productivity and site (geographic and climatic) variables. In fitting the CART models, vegetation and site variables were used to split the data into increasingly homogenous subsets, using binary recursive algorithms developed in R language (http://www.r-project.org). The geographic and climatic classifying variables included latitude, longitude, elevation, temperature and precipitation. Within each homogenous subset (also known as node), sample points were randomly divided into reference and target points. Following that, imputation was carried out within each node. The reference points were used to develop a similarity function to select a neighbour point within each node. To quantify similarity a distance measure, which was calculated as one minus the proportion of trees where a target observation is in the same terminal node as a reference observation, was used. Following that, within each node, the target points represented unsampled locations without biomass and productivity values. The number of neighbours (k) considered for predictions of the target unit can be set to any number of reference points. In this study, we set the number of reference points (k) to 1 and 5.

For NN methods, prediction standard errors are computed using cross-validation (Ohmann and Gregory, 2002; Temesgen *et al.*, 2003; Stage and Crookston, 2007; Eskelson *et al.*, 2009b). Cross-validation makes predictions for sites that already have values, where each sample is removed one at a time, and the rest of the sample is used to predict the one that was removed. The idea is to use in-sample averaged squared errors between the predicted and observed values to serve as a global estimator of squared errors when out-of-sample. Let the NN prediction standard error be estimated as (after Ver Hoef and Temesgen 2013),

$$\hat{\zeta} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{\theta}_i - \theta_i)^2}$$

where $\hat{\theta}_i$ is the cross-validation prediction of θ_i for $i = 1,...,n$ sample values. Assuming prediction errors are normally distributed, 90 per cent prediction intervals are formed as $\hat{\theta}_i \pm 1.645\hat{\zeta}$ for $j = n+1,..., n+m$ out-of-sample values, where m is the sample size of the target dataset. Note that $\hat{\zeta}$ is constant for all sites.

For the standard error (SE) of estimating a total ($\hat{T}$), one can use the idea of classical sampling theory, e.g. (Thompson, 1992), where $\hat{\zeta}$ replaces the standard error (after Ver Hoef and Temesgen 2013),

$$\mathrm{SE}(\hat{T}) = \hat{\zeta}\sqrt{(n+m)m}$$

Spatial linear model

Heuristically, the LM and SLM can be envisioned as follows. Suppose that we develop a multiple regression model between a response variable and covariates; this is the fitted LM. Predictions for samples with covariates, but without the response variable, are then based on the estimated regression coefficients applied to the covariate values in that sample. The SLM uses additional information. For example suppose we notice that, when making a prediction at an unsampled location, the residuals tend to be positive between the LM fit and the actual values at nearby sampled locations. Then it makes sense to adjust a prediction to some value that is also above its LM fit. The amount of that adjustment depends on the estimated spatial covariance model, which also provides the estimated variance for the

prediction through a set of formulas that can be found in Ver Hoef and Temesgen (2013) and references therein, and many other places.

One problem in making the comparison is that GNN and RF are algorithms that make no assumption about how spatial data were created; they only assume a fixed surface. The SLM, on the other hand, is based on the idea that data are a realization of a spatial stochastic model. By conditioning on realizations under a SLM, and hence adopting a fixed surface perspective, GNN, RF and SLM can be evaluated in a common framework. Foresters and forest managers are interested in the global performance of predictive or imputation methods across a management region, among management regions, and across time. Supporting that practice, the comparison of prediction or imputation methods using global computations of average bias and RMSPE is described in the following section.

Performance measures

Following the performance measures used in Ver Hoef and Temesgen (2013), we evaluated the predictive performance of RF, GNN and SLM for both individual point predictions and totals over all points. For each of the seven prediction methods, the random separation of the data into target *vs* reference points was repeated 500 times for the three performance measures. In general, let θ_i be the true, known values from the dataset, and let $\hat{\theta}_i$ be a prediction, either for a specific point or for a total (after Ver Hoef and Temesgen 2013).

(1) RMSPE: root-mean-squared prediction error, measures how close the estimates are to the true values.

$$\text{RMSPE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \theta_i)^2}$$

where m is the number of simulated values and MSPE is RMSPE squared. A smaller value of RMSPE indicates that predictors are closer to true values.

(2) SRB: signed relative bias. Absolute bias is meaningless, so it is expressed as a fraction of the variability. It is well known that MSPE = bias^2 + variance. The sign of bias is also informative, so signed relative bias as a fraction of variability, and is computed as,

$$\text{SRB} = \text{sign}(\tau)\sqrt{\frac{\tau^2}{\text{MSPE} - \tau^2}}$$

where

$$\tau = \frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \theta_i)$$

and sign (τ) is the sign positive or negative of τ. A smaller absolute value of SRB has smaller bias, a negative sign indicates under-prediction, and a positive sign indicates over-prediction.

(1) PIC90: 90 per cent prediction interval coverage. This measures how well uncertainty is being estimated. For many predicted values, or over many simulations, a prediction interval should cover the true value with the claimed percentage. For normally distributed estimators, the prediction interval coverage is

$$\text{PIC90} = \frac{1}{m}\sum_{i=1}^{m} I((\hat{\theta}_i - 1.645\widehat{\text{SE}}(\hat{\theta}_i)))\theta_i \,\&\, \theta_i(\hat{\theta}_i + 1.645\widehat{\text{SE}}(\hat{\theta}_i))$$

where $\hat{\theta}_i$ is the ith predictor, SE ($\hat{\theta}_i$) is the estimated standard error of $\hat{\theta}_i$, and θ_i is the true value. PIC90 should be near 0.90, if prediction intervals are properly estimated. It is also possible to compute PIC95 by replacing 1.645 with 1.96 in the formula above, in which case PIC95 should be near 0.95.

Results and discussion

Data summary

The FIA data collected in Oregon and Washington showed that both PMAI and DRYBOT were negatively correlated with elevation, but positively correlated with temperature, precipitation, CMI and percent composition of WH (Table 1). The data depicted relationship between distance and semivariance (Figure 2).

Predictions

For both the simulated and forestry data, the predictive abilities of the seven methods substantially differed in terms of RMSPE, SRB and prediction interval coverage. The performance of these methods varied by the type of distribution and variables examined. Except for the lognormal distribution, SLM resulted in the smallest RMSPE. For predicting biomass and potential productivity, SLM resulted in the lowest RMSPE and reasonable prediction interval coverage.

Simulated data

For the Gaussian simulated data, the SLM had the lowest RMSPE, for both point and total predictions, as expected (Table 2). Not only was it lowest, it was dramatically lower than any other predictor. The data were simulated with a high amount of autocorrelation, so this demonstrates how much better SLM can be in that case. When compared with RF1 and GNN1 (the two commonly used NN methods for prediction and mapping of attributes of western forests), SLM reduced RMSPE by 68.3 and 78.4 per cent for the point predictions and by 60.2 and 75.9 per cent for the total predictions, respectively. SLM was also noticeably better than LM (linear model assuming independence), with reduced RMSPE of 37.7 and 35.8 per cent for point- and block prediction, respectively.

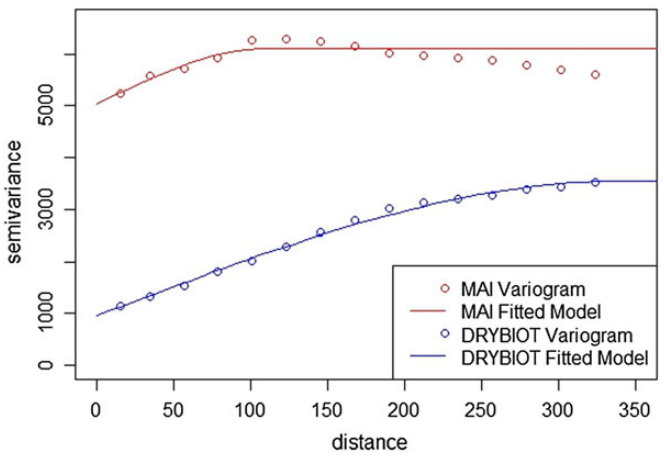

Figure 2 Variograms for the PMAI and dry biomass, fitted model using the spherical model.

Table 2 Performance summaries for the 2000 Gaussian simulated spatial datasets

Gauss	Point/total	RF1	RF5	bestNN	GNN1	GNN5	LM	SLM
RMSPE	Point	7.717	7.714	4.476	11.300	9.100	3.922	2.443
RSB		−0.002	−0.002	−0.001	0.036	0.044	0.001	0.006
PIC90		0.703	0.704	0.880	0.875	0.864	0.897	0.894
RMSPE	Total	223.4	222.9	154.2	370.0	320.0	138.7	89.0
RSB		−0.020	−0.018	−0.012	0.345	0.401	0.008	0.052
PIC90		0.784	0.784	0.882	0.906	0.858	0.882	0.892

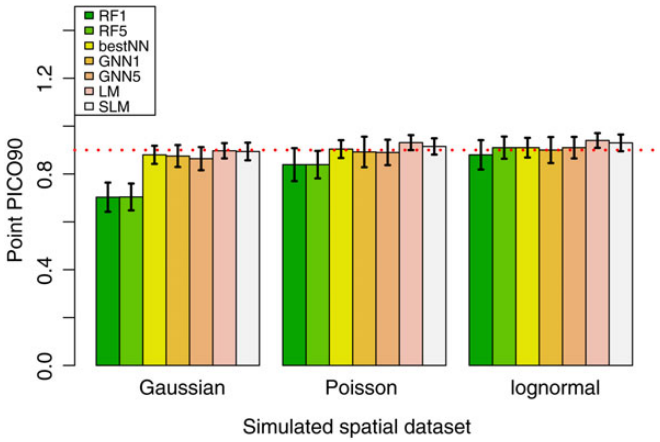

Figure 3 Prediction interval coverage (90%) for point by simulated spatial dataset.

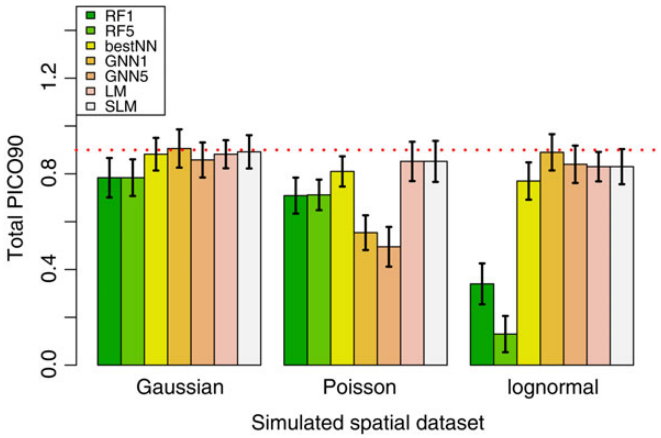

Figure 4 Prediction interval coverage (90%) for total by simulated spatial dataset.

All prediction methods were essentially unbiased for both point and total predictions. Except for RF1 and RF5, prediction interval coverage was near 0.90 for all point predictions. It appears that the RF1 and RF5 had extremely low interval coverage for both point and total estimates (0.78). GNN5 had relatively low coverage (0.86) as well (Table 2).

While there was no sizable difference between RF1 and RF5, GNN5 performed much better than GNN1. When compared with GNN5, RF1 and bestNN reduced RMSPE by a respective 15.2 and 42.0 per cent for point predictions and by 30.3 and 31.0 per cent for total predictions. However, RF1 and RF5 had low prediction interval coverage for both point (0.703) and total (0.784) predictions (Figures 3 and 4).

For the Poisson-simulated data, SLM had, again, the lowest RMSPE, for both point and total prediction (Table 3). When compared with RF1 and GNN1, SLM reduced RMSPE by 28.2 and 31.1 per cent for the point and by 16.3 and 43.1 per cent for total predictions, respectively. The performance difference declined when the number of NNs increased from 1 to 5. Compared with GNN5, SLM reduced RMSPE by 14.9 and 36.1 per cent for point and total predictions, respectively. Except for GNN1 and GNN5, all of the methods appeared to be unbiased for point prediction, with generally valid confidence interval coverage. There appeared to be some bias among the RF and GNN methods for predicting totals, and their confidence interval coverage was low, at 0.71 and 0.55, respectively. The 0.852 prediction interval coverage for the LM and SLM was also somewhat low, and this simulation was SLM's poorest performance on that measure.

Although there was also no sizable difference between RF1 and RF5 for the Poisson-simulated data, GNN5 performed much better than GNN1. For point estimate, GNN5 reduced RMSPE by 18.4 per cent when compared with RF5.When compared with GNN5, RF1 and bestNN reduced RMSPE by 23.1 and 3.0 per cent for total predictions. However, both RF1 and RF5 had low prediction interval coverage (0.703). Surprisingly, the prediction interval coverage for totals was extremely low for GNN (0.5) and very poor for RF (0.7).

The performance measures for lognormal simulated data are presented in Table 4. For both point and total prediction, the bestNN had the lowest RMSPE. When compared with SLM, bestNN reduced RMSPE by 19.9 per cent for point estimates and by 15.8 per cent for total predictions, but its PIC90 was very short when predicting total. It had prediction interval coverage of 0.77. This was a surprising result and can be partly attributed to the skewed distribution depicted by lognormal distribution.

For lognormal simulated data, except for RF1 and RF5, all of the methods appeared to be unbiased for point prediction, with generally valid confidence interval coverage. For RF1 and RF5, there appeared to be some bias among all methods for predicting the total. Their confidence interval coverage was dismally low, at 0.34 and 0.13 (Figure 5). Also, the 0.83 total prediction interval coverage for the LM and SLM was a bit low, and this simulation was its poorest performance on that measure.

While there was also no sizable difference between RF1 and RF5 for the lognormal simulated data, GNN5 performed much better than GNN1. For total prediction, GNN5 reduced RMSPE by 142.3 per cent, when compared with RF5.When compared with GNN5,

Table 3 Performance summaries for 2000 Poisson-simulated spatial datasets

Poisson	Point/total	RF1	RF5	bestNN	GNN1	GNN5	LM	SLM
RMSPE	Point	6.29	6.28	5.27	6.54	5.30	5.32	4.51
RSB		−0.052	−0.052	−0.033	0.242	0.290	−0.002	−0.001
PIC90		0.839	0.839	0.904	0.892	0.890	0.931	0.915
RMSPE	Total	309.0	311.6	297.1	455.0	405.0	299.7	258.7
RSB		−0.157	−0.161	−0.142	0.571	0.560	0.010	0.015
PIC90		0.709	0.712	0.810	0.554	0.495	0.852	0.852

Table 4 Performance summaries for 2000 lognormal simulated spatial datasets

Lognormal	Point/total	RF1	RF5	bestNN	GNN1	GNN5	LM	SLM
RMSPE	Point	1.35	1.26	1.00	1.50	1.30	1.30	1.19
RSB		−0.188	−0.280	−0.047	−0.027	−0.040	−0.008	−0.004
PIC90		0.88	0.91	0.91	0.90	0.91	0.94	0.93
RMSPE	Total	82.0	106.6	36.5	49.0	44.0	45.8	43.4
RSB		−2.074	−2.798	−0.492	−0.235	−0.411	−0.110	−0.067
PIC90		0.34	0.13	0.77	0.89	0.84	0.83	0.83

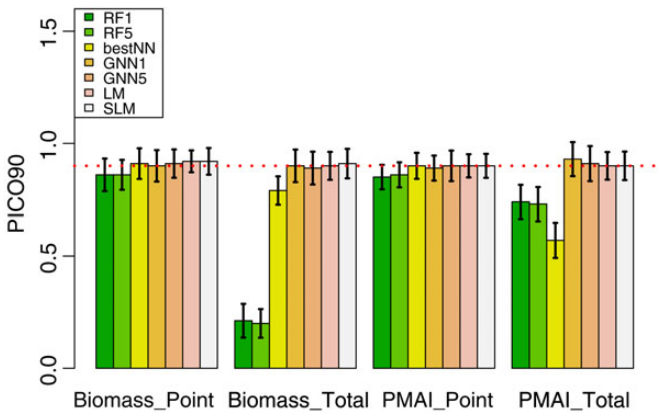

Figure 5 Point and total prediction interval coverage (90%) for biomass and productivity.

RF1 and bestNN reduced RMSPE by 2.8 and 25.6 per cent, respectively, for total predictions. However, RF1 and RF5 had low prediction interval coverage. Among the methods examined, only GNN1 had valid prediction interval coverage for both point and total predictions (Figure 3 and 4).

SLM outperformed LM for Poisson and lognormal distributions for both point and total estimates. When compared with GNN5, LM reduced RMSPE by 65.3 per cent for point and 62.5 per cent for total predictions for Gaussian; 18.7 per cent for point and 34.1 per cent for total predictions for Poisson; and 13.2 per cent for point and 6.5 per cent for total predictions for lognormal distributions (Tables 2–5). In addition to high RMSPE, GNN1 and GNN5 had the poorest prediction interval coverage, except for the lognormal distribution. Based on these results, we do not suggest the use of GNN, either for point or total predictions, when data have Gaussian and Poisson distributions.

Forestry data

The performance measures for resampling PMAI are presented in Table 5. For point prediction, SLM, followed by the LM and bestNN, resulted in the smallest RMSPE. For predicting total, SLM again had the lowest RMSPE. SLM reduced RMSPE by 31.4 and 26.9 per cent over RF1 and GNN1 (Table 5). There appeared to be some bias for the bestNN and RF methods. All prediction intervals were reasonable for the point estimates. Prediction interval coverage was too short for RF1, RF5 and bestNN (Figure 5).

The performance measures for resampling DRYBIOT data are presented in Table 6. For point prediction, SLM reduced RMSPE by 25.0 and 28.9 per cent over RF1 and GNN1, respectively. The performance difference declined when RF5 and GNN5 were considered. SLM reduced RMSPE by 9.2 per cent over both RF5 and GNN5. Point prediction appeared unbiased for all methods. Prediction interval coverage was quite good for all methods. For predicting totals, there appeared to be some bias for the RF and GNN methods, and prediction intervals were extremely short, as well as small, for RF. SLM reduced RMSPE for predicting total DRYBIOT by 66.6 and 25.1 per cent over RF1 and GNN1, respectively.

SLM and LM outperformed both GNN5 and bestNN for both point and total prediction of PMAI and DRYBIOT. Except for point biomass prediction, GNN5 outperformed RF5 in predicting both biomass and PMAI. Although there were minor differences between RF1 and RF5 for predicting PMAI and biomass, GNN5 resulted in lower RMSPE than did GNN1. Increasing the number of reference points effectively shifted the prediction toward the sample mean, or altered the shape of the distribution of predictions toward normal, which is unrealistic when the distribution is non-normal or skewed.

Table 5 Performance summaries for 500 resampling of the forest productivity data

PMAI	Point/total	RF1	RF5	BestNN	GNN1	GNN5	LM	SLM
RMSPE	Point	2.9	2.9	2.3	3.3	2.6	2.4	2.2
RSB		0.023	0.024	0.048	0.004	0.004	−0.003	−0.002
PIC90		0.846	0.845	0.904	0.891	0.899	0.904	0.902
RMSPE	Total	325.4	323.2	404.6	305.2	257.8	255.6	223.1
RSB		0.746	0.732	1.434	0.113	0.104	−0.097	−0.05
PIC90		0.74	0.728	0.572	0.932	0.906	0.902	0.894

Table 6 Performance summaries for 500 resampling of the forest biomass data

Biomass	Point/total	RF1	RF5	bestNN	GNN1	GNN5	LM	SLM
RMSPE	Point	89.5	89.5	68.5	94.3	73.9	68.3	67.1
RSB		−0.082	−0.083	−0.025	−0.011	−0.003	−0.001	−0.002
PIC90		0.859	0.86	0.914	0.899	0.906	0.922	0.918
RMSPE	Total	21 106	21 291	8728	9411	7429	7244	7048
RSB		−2.3	−2.321	−0.63	−0.303	−0.086	−0.044	−0.053
PIC90		0.212	0.2	0.792	0.902	0.888	0.904	0.908

In predicting stand volume and biomass in a temperate forest, Latifi and Koch (2012) reported that RF resulted in lower RMSPE (by 3.5 per cent, on average) than the MSN method, but higher bias than MSN (average bias of 5.13 per cent with RF, compared with 2.44 per cent with MSN for stem volume). Other authors have also reported the general superiority of RF over MSN and *k*-NN. Examples include Vauhkonen *et al.* (2010) for estimating single-tree attributes and Hudak *et al.* (2008) for predicting basal area and tree counts. However, these studies neither split the data in their examinations (simulations) more than once, nor examined the prediction interval coverage. Hence, the inference and assertion might be easily influenced by random sampling of data.

In contrast to the above studies, McInerney and Nieuwenhuis (2009) reported that *k*-NN outperformed RF in predicting volume and basal area per hectare. Eskelson *et al.* (2009a) also concluded that MSN is better suited for predicting mean annual increment than is RF. Breidenbach *et al.* (2010) found that RF outperformed MSN when predicting stand volume in central Europe. In this study, RF resulted in the highest bias in predicting total productivity and biomass. The bias in RF was more than twice as high as that of the GNN methods. The high bias is one of the reasons for the poor performance of RF in this study. Bias in RF has been previously reported by Chen *et al.* (2004) and Latifi and Koch (2012).

Chen *et al.* (2004) and Strobl *et al.* (2007), respectively, asserted that RF does not perform well for imbalanced data and for correlated covariates. In contrast, Prasad *et al.* (2006) asserted that RF is efficient for high-dimensional data. Our study does not bear out that assertion. RF and GNN had substantial bias for both the simulated data and the FIA data. RF and GNN entirely failed to project or relate the references to the target units. They had dismal prediction interval coverage for predicting both total dry biomass and forest productivity.

Hudak *et al.* (2008) reported that MSN performed better than GNN when predicting basal area and stem counts per hectare using laser imaging detection and ranging (LiDAR) attributes. In this study, RF and GNN prediction methods performed poorly. The dismal performances of RF and GNN are disconcerting, and their use might lead to incorrect forest management decisions. Because of their poor prediction interval coverage, we do not suggest the use of RF and GNN for predicting PMAI and DRYBIOT across the landscape.

Performance measures and their patterns were similar for RF and GNN. SLM outperformed GNN and RF when bias, RMSPE and PIC90 were considered. Both RF and GNN select NN based on non-spatial variables and the weight matrix (W in equation 1) is not directly related to the spatial distances between observations. On the other hand, SLM bases prediction on spatial nearness and on the realized residuals in the neighbourhood. The improvement of the prediction of PMAI and biomass by using SLM suggests that the UTM Easting and Northing may have captured the warm-dry to cool-moist, north–south and east–west gradients that span Oregon and Washington. The performance of SLM might be attributed to its effective incorporation of the UTM coordinates. In contrast to our findings, Pierce *et al.* (2009) working on a fire model, found that kriging performed poorly when compared with GNN.

Bias was lowest for LM and SLM in predicting both point and total forest DRYBIOT and PMAI. Unlike LM and SLM, RF and GNN do not minimize for error and also include distance component imputation error. Hence, they almost always result in greater error than do LM and SLM (Ver Hoef and Temesgen 2013).

RMSPE and bias of SLM were greater for the simulated data than for the FIA data because the forestry data exhibited lower variability than did the simulated data. Unlike the parametric LM and SLM approaches, GNN and RF methods did not require distributional assumptions in order to predict DRYBIOT and PMAI for non-sampled locations. Although NN provides the advantage of predicting forest biomass and PMAI in one step, the results from this study indicated that SLM generally performed better at this than did GNN and RF. SLM allows spatial modelling of forest biomass and PMAI, which are useful in assessing forest resources.

The approaches examined in this study can be used to estimate forest biomass and productivity from map labels or aerial attributes and can used in other applications, such as landscape modelling or 3D visualization (Magauhye, 1998; McCarter *et al*., 1998).

Our results can be compared with those reported in Pierce *et al*. (2009) and Raty and Kangas (2012), where SLM is mathematically equivalent to their universal kriging and kriging-with-external-drift, respectively. In Pierce *et al*. (2009), the SLM performed well compared with GNN, but not as consistently better as our results. However, parameter estimation likely differed in the studies, as they do not specify if they used the REML option when they fit variograms using the GSTAT package. Raty and Kangas (2012) estimate the SLM by first fitting a linear model assuming independence, and then computing and fitting a semivariogram on residuals. It is worth to note that both studies do not give a standard error estimator of point-wise predictions. The use of REML for the SLM, as described in Ver Hoef and Temesgen (2013), estimates the fixed effects assuming correlated residuals, and is expected to be more efficient.

In summary, this article set out to compare RF, GNN and SLM for predicting totals or averages for DRYBIOT and PMAI in the PNW forests. Our simulations of synthetic data and resamplings of FIA data are not exhaustive; however, for the criteria that we chose (RMSPE, signed relative bias and prediction interval coverage), the results clearly favour SLM in general. We simulated data under conditions that should severely test the SLM method. In all cases, even with mis-specified covariance models, mis-specified linear models (including non-significant covariates and ignoring significant ones), zero-inflated count data, lognormal data and skewed FIA data, the SLM performed better than RF and GNN and generally provided valid inference with little bias, and prediction intervals that contained the true values the correct proportion of time.

Conclusions

Among the methods considered, substantial differences were found in the abilities for predicting DRYBIOT and PMAI. As a result, the seemingly divergent parametric and non-parametric approaches resulted in different point and total predictions. SLM outperformed GNN and RF in terms of accuracy and precision because SLM localizes the relation between the response variables and covariate in both the geographical and variable space. Unlike GNN, SLM also accounts for the spatial structure of the data and minimizes differences between observed and predicted values.

The potential applications for SLM are numerous. The SLM can be used to extrapolate any response variable collected at sample locations across the landscape and to map probability surfaces for prediction or errors, which, in turn, provide a higher level of confidence for their uses. However, like most other methods, the accuracy of both the SLM, RF and GNN methods depends on the representativeness of the sample, as well as the similarity of the target and reference points.

We foresee future work in several directions. Examination of multivariate SLM to preserve the covariance among multiple response variables at new locations is warranted. In addition to considering the spatial structure, there is a need to relate forest biomass and PMAI with stand density and other confounding factors. Hence, the examination of spatial linear mixed model, spatial generalized linear model (Diggle *et al*., 1998) and Bayesian spatial regression models (Finley *et al*., 2008, 2011, 2013) using LiDAR and multispectral imagery that provide wall-to-wall sets of predictor variables which describe stand structure and stand density might improve the quality of predictions.

Recently Bayesian spatial regression models have been used to fully incorporate uncertainty of selected forest attributes in the spatial covariance parameters and propagate uncertainty through to prediction intervals (Finley *et al*., 2008, 2011, 2013). These approaches also provide a framework for fitting various types of models. Future studies or activities that consider spatial prediction would benefit from further comparative investigation of Bayesian spatial regression models and SLM.

Acknowledgements

We gratefully acknowledge the support provided by the Forest Inventory and Analysis program, Pacific Northwest Research Station, United States Forest Service.

Conflict of interest statement

None declared.

Funding

This project received financial support from the National Marine Mammal Laboratory, NOAA-NMFS Alaska Fisheries Science Center. Reference to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA-NMFS.

References

Aleong, J., Parker, B.L., Skinner, M. and Howard, D. 1991 Analysis of Thrips Distribution: Application of Spatial Statistics and Kriging. Toward Understanding Thysanoptera: Proceedings, International Conference on Thrips, *February 21–23, 1989, Burlington, Vermont, USA*. U.S. Dep. of Agriculture, Forest Service, Northeastern Forest Experimental Station, Gen. Tech. Rep. NE-147.

Bolthausen, E. 1982 On the central limit theorem for stationary mixing random fields. *Ann. Probab.* **10**, 1047–1050.

Breidenbach, J., Nothdurft, A and Kandler, G. 2010 Comparison of nearest neighbor approaches for small area estimation of tree species-specific forest inventory attributes in central Europe using airborne laser scanner data. *Eur. J. For. Res.* **129**, 833–846.

Breiman, L. 2001 Random forests. *Mach. Learn.* **45**, 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. 1984 *Classification and Regression Trees*. Wadsworth and Brooks Advanced Books and Software, Monterey, CA, 358 p.

Chen, C., Liaw, A. and Breiman, L. 2004 *Using Random Forest to Learn Imbalanced Data*. Department of Statistics, University of California Berkeley, Technical Report: 666.

Cressie, N. 1985 Fitting models by weighted least squares. *J. Int. Assoc. Math. Geol.* **17**, 563–586.

Cressie, N. 1990 The origins of Kriging. *Math. Geol.* **22**, 239–252.

Cressie, N.A.C. 1993 *Statistics for Spatial Data*. Wiley, pp. 69–70, 151–155.

Crookston, N.L. and Finley, A.O. 2008 yaImpute: an R Package for kNN imputation. *J. Stat. Softw.* **23**, 1–16.

Czaplewski, R. 1999 Forest survey sampling designs: a history. *J. Forestry*. **97**, 4–10.

Diggle, P.J., Tawn, J.A. and Moyeed, R.A. 1998 Model-based geostatistics disc: P326–350. *J. R. Stat. Soc. Ser. C Appl. Stat.* **47**, 299–326.

Eskelson, B.N.I., Temesgen, H. and Barrett, T. 2009a Estimating current forest attributes from paneled inventory data using plot-level imputation: a study from the Pacific Northwest. *For. Sci.* **55**, 64–71.

Eskelson, B.N.I., Barrett, T.M. and Temesgen, H. 2009b Imputing mean annual change and estimating current forest attributes. *Silva Fenn.*. **43**, 649–658.

Finley, A.O. and McRoberts, R.E. 2008 Efficient *k*-nearest neighbor searches for multi-source forest attribute mapping. *Remote Sens. Environ.* **112**, 2203–2211.

Finley, A.O., Banerjee, S., Ek, A.R. and McRoberts, R.E. 2008 Bayesian multivariate process modeling for prediction of forest attributes. *J. Agric. Biol. Environ. Stat.* **13**, 60–83.

Finley, A.O., Banerjee, S. and MacFarlane, D.W. 2011 A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas. *J. Am. Statist. Assoc.* **106**, 31–48.

Finley, A.O., Banerjee, S., Cook, B.D. and Bradford, J.B. 2013 Hierarchical Bayesian spatial models for predicting multiple forest variables using waveform LiDAR, hyperspectral imagery, and large inventory datasets. *Int. J. Appl. Earth Obs.,* **22**, 147–160.

Gagliasso, D., Hummel, S. and Temesgen, H. 2014 A comparison of selected parametric and non-parametric imputation methods for estimating forest biomass and basal area. *Open J. For.*. **4**, 42–48.

Gunnarsson, F., Holm, S., Holmgren, P. and Thuresson, T. 1998 On the potential of kriging for forest management planning. *Scand. J. For. Res.* **13**, 235–237.

Hanson, E., Azuma, D.L. and Hiserote, B.A. 2002 Site Index Equations and Mean Annual Increment Equations For Pacific Northwest Research Station Forest Inventory and Analysis Inventories, 1985–2001. USDA Forest Service Research Note PNW-RN-533.

Harville, D.A. 1977 Maximum likelihood approaches to variance component estimation and to related problems C/R: P338–340. *J. Am. Stat. Assoc.,* **72**, 320–338.

Houghton, R.A., Butman, D., Bunn, A.G., Krankina, O.N., Schlesinger, P. and Stone, T.A. 2007 Mapping Russian forest biomass with data from satellites and forest inventories. *Environ. Res. Lett.* **2**, 045032.

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E. and Falkowski, M.J. 2008 Nearest neighbor Imputation of Species-level, Plot-scale Forest Structure Attributes from LiDAR data. *Remote Sens. Environ.* **112**, 2232–2245. Corrigendum in Remote Sensing of Environment 113, 289–290.

Kheir, B.R., Greve, M.H., Bøcher, P.K., Greve, M.B. and Larsen, R. 2010 Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: the case study of Denmark. *J. Environ. Manage.* **91**, 1150–1160.

Koistinen, P., Holmström, L. and Tomppo, E.O. 2008 Smoothing methodology for predicting regional averages in multi-source forest inventory. *Remote Sens. Environ.* **112**, 862–871.

Lappi, J. 2001 Forest inventory of small areas combining the calibration estimator and a spatial model. *Can. J. For. Res.* **31**, 1551–1560.

Latta, G., Temesgen, H., Adams, D.A. and Barrett, T. 2010 Analysis of potential impacts of climate change on forests of the United States Pacific Northwest. *For. Ecol. Manage.* **259**, 720–729.

Latifi, H. and Koch, B. 2012 Evaluation of most similar neighbour and random forest methods for imputing forest inventory variables using data from target and auxiliary stands. *Int. J. Remote Sens.* **33**, 6668–6694.

LeMay, V. and Temesgen, H. 2005 Comparison of nearest neighbor methods for estimating basal area and stems per ha using aerial auxiliary variables. *For. Sci.* **51**, 109–119.

LeMay, V., Maedel, J. and Coops, N.C. 2008 Estimating stand structural details using variable-space nearest neighbour analyses to link ground data, forest cover maps, and landsat imagery. *Remote Sens. Environ.* **118**, 2578–2591.

Lemercier, B., Lacoste, M., Loum, M. and Walter, C. 2012 Extrapolation at regional scale of local soil knowledge using boosted classification trees: a two-step approach. *Geoderma.* **171–172**, 75–84.

Liaw, A. and Wiener, M. 2002 Classification and regression by randomForest. *R News.* **2**(3), 18–22.

Magauhye, R. 1998 Techniques for visualizing the appearance of forestry operations. *J. For.* **96**, 9–14.

McCarter, J.B., Wilson, J.S., Baker, P.J., Moffett, J. and Oliver, C.D. 1998 Landscape management through integration of existing tools and emerging technologies. *J. For.* **96**, 17–23.

McInerney, D.O. and Nieuwenhuis, M. 2009 A comparative analysis of kNN and decision tree methods for the Irish national forest inventory. *Int. J. Remote Sens.* **30**, 4937–4955.

McRoberts, R., Nelson, M. and Wendt, D. 2002 Stratified estimation of forest area using satellite imagery, inventory data, and the *k*-nearest neighbors technique. *Remote Sens. Environ.* **82**, 457–468.

McRoberts, R., Tomppo, E.O., Finley, A.O. and Heikkinen, J. 2007 Estimating areal means and variances of forest attributes using the *k*-nearest neighbors technique and satellite imagery. *Remote Sens. Environ.* **111**, 466–480.

Moeur, M. and Hershey, R.R. 1999 Preserving spatial and attribute correlation in the interpolation of forest inventory data. In Lowell, K. and Jaton, A. (eds), *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources* (pp. 419–429). Ann Arbor.

Moeur, M. and Stage, A. 1995 Most similar neighbor: an improved sampling inference procedure for natural resource planning. *For. Sci.* **41**, 337–359.

Ohmann, J. and Gregory, M. 2002 Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. *Can. J. For. Res.* **32**, 725–741.

Packalen, P. and Maltamo, M. 2007 The k-MSN method for the prediction of species specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sens. Environ.* **109**, 328–341.

Patterson, H.D. and Thompson, R. 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika.* **58**, 545–554.

Pierce, K.B., Ohmann, J.L., Wimberly, M.C., Gregory, M.J. and Fried, J.S. 2009 Mapping wildland fuels and forest structure for land management: a comparison of nearest neighbor imputation and other methods. *Can. J. For. Res.* **39**, 1901–1916.

Prasad, A., Iverson, L. and Liaw, A. 2006 Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* **9**, 181–199.

Putter, H. and Young, G.A. 2001 On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli,* **7**, 421–438.

Raty, M. and Kangas, A. 2012 Comparison of *k*-msn and kriging in local prediction. *For. Ecol. Manage.* **263**, 47–56.

Roesch, F. and Reams, G. 1999 Analytical alternatives for an annual inventory system. *J. Forestry.* **97**, 44–48.

Stage, R. and Crookston, N.L. 2007 Partitioning error components for accuracy assessment of near-neighbour methods of imputation. *Forest Sci.* **53**, 62–72.

Stein, M.L. 1988 Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Stat.* **16**, 55–63.

Strobl, C., Boulesteix, A.L., Kneib, T., Zeileis, A. and Hothorn, T. 2007 Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* **25**, 1471–2105.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T. and Zeileis, A. 2008 Conditional variable importance for random forests. *BMC Bioinformatics* **307**, 1471–2105.

Svetnik, V., Liaw, A., Tong, C., Cullberson, C., Sheridian, R.P. and Feuston, B.P. 2003 Random forest: a classification and regression tool for compound classification and QSAR modelling. *J. Chem. Inf. Comput. Sci.* **43**, 1947–1958.

Temesgen, H. 2003 Estimating tree-lists from aerial information: a comparison of parametric and most similar neighbor approaches. *Scand. J. For. Res.* **18**, 279–288.

Temesgen, H., LeMay, V.M., Marshall, P.L. and Froese, K. 2003 Imputing tree-lists from aerial attributes for complex stands of British Columbia. *For. Ecol. Manage.* **177**, 277–285.

Thompson, S.K. 1992 *Sampling*. John Wiley & Sons, New York.

Vauhkonen, J., Korpela, I., Maltamo, M. and Tokola, T. 2010 Imputation of singletree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sens. Environ.* **114**, 1263–1276.

Ver Hoef, J.M. and Temesgen, H. 2013 A comparison of the spatial linear model to nearest neighbor (*k*-NN) methods for forestry applications. *PLoS ONE* **8**: e59129. doi:10.1371/journal.pone.0059129

Yu, X., Hyypa, J., Vastarnta, M., Holopainen, M. and Viitala, R. 2011 Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *Int. J. Photogramm.* **66**, 28–37.

Zimmerman, D.L. and Cressie, N.A.C. 1992 Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Ann. Inst. Stat. Math.* **44**, 27–43.