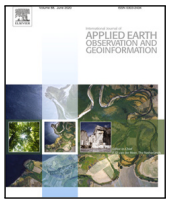
International Journal of Applied Earth Observation and Geoinformation 146 (2026) 105104

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Large-scale forest resource mapping with spatial gaps in the training data: Comparison of different modeling approaches

Andras Balazs [a,*], Jukka Miettinen [b], Mats Nilsson [c], Johannes Breidenbach [d], Timo P. Pitkänen [a], Mari Myllymäki [a]

[a] Natural Resources Institute Finland (Luke), Latokartanonkaari 9, Helsinki, 00790, Finland
[b] VTT Technical Research Centre of Finland, P.O. Box 1000, Espoo, 02044, Finland
[c] Swedish University of Agricultural Sciences (SLU), Skogsmarksgränd, Umeå, 90183, Sweden
[d] Norwegian Institute of Bioeconomy Research (NIBIO), Høgskoleveien 8, Ås, 1433, Norway

ARTICLE INFO

Keywords:
Biomass map
Remote sensing
Sentinel-2
Missing reference data
Random forest
k-NN
Multi-layer perceptron

ABSTRACT

Forest attribute maps are essential for supporting local decision-making regarding forest resource use. Such maps are produced by combining remote sensing and field data through various modeling approaches. When mapping across large areas, spatial gaps in field data used for model training are common. Our study evaluates the performance of three methods—*k*-Nearest Neighbor (*k*-NN), Random Forests (RF), and Multi-Layer Perceptron (MLP)—for forest resource mapping across Norway, Sweden, and Finland in an experimental setup with respect to availability of field data around the target area. Models were trained with sample plot sizes (N) ranging from 100 to 3000. RF consistently produced the most accurate predictions in terms of relative bias and RMSE. While spatial gaps in the training data (radius: 7–141 km) affected %RMSE of broad-leaved above ground biomass (AGB), they had minimal impact on %RMSE of both local and country-level predictions of total AGB and volume. For RF with $N = 3000$, %RMSE of total AGB ranged between 53%–55% in Finland and Sweden, and 70%–72% in Norway across gap sizes. However, %bias increased for local predictions across the whole study region with larger gaps: RF with $N = 500$ showed bias of −12%–12% (7 km gap) and −17%–28% (78 km gap). Similarly, country-level %bias of total AGB for Norway increased from −1.7% to −3.7% with larger gaps. In conclusion, spatial gaps in training data can significantly affect bias in predictions. Therefore, forest attribute maps should always be accompanied by metadata describing the training data used.

1. Introduction

Forests store huge carbon stocks, possess high climate mitigation potential (Hetemäki et al., 2022), and play an important role for biodiversity (Hunault-Fontbonne and Eyvindson, 2023). Climate-smart decision-making requires accurate and up-to-date information about local forest resources. High-resolution forest resources maps can be produced by combining remote sensing and field data available particularly from national forest inventories (NFIs) (e.g. Mäkisara et al., 2022; Nilsson et al., 2017; Hauglin et al., 2021). Due to proposed EU regulations (e.g. COM, 2023b,a), there is an increasing need for such maps also across countries, which brings new aspects to consider in comparison to studies done more locally or nationally. This paper focuses on how spatial gaps in the coverage of field plots used to train various mapping models influence map accuracy. Such gaps are likely when mapping is performed at large scales (Miettinen et al., 2025).

Models are required for linking ground truth and remotely sensed data. There are several different approaches that can be used ranging from parametric models to non-parametric models, machine learning methods and deep learning. One of the most popular and widely used method for large scale forest mapping in Europe is the *k*-Nearest Neighbor (*k*-NN) method (Tomppo and Halme, 2004; Chirici et al., 2016). A key benefit, and one reason for the popularity of the *k*-NN method is that consistent predictions can be straightforwardly obtained simultaneously for all variables of interest. For example, seemingly unrelated regression (SUR) (Fiebig, 2003) and Gaussian process regression (Rasmussen and Williams, 2005) also allow multivariate modeling, but generally require a larger modeling effort. Machine learning techniques, such as neural networks, boosted regression trees or random forests (RF), are also widely used in remote sensing applications (e.g. Lourenço, 2021; Pohjankukka et al., 2018; Hauglin et al., 2024). Some

* Corresponding author.
E-mail addresses: andras.balazs@luke.fi (A. Balazs), Jukka.Miettinen@vtt.fi (J. Miettinen), mats.nilsson@slu.se (M. Nilsson), johannes.breidenbach@nibio.no (J. Breidenbach), timo.p.pitkanen@luke.fi (T.P. Pitkänen), mari.myllymaki@luke.fi (M. Myllymäki).

https://doi.org/10.1016/j.jag.2026.105104
Received 30 May 2025; Received in revised form 15 December 2025; Accepted 8 January 2026
Available online 9 January 2026

of these methods have been developed and implemented for the multivariate case, i.e., simultaneous modeling of multiple forest attributes. In addition to the *k*-NN method, we tested multivariate RF (Ishwaran et al., 2008) and Multi-Layer Perceptron (MLP) (Abadi et al., 2015), a shallow neural network suitable to be used with satellite data. The set of these three models represent commonly used approaches in forest attribute mapping (e.g. Lourenço, 2021; Zhang et al., 2023).

Different modeling approaches have been compared in several studies. For example, Balazs et al. (2022) tested *k*-NN and different neural networks for predicting volume of growing stock, stand mean height and mean diameter, for each variable separably. In their study using airborne laser scanning (ALS) and field data from Central Finland, the neural networks outperformed the benchmark *k*-NN method by a slight margin, and they called for further investigations, e.g., with higher ALS point density. Moisen and Frescino (2002) compared several approaches with satellite and NFI data from the United States and found only little differences between them when using real data for model training. Brosofske et al. (2014) also reviewed regression, *k*-NN, artificial neural networks, decision trees, and ensembles such as RF and found that none of these methods was superior for predicting forest inventory attributes. Thus, there is a tendency to find little differences between model performances. However, most studies compared the models in local or national situations and with training data available homogeneously across the study region.

Large area mapping of forest inventory attributes brings additional challenges to the prediction methods used. While it would be preferable that the field data available for model training are homogeneously distributed in the region of interest, this may often not be the reality and there may be areas (spatial gaps) with no ground truth data. Using models to predict to such areas is a type of interpolation or extrapolation, which is generally not recommendable. For example, Mitchard et al. (2013) concluded that a good quality map requires good quality field data drawn from across the spatial extent and ecological variability of the prediction area. Meyer and Pebesma (2021, 2022) also stressed the spatial distribution of training data as a prerequisite for applicable maps, see also discussion in Kangas et al. (2023). However, using models trained with data from other regions is often the only way to make predictions for areas that are missing training data. Thus, the models are evidently used in this manner to produce large-scale wall-to-wall maps (Kangas et al., 2018) and their performance should be understood in this context.

The objective of this work was to compare different models for the production of large-scale forest attribute maps. Our main aim was to compare the performance of the tested methods with respect to data availability in the vicinity of the target area. The models were trained using varying numbers of the nearest available field plots, allowing us to evaluate performance in relation to the amount of training data. We also tested different sizes of areas without training data and hypothesized that map accuracy would deteriorate as the size of these areas increase. Furthermore, we expected that differences in model performance observed in the previously mentioned studies could become more pronounced under such conditions. We compared three approaches, *k*-NN, RF, and MLP, all of which are able to produce consistent predictions simultaneously for multiple forest attributes. We used NFI field data and Sentinel-2 satellite mosaics from Norway, Sweden, and Finland. The field data had a good coverage across the whole Nordic region. This allowed us to construct artificial gaps and to use the field data in the gaps to evaluate model performances. We evaluated the model performance at country level as well as locally. The study highlights that the quality of predictions decreases with regard to prediction bias when interpolation or extrapolation is required.

2. Material

2.1. Field data

We used Norwegian, Swedish, and Finnish NFI plot level data from years 2020–2021 (Fig. 1). The NFI plots in each country are distributed according to a probability sampling design (Särndal et al., 1992). The main forest type in these countries is boreal forest in high, middle and southern boreal subzones. Scots pine and Norway spruce are the most dominant tree species. A total of 4701, 6482 (1360 clusters), and 16 558 (3010 clusters) sample plots were available from Norway, Sweden, and Finland, respectively. Response variables calculated for sample plots were harmonized volume (m^3/ha) and above ground biomass (t/ha) of total growing stock, conifer and broad-leaved species. Sample plot locations were available for modeling in one kilometer precision due to data policy restrictions. However, precise plot locations were used to calculate remote sensing features.

The Norwegian NFI plots were fixed circular plots with a radius of 8.92 m and trees with a diameter at breast height (dbh) ≥ 5 cm were measured. Smaller trees were assessed at four circular subplots located 5 m from the plot center in the cardinal directions with a radius of 1.3 m (Breidenbach et al., 2020). For this study, sample plots whose plot center was located inside forest and other wooded land by the FAO's definition (FAO, 2023) were available. The Norwegian sample plots are positioned according to a sampling grid with different sizes in three strata: 3 × 3 km in the productive lowland region, 3 × 9 km in the mountain region, 9 × 9 km in the northern Finnmark region.

The Swedish NFI plots used in this study were concentric circular plots and trees with a dbh ≥ 10 cm were measured within a radius of 7 m. Trees with 4 cm ≤ dbh < 10 cm were measured on a 3.5 m radius plot, and trees with a dbh < 4 cm were measured on two 0.5 m radius plots (Fridman et al., 2014). The criteria for Swedish NFI sample plots to be included in our study was that at least half of the plot's area was within FAO's forest land-use class. In the Swedish NFI sample plots are located in rectangular clusters, whose side length varies between 300 and 1800 m and the distance between sample plots within the clusters varies between 300 and 600 m form south to north. Clusters are distributed in 5 strata covering the entire country with increasing sampling intensity from north to south and from higher to lower elevations (Persson et al., 2017).

The Finnish NFI plots were concentric circular plots, trees with a dbh ≥ 9.5 cm were measured within a radius of 9 m, trees with 4.5 cm ≤ dbh < 9.5 cm were measured within a radius of 4 m, and smaller trees with height >1.3 m from a relascope plot with factor 1.5 (Korhonen et al., 2024). The inclusion criteria for Finnish sample plots was that the plot's center was within forest land according to FAO's definition. Additionally, plots within other forestry land defined by the Finnish land-use/land-cover classification system were also included. The basis of the sampling design was the grid of permanent clusters established in 1996. In this study 4 strata out of the 6 covering Finland was included, as Northernmost Lapland and Åland were not sampled in 2020–2021. Grid spacing was 12 km in southern Finland, 14 km in Central and the southern part of Northern Finland, and 20 km in southern Lapland. Between these already existing permanent clusters one temporary and 3 permanent clusters were additionally established and measured. The number of sample plots per cluster was ranging from 8 to 11 and distances of neighboring plot centers were between 250 m and 450 m. Both, the number of plots per cluster and plot distances varied by stratum (Korhonen et al., 2024).

Plot level volume and biomass calculations for all sample plots were carried out using stem volume harmonization as described by Gschwantner et al. (2019) to avoid discrepancies in growing stock statistics between countries. Country-wise statistics of the field data are shown in Table 1.

2.2. Remotely sensed data

Annual growing season cloud free Sentinel-2 composite images from 2020 and 2021 were used as remote sensing dataset. The image compositing was conducted with an approach developed by Terramonitor and described in detail by Miettinen et al. (2021). To build snow-free growing season composite images, imagery between 15th of June and

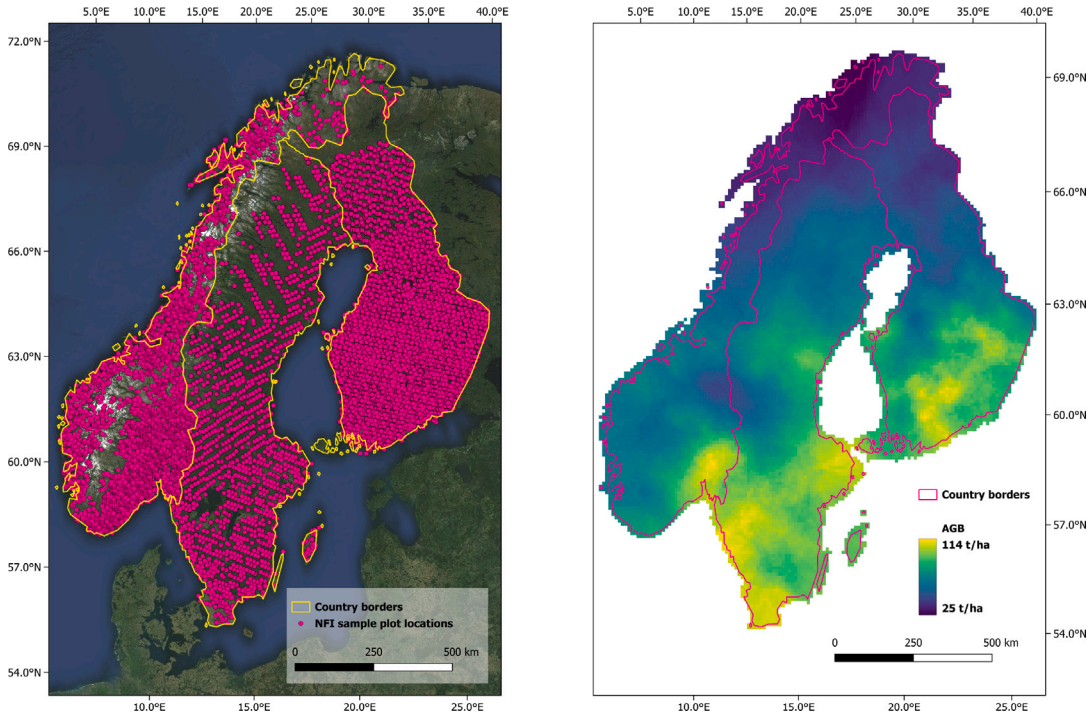


Fig. 1. Approximate locations of NFI field plots (left), and total AGB map of the study region at 10 km resolution derived from field observations (right). The background of the left figure features Sentinel-2 cloudless mosaic (EOX IT Services GmbH, 2020). Biomass was computed for each 10 km grid cell as the average of the nearest 300 sample plots.

Table 1
Country-wise statistics of volume (m^3/ha) and AGB (t/ha) of total growing stock, conifers, and broad-leaved species (n: number of sample plots per country).

	V_{tot}	V_{con}	V_{bl}	AGB_{tot}	AGB_{con}	AGB_{bl}
Norway ($n = 4701$)						
Mean	102.5	77.8	24.8	67.3	49.2	18.2
SD	111.8	107.6	40.9	67.2	63.1	30.2
Max	950.4	950.4	476.4	545.7	545.7	335.1
Sweden ($n = 6482$)						
Mean	148.1	117.0	31.1	82.8	63.6	19.1
SD	123.7	113.0	64.1	60.2	54.9	33.8
Max	884.8	874.8	822.3	428.8	410.7	355.6
Finland ($n = 16\,558$)						
Mean	131.3	105.2	26.1	75.2	58.1	17.1
SD	103.9	95.6	44.8	55.9	49.4	28.6
Max	922.3	922.3	635.5	525.4	437.9	393.5

15th of September were used. The final composite image pixels were weighted averages of the available cloud-free observations. The weights of the observations were based mainly on haziness and shadows (Miettinen et al., 2021). The final composite images included seven spectral bands (B02 Blue 0.49 µm central wavelength, B03 Green 0.56 µm, B04 Red 0.67 µm, B05 Red Edge 1 0.71 µm, B08 NIR 0.84 µm, B11 SWIR 1.61 µm and SWIR 2.19 µm). These bands were selected based on earlier results on optimal set of bands for forest variable prediction (Astola et al., 2019; Miettinen et al., 2021). All bands were resampled to 10 m spatial resolution using nearest neighbor resampling.

In addition to spectral bands, the composite images included a quality band. The composite quality band was calculated per pixel using the formula $P = 1 - \prod_{i=1}^{n}(1 - p_i)$ where the probability of a good observation, p_i, was derived from the weight of observation i. High composite quality band values indicate that the observations that were available to create the composite image at the corresponding location included at least one high quality observation (Miettinen et al., 2021). The quality band was used to select the spectral vector for plots that were covered by several images due to the overlap of adjacent Sentinel-2 tiles. In case of multiple spectral vectors available for one plot, the observation with the highest quality value was selected.

3. Methods

3.1. Sentinel-2 features

Sentinel-2 band values were calculated for each sample plot as area weighted means of pixel values using the R package terra (Hijmans, 2024). Pixels overlapping a 5.64 m radius circle (covering an area of 100 m^2) around sample plot centers were considered in the calculation, and pixel values were weighted by the proportion of the pixel area within the circle (Contributors, 2024; Miettinen et al., 2025). In addition to band-wise reflectance features, we explored the potential benefits of incorporating features such as spectral band ratios and vegetation indices. While we observed a slight improvement in the relative root mean squared error (%RMSE) for broad-leaved forests, the %RMSE values for all other response variables deteriorated. Based on these findings, we decided to limit our feature set to reflectance values, which provided a more balanced and consistent performance across variables. This approach is also consistent with the methodology used in the operational Multi-Source National Forest Inventory of Finland (Mäkisara et al., 2022).

3.2. Models

Modeling was carried out using three different methods: k-Nearest Neighbor, Multivariate Random Forest, and Multilayer Perceptron.

3.2.1. k-Nearest Neighbor (k-NN)

The k-Nearest Neighbor method has been widely used in forest inventory applications (Chirici et al., 2016). In k-NN, a prediction to a target unit is computed as a weighted average of the response variable values of the k nearest training data units. The k nearest units are

determined by a distance metric in the q-dimensional feature space, with q equal to the number of features. In this study, the features were equal to the Sentinel-2 spectral band values described in Section 2.2, the distance metric was the Euclidean distance, and equal weights were given to all seven features. Calculations were carried out in R statistical software using the nabor package (Elseberg et al., 2012). Utilizing the Finnish field dataset, we evaluated various k values (3, 4, 5, and 6) by predicting the response variables and choosing the k value that minimized the total RMSE score. The best fit for k was found to be 5 which was used in all analysis that follow.

3.2.2. Multivariate Random Forest (RF)

Random Forest is a method from the family of ensemble learning. Random Forests are an ensemble of bagged decision trees introduced by Breiman (1996). We utilized the R package randomForestSRC (Ishwaran and Kogalur, 2025), which enables the application of Random Forests for multivariate regression tasks. The response variable given to randomForestSRC in our study was six-dimensional including volume and AGB of total growing stock, conifers and broad-leaved species. The explanatory variables were the seven features, i.e., Sentinel-2 bands. Similarly to the k-NN method, we used the Finnish dataset to test values ranging from 1 to 7 for the number of variables randomly selected at each node (mtry) and values 50, 100 and 300 for the number of trees in the forest (ntree). The values tested for ntree were relatively low, which was a conscious choice due to the computationally intensive nature of the study design. As a result of the test runs, the value 2 was chosen for mtry, and 100 for ntree.

3.2.3. Multi-Layer Perceptron (MLP)

MLPs are simple, fully connected neural networks. In this study MLPs were implemented in Python language using TensorFlow (Abadi et al., 2015). MLPs included a single hidden layer, with the number of neurons matching the count of Sentinel-2 bands utilized. MLPs were set to minimize mean squared error (MSE) of the response variables. The initial learning rate of 0.01 was lowered after each epoch exponentially in a way that it was 0.001 after the maximum number of epochs. Epoch refers to one complete pass through the entire training dataset. The training data was fed to the MLPs in minibatches of 32. The batch size was selected after testing the values 8, 16, 32, and 64 on the Finnish data. Other parameters like number of layers, number of neurons, and learning rate were not tuned due to the high number of individual models to be trained (see Section 3.3). We aimed for a design that is light-weight, but is able to compete with the other two methods.

Since neural networks are prone to overfitting, we withheld 20% of the training data for model validation during training. After each training iteration the MSEs of predicted response variables were calculated using the initially withheld data. Training was stopped when the sum of validation MSEs did not improve after 30 validation rounds. The best performing model weights in terms of validation MSE were saved. MLPs were trained for 200 epochs with at least 1000 sample plots in the training dataset, and below 1000 plots for 400 epochs. During one training epoch the training data was fed to the MLP in minibatches (training step). Model weights were updated between each training step. At the end of each epoch a validation round was carried out and the model was saved if validation results were better than in earlier epochs. In order to obtain consistent results for all response variables, conifer and broad-leaved biomass predictions were derived within the MLPs from conifer and broad-leaved volume predictions by means of least squares linear regression models. Models for conifer and broad-leaved species were fitted separately on the training data before training/validation split with biomass as response and volume as explanatory variable. Parameters of the regression models were then used within the MLPs to calculate biomass predictions for conifer and broad-leaved species from volume predictions, and total biomass predictions were calculated by the sum of these. Finally all response variable predictions were joined and error metrics calculated. This approach was found to be suitable due to the high correlation between AGB and volume. In some cases MLPs did not converge likely due to small training sample size. In such cases model fitting was repeated with randomly re-initialized model weights.

3.3. Experimental setup

To evaluate model performance under varying conditions of training data availability, we designed an experiment using a 10 km × 10 km grid, aligned with the INSPIRE grid and using the ETRS89 Lambert Azimuthal Equal Area coordinate reference system. Each 10 km × 10 km grid cell represents a localized prediction area for which a separate model is fitted. The 10 km grid was chosen to allow the model fit to adapt smoothly to local conditions, while balancing spatial resolution with field data availability and computational feasibility. Importantly, sample plots located within a given grid cell were excluded from the training data used to fit the model for that cell. Instead, these plots were reserved for validation, allowing us to assess prediction accuracy independently of the training data. This design simulates the scenario where field data are unavailable within the target mapping area, and models must rely on data from surrounding regions. For model training, we selected the N nearest sample plots located outside a circular gap centered on each grid cell. The radius R of this gap was defined as $R = \sqrt{(2B^2)}/2$, where B is a multiple of the grid cell side length (10, 30, 50, 110, 150, 200 km), resulting in approximate radii of 7, 21, 35, 78, 106, and 141 km. This setup allowed us to test the impact of increasing distances between training data and prediction locations. We trained models using six different values of N (100, 200, 300, 500, 1000, and 3000 plots). With each value of N, we considered each value of R, resulting in 36 training configurations in total. This factorial design allowed us to isolate the effects of spatial gap size from the effects of training sample size and to evaluate their individual and combined influence on model performance. For each configuration, localized models were fitted using k-NN, RF, and MLP methods. In total, over 206,000 models were trained per method across 5736 grid cells. An illustration of the experimental design using real data is provided in Figs. 2 and 3.

3.4. Evaluation of model performances and validation metric maps

Each fitted model was used to predict forest attributes at the plot locations within its corresponding 10 km × 10 km grid cell. These plots, which were excluded from model training (see Section 3.3), served as independent validation data. This approach allowed us to assess model performance in a realistic extrapolation scenario, where predictions are made for areas lacking local training data. For each sample plot in the study region (Fig. 1), we obtained both the observed value y_i and the predicted value $\hat{y}_i$ for all response variables and modeling methods. Validation metrics – relative RMSE and relative bias – were calculated using the following formulas:

$$\%\text{RMSE} = \frac{1}{\bar{y}} \sqrt{\frac{\sum_{i=1}^{N_a} (y_i - \hat{y}_i)^2}{N_a}} \tag{1}$$

$$\%\text{bias} = \frac{1}{\bar{y}} \frac{\sum_{i=1}^{N_a} (y_i - \hat{y}_i)}{N_a} \tag{2}$$

where N_a is the number of validation plots in the area and $\bar{y}$ is the mean of the observed values. We computed %RMSE and %bias at country level, at sub-country level areas in the three countries, and for neighborhoods of all 10 km × 10 km grid cells. Here the neighborhood of a grid cell was the area covered by 300 sample plots nearest to the grid cell center, i.e. the validation metrics for each grid cell were computed from the 300 plots. This multi-scale evaluation enabled us to assess both general model performance and localized prediction accuracy, highlighting how spatial gaps in training data influence mapping outcomes.

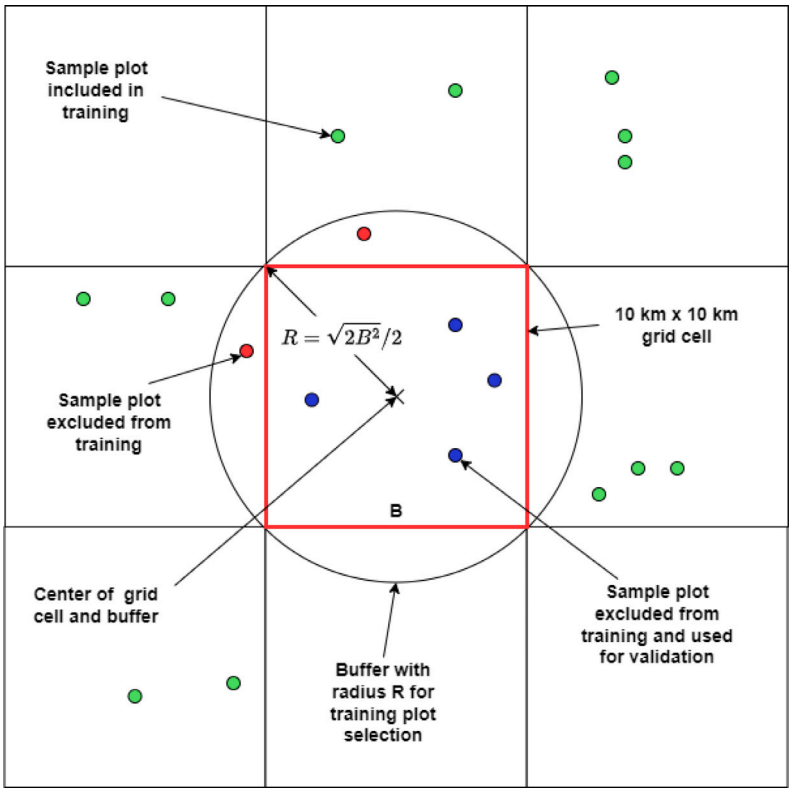


Fig. 2. Graphical representation of the experimental setup. The red square represents a grid cell of size 10 km × 10 km. (Note: in this case B and the grid cell size are both 10 km.)

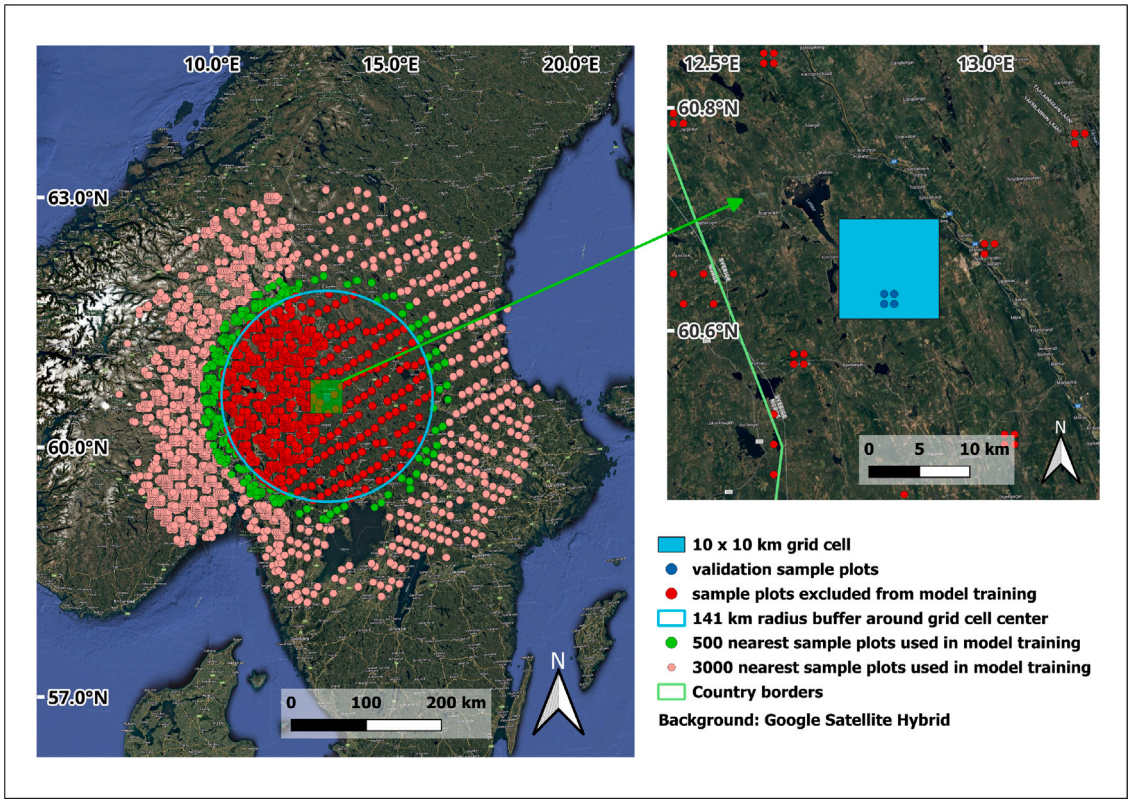
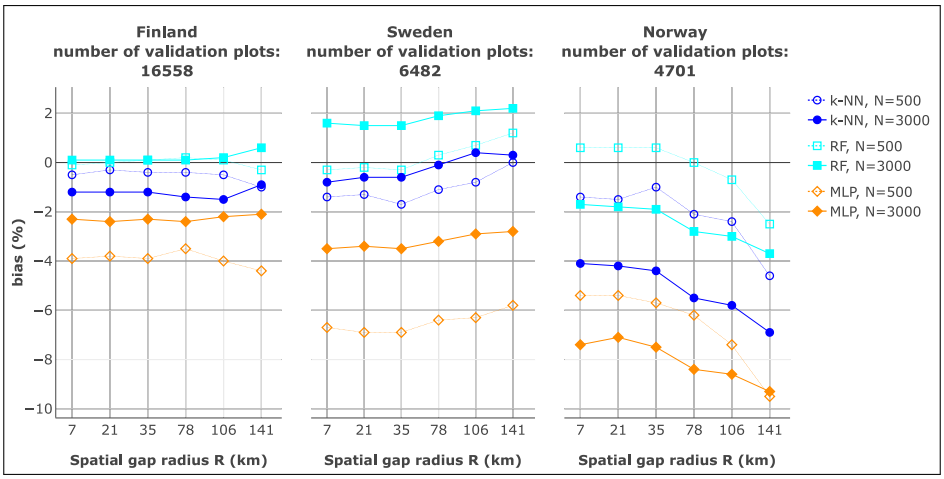
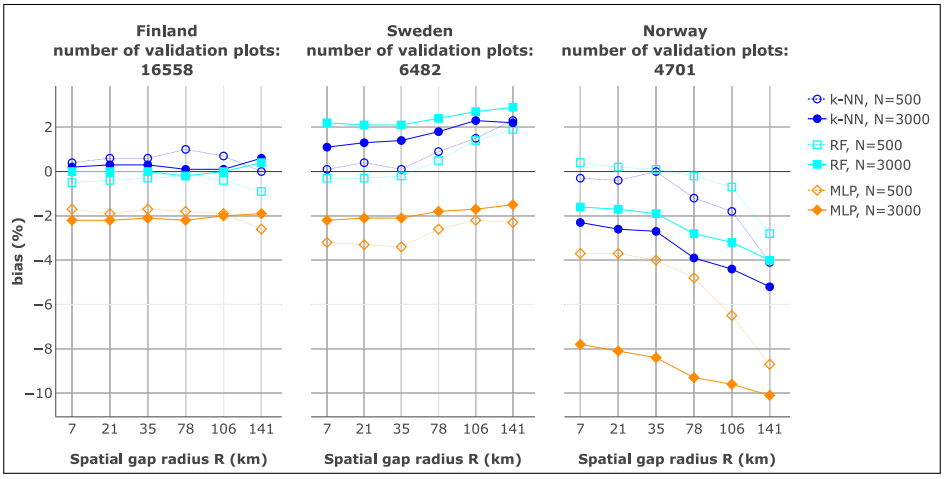


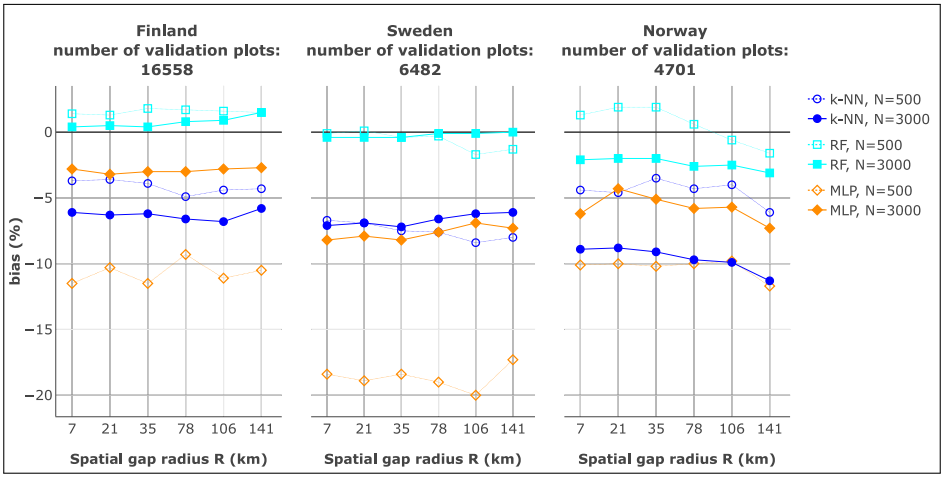
Fig. 3. Example of the research design with $R \approx 141$ km, $N = 500$ and $N = 3000$.



(a) AGB of total growing stock



(b) AGB of conifers



(c) AGB of broad-leaved

Fig. 4. %bias of AGB of total growing stock (a), conifers (b) and broad-leaved species (c) for Finland, Sweden, and Norway with the number of training plots (N) equal to 500 and 3000.

4. Results

Below we describe in detail the simulation study results. Figures S.1–S.10 can be found in the Supplementary Material.

4.1. Country-wise results

We first inspected country-specific %bias and %RMSE with respect to the number of sample plots used for training N (Supplementary

Material Figures S.1, S.2, S.3, and S.4). While the other methods were less affected, N had a significant impact on MLP predictions. With increasing N %bias steeply decreased in absolute terms. The same effect could be observed for %RMSE particularly in the case of broad-leaved AGB and volume (Figure S.2c and S.4c). In most of these cases MLP achieved %bias and %RMSE scores comparable to k-NN and RF with $N = 500$ and above. Based on these results we have selected $N = 500$ and 3000 to further examine the impact of spatial gap size on model performances.

Results of %bias and %RMSE for total, conifer, and broad-leaved AGB with respect to spatial gap radius R are shown in Figs. 4 and 5. RF and k-NN led to similar %bias for total and conifer AGB. MLP predictions were the most biased, regardless of R and for all countries. RF models produced the least biased broad-leaved AGB predictions. In Finland and Sweden R had no significant impact on the %bias of biomass variables. However, in Norway AGB predictions' %bias increased in absolute terms with increasing spatial gap size. Regarding the volume predictions, there were no significant differences between the tested methods for total and conifer volume predictions (Figure S.5). However, for broad-leaved volume, RF and MLP (with $N = 3000$) performed markedly better than k-NN. R had no visible effect on the results.

In terms of %RMSE, RF produced the best results throughout biomass variables and countries with MLP being the second best method (with $N = 3000$) (Fig. 5). R had a significant impact on %RMSE only in the case of broad-leaved AGB, resulting in higher %RMSE values with increasing gap radius in all countries. %RMSE of volume predictions showed a very similar pattern compared to AGB (Figure S.6). For both broad-leaved biomass and volume, MLP produced significantly more accurate predictions in terms of %RMSE with $N = 3000$ compared to $N = 500$.

4.2. Local results

In addition to country-wise results, we inspected the spatial distributions of %RMSE and %bias across the whole study region by raster maps. Recall that these evaluation metric values were computed from the sample plot data, which were not used for model training (see Section 3.4). We applied adaptive smoothing of the metrics with the %bias and %RMSE computed for each grid cell from the nearest 300 sample plots. In the following, we present AGB maps created using RF models. Corresponding maps produced by k-NN (Figures S.7 and S.8) and MLP (Figures S.9 and S.10) can be found in the Supplementary Material. The k-NN and MLP maps produced with different R and N values exhibited very similar patterns to RF maps.

Raster maps show high variation between and within countries in terms of local %RMSE with largest values found in Norway (Fig. 6). The effects of the number of plots used for training the models, N (top vs. bottom of Fig. 6) and the spatial gap radius R (left vs. right of Fig. 6) on the %RMSEs were minor: the minimum and maximum values of all four maps of Fig. 6 were between 40%–42% and 94%–98%, respectively.

Maps of %bias also display a similar spatial pattern regardless of N and R (Fig. 7). However, both N and R had a significant effect on the magnitude of variation. With $R \approx 78$ km and $N = 3000$ (Fig. 7(b)), %bias ranged from −21% to 27%, whereas in case of $R \approx 7$ km and $N = 500$ (Fig. 7(c)) the value range was reduced to −12%–12%. When comparing maps using other parameter combinations to the case of $R \approx 7$ km and $N = 500$ (Fig. 7(c)), one can see regional differences on how N and R affected %bias and %RMSE. Higher number of sample plots used for model training with the same spatial gap size (Fig. 7(a)) produced more biased results than the increased spatial gap size and the same number of training plots in Norway and Sweden (Fig. 7(d)). On the other hand, increasing the spatial gap radius led to higher %bias in Finland than with increased N. %bias values ranged from −18% to 24% and from −17% to 28% for maps shown in Figs. 7(a) and 7(d), respectively.

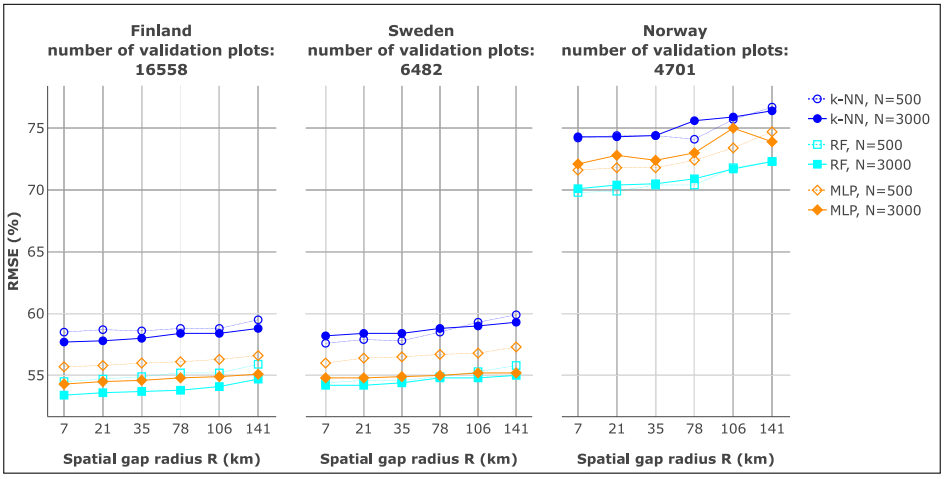
We also investigated some of the bias-hotspots visible in Fig. 7(b). Our hypothesis was that the explanation for high positive or negative %bias was the difference in the amount of biomass at these locations and the surrounding area further away from them. Fig. 8 shows three hotspots with high positive %bias (from 26% to 27%) and corresponding distributions of observed total AGB calculated from the nearest 100, and 3000 sample plots excluding the nearest 100. Fig. 9 on the other hand displays three locations with high negative %bias (from −20% to −16%) and corresponding distribution curves. Hotspot locations are also displayed over observed total AGB maps for reference. As expected, the areas with high positive %bias had less plots with high biomass, while the biomass distribution in areas with high negative %bias was shifted towards larger values in comparison to the 3000 reference plots. This at least partly explains why the predictions were too high in the former and too low in the latter case.

We further inspected the %biases of the tested methods in the neighborhood of high positive and negative %bias hotspot areas, respectively (Figs. 10 and 11). Validation plots were selected around the locations displayed in Figs. 8(a) and 9(a), numbers 1, 2, and 3 corresponding to the locations in Norway, Sweden, and Finland, respectively. This way we were able to compare the performances of the methods locally. The results indicate that the spatial gap radius (R) had a significant effect on the %bias of estimates with the number of validation plots being 100 or 300. On the other hand, the size of spatial gaps had a less marked effect when tested with 3000 validation sample plots. Furthermore, estimates produced using models trained with 500 sample plots were less biased compared to models with $N = 3000$. Generally, we can say that there was no clearly superior method in terms of %bias when looking at the results of bias-hotspot areas.

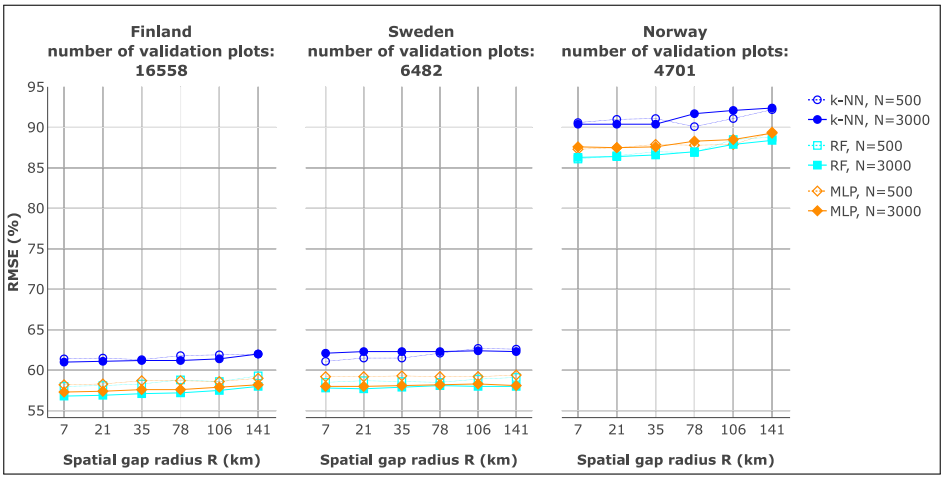
5. Discussion and conclusions

The aim of this study was to investigate how missingness of field data for training models affects map accuracy and to compare the performance of k-NN, RF and MLP methods when such missingness occurs. Map accuracy was expected to decrease when no field data were available in the neighborhood of the area of interest, and we also expected to observe larger differences between the methods in such cases. As expected, increasing the radius of spatial gaps, i.e. the area with no field data available for training the models for mapping, generally led to decreased map accuracy in terms of %bias. The effect of the gap radius was minor in terms of %RMSE for the total AGB and volume predictions. However, broad-leaved predictions, which had higher %RMSEs overall, further suffered from increasing spatial gaps (Figs. 5 and S.6). Larger effects were observed in terms of %bias of AGB. Here, the country-level %bias clearly increased with respect to the size of spatial gaps in Norway (Fig. 4). Further, we could see that more locally the magnitude of %bias increased with spatial gaps across the whole studied Nordic region (Fig. 7). We did not find any clear differences between the tested modeling approaches' behavior with respect to the spatial gap radius. However, RF was generally the best approach both in terms of %RMSE and %bias and regardless of the number of training plots and the radius of spatial gap used.

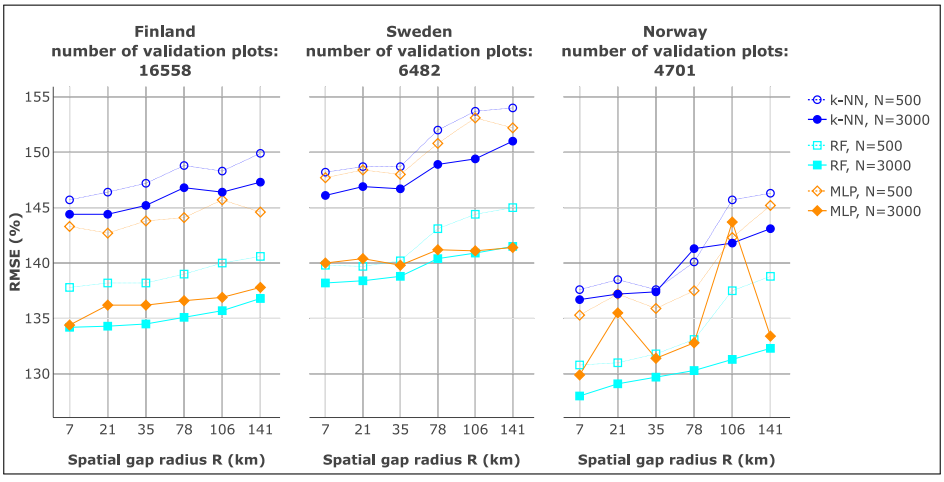
The amount of training data N also affected the results. While k-NN and RF were not that sensitive to N in terms of country-level %RMSE and %bias, MLP required a large number of field plots for training (in our study $N = 3000$). This could be expected as sufficient amount of training data is needed for neural networks to converge (Alwosheel et al., 2018; Baum and Haussler, 1988). Additionally, MLP showed an unexpected behavior: MLP's volume predictions were less biased than AGB predictions (Figs. S.5a vs. 4(a)). This could be due to the way AGB predictions were calculated from volume predictions using linear models, even though the fit of the linear models was rather good ($R^2 > 0.95$ on country level). Future studies could test other possible ways to produce consistent predictions with MLP simultaneously for a set of forest attributes.



(a) AGB of total growing stock



(b) AGB of conifers



(c) AGB of broad-leaved

Fig. 5. %RMSE of AGB of total growing stock (a), conifers (b) and broad-leaved species (c) for Finland, Sweden, and Norway with the number of training plots (N) equal to 500 and 3000.

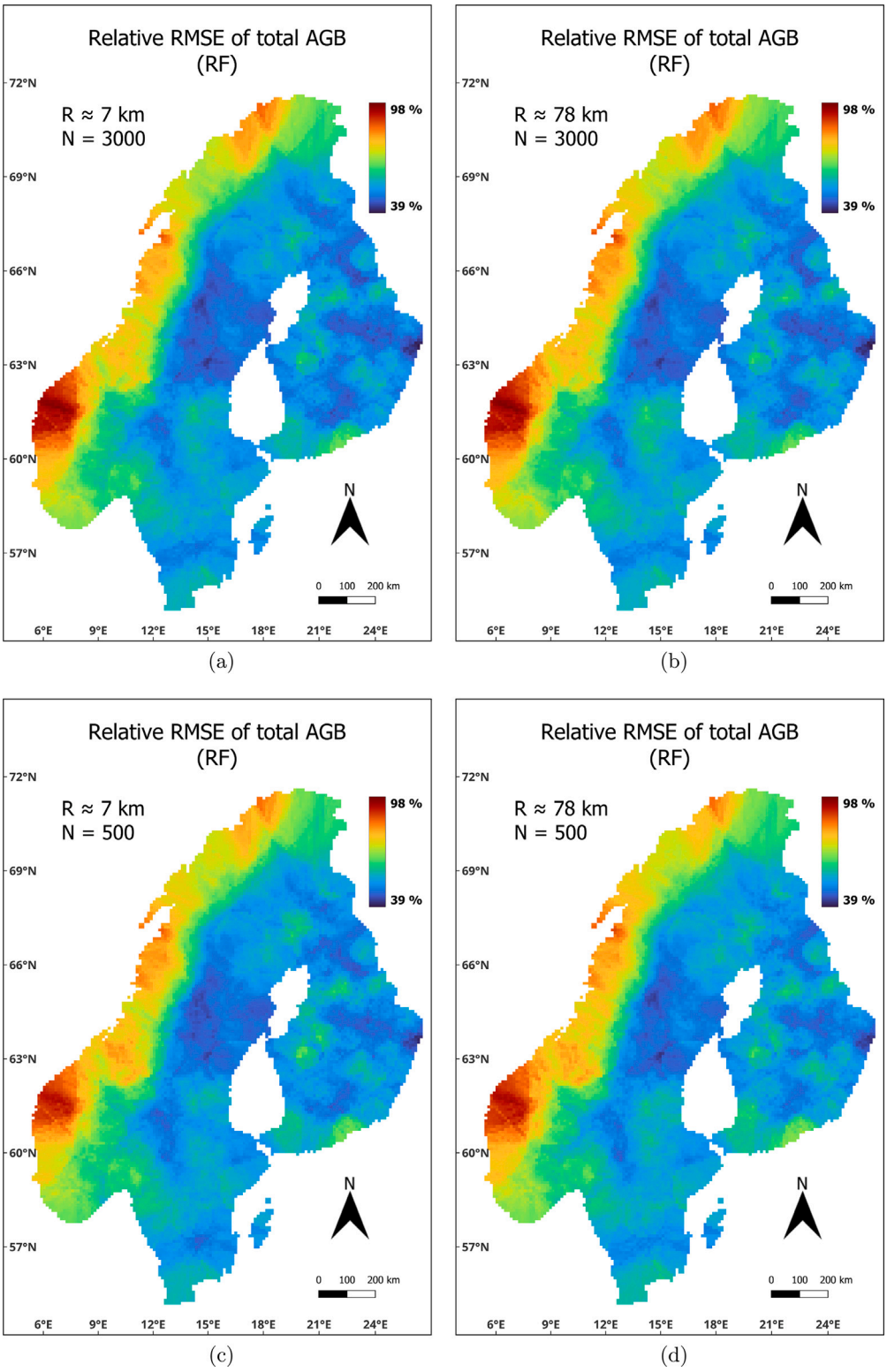


Fig. 6. Maps showing spatial variation of %RMSE at 10 km resolution. Pixel values were derived from RF models that were trained using different combinations of spatial gap radius (R) and number of training plots (N). Note: the applied value range of 39%–98% is only for visualization purposes, actual range of pixel values of individual maps might differ.

In our country-wise comparison, Norway's results clearly stood out from those of Finland and Sweden. Overall, %RMSE by RF was larger in Norway than in Sweden and Finland: the country-level %RMSE values for total AGB were between 70% and 72% for Norway, and 53%–55% for Sweden and Finland. This was partly attributed to the lower average volume and AGB values in Norway (see Table 1), and similar differences between the countries have been reported before (Puliti et al., 2020; Persson et al., 2021; Pitkänen et al., 2024). Absolute RMSE values for AGB were in comparable ranges across all countries. For example, absolute RMSE results for total AGB with $R \approx 7$ km and

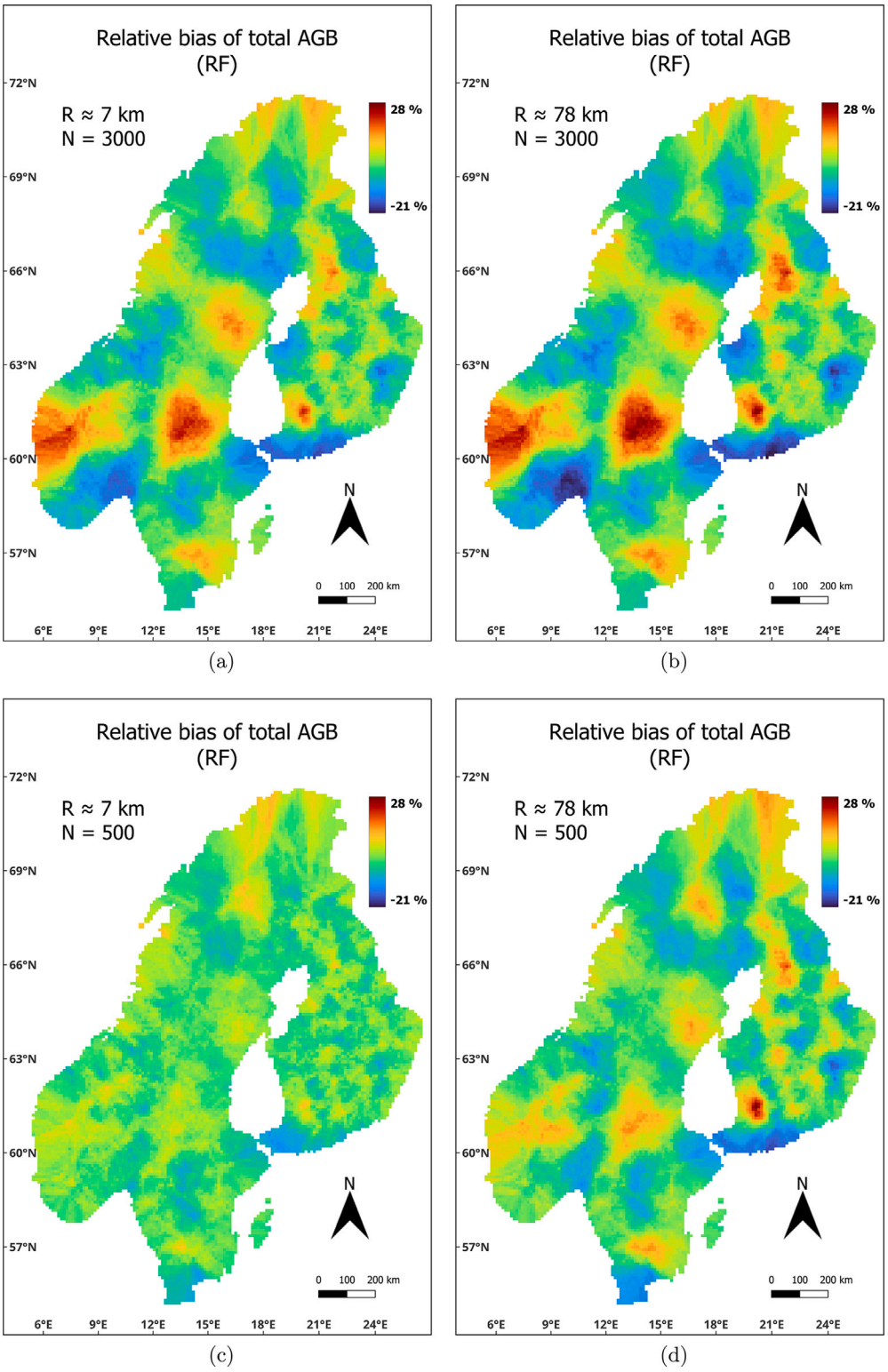
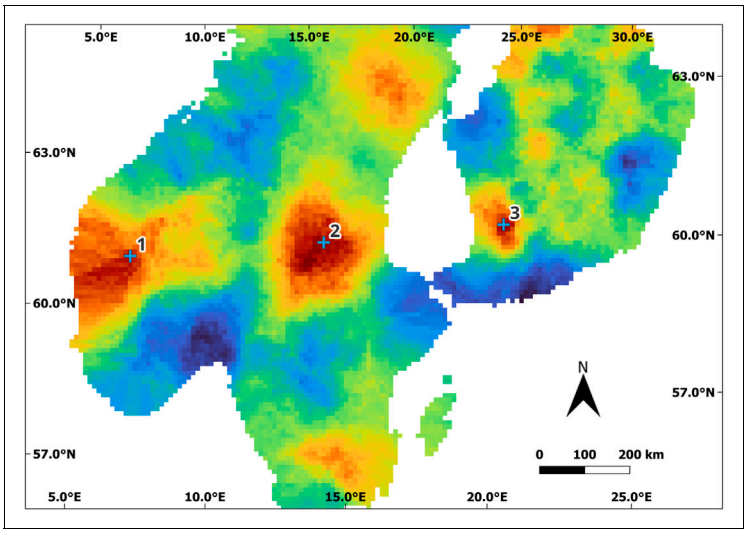


Fig. 7. Maps showing spatial variation of %bias at 10 km resolution. Pixel values were derived from RF models that were trained using different combinations of spatial gap radius (R) and number of training plots (N). Note: the applied value range of −21%–28% is only for visualization purposes, actual range of pixel values of individual maps might differ.

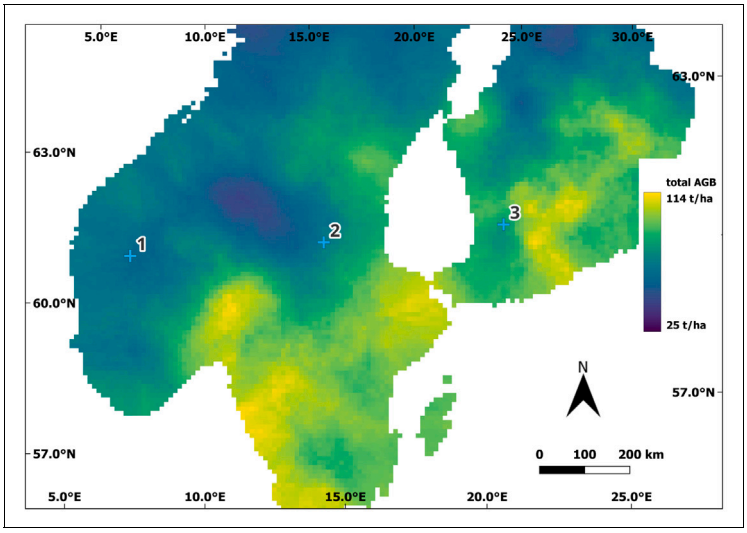
N = 500 were 41.0, 45.1, and 47.0 t/ha for Finland, Sweden, and Norway, respectively.

While we are not aware of publications that would have considered spatial gaps in the training data and be both thematically and geographically relevant, there are several studies that have reported %RMSE results in parts of our study region. Our mean %RMSE results align well with the studies from the literature in Finland, Sweden and Norway. All numerical results of the present study reported below are for spatial gap size of 7 km.

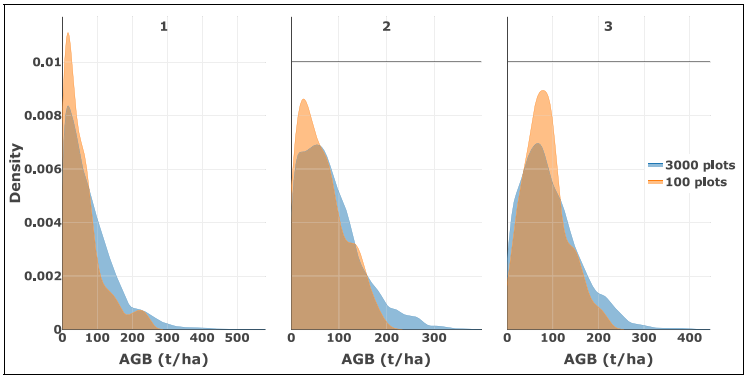
In Finland, Pitkänen et al. (2024) estimated AGB and volume using various mosaicing techniques of Sentinel-2 imagery and the k-NN method over approximately 75 000 km^2. Their best approach yielded



(a)



(b)

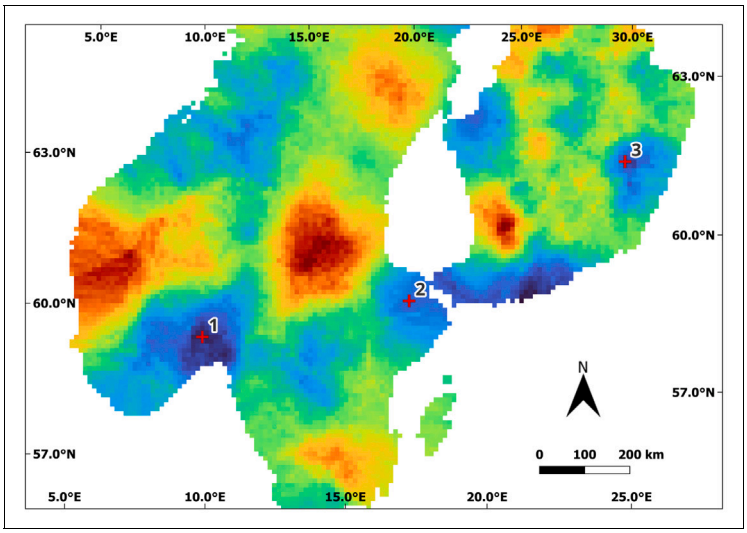


(c)

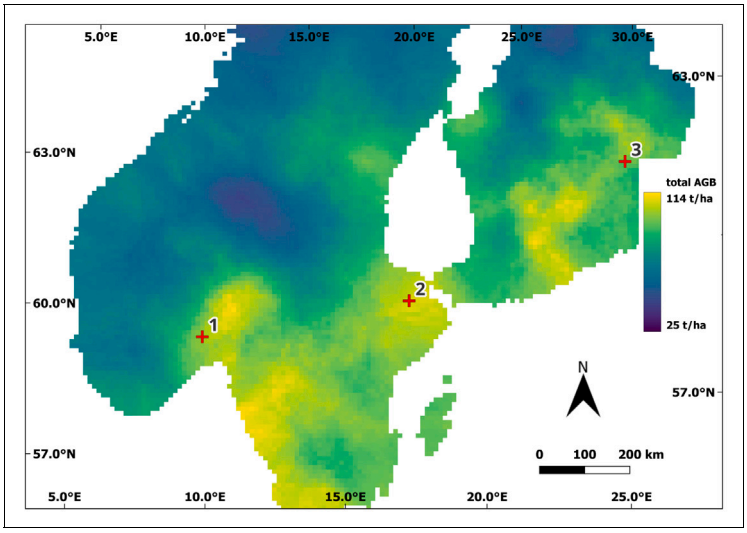
Fig. 8. Map indicating high %bias hotspots (method: RF, $R \approx 78$ km, and $N = 3000$) (a), hotspot locations over map of observed total AGB (b), and distributions of observed total AGB for each hotspot using the nearest 100 and 3000 sample plots (c).

a %RMSE of 56.2% for total volume of growing stock, whereas in our study, k-NN produced a relative RMSE of 61.1% with $N = 3000$ (56.5% using RF). For deciduous species, their best %RMSE was 131.4%, while our k-NN model resulted in 148.3% with $N = 3000$ (137.8% using RF). Tuominen et al. (2017) tested various feature combinations for forest variable estimation using the k-NN method in Finland and reported %RMSEs of 61.0% for total volume and 162.1% for volume of broad-leaved species, using bands 1–7 of Landsat 8 imagery.

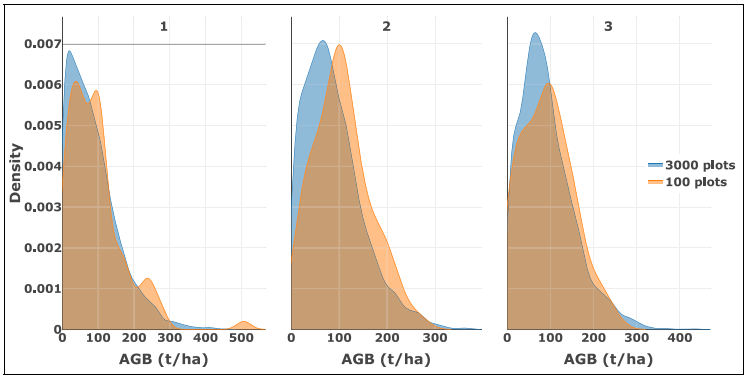
In Norway, Puliti et al. (2020) reported a %RMSE of 80.7% for total AGB in Norway when using only Sentinel-2 reflectance values and RF.



(a)



(b)

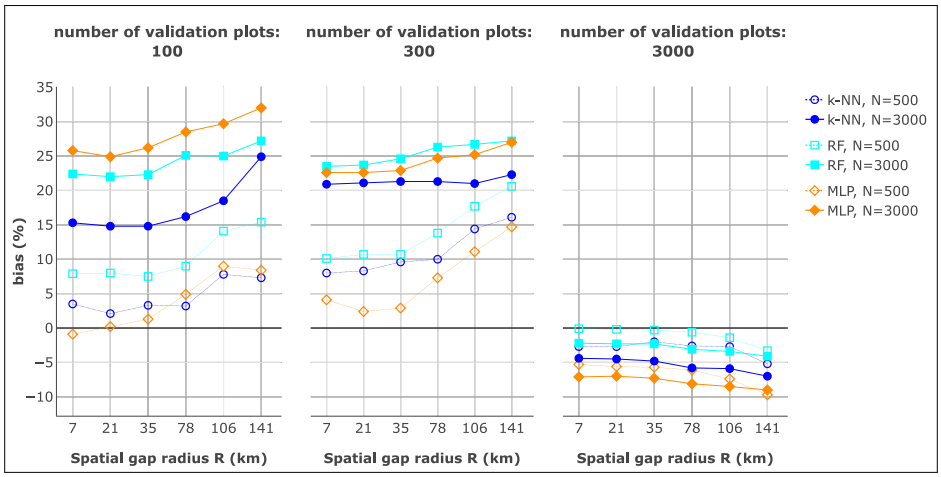


(c)

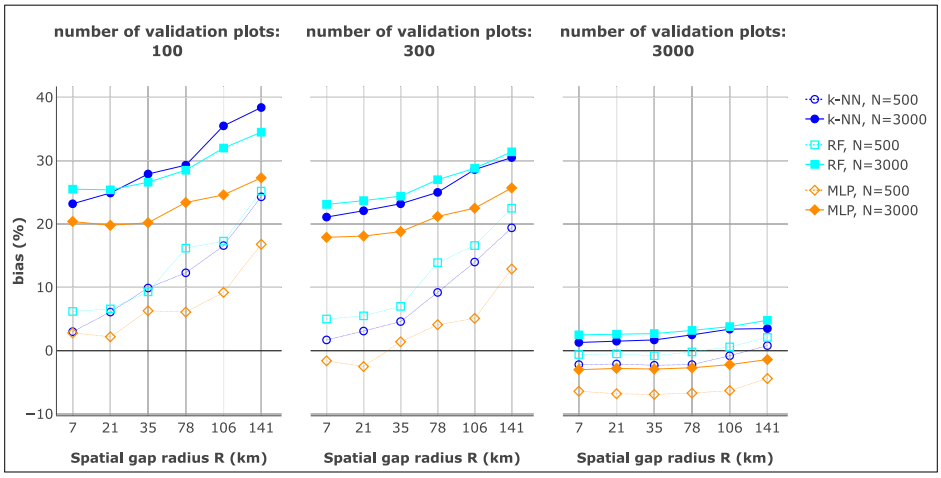
Fig. 9. Map indicating low %bias hotspots (method: RF, $R \approx 78$ km, and $N = 3000$) (a), hotspot locations over map of observed total AGB (b), and distributions of observed AGB for each hotspot using the nearest 100 and 3000 sample plots (c).

By combining Sentinel-2 data with a 2-meter-resolution canopy height model derived from ArcticDEM imagery, they achieved a reduced RMSE of 72.8%, which is consistent with our RF-based country-level result for Norway (70.1% with $N = 3000$). The authors also documented a %bias of −2.8%, slightly higher (in absolute terms) than the −1.7% observed in our study.

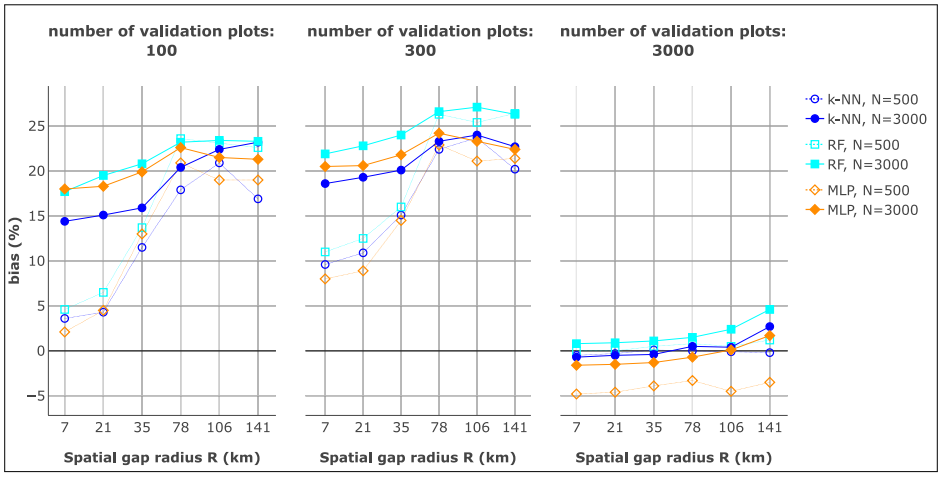
Further, a regional study by Fazakas et al. (1999) in central Sweden, using only Landsat TM reflectance values and the k-NN method,



(a) Norway



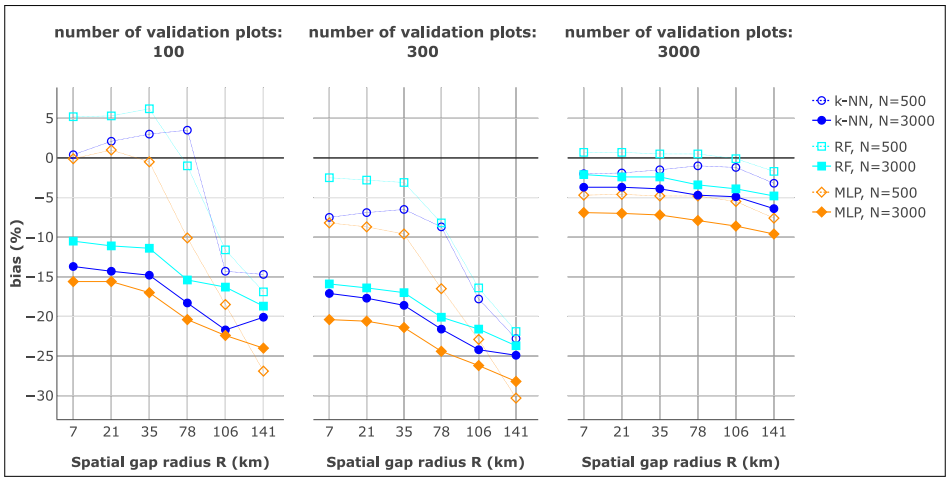
(b) Sweden



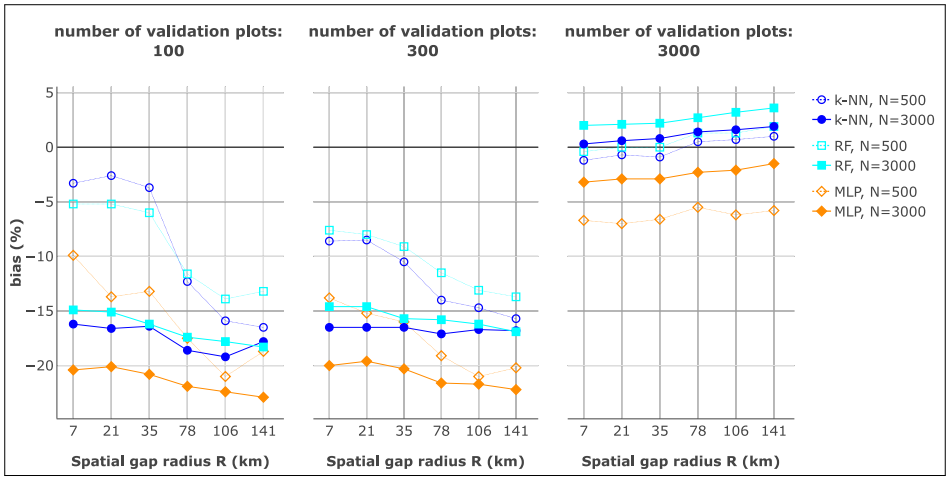
(c) Finland

Fig. 10. %bias of total AGB using various amounts of validation plots around high positive %bias hotspot areas in Norway, Sweden, and Finland with respect to spatial gap radius (R) and with $N = 500$ or $N = 3000$.

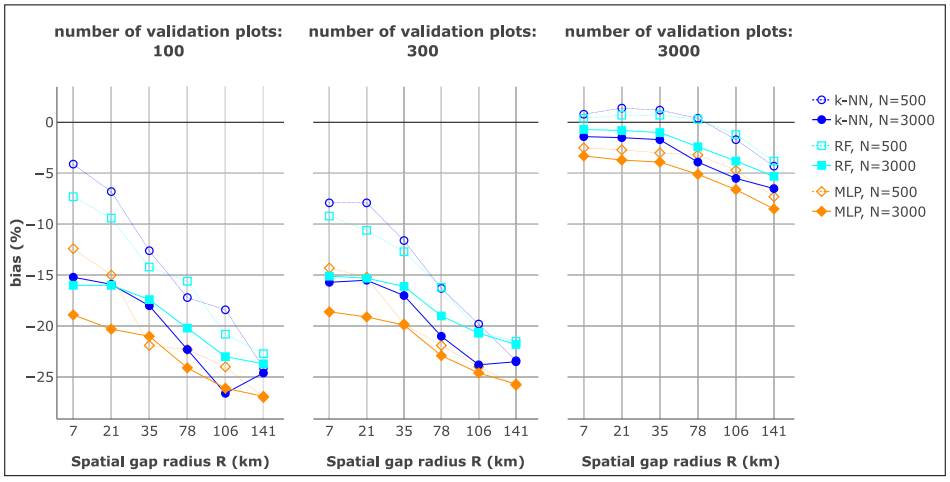
resulted in approximately 66% relative RMSE for both total AGB and volume, validated with Swedish NFI plots. In our study, %RMSE scores for Sweden using k-NN with $N = 500$ were 57.6% and 66.7% for total AGB and volume, respectively. In a study covering 70% of Sweden's forest area, Persson et al. (2021) combined TanDEM-X and Sentinel-2 data to predict tree volume and AGB using k-NN. They reported a plot-level %RMSE of 68.0% for total volume using Sentinel-2 data alone, which closely matches our result of 66.7%. When combining TanDEM-X and Sentinel-2 data, %RMSE decreased to 59.3%, indicating that the use of multiple sensors can enhance prediction accuracy.



(a) Norway



(b) Sweden



(c) Finland

Fig. 11. %bias of total AGB using various amounts of validation plots around high negative %bias hotspot areas in Norway, Sweden, and Finland with respect to spatial gap radius (R) and with $N = 500$ or $N = 3000$.

Besides %RMSE being highest in Norway, the effect of missingness of field plots in the vicinity of the area where the predictions were produced had the highest effect on %bias in Norway (Fig. 4). This likely results from more abrupt changes in the landscape in Norway in comparison to that in Sweden and Finland. Furthermore, due to the shape of Norway, in case of bigger gap sizes the field data used for model training was partly dominated by Swedish sample plots, which tend to have higher biomass than Norwegian plots, given the same features. In our study, always the nearest N sample plots were used for training the models, with the specified spatial gap. Future research

may investigate if restricting the plot selection by other meaningful conditions, such as altitude, could make them more representative for the targeted conditions.

Regarding local results, it is interesting to note how similar the local %RMSEs of total biomass behaved with respect to both the number of training plots (N) and the gap radius (R) (Fig. 6). Possibly, even if the most localized collection of plots is used for training, they extend spatially far enough to include the "full range" of variation, i.e. the amount of field data is sufficient even in this case. More generally, if local conditions are very variable, decreasing the radius for selecting plots may not help much for minimizing the %RMSE. On the other hand, the relative local bias for different N and R differed substantially from each other (Fig. 7). The smallest %bias was observed when a small number of nearby plots was used (Fig. 7(c)) - this was the case even in the most "difficult" areas with high %RMSE. This may suggest that the biggest advantage of localizing the selection is to get better estimates on the distribution and the mean value, while not helping much at the level of a single plot or in small area estimation. Another noteworthy aspect involves the areas where biomass was substantially over- or underestimated (Figs. 8 and 9). Our study shows that notably higher biases often arise at biomass gradients: negative biases occur in high-biomass areas adjacent to lower biomass, while positive biases appear where low-biomass zones are surrounded by higher biomass. When either the spatial exclusion gap is extended, or N is increased, the distribution of the training plots shifts further from the characteristics of the target region, and results become more biased (Fig. 7).

Our study demonstrated that spatial gaps in field data and the resulting need for interpolation or extrapolation can reduce map accuracy, particularly by increasing %bias. This outcome was expected and aligns with previous studies that have already called for methods to estimate the area of applicability of spatial prediction models (Meyer and Pebesma, 2021, 2022). Therefore, it would be preferable to have training data available from the entire target area. As this evidently does not always hold, map products should always include metadata describing the spatial coverage of the field data used in their production.

CRediT authorship contribution statement

Andras Balazs: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Jukka Miettinen:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Mats Nilsson:** Writing – review & editing, Writing – original draft, Data curation. **Johannes Breidenbach:** Writing – review & editing, Writing – original draft, Data curation. **Timo P. Pitkänen:** Writing – review & editing, Writing – original draft. **Mari Myllymäki:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research leading to these results has received funding from the European Union Horizon Europe (HORIZON) Research & Innovation programme under the Grant Agreement no. 101056907 (PathFinder). AB, TP and MM did their work under the Research Council of Finland's flagship ecosystem for Forest-Human–Machine Interplay–Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) (Grant number 357909).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jag.2026.105104.

Data availability

The authors do not have permission to share data.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/.

Alwosheel, A., van Cranenburgh, S., Chorus, C.G., 2018. Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis. J. Choice Model. 28, 167–182. http://dx.doi.org/10.1016/j.jocm.2018.07.002.

Astola, H., Häme, T., Sirro, L., Molinier, M., Kilpi, J., 2019. Comparison of sentinel-2 and landsat 8 imagery for forest variable prediction in boreal region. Remote Sens. Environ. 223, 257–273. http://dx.doi.org/10.1016/j.rse.2019.01.019.

Balazs, A., Liski, E., Tuominen, S., Kangas, A., 2022. Comparison of neural networks and k-nearest neighbors methods in forest stand variable estimation using airborne laser data. ISPRS Open J. Photogramm. Remote. Sens. 4, 100012. http://dx.doi.org/10.1016/j.ophoto.2022.100012.

Baum, E., Haussler, D., 1988. What size net gives valid generalization?. In: Touretzky, D. (Ed.), Adv. Neural Inf. Process. Syst. Morgan-Kaufmann, pp. 81–90.

Breidenbach, J., Granhus, A., Hylen, G., Eriksen, R., Astrup, R., 2020. A century of national forest inventory in Norway – informing past, present, and future decisions. For. Ecosyst. 7, 46. http://dx.doi.org/10.1186/s40663-020-00261-0.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. http://dx.doi.org/10.1007/BF00058655.

Brosofske, K.D., Froese, R.E., Falkowski, M.J., Banskota, A., 2014. A review of methods for mapping and prediction of inventory attributes for operational forest management. For. Sci. 60, 733–756. http://dx.doi.org/10.5849/forsci.12-134.

Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E.O., Waser, L.T., Travaglini, D., McRoberts, R.E., 2016. A meta-analysis and review of the literature on the k-nearest neighbors technique for forestry applications that use remotely sensed data. Remote Sens. Environ. 176, 282–294. http://dx.doi.org/10.1016/j.rse.2016.02.001.

COM, 2023a. ANNEXES to the proposal for a regulation of the european parliament and of the council on a monitoring framework for resilient European forests. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52023PC0728&qid=1707223205537.

COM, 2023b. Proposal for a regulation of the European parliament and of the council on a monitoring framework for resilient European forests. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52023PC0728.

Contributors, n., 2024. nFIESTA wiki home. https://gitlab.com/nfiesta/pathfinder_demo_study/-/wikis/home (Accessed: 2025-02-19).

Elseberg, J., Magnenat, S., Siegwart, R., Nüchter, A., 2012. Comparison of nearest-neighbor-search strategies and implementations for efficient shape registration. J. Softw. Eng. Robot. JOSER 3, 2–12, https://robotik.informatik.uni-wuerzburg.de/telematics/download/joser2012.pdf.

EOX IT Services GmbH, 2020. Sentinel-2 cloudless. https://s2maps.eu.

FAO, 2023. Global forest resources assessment - FRA 2025 - terms and definitions. https://openknowledge.fao.org/items/43aca42c-3ead-4747-9a48-0b20ae248a5e.

Fazakas, Z., Nilsson, M., Olsson, H., 1999. Regional forest biomass and wood volume estimation using satellite data and ancillary data. Agric. For. Meteorol. 98–99, 417–425. http://dx.doi.org/10.1016/S0168-1923(99)00112-4.

Fiebig, D.G., 2003. Seemingly unrelated regression. In: A Companion To Theoretical Econometrics. John Wiley & Sons, Ltd, pp. 101–121. http://dx.doi.org/10.1002/9780470996249.ch6, chapter 5.

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., Ringvall, A.H., Ståhl, G., 2014. Adapting national forest inventories to changing requirements – the case of the Swedish national forest inventory at the turn of the 20th century. Silva Fenn. 48, http://dx.doi.org/10.14214/sf.1095.

Gschwantner, T., Alberdi, I., Balazs, A., Bauwens, S., Bender, S., Borota, D., Bosela, M., Bouriaud, O., Cañellas, I., Donis, J., Freudenschuß, A., Hervé, J.C., Hladnik, D., Jansons, J., Kolozs, L., Korhonen, K.T., Kucera, M., Kulbokas, G., Kuliešis, A., Lanz, A., Lejeune, P., Lind, T., Marin, G., Morneau, F., Nagy, D., Nord-Larsen, T., Nunes, L., Pantić, D., Paulo, J.A., Pikula, T., Redmond, J., Rego, F.C., Riedel, T., Saint-André, L., Šebeň, V., Sims, A., Skudnik, M., Solti, G., Tomter, S.M., Twomey, M., Westerlund, B., Zell, J., 2019. Harmonisation of stem volume estimates in European national forest inventories. Ann. For. Sci. 76, 1–23. http://dx.doi.org/10.1007/s13595-019-0800-8.

Hauglin, M., Puliti, S., Schumacher, J., Rahlf, J., Breidenbach, J., 2024. Mapping tree species proportions in boreal forest with Sentinel2 time series data and a convolutional neural network. http://dx.doi.org/10.5281/zenodo.13236952, Preprint at Zenodo.

Hauglin, M., Rahlf, J., Schumacher, J., Astrup, R., Breidenbach, J., 2021. Large scale mapping of forest attributes using heterogeneous sets of airborne laser scanning and national forest inventory data. For. Ecosyst. 8, 65. http://dx.doi.org/10.1186/s40663-021-00338-4.

Hetemäki, L., Kangas, J., Peltola, H. (Eds.), 2022. Forest bioeconomy and climate change. In: Managing Forest Ecosystems, vol. 42, Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-030-99206-4.

Hijmans, R.J., 2024. Terra: Spatial data analysis. http://dx.doi.org/10.32614/CRAN.package.terra, r package version 1.7-83.

Hunault-Fontbonne, J., Eyvindson, K., 2023. Bridging the gap between forest planning and ecology in biodiversity forecasts: A review. Ecol. Indic. 154, 110620. http://dx.doi.org/10.1016/j.ecolind.2023.110620.

Ishwaran, H., Kogalur, U.B., 2025. Randomforestsrc: Fast unified random forests for survival, regression, and classification (RF-SRC). http://dx.doi.org/10.32614/CRAN.package.randomForestSRC, institution: Comprehensive R Archive Network Pages: 3.3.3.

Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S., 2008. Random survival forests. Ann. Appl. Stat. 2, 841–860. http://dx.doi.org/10.1214/08-AOAS169.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K.T., Maltamo, M., Nilsson, M., Nord-Larsen, T., Næsset, E., Olsson, H., 2018. Remote sensing and forest inventories in nordic countries – roadmap for the future. Scand. J. For. Res. 33, 397–412. http://dx.doi.org/10.1080/02827581.2017.1416666.

Kangas, A., Myllymäki, M., Mehtätalo, L., 2023. Understanding uncertainty in forest resources maps. Silva Fenn. 57, http://dx.doi.org/10.14214/sf.22026.

Korhonen, K.T., Räty, M., Haakana, H., Heikkinen, J., Hotanen, J.P., Kuronen, M., Pitkänen, J., 2024. Forests of Finland 2019–2023 and their development 1921–2023. Silva Fenn. 58, http://dx.doi.org/10.14214/sf.24045.

Lourenço, P., 2021. Biomass estimation using satellite-based data. In: Gonçalves, A.C., Sousa, A., Malico, I. (Eds.), Forest Biomass. IntechOpen, Rijeka, http://dx.doi.org/10.5772/intechopen.93603, chapter 3.

Mäkisara, K., Katila, M., Peräsaari, J., 2022. The Multi-Source National Forest Inventory of Finland — Methods and Results 2017 and 2019. Technical Report 90/2022, Natural Resources Institute Finland, Helsinki, http://urn.fi/URN:ISBN:978-952-380-538-5.

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12, 1620–1633. http://dx.doi.org/10.1111/2041-210X.13650.

Meyer, H., Pebesma, E., 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. Nat. Commun. 13, 2208. http://dx.doi.org/10.1038/s41467-022-29838-9.

Miettinen, J., Breidenbach, J., Adame, P., Adolt, R., Alberdi, I., Antropov, O., Arnarsson, O., Astrup, R., Berger, A., Bogason, J., Chirici, G., Corona, P., D'Amico, G., Fejfar, J., Fischer, C., Gohon, F., Gschwantner, T., Hertzler, J., Koma, Z., Korhonen, K.T., Krajnc, L., Latte, N., Lejeune, P., McCullagh, A., Mionskowski, M., Moreno-Fernández, D., Myllymäki, M., Nilsson, M., Perin, J., Pitkänen, J., Redmond, J., Riedel, T., Schumacher, J., Seitsonen, L., Sirro, L., Skudnik, M., Snorrason, A., Sroga, R., Traub, B., Traustason, B., Westerlund, B., Wurpillot, S., 2025. Pan-European forest maps produced with a combination of earth observation data and national forest inventory plots. Data Brief 60, 111613. http://dx.doi.org/10.1016/j.dib.2025.111613.

Miettinen, J., Carlier, S., Häme, L., Mäkelä, A., Minunno, F., Penttilä, J., Pisl, J., Rasinmäki, J., Rauste, Y., Seitsonen, L., Tian, X., Häme, T., 2021. Demonstration of large area forest volume and primary production estimation approach based on Sentinel-2 imagery and process based ecosystem modelling. Int. J. Remote Sens. 42, 9467–9489. http://dx.doi.org/10.1080/01431161.2021.1998715.

Mitchard, E.T., Saatchi, S.S., Baccini, A., Asner, G.P., Goetz, S.J., Harris, N.L., Brown, S., 2013. Uncertainty in the spatial distribution of tropical forest biomass: A comparison of pan-tropical maps. Carbon Balance Manag. 8, 10. http://dx.doi.org/10.1186/1750-0680-8-10.

Moisen, G.G., Frescino, T.S., 2002. Comparing five modelling techniques for predicting forest characteristics. Ecol. Model. 157, 209–225. http://dx.doi.org/10.1016/S0304-3800(02)00197-7.

Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J., Egberth, M., Larsson, S., Nilsson, L., Eriksson, J., Olsson, H., 2017. A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the national forest inventory. Remote Sens. Environ. 194, 447–454. http://dx.doi.org/10.1016/j.rse.2016.10.022.

Persson, H.J., Jonzén, J., Nilsson, M., 2021. Combining TanDEM-X and Sentinel-2 for large-area species-wise prediction of forest biomass and volume. Int. J. Appl. Earth Obs. Geoinf. 96, 102275. http://dx.doi.org/10.1016/j.jag.2020.102275.

Persson, H.J., Olsson, H., Soja, M.J., Ulander, L.M.H., Fransson, J.E.S., 2017. Experiences from large-scale forest mapping of Sweden using TanDEM-X data. Remote. Sens. 9, 1253. http://dx.doi.org/10.3390/rs9121253.

Pitkänen, T.P., Balazs, A., Tuominen, S., 2024. Automatized Sentinel-2 mosaicking for large area forest mapping. Int. J. Appl. Earth Obs. Geoinf. 127, 103659. http://dx.doi.org/10.1016/j.jag.2024.103659.

Pohjankukka, J., Tuominen, S., Pitkänen, J., Pahikkala, T., Heikkonen, J., 2018. Comparison of estimators and feature selection procedures in forest inventory based on airborne laser scanning and digital aerial imagery. Scand. J. For. Res. 33, 681–694. http://dx.doi.org/10.1080/02827581.2018.1482955.

Puliti, S., Hauglin, M., Breidenbach, J., Montesano, P., Neigh, C.S.R., Rahlf, J., Solberg, S., Klingenberg, T.F., Astrup, R., 2020. Modelling above-ground biomass stock over Norway using national forest inventory data with ArcticDEM and Sentinel-2 data. Remote Sens. Environ. 236, 111501. http://dx.doi.org/10.1016/j.rse.2019.111501.

Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Processes for Machine Learning. The MIT Press, https://direct.mit.edu/books/monograph/2320/Gaussian-Processes-for-Machine-Learning.

Särndal, C.E., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer Science & Business Media, Google-Books-ID: ufdONK3E1TcC.

Tomppo, E., Halme, M., 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: A genetic algorithm approach. Remote Sens. Environ. 92, 1–20. http://dx.doi.org/10.1016/j.rse.2004.04.003.

Tuominen, S., Pitkänen, T., Balazs, A., Kangas, A., 2017. Improving finnish multi-source national forest inventory by 3D aerial imaging. Silva Fenn. 51.

Zhang, C., Liu, Y., Tie, N., 2023. Forest land resource information acquisition with Sentinel-2 Image Utilizing Support Vector Machine, K-nearest neighbor, random forest, decision trees and multi-layer perceptron. Forests 14, http://dx.doi.org/10.3390/f14020254.