Appendix A.

1. Material

The selected study area covers ca. 25694 km^2 across Southern Sweden (Figure 1) and is characterized by a coniferous dominated forest types. The dominant species are Norway spruce (Picea abies (L.) Karst.) and Scots pine (Pinus sylvestris L.), while silver birch (Betula pendula Roth.), downy birch (Betula pubescens Ehrh.), and aspen (Populus tremula)) are the most common minority species.

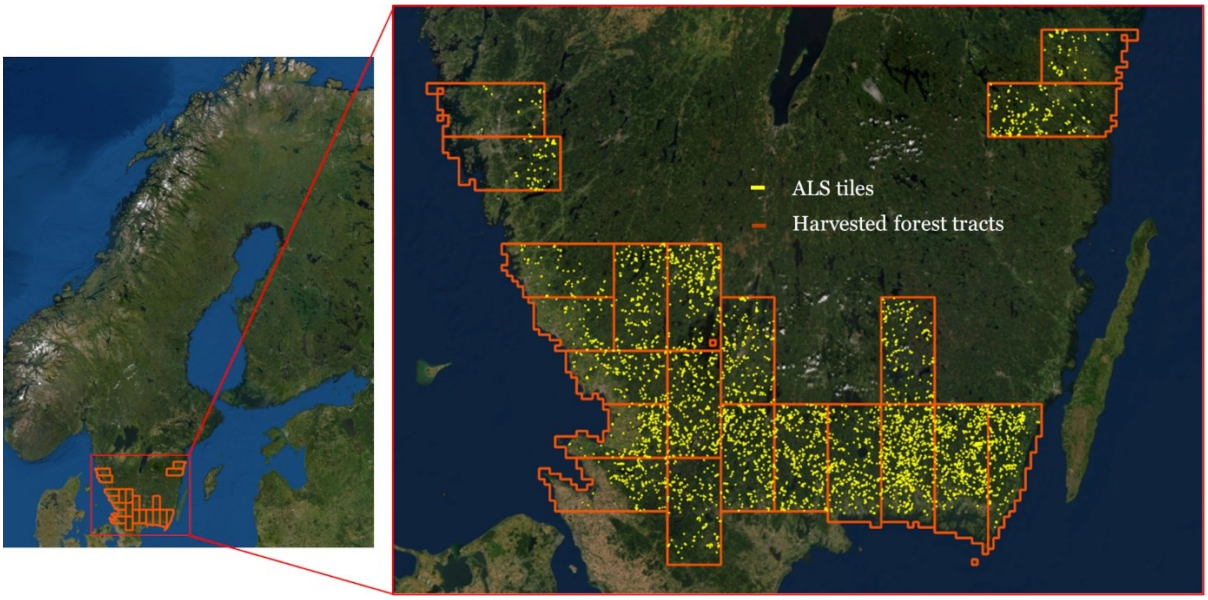

Figure 1 Study area covering with the yellow polygons indicating the locations of harvested forest tracts. The delineation of the airborne laser scanning datasets is in orange color.

1.1 Field data

The input field data was compiled from the harvester production files between 2018-2021 provided by the forest owners' association *Södra*. The tree lists containing detailed stem measurements, volume and height estimates, species, and quality. The locations were inferred from the GNSS positioning of the harvester at the felling cut time stamp. The forest tracts that were not fully covered by the remote sensing data, contained mixed logging forms such as thinnings and final cuts as well harvests along roads were filtered out. The final dataset for analyses contained 3826 forest tracts distributed as presented in Table 1. A spatial representation of the harvester data collected on a forest tract is shown in Figure 2.

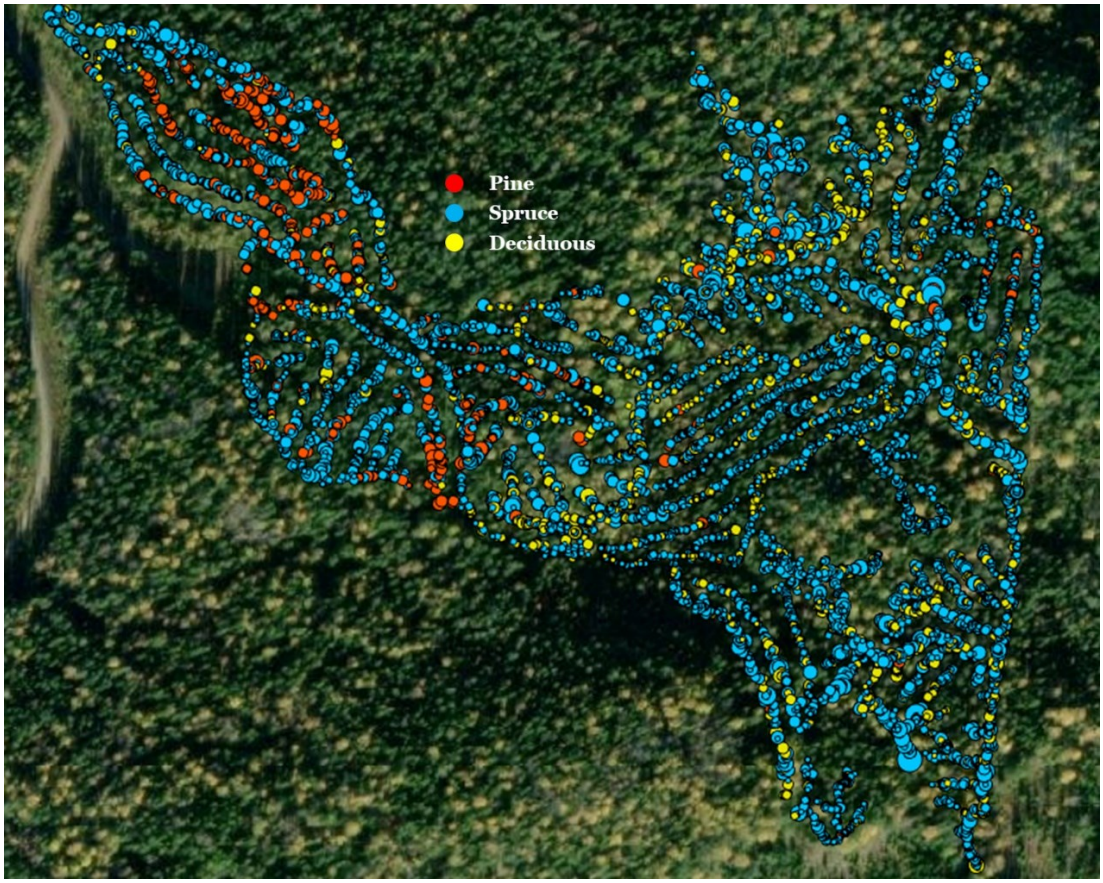

Figure 2 The dots represent the harvester positions registered in the production files (.hpr) overlayed on high-resolution ESRI background aerial map. The machine coordinates are inferred to each tree to produce georeferenced lists containing quantitative and qualitative information on the felled trees. The dot size is correlated with the tree DBH, and the color indicates the tree species.

1.2 Remote sensing data

1.2.1 Satellite imagery

The use of Earth Observation (EO) data, specifically multispectral satellite imagery, is a crucial tool for accurate, large-scale tree species mapping. The European Space Agency's Sentinel-2 mission (henceforth denoted S2) provides access to free and open multispectral satellite data for land monitoring at a high resolution in the visible, near-infrared, and shortwave spectrum, including the red-edge region, which is well-suited for vegetation mapping.

Google Earth Engine (GEE) is a powerful solution for processing and analyzing large amounts of satellite imagery data from scattered regions, such as forest tracts planned for harvesting. GEE is a mature technology that offers access to a wide range of image catalogs and is relatively easy to use through high-abstraction APIs. Applications using GEE on global EO satellite data sets cover a wide range of scientific domains, and the platform allows for sharing of custom scripts among users.

In this project, GEE is used to produce a 'cloud-free' image mosaic, using the algorithm described by Schmitt et al. (2019). The algorithm calculates various pixel- and image-level quality scores from a collection of Sentinel-2 images acquired at different time points to produce a cloud-free image

mosaic. The algorithm differentiates between various types of clouds and attempts to remove cloud shadow areas using image morphology. The original algorithm was adapted to fit closely to the project's needs and to produce outputs that are relevant only for the analyzed forest tracts. Previous research indicates that using multitemporal Sentinel-2 imagery for forest inventories can result in moderate to low accuracy improvements. In this project, the Sentinel-2 image collection was ingested in 2017 during three time intervals along the vegetation season: May-June, June-August and September-October.

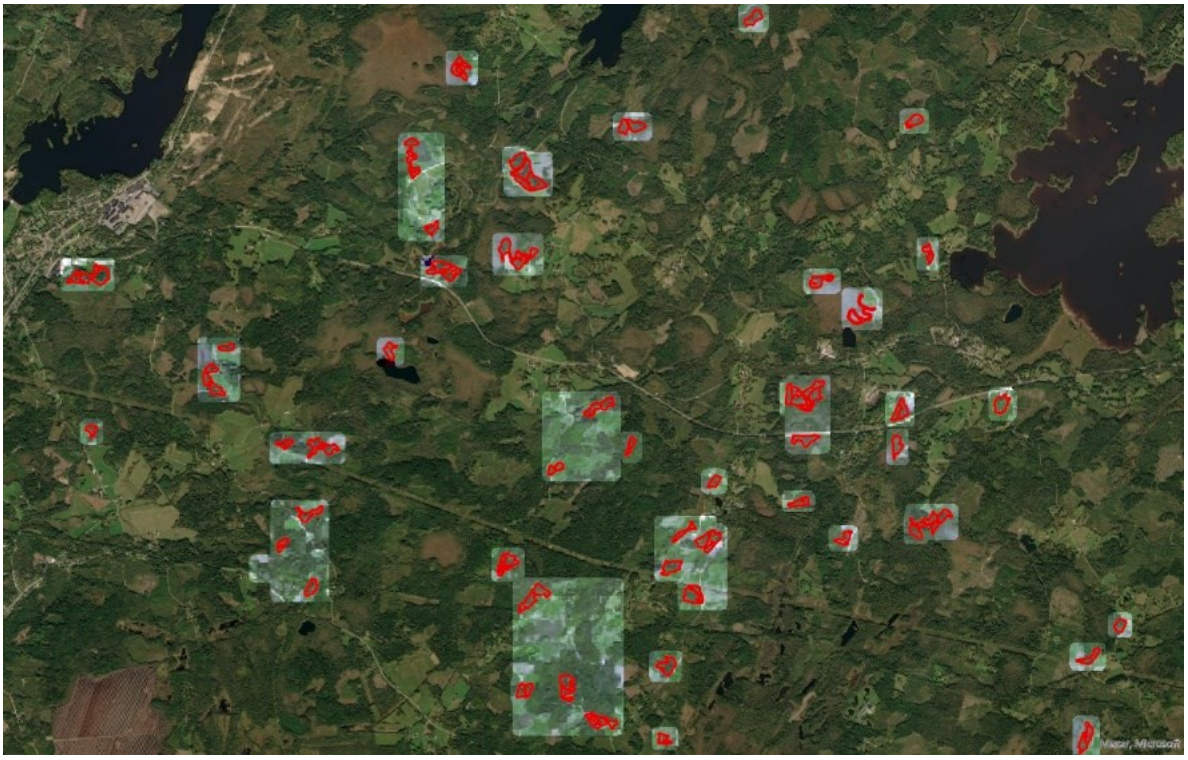

Figure 3 Sentinel-2 mosaics for May-June 2019 for the harvest objects in the study area (red polygons) overlayed on high-resolution ESRI background aerial map.

1.2.2 Airborne laser scanning data

Airborne Laser Scanning (ALS) is a remote sensing technique that uses a laser scanner mounted on an aircraft to collect highly accurate 3D point cloud data of the earth's surface. This data can be used to create detailed maps of the terrain, vegetation, and other features of an area, including tree species composition. The national ALS data survey carried out by the Swedish mapping, cadastral, and land registration authority (Lantmäteriet) is an important resource for forest management and planning. It is conducted with an updating frequency of about seven years and covers approximately 75% of Sweden's area (around 350,000 km2). The on-going acquisition campaign provides a point density of ca. 1-2 points m2 with up to five returns per pulse classified as ground, water, low or high point or unclassified. In forested areas, it is assumed that the unclassified returns are the vegetation hits. The ALS data is downloaded from Lantmäteriet as compressed files in the .LAZ-format, which is a popular file format for storing and sharing large point cloud data sets (Anon 2023c).

The 3D point clouds collected by the ALS survey can be used to create various products such as Canopy Height Models (CHM) at 1m resolution, which are digital models of the canopy height of a forest. These models can be used to estimate the amount of biomass, carbon sequestration and timber volume in the forest.

The ALS measurements provided by Lantmäteriet contain up to 5 returns per pulse as well as calibrated intensity values. For this study, the last return, last of many and single returns (1 to 3)

were included on the analyses. Using multiple discrete ALS returns may help describing the canopy structure at different heights, which can be used to estimate the volume of the tree species, and the number of trees in the forest. An example of an 3D ALS point cloud is shown in Figure 4.

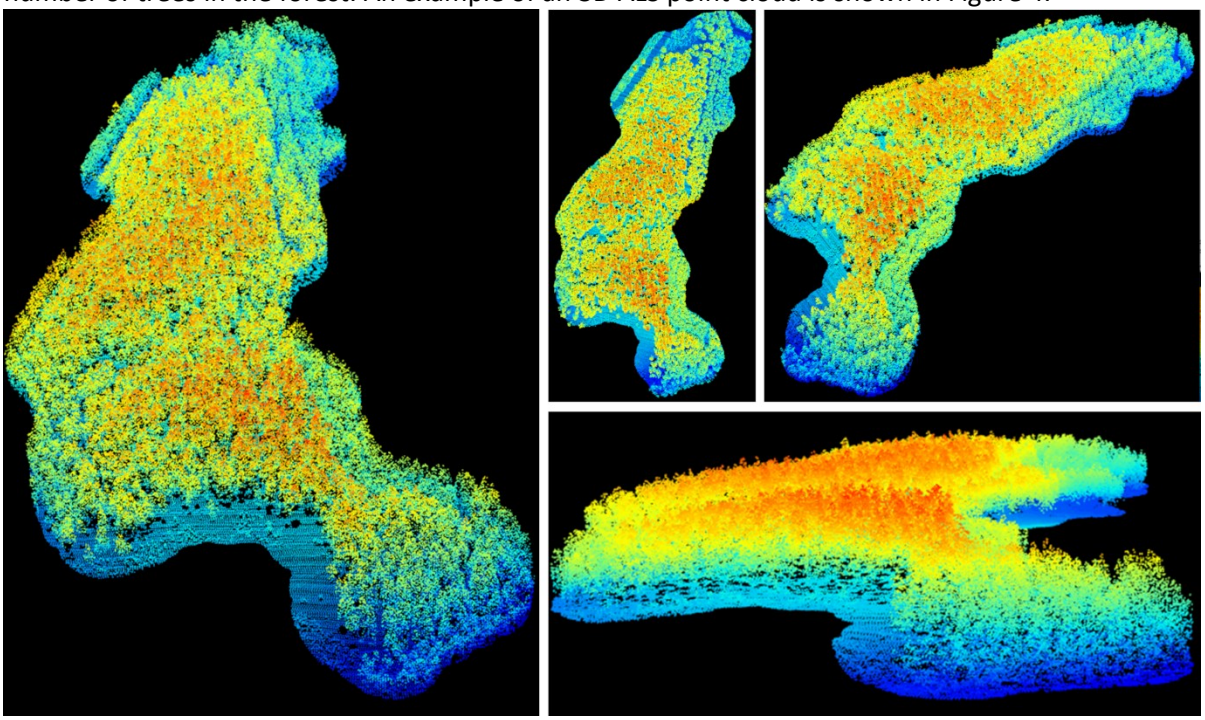

Figure 4 3D point cloud data on a forest tract visualized from different viewing angles.

1.2.3 Aerial imagery

Very high resolution (VHR) orthophotos produced from 8 bit digital aerial imagery (RGB + NIR channels) acquired during the leaf-on season in 2016 were obtained from Lantmäteriet Anon (2023d). The images 0.25m spatial resolution and are radiometrically processed to obtain standardized spectral signatures. Nevertheless, color unbalance may occasionally occur due to changing condition at different acquisition dates.

1.3 Cartographic products

The land cover thematic maps 'Nationella Marktäckedata' (NMD) are developed by the Swedish Environmental Protection Agency with an expected overall accuracy of 74% on areas > 1ha, and a minimum mapping unit of 0.01 ha. The NMD classes for productive forest relevant for the study area were Pine forests (>70% Pine), Spruce forests (>70% Spruce), mixed coniferous forest (70% mixed coniferous), mixed coniferous/deciduous forests (no category >70%) and deciduous forest (>70% Deciduous). The NMD classes were hot-encoded (0/1) to be used as numeric auxiliaries by the machine learning algorithms. The maps can be downloaded in raster format at 10 m spatial resolution via web services (Anon 2023b).

The SLU Forest Map (SLU) product provides a wide range of forest attributes estimates in raster format at 12.5 x. 12.5 m spatial resolution covering large parts (but not entirely) of Sweden's productive forest area Anon (2023a). The maps are produced by combining National Forest Inventory field data, Sentinel-2 imagery and canopy height models provided by Lantmäteriet. The attributes of interest in the project consists of raster maps containing tree species specific volume estimates (m^3 ha^{-1}). SLU maps were last updated in 2015, but new improved versions are expected to become available in the near future.

1.4 Stem price models

The monetary calculations were preformed using stem price lists for the main tree species were compiled by the Skogforsk specialists using industry data from 2020 that are valid for our study area. The prices for Spruce and Pine trees refer to round wood, while the pulpwood from coniferous trees and all deciduous trees were aggregated into a common price category. We resorted to this simplification in order to eliminate the uncertainties related to assigning tree-level quality attributes from the assortment lists in the harvester data to the entire stem.

Table 1 Tree species specific stem price list (SEK/m³ underbark) by breast height diameter (DBH) classes

Tree species	DBH class (mm)																					
	80	100	120	140	160	180	200	220	240	260	280	300	320	340	360	380	400	420	440	460	480	500
Spruce	270	270	270	270	331	366	399	410	420	430	435	435	435	435	440	440	445	445	445	445	445	445
Pine	270	270	270	270	310	374	395	411	415	420	425	425	430	430	430	430	430	435	435	423	435	435
Deciduous	250	250	250	250	250	250	250	250	250	250	250	250	250	250	250	250	250	250	250	250	250	250

2. Methods and results

The data processing and for the proposed approach is summarized in Figure 5.

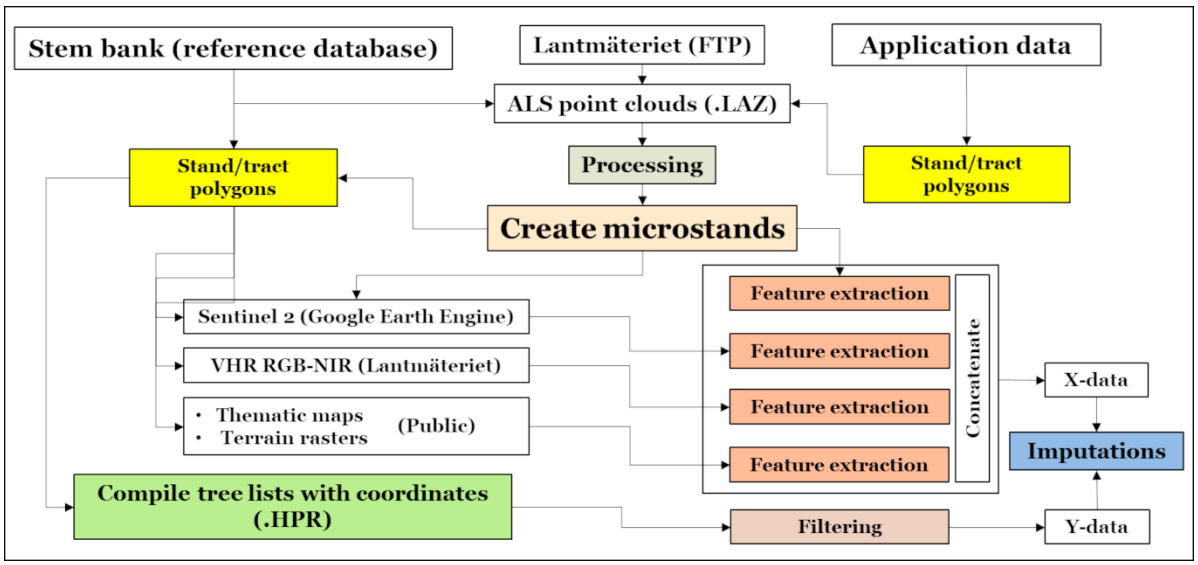

Figure 5 The proposed workflow for data ingestion and imputations.

2.1 Microstand delineation

In order to accurately map the tree species composition in a harvested area, the harvested areas are divided into smaller units called microstands. Microstands are defined as small plots of land that have similar characteristics such as tree species composition, age, and size. The delineation of microstands is done using image segmentation techniques applied to the Canopy Height Models (CHM) rasters produced from the first and first-of-many ALS returns.

The harvested areas were delineated into microstands using image segmentation techniques applied to the CHM rasters. One common technique used is the Adaptive Simple Linear Iterative Clustering (SLICO) algorithm (Achanta et al 2012) applied to the normalized CHM. SLICO is an image segmentation algorithm that groups the image pixels into sets (or superpixels, in the image processing parlance) that are homogeneous with regard to size and compactness and follow relatively well the local image content such as image boundaries and object contours (Achanta et al 2012, Li et al 2021). Images with complex structures can be thus efficiently represented using a smaller number of primitives, as each superpixel is supposed to encapsulate meaningful content. The harvest objects were represented as undirected acyclic graphs, with the superpixels in the nodes and the adjacency relationships in the edges, followed by spatial clustring of the graph nodes using the walktrap community finding algorithm (Pons & Latapy 2005).

In order to fine-tune the microstand delineation, a genetic algorithm (Scrucca 2013, 2017) was used for weighting the graph edges in order to optimize a multi-criteria cost function that balances the sizes distribution and shapes of the delineated microstands. A genetic algorithm is a heuristic optimization method that is inspired by the process of natural selection. It is used to fine-tune the microstand delineation process by weighting the graph edges in order to optimize a multi-criteria cost function that balances the size distribution and shapes of the microstands.

The genetic algorithm works by creating an initial population of solutions, which are represented as sets of weights for the graph edges. These solutions are then evaluated based on how well they meet the criteria defined in the cost function. The best solutions are then selected and used to create a new population of solutions through a process of reproduction, mutation, and crossover:

- during reproduction, the best solutions are chosen and used to create new solutions.
- mutation is the process of randomly changing the values of the weights in a solution.
- crossover is the process of combining the information from two solutions to create a new one.

These processes are designed to mimic the natural process of evolution and are used to improve the quality of the solutions over time. The genetic algorithm is run iteratively, with each iteration producing a new population of solutions, and the process continues until a satisfactory solution is found, or a predefined stopping criterion is met. The final solution is the set of weights for the graph edges that produce the best microstands according to the cost function mentioned above. Outputs of the microstand delineation algorithm are shown in Figure 6.

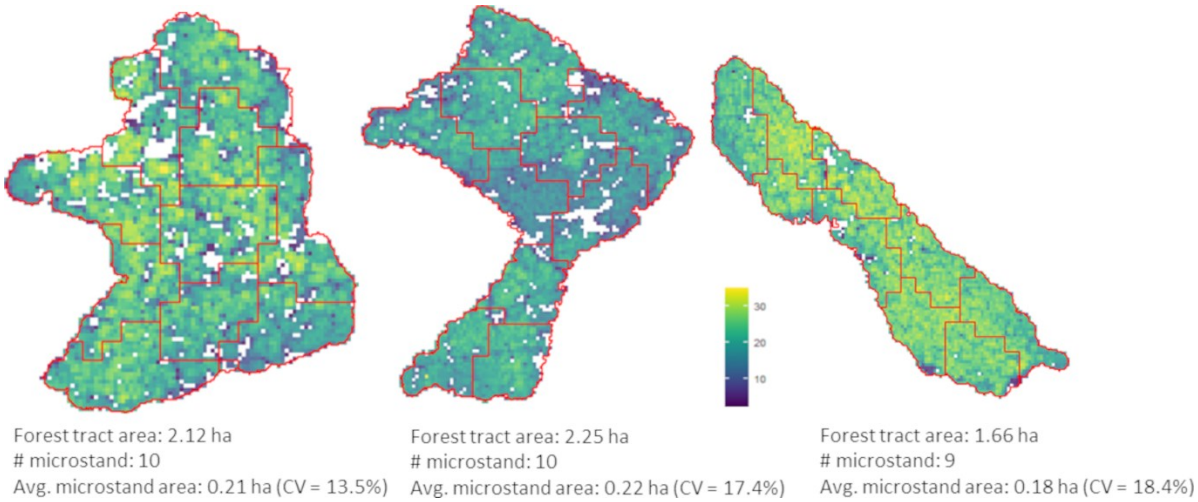

Figure 6 Microstand delineation example for three forest tracts. The red polygons represent the microstand borders, and the white (empty) areas within the forest tracts occur where the canopy height model heights are < 1 m. The statistics show the forest tract (ha), the average microstand areas (ha) and their variation coefficient (CV%), and the resulted number of microstands.

2.2 Assessing the imputation approaches.

Due to their relative simplicity and the ability to handle multivariate in-situ observations, the nearest neighbor imputations have become a popular predictive approach in support to operational forest planning (Eskelson et al 2009, Söderberg et al 2017, Söderberg et al 2018, Söderberg et al 2021). The current approach to forest recovery predictions adopted by Swedish forest companies is the k-Most Similar Neighbor (kMSN) imputations (Moeur & Stage 1995, Packalén & Maltamo 2007), with the number of neighbors set to k=5, as suggested by Söderberg et al. (2021). The predictions on target observations were obtained as simple averages of the k-closest ground-truth observations from the reference dataset. The computations were performed using the R-package 'yaImpute' (Crookston & Finley 2007). The k-MSN imputations were run at microstand level, and the results were aggregated at stand level as weighted averages, with the weights proportional to the areas of the target microstands. The main output of the kMSN is a tree lists describing the expected forest structure in a given forest tract that can be used for optimizing the operational planning and logistics.

The kMSN algorithm was assessed against other popular multivariate methods that can be used for product recovery predictions, namely Partial Least Squares regression (PLSR) and Multivariate Random Forests regression (MRFR). Both PLSR and MRFR are popular machine learning methods that are well understood, robust to noise and can handle a large numbers of correlated predictors.
PLSR and MRFR cannot directly predict the tree lists required for yield and product recovery calculations, but instead they can predict multiple forest stand attributes simultaneously. Thus, a two-steps procedure was employed, where first 11 forest microstand-level attributes (total and species specific volumes, mean height and diameter, total stem number, skewness and kurtosis of the height and diameter distribution of the trees) were predicted using PLSR and MRFR, and then simple nearest neighbor imputations was used to find the 5 closest Euclidean distance. The PLSR computations were performed using the implementation available in the 'pls'-package (Mevik & Meinshausen 2007), and MRFR was run using the 'randomForestSRC'-package (Ishwaran & Kogalur 2007, Ishwaran et al 2008, Ishwaran & Kogalur 2022) of the R statistical software (R Core Team 2022)

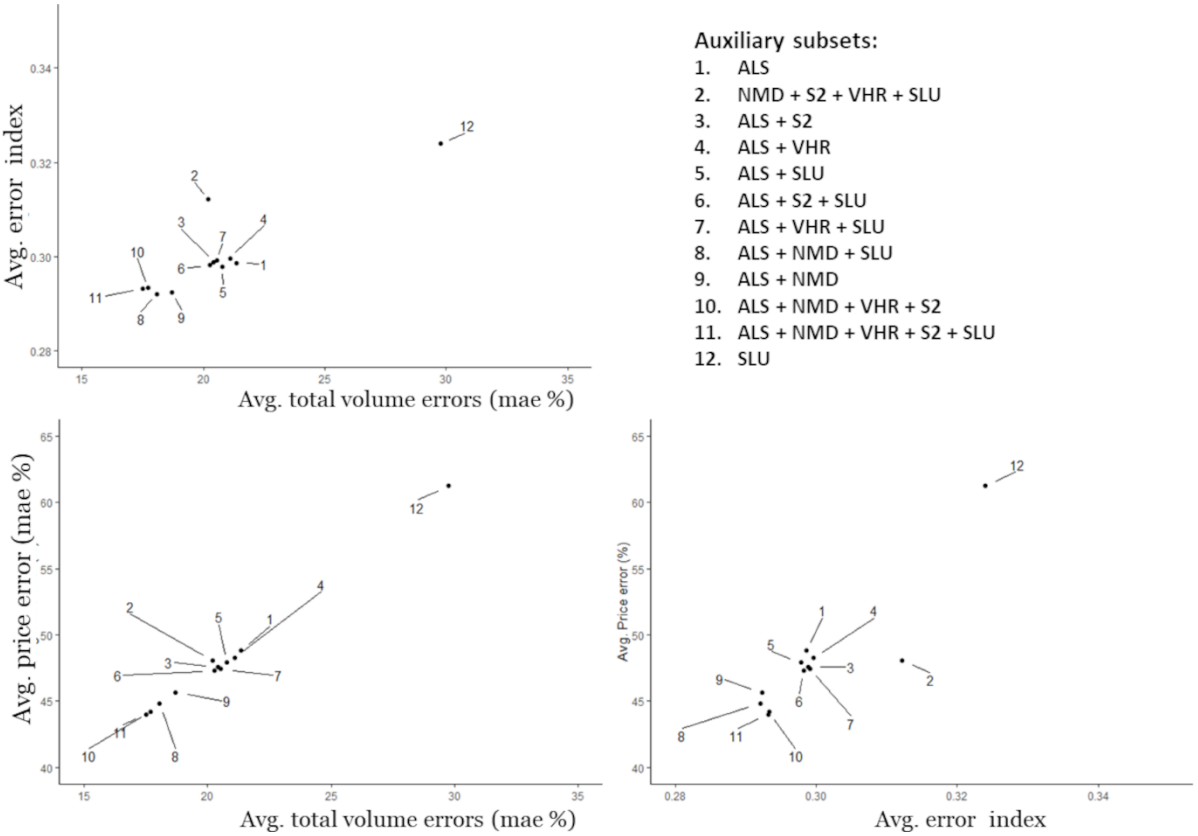

Figure 7 Imputation errors (for total volume, error index and total monetary value) by auxiliary subsets averaged over all prediction methods. The subset no.8 consisting of microstand-level ALS-based predictors in combination with hot-encoded NMD majority classes and species-specific volume SLU estimates from 2015 was deemed as a feasible trade-off between accuracy, computation efficiency and data requirements.

An example for the prediction accuracy for multivariate k-MSN imputations using the most feasible subset of predictors (i.e., subset #8) is presented in Figure 8.

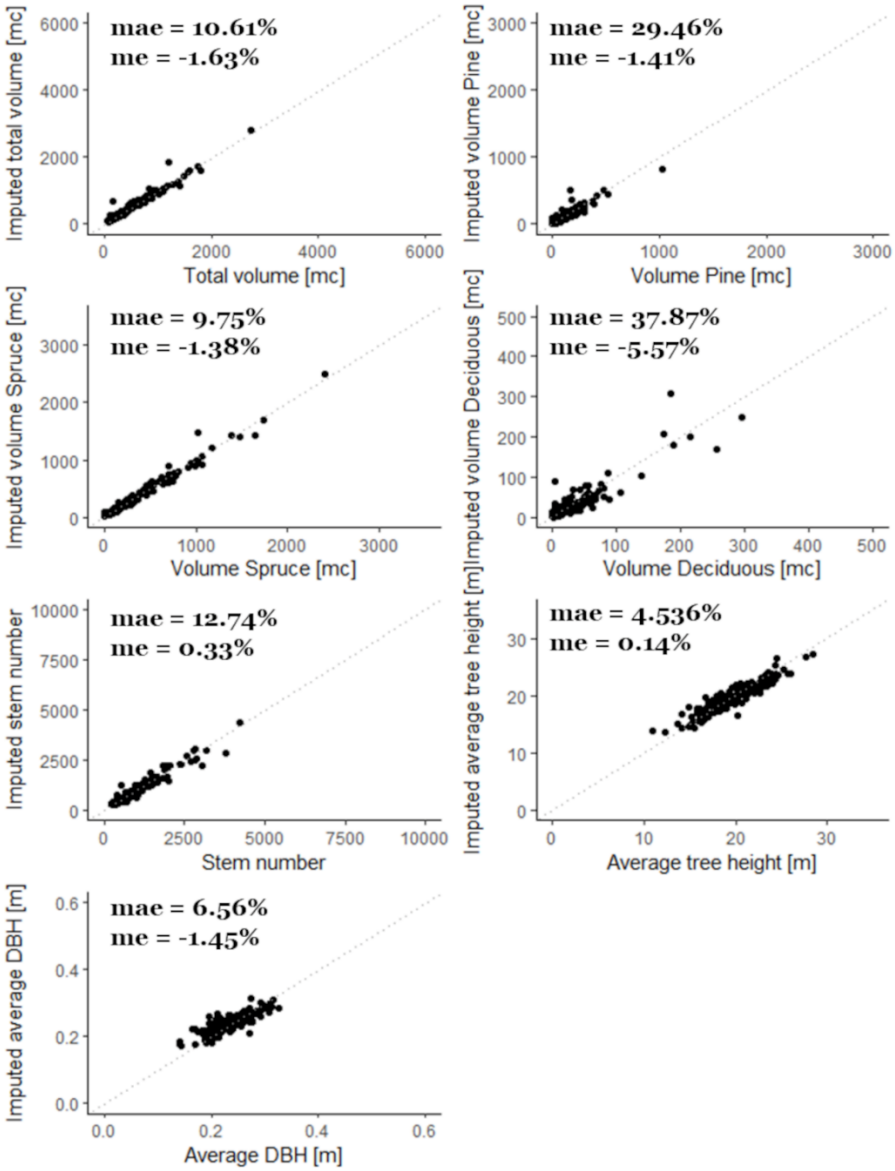

*Figure 8 The accuracy for the main forest state attributes in terms of mean absolute error (**mae**) and mean error (**me**) in percentage relative to the ground-truth values compiled from harvester production files. The example pertains to a random validation dataset produced during the data splitting procedure.*

Table 2 Prediction accuracy in terms of percentage mean absolute error for total and species specific volume and monetary loss at forest tract level. The numbers in parentheses represent the spreading (i.e., standard deviations in percentage points) of the error estimates during the simulation studies.

Auxiliary Subset	Method	Errors (%)										
		Volume				Stem number	Mean height	Mean DBH	Monetary loss			
		Total	Pine	Spruce	Deciduous				Total	Pine	Spruce	Deciduous
ALS NMD SLU (#8)	k-MSN	9.19	27.77	9.64	42.54	14.62	4.15	6.25	25.94	53.07	27.48	64.64
		4.82	*9.77*	*5.59*	*13.89*	*6.42*	*4.66*	*5.79*	*1.95*	*6.61*	*3.16*	*9.21*
	PLSR	9.93	32.01	10.63	46.47	14.79	4.23	6.21	27.14	53.25	29.43	63.00
		6.33	*8.47*	*4.74*	*7.3*	*6.63*	*5.64*	*5.88*	*2.53*	*5.02*	*2.99*	*5.72*
	MRFR	10.96	35.40	12.68	52.86	17.14	4.84	7.08	30.43	59.76	32.95	70.44
		5.55	*8.38*	*5.09*	*9.56*	*6.65*	*4.70*	*5.89*	*2.81*	*6.61*	*3.14*	*7.16*

2.4 Diameter distribution predictions

The DBH distributions were quantized in 2 cm classes, with the mid-class values between 6 to 60 cm. The prediction quality is quantified using the Error Index (EI) described in Packalén & Maltamo (2008) as:

$$EI = \sum_{i \in S_k} 0.5 \left| \frac{\hat{n}_i}{\hat{N}} - \frac{n_i}{N} \right|, EI \in [0,1]$$

where N and $\hat{N}$ are the ground-truth and imputed total number of trees in a forest tract, and n_i and $\hat{n}_i$ are the ground-truth and total number of trees in the i*th* DBH class. *EI* = 0 for perfect distribution match, and *EI* = 1 when the imputed and true distributions are completely different.

The overall errors in distribution imputations at forest tract level Figure xxx (a) are in general much below (ca. 0.13) the lowest reported values in literature (ca. 0.15-0.20). The species- specific EI values are in general much higher, being in general between 0.4-0.6 (Table 4) According to our knowledge, reference error index values for species specific diameter distribution predictions are not available in the literature.

An example of very good fit for the overall and species-specific diameter distribution predictions at forest tract level according to the EI values is shown in Figure 9, and an average result is presented in Figure 10.

Table 3 Average error indexes describing the imputation accuracy for DBH distributions. The numbers in parentheses represent the spreading (i.e., standard deviations in percentage points) of the error estimates during the simulation studies.

Auxiliary subset	Method	Error index			
		Total	Pine	Spruce	Deciduous
ALS NMD SLU (#8)	k-MSN	0.137	0.465	0.162	0.347
		2.665	*6.131*	*5.641*	*2.984*
	PLSR	0.143	0.464	0.178	0.348
		2.744	*5.435*	*7.189*	*2.09*
	MRFR	0.165	0.502	0.202	0.388
		3.112	*5.9*	*7.724*	*3.181*

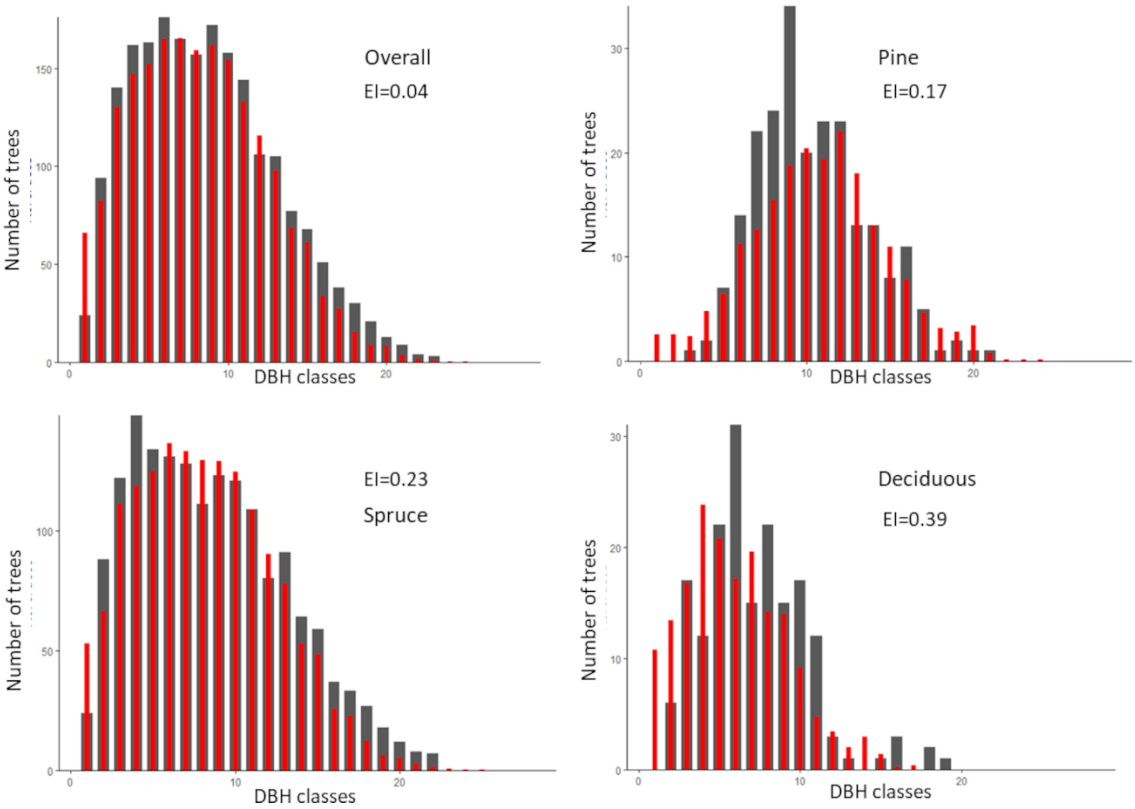

Figure 9 Overall and species-specific diameter distribution predictions at forest tract level considered to be of high quality. The ground-truth distributions are grey color, and the imputed distribution are in red.

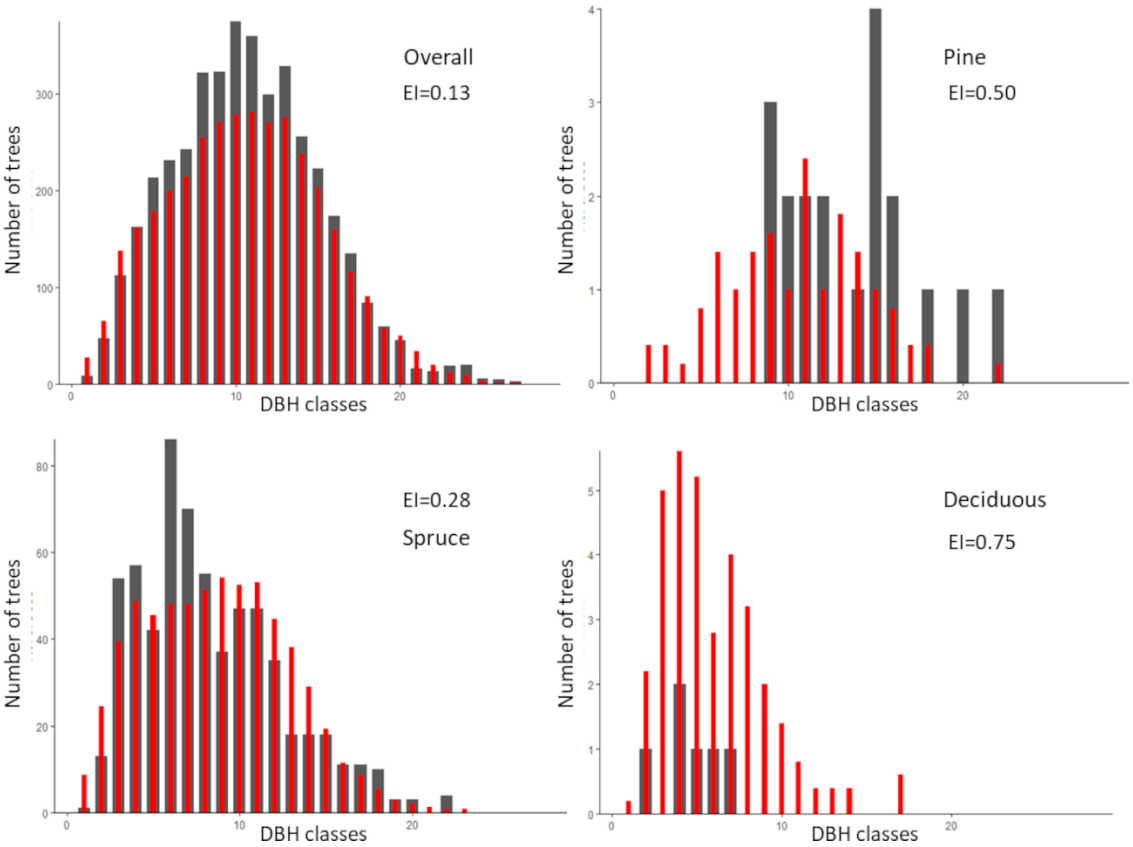

Figure 10 Overall and species-specific diameter distribution predictions at forest tract level considered to be of average quality. The ground-truth distributions are grey color, and the imputed distribution are in red.

2.5 Uncertainty assessment for k-MSN imputations

The uncertainty assessment focused solely on the total and species-specific volume imputations aggregated at forest tract level. The approach combines the prediction interval construction algorithm via split conformal inference described in Lei et al. (2018, § 2.4) with the quantile random forest regression algorithm implemented in the 'randomForestSRC'- package (Greenwald & Khanna 2001, Meinshausen 2006) of the R statistical software (R Core Team 2022). The general numerical simulation methodology follows Zhang et al (2020). The four coverage probability types characterizing the prediction intervals described in Zhang et al (2020, §3) are the following:

- $Type\ 1$: $\mathbb{P}[Y \in \mathrm{I}_{\alpha}(\boldsymbol{X}, \mathrm{C}_n)]$ $(marginal\ coverage)$
- $Type\ 2$: $\mathbb{P}[Y \in \mathrm{I}_{\alpha}(\boldsymbol{X}, \mathrm{C}_n)|\mathrm{C}_n]$ $(conditional\ coverage\ given\ \mathrm{C}_n)$
- $Type\ 3$: $\mathbb{P}[Y \in \mathrm{I}_{\alpha}(\boldsymbol{X}, \mathrm{C}_n)|\boldsymbol{X} = x]$ $(conditional\ coverage\ given\ \boldsymbol{X} = x)$
- $Type\ 4$: $\mathbb{P}[Y \in \mathrm{I}_{\alpha}(\boldsymbol{X}, \mathrm{C}_n)|\mathrm{C}_n, \boldsymbol{X} = x]$ $(conditional\ coverage\ given\ \boldsymbol{X} = x\ and\ \mathrm{C}_n)$

where I_{α} is the $1 - \alpha$ prediction interval, **Y** is the ground truth value of a certain forest attribute, and C_n is the calibration dataset containing *n* independent observations (i.e., forest tracts) and **X** is the set of predictors (i.e., predicted volumes) used for controlling the heteroscedastic errors. The naive predictions intervals were constructed at Student-t intervals using the root mean square error of the fitted residuals, for each of the estimated forest tract attributes. The nominal coverage probabilities of 0.50, 0.68, 0.75, 0.90 and 0.95 were considered for the assessment. The performance of both inferential methods is presented in Figure 11. The results indicate that the conformal inference approach produced coverage rate estimates that are very close to the nominal levels, with the conditional coverage types being in general more efficient.

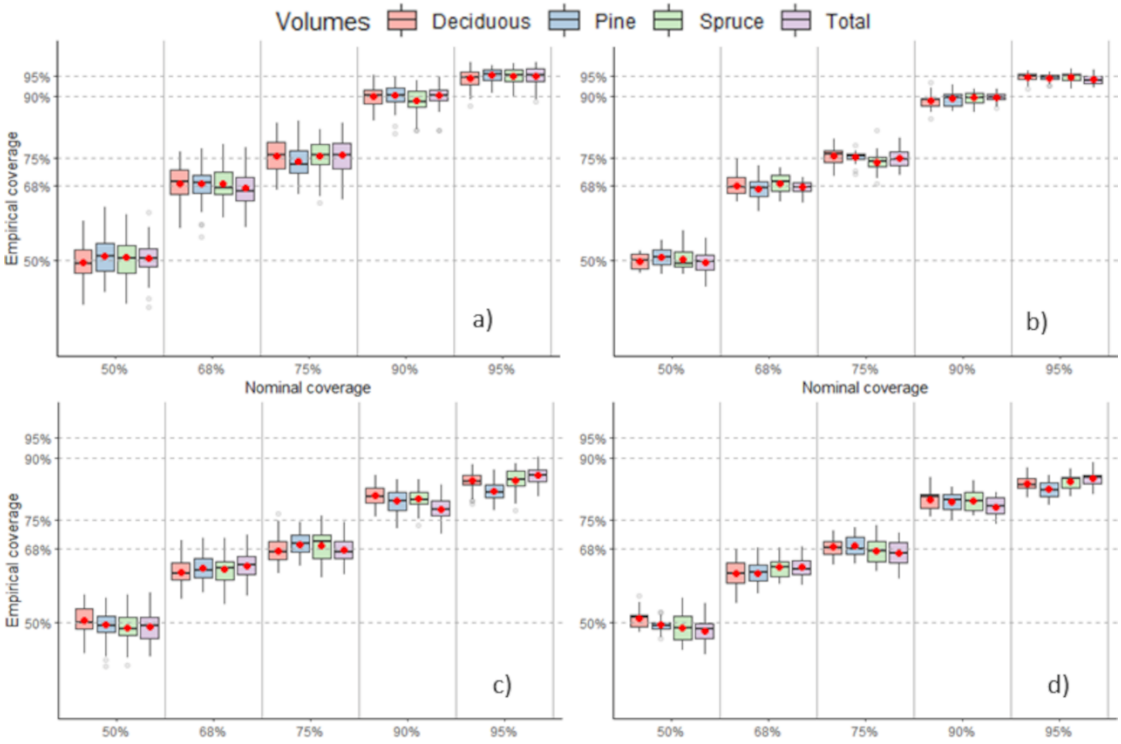

Figure 11 Empirical versus nominal prediction interval coverages. Type 1 and 2 coverage rate estimates resulted from conformal predictions are shown in Figure 11 a, and the results for types 3 and 4 in Figure 11 b. The corresponding coverage rate estimates using naive are shown in Figure 11 c and d, respectively. The boxplots indicate the coverage rates, and the conditional coverage rate estimates are represented by the red dots.

Table 4 Empirical coverage rates (type 1 and 2) for conformal and naive prediction intervals. The number in italics indicate the spreading (percentage points standard deviation) during the simulation studies.

Method	Property	Attribute	100(1-α) Prediction intervals				
			50	68	75	90	95
			Types 1 and 2 coverage rates				
Naive inference (RMSE)	Coverage	Total	0.65	0.79	0.83	0.92	0.95
			5.13	*2.52*	*3.07*	*1.70*	*1.26*
		Pine	0.74	0.84	0.87	0.93	0.95
			4.24	*3.64*	*2.36*	*2.03*	*1.27*
		Spruce	0.66	0.80	0.84	0.92	0.95
			4.25	*3.03*	*2.72*	*1.63*	*1.76*
		Deciduous	0.74	0.84	0.87	0.93	0.95
			3.22	*3.43*	*2.63*	*1.75*	*1.53*
	Median width	Total	13.34	19.67	22.75	32.53	38.76
		Pine	85.10	116.41	129.60	168.55	192.10
		Spruce	13.96	20.58	23.81	34.05	40.57
		Deciduous	104.32	140.61	155.95	200.88	228.47
Conformal Inference	Coverage	Total	0.50	0.69	0.76	0.90	0.95
			7.91	*6.71*	*5.31*	*2.48*	*2.20*
		Pine	0.51	0.69	0.76	0.90	0.95
			8.82	*5.43*	*5.36*	*2.77*	*1.51*
		Spruce	0.51	0.69	0.76	0.90	0.95
			8.66	*5.86*	*4.94*	*2.89*	*2.05*
		Deciduous	0.49	0.67	0.74	0.89	0.94
			9.05	*6.21*	*5.34*	*2.74*	*2.39*
	Median width	Total	8.63	13.31	15.95	24.05	30.28
		Pine	50.11	67.19	74.88	96.71	114.02
		Spruce	9.37	14.40	16.86	25.24	31.49
		Deciduous	58.08	82.07	91.33	121.06	144.55

Table 5 Empirical coverage rates (type 1 and 2) for conformal and naive prediction intervals. The number in italics indicate the spreading (percentage points standard deviation) during the simulation studies.

Method	Property	Attribute	100(1-α) Prediction intervals				
			50	68	75	90	95
			Types 3-4 coverage rates				
Naive inference (RMSE)	Coverage	Total	0.66	0.79	0.84	0.92	0.95
			4.89	*2.77*	*2.14*	*1.67*	*1.29*
		Pine	0.74	0.84	0.87	0.93	0.95
			3.12	*3.43*	*2.52*	*1.80*	*1.66*
		Spruce	0.67	0.80	0.84	0.92	0.95
			4.55	*2.53*	*2.86*	*2.12*	*1.52*
		Deciduous	0.74	0.84	0.87	0.93	0.95
			4.09	*2.72*	*1.92*	*1.58*	*1.95*
	Median width	Total	13.51	19.91	23.03	32.93	39.24
		Pine	84.44	115.62	128.51	167.48	191.31
		Spruce	14.03	20.69	23.93	34.22	40.78
		Deciduous	104.04	139.39	154.80	199.80	227.16
Conformal Inference	Coverage	Total	0.49	0.67	0.74	0.89	0.94
			5.74	*2.29*	*3.38*	*1.46*	*1.35*
		Pine	0.51	0.69	0.75	0.90	0.94
			5.13	*4.04*	*2.24*	*2.13*	*1.03*
		Spruce	0.50	0.68	0.75	0.89	0.95
			5.23	*3.83*	*3.94*	*1.84*	*1.37*
		Deciduous	0.50	0.68	0.75	0.90	0.95
			3.83	*4.07*	*3.22*	*2.28*	*1.26*
	Median width	Total	8.68	13.31	15.71	24.00	30.27
		Pine	52.50	70.10	77.99	100.61	116.30
		Spruce	9.40	14.31	16.75	25.08	31.26
		Deciduous	58.97	83.09	93.36	123.56	147.28

Conditional prediction intervals (type 4) for total volume and species specific volumes are presented in Figures 12- 15. The black dots represent the predicted versus ground-truth forest tract volumes, and the vertical black lines are the associated conformal prediction intervals with nominal coverage error of 68%. The 1:1 line between ground-truth and predicted total volumes level is shown in red. The prediction intervals are valid if they are crossing the 1.1 line. The left-side panels show the resulted prediction intervals for all the forest tracts in the independent validation dataset. Zoomed views of the rectangular regions marked with dark gray is provided in the right-panels.

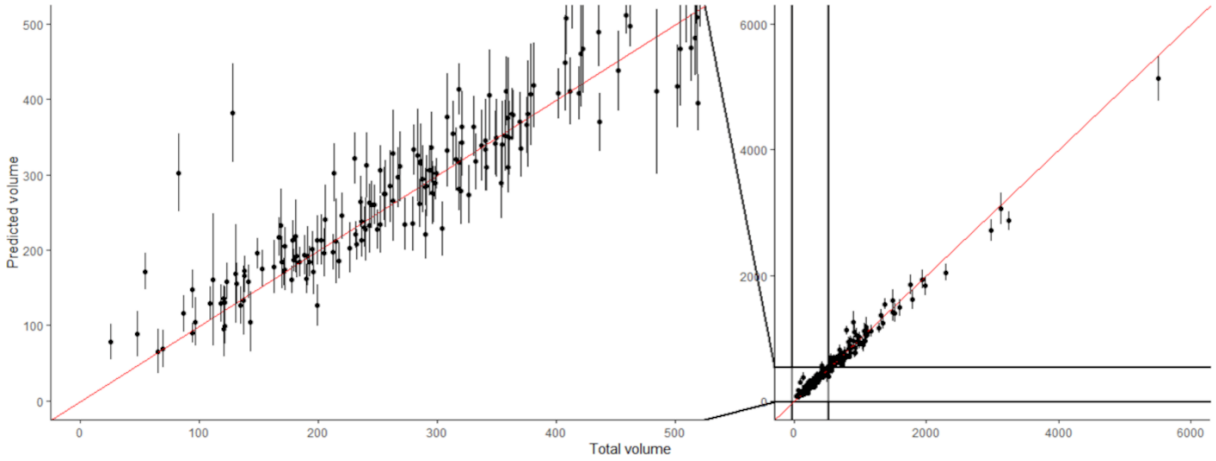

Figure 12 Conformal prediction intervals for total volumes

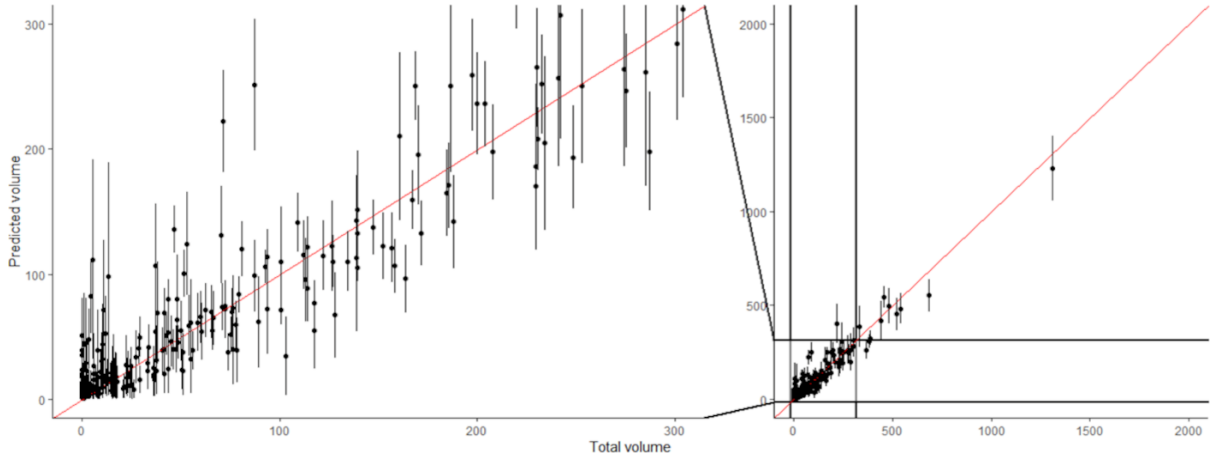

Figure 13 Conformal prediction intervals for total Pine sp. volumes

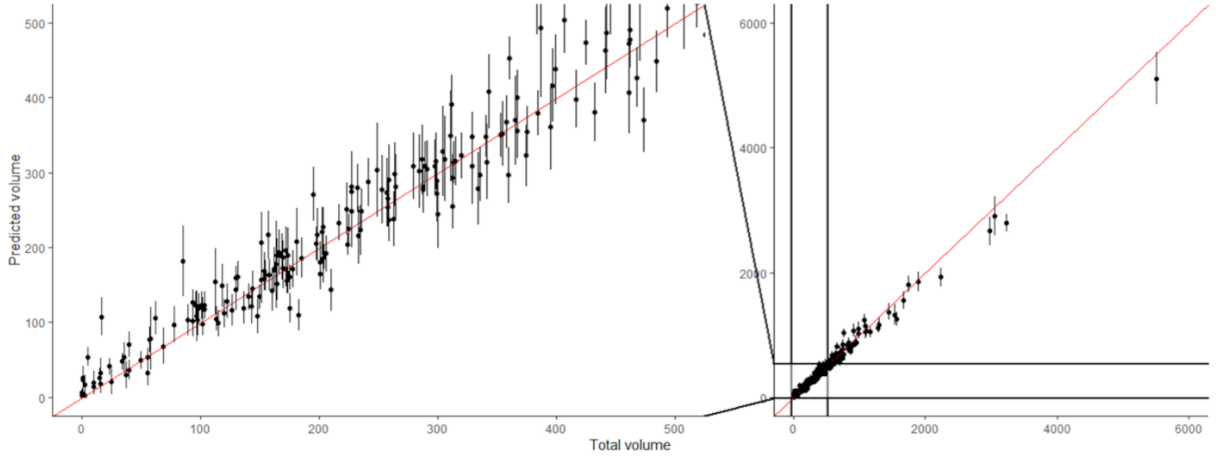

Figure 14 Conformal prediction intervals for total Spruce volumes

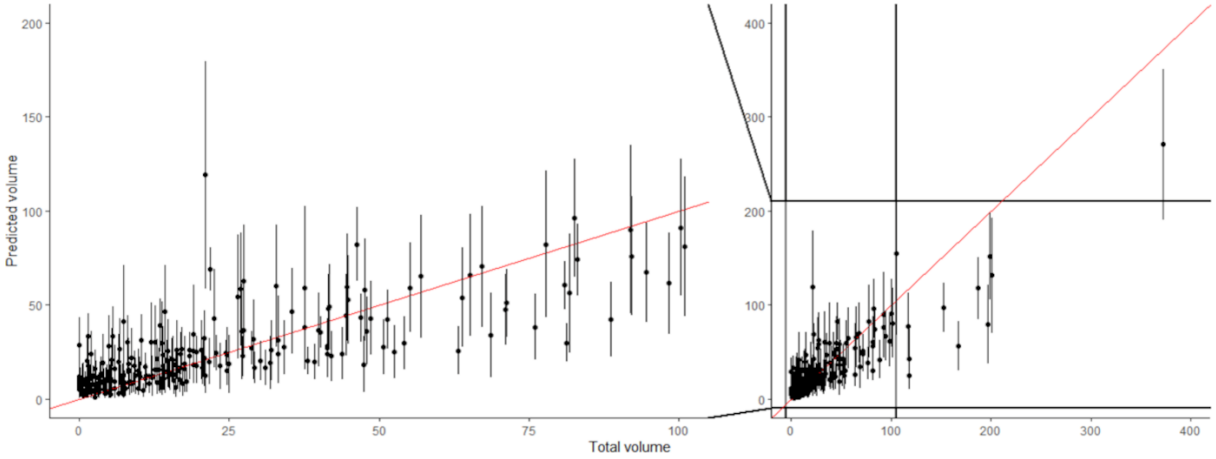

Figure 15 Conformal prediction intervals for total Deciduous sp. volumes

3. References

Achanta R, Shaji A, Smith K, Lucchi A, Fua P & Süsstrunk S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Mach Intelligence,34, 2274-82. doi: 10.1109/TPAMI.2012.120.

Anon (2023a). SLU Forest Map, Dept. of Forest Resource Management, Swedish University of Agricultural Sciences. https://www.slu.se/centrumbildningar-och-projekt/riksskogstaxeringen/statistik-om-skog/slu-skogskarta/ (last accessed 28 January 2023)

Anon (2023b). Swedish Environmental Protection Agency: Nationella Marktäckedata. https://www.naturvardsverket.se/verktyg-och-tjanster/kartor-och-karttjanster/nationella-marktackedata/ladda-ner-nationella-marktackedata/ (last accessed 28 January 2023)

Anon (2023c). Lantmäteriet: Laser data Download, forest. https://www.lantmateriet.se/sv/Kartor-och-geografisk-information/geodataprodukter/produktlista/laserdata-nedladdning-skog/#steg=4 (last accessed on 28 February 2022)

Anon (2023d). Lantmäteriet: Digital aerial photographs. https://www.lantmateriet.se/globalassets/geodata/geodataprodukter/flyg--och satellitbilder/e_pb_dig_flygb.pdf (last accessed on 28 February 2022)

Crookston NL & Finley AO (2007). yaImpute: An R Package for k-NN Imputation. Journal of Statistical Software 23, 1-16.

Eskelson BNI, Temesgen H, Lemay V, Barrett TM, Crookston NL & Hudak AT (2009) The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases, Scandinavian Journal of Forest Research, 24:3, 235-246, DOI: 10.1080/02827580902870490

Greenwald M & Khanna S (2001). Space-efficient online computation of quantile summaries. Proceedings of ACM SIGMOD, 30(2):58-66.

Ishwaran H & Kogalur UB (2007). Random survival forests for R. R News 7(2), 25--31.

Ishwaran H, Kogalur UB, Blackstone EH & Lauer MS (2008). Random survival forests. Ann. Appl. Statist. 2(3), 841--860.

Ishwaran H & Kogalur UB (2022). Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC), R package version 3.1.1.

Lei J, G'Sell M, Rinaldo A, Tibshirani RJ & Wasserman L (2018). Distribution-free Predictive Inference for Regression. Journal of the American Statistical Association, 113, 1094–1111

Li H, Jia Y, Cong R, Wu W, Kwong STW & Chen C (2021). Superpixel Segmentation Based on Spatially Constrained Subspace Clustering. IEEE Transactions on Industrial Informatics, 17, 7501-7512, doi: 10.1109/TII.2020.3044068.

Meinshausen N (2006) Quantile regression forests. Journal of Machine Learning Research, 7:983-999.

Mevik B-H & Wehrens R (2007). The pls package: Principal component and partial least squares regression in R. Journal of Statistical Software, 18(2):1–24, 2007.

Moeur M & Stage A.R (1995). Most similar neighbor: An improved sampling inference procedure for natural resource planning. Forest Science, 41, 337-359.

Packalén P & Maltamo M (2007). The k-msn method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment*, 109, 328–341.

Pons P & Latapy M (2005). Computing communities in large networks using random walks. https://arxiv.org/abs/physics/0512106.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schmitt M, Hughes LH, Qiu C & Zhu XX (2019). Aggregating cloud-free Sentinel-2 images with Google Earth Engine. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 142-152.

Scrucca L (2013). GA: A Package for Genetic Algorithms in R. Journal of Statistical Software, 53(4), 1-37. https://doi.org/10.18637/jss.v053.i04.

Scrucca L (2017). On some extensions to GA package: hybrid optimisation, parallelisation and islands evolution. The R Journal, 9/1, 187-206. https://doi.org/10.32614/RJ-2017-008.

Söderberg J, et al (2017). Utvärdering av utbytesprognoser med skogliga laserskattningar och skördardata - Evaluation of yield forecasts produced by forest laser estimations and harvester data. Work report 937-2017. Uppsala, Skogforsk.

Söderberg J, Möller J & Willén E (2018). Evaluation of yield prediction with harvester data. Work Report 981-2018, Skogforsk (in Swedish).

Söderberg J, Wallerman J, Almäng A, Möller J & Willén E (2021). Operational prediction of forest attributes using standardised harvester data and airborne laser scanning data in Sweden. Scandinavian Journal of Forest Research, 36:4, 306-314, DOI: 10.1080/02827581.2021.1919751

Zhang H, Zimmerman J, Nettleton D & Nordman DJ (2020). Random Forest Prediction Intervals. The American Statistician, 74:4, 392-406, DOI:10.1080/00031305.2019.158528