Title:

Assessing the importance of detailed forest inventory information using stochastic programming

Authors: Olha Nahorna[1*], Lennart Noordermeer[1], Terje Gobakken[1], Kyle Eyvindson[1]

Affiliations:

[1] Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, NMBU, P.O. Box 5003, NO-1433 Ås, Norway;

*Corresponding author: olha.nahorna@nmbu.no

Abstract

Errors in forest inventory data can lead to sub-optimal management decisions and dramatic economic losses. Forest inventory approaches are typically evaluated by their levels of precision and accuracy; however, this overlooks the specific usefulness of the data in decision-making. By evaluating the value of information (VoI), we can assess the usefulness of the data for specific decision-making problems. We evaluated the VoI through stochastic programming for four airborne laser scanning-based inventory approaches. The stochastic programming model explored the trade-off between the maximal net present value and the minimal conditional value at risk of meeting specified periodic income targets. We evaluated a range of periodic targets and risk aversion preference levels. To compare the performance of the inventory approaches, we used a reference dataset that was acquired using a forest harvester with precise positioning. For a wide range of the trade-offs, inventory approaches with higher-quality information provided the best overall performance. If only one of the extreme objectives was desired, less precise inventory approaches were sufficient to produce high-quality solutions.

Keywords: forest planning, value of information, stochastic programming, uncertainty, risk management, forest inventory, data quality

1. Introduction

Forest planning is a multifaceted process that involves designing and proposing alternatives for implementing strategies and operations aimed at managing forest resources. The traditional aim has been to meet industrial needs for forest products while ensuring that silvicultural practices are carried out in a sustainable, cost-efficient, and responsible manner (Bettinger et al. 2017). This aim has expanded to include more varied benefits, including biodiversity, carbon storage, and protecting sensitive ecosystems (Hunault-Fontbonne and Eyvindson 2023). Achieving a balance between these objectives is a complicated task, requiring the use of complex data and mathematical models.

High-quality data are essential for accurate forecasts of a forest's development; however, the expense of these data should be carefully considered when evaluating inventory data needs. All data include some level of uncertainty, as perfect information about the standing trees is nearly impossible to obtain. In addition to data uncertainty, other significant sources of uncertainty include inaccuracies in growth and yield models, market volatility, and natural disturbances (Pasalodos-Tato et al. 2013).

The study of inventory errors has been of particular interest to foresters, with detailed analyses starting in the late 20th century. Uncertainties in the initial inventory data can substantially impact the accuracy of the predictions of the growth and yield models (Gertner and Dzialowy 1984), leading to sub-optimal management decisions with the potential for large economic losses. To limit the impact of this uncertainty, decision-makers often opt to obtain forest information with the greatest accuracy, to minimize the losses. However, data accuracy alone does not indicate the extent to which a given inventory approach is useful for decision-making (Ketzenberg et al. 2007). To effectively evaluate the usefulness of accurate inventory data in forest planning, decision-makers should conduct a value of information (VoI) assessment.

The value of forest information has been defined as the difference between the expected values of a management decision made with and without additional information (Kangas, 2010). By evaluating the VoI obtained using alternative forest inventory approaches, we can quantify which inventory approach is the most efficient for the specific forest planning use case.

The most common method used to compare the VoI of inventory approaches is the so-called cost-plus-loss (CPL) analysis. This method quantifies the total inventory cost as the sum of the direct inventory costs and the losses that occur due to sub-optimal decisions. Consequently, the inventory method with the lowest total cost is identified as the most favorable (Burkhart et al. 1978). In early forestry CPL analyses, Eid et al. (2004) compared forest inventory approaches based on laser scanning and photo-interpretation. The results demonstrated that despite the higher direct cost of laser scanning, it led to improved decision-making and reduced the total cost compared to lower-cost photo-interpretation data. Later studies compared expected losses from different inventory approaches, e.g., stand-wise visual inventory and airborne laser scanning (ALS) (Mäkinen et al. 2010), inventory approaches that rely on the use of ALS, satellite data, or their combination (Duvemo et al. 2007), and ALS and digital aerial photogrammetry (Kangas et al. 2018).

Bergseng et al. (2015) applied CPL analysis to identify expected losses when using data obtained from four ALS-based inventory approaches. These inventory approaches were area-based approaches (ABAs) to calculate the mean values and diameter distributions (ABA-MV and ABA-DD, respectively) of forest stands, the individual tree crown (ITC) approach, and the semi-ITC approach. In the ABA, regression models fitted on a sample of field plots with corresponding ALS metrics are used to predict forest attributes over a grid tessellating the inventory area (Næsset 2002). In ABA-MV, predicted forest attributes are then summarized to mean stand values, whereas in ABA-DD, diameter distributions are obtained (Gobakken and Næsset 2004, 2005). In the ITC approach, tree-level information is obtained by delineating ITC segments from ALS data (Hyyppä and Inkinen 1999). The semi-ITC approach mitigates systematic errors in the ITC approach that arise from segmentation errors, by allowing tree crown segments to contain single, multiple, or no trees (Breidenbach et al. 2010). The results from the CPL analysis (Bergseng et al. 2015) demonstrated that ABA-DD was a favorable inventory approach that resulted in the smallest losses. The ITC and semi-ITC inventory approaches avoided large losses as they both tended to detect the largest and most valuable trees; however, their inventory costs were substantially larger than for ABA methods. ABA-MV demonstrated the largest losses among the studied approaches, resulting from sub-optimal management decisions.

Although CPL analysis has been shown to be effective in evaluating the VoI obtained with different inventory approaches, the method is not without limitations. Forest-focused CPL studies have only evaluated the economic impact of the data quality. This assumes that the utility of a decision-maker is expressed solely through the net present value (NPV) maximization problem. Moreover, the reference data used for comparing inventory approaches were typically assumed to be free of error.

An alternative way of quantifying the VoI is using stochastic programming. Stochastic programming allows the incorporation of various sources of uncertainty into the development of forest management plans, as well as the specific formulation of the objective function (Birge and Louveaux 2011; King and Wallace 2012). Earlier forest-specific studies using stochastic programming have demonstrated how the VoI can be evaluated by quantifying the difference in the objective function values resulting from various model formulations (Eyvindson and Cheng 2016; Eyvindson and Kangas 2016; Eyvindson et al. 2017). These studies primarily focused on the evaluation of the VoI derived from using different optimization model formulations while utilizing the same inventory data.

We explored the difference in the VoI obtained by using inventory data acquired from a variety of inventory approaches, applied to the same stochastic programming optimization problem. The optimization problem aimed to maximize the NPV while minimizing the conditional value at risk (CVaR) of not achieving specific income targets. The use of this CVaR formulation can be interpreted as an income even-flow requirement that strives to ensure a relatively consistent income over the planning horizon. Providing an even flow of income (or timber) enables stable economic conditions for forest owners and communities that rely on the forestry sector.

The research objectives of this study were twofold. The first main research objective was to develop a stochastic programming-based framework for the comparison of the VoI of four inventory approaches with different costs: ABA-DD, ABA-MV, ITC, and semi-ITC. When designing the framework, we addressed three specific sub-objectives: (1.1) to account for uncertainty in all of the inventory datasets and the reference inventory data; (1.2) to expand the VoI problem formulation to incorporate multiple objectives; and (1.3) to quantify the impact of the VoI with respect to the decision-maker's risk-aversion preferences. The second main research objective was to evaluate when accurate forest inventory data are cost-efficient and when less accurate data may be sufficient for specific problems.

To enhance the readability of this paper, we have included a comprehensive list of abbreviations in Table 1.

Table 1. List of abbreviations used in the paper.

Abbreviation	Definition
ABA	Area-based approach
ABA-DD	Area-based approach diameter distribution
ABA-MV	Area-based approach mean values
ALS	Airborne laser scanning
CPL	Cost-plus-loss
CVaR	Conditional value at risk
EVPI	Expected value of perfect information
ITC	Individual tree crown
kNN	k nearest neighbors
NPV	Net present value
VaR	Value at risk
VoI	Value of information
VoII	Value of improved information

2. Materials and methods

2.1. Study overview

Forest inventory data were obtained using four ALS-based inventory approaches: the area-based mean values (ABA-MV), area-based diameter distribution (ABA-DD), individual tree crown (ITC), and semi- individual tree crown (semi-ITC). Data collected by the harvester during harvesting operations were used as a reference dataset, as these data provided high-quality on-site information. All data were collected from the municipalities of Etnedal, Nord-Aurdal, Sør-Aurdal, and Nordle Land in southern Norway.

To evaluate the differences between the VoI provided with the four studied inventory approaches, we designed the methodology presented in Fig. 1. The input data used to simulate the development of the forest under uncertainty included forest inventory information, estimates of the inventory uncertainty, and a set of management alternatives to be applied. Through the simulation process, we generated multiple simulations for each stand, considering different levels of uncertainty. We then applied Monte Carlo random sampling to create 1000 unique scenarios to represent different realizations of the possible initial inventory conditions of the forest holding. These scenarios were fed into a stochastic optimization model to assess how risk preferences affect the output. To compare the various inventory approaches, we evaluated the outcomes of the obtained solutions against the reference dataset and calculated the VoI for each solution.

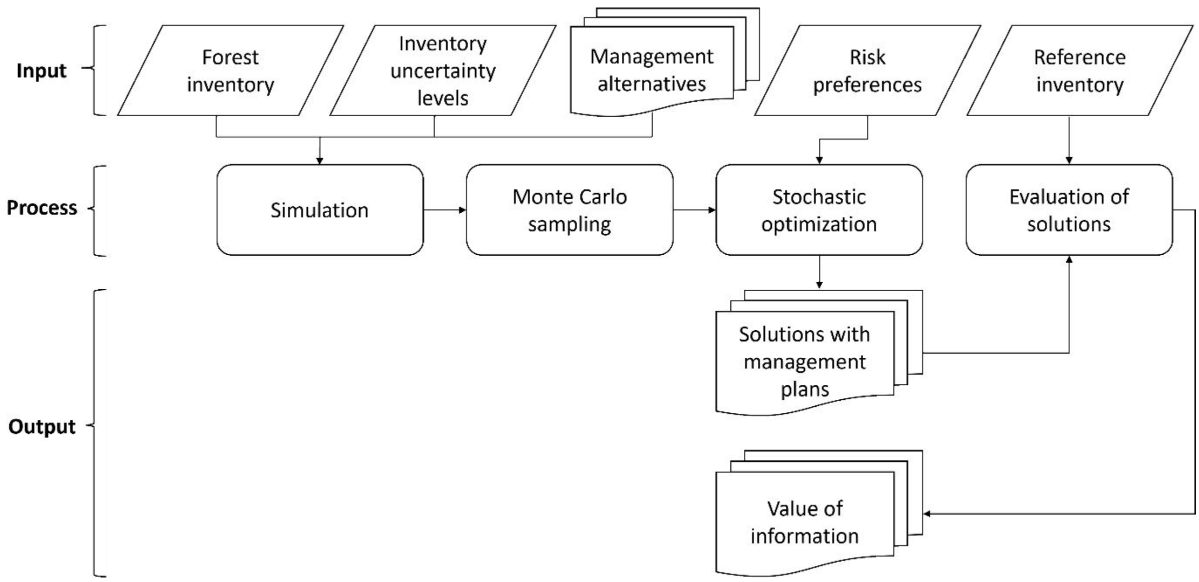

Fig. 1. Methodology flowchart used to assess the value of information of inventory approaches with varying properties of uncertainty.

2.2. Inventory data

We used forest inventory data that applied four ALS-based inventory approaches (ABA-MV, ABA-DD, ITC, semi-ITC), using data collected by four ALS surveys acquired in 2013, 2016, 2017, and 2019 using different instruments and acquisition parameters. The reference data used in VoI assessment were collected by a single-grip Komatsu 931XC harvester with precise positioning of the harvester head. In total, 131375 trees from 49 stands were recorded, with Norway spruce accounting for 89%, Scots pine for 6%, and deciduous trees, mainly birch, for 5%. For details on the collected harvester data, see section A1 in the Appendix.

For the ABA-DD, ITC, and semi-ITC approaches, we followed the data processing methodology described in Noordermeer et al. (2023). Tree crowns were segmented from the ALS data and harvested areas were delineated from the harvester data (for details, see section A2.1. in the Appendix). The polygons of harvested areas were tessellated using a regular grid of 250 m^2, and grid cells that were completely located within the polygons were selected as observations for the ABA (Fig. 2A). The polygons of segmented tree crowns that intersected ABA grid cells were then used as observations for the ITC approach. For each segment, we used the harvested tree within the segment for which the coordinates of the treetop (x,y,z) in the harvester data were nearest to the coordinates of the treetops detected in the ALS data (Fig. 2C). In this way, we established the closest spatial match in three dimensions between trees registered in the harvester data and those detected in the ALS data. For the semi-ITC approach, the methods proposed by Breidenbach et al. (2010) were used. The same tree crown segments that intersected ABA grid cells were used as observations; however, multiple harvested trees were allowed for a given tree crown segment. As a result, some semi-ITC segments were empty, some contained a single tree, and some contained multiple trees (Fig. 2D). Tree lists were then compiled for all harvester observations, and for the ABA-MV specifically, the number of stems within the grid cell (N), and the diameter and height of the basal area median tree (Dgm and Hgm, respectively) were computed.

For the ABA-DD, ITC, and semi-ITC approaches, tree lists were imputed for each target observation using the harvester data, based on the k nearest neighbors method (kNN, McRoberts et al. 2015) according to selected ALS metrics (for details, see section A3 in the Appendix). For the ABA-MV approach, the mean values of N, Dgm, and Hgm were correspondingly imputed (Fig. 2B). The ALS metrics were selected using the leaps package (Lumley 2004), based on linear models with the tree height and volume as response variables. The number of neighbors was selected using the caret package (Kuhn 2008) by fitting kNN models with the tree height and volume as response variables. The value of k that minimized the root mean square error was selected. Finally, kNN models were trained using the yaImpute package (Crookston and Finley 2008) in R, with the selected ALS metrics as predictors and default parameters (for details, see section A2.2. in the Appendix).

For the ABA-MV approach, tree-list information was created using diameter distribution models. Specifically, the forest simulator applied a diameter distribution model proposed by Kangas and Maltamo (2000) if basal area information was present; otherwise, a model proposed by Kilkki et al. (1989) was applied for spruce, and a model proposed by Siipilehto (1999) was applied for other species. One important note is that depending on the inventory approach used, there can be variations in the modeled errors due to the method used to construct the tree list. Quantifying the impact on the overall error is challenging as it differs at a stand level and is also influenced by the distribution of tree sizes within the stand.

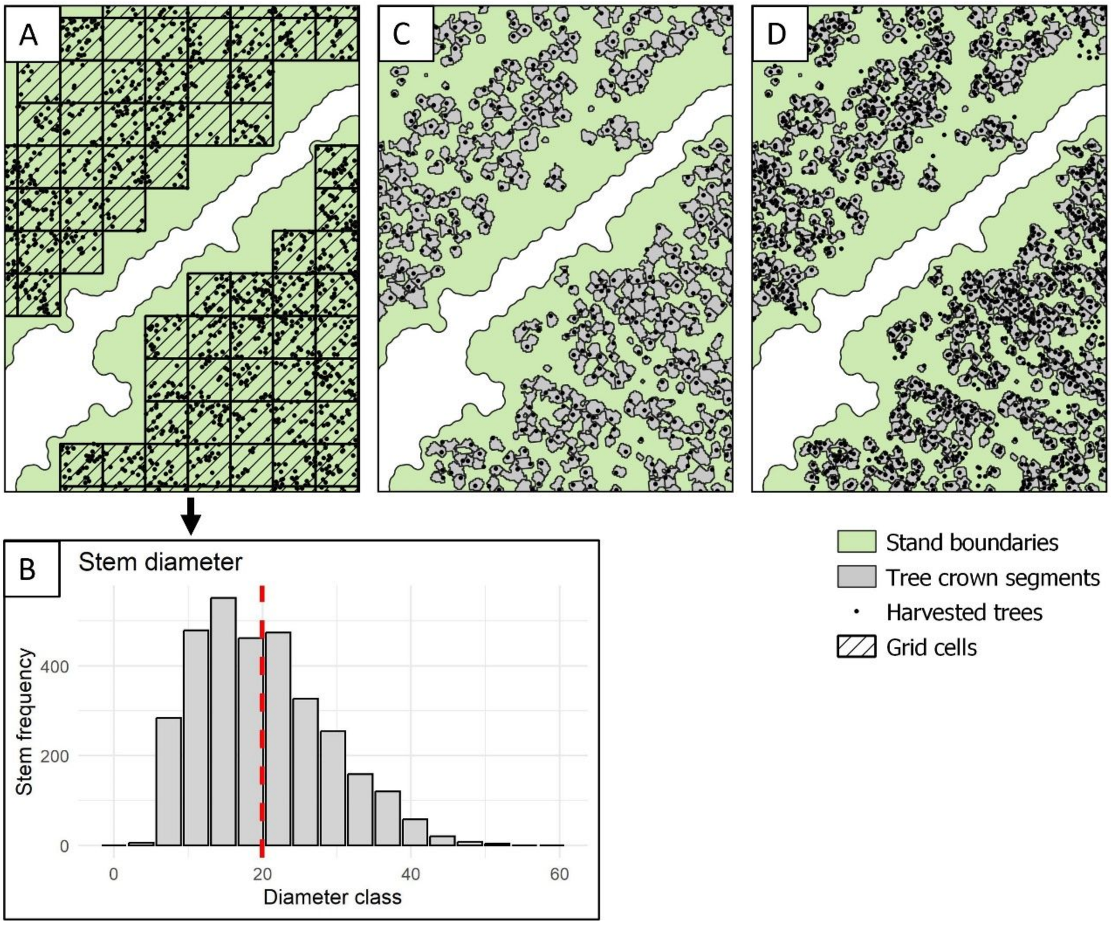

Fig. 2. The inventory approaches used in the study: (A) area-based approach (ABA), where plot-level reference data are linked to ALS data in statistical models, and the models are then used for prediction over a grid tessellating the inventory area; (B) in area-based mean values (ABA-MV), predicted forest attributes are summarized to mean stand values, and in area-based diameter distribution (ABA-DD) diameter distributions are obtained; (C) individual tree crown (ITC) approach, where reference data on individual trees are linked to ALS data computed for segmented tree crowns; and (D) semi-individual tree crown (semi-ITC) approach, which allows tree crown segments to include one, multiple, or no trees.

2.3. Simulation process

The inventory data from the four inventory approaches and the tree information from the reference harvester data were used as input for the simulation. Inventory errors were introduced systematically, sampling across the distribution of the errors. For each of the inventory approaches, we assumed the errors of the inventory of the number of trees and height of the trees (or stratum for the ABA-MV case) to be normally distributed around the mean of predicted values with a standard deviation of 20%. To address sub-objective 1.1, we assumed the information on the trees from the reference data to be more accurate than the outputs of the four inventory approaches but not perfect. To reflect this, we assumed the error of height of the trees in the reference data to be normally distributed around the mean with a standard deviation of 5%. To forecast the future development of the forest, we used the forest simulator SIMO (Rasinmäki et al. 2009). The SIMO simulator is open-source software that can utilize a wide variety of input data. For this application, both stand-level and tree-level data were used. With tree-level data, a tree list is directly imputed to the simulator, whereas with stand-level data, a tree list is constructed based on diameter distribution assumptions before starting the simulation. For this application, forest development was simulated for 50 years, with 10 five-year periods, and a branching approach similar to that of Siitonen (1993) was applied to construct a large variety of management schedules.

The same simulation process was applied to all five datasets, incorporating uncertainty using a systematic approach. This process produced multiple sets of simulations with different levels of uncertainty for each stand, each containing an identical set of management schedules. The simulated data were used as the input when constructing scenarios for the optimization process. For each scenario, the Monte Carlo random sampling approach was applied for each stand’s sets of simulations, resulting in a set of 1000 scenarios. Each scenario was represented by a randomly selected simulation for each stand. According to Eyvindson and Kangas (2015), the generated number of scenarios is more than sufficient to represent the stochastic forest management problem.

2.4. Stochastic programming optimization model

To evaluate the VoI obtained from the four inventory approaches, we utilized a standard stochastic programming optimization model. Following sub-objective 1.2, the model aims to maximize the expected net present value ($\mathbb{E}(NPV)$) while simultaneously minimizing the conditional value at risk (CVaR) of not achieving the periodic target incomes. The CVaR is a measure of downside risk. The CVaR measures the mean of the losses that exceed the value at risk (which measures the upper quantile of potential losses; Duffie and Pan 1997). The CVaR can be easily linearized (Rockafellar and Uryasev 2000), which simplifies its incorporation into a stochastic programming model. In this formulation, the CVaR minimization objective acts as a soft constraint that helps to ensure income even-flow over the periods. Soft constraints in stochastic programming can be violated by the model when necessary to ensure the feasibility of the solution. The overall formulation of the model represents one of the most common forest management problems, with the goal of maximizing the NPV while ensuring an even-flow constraint (Eyvindson and Kangas 2016). The designed stochastic programming model presents an updated formulation of an earlier model presented by Eyvindson and Cheng (2016):

Objective function:

$$Max\left(\lambda\frac{\mathbb{E}(NPV)-\ NPV_{*}}{NPV^{*}-NPV_{*}}-(1-\lambda)\frac{\sum_{t=1}^{T}CVaR_{t}}{CVaR^{*}}\right) \quad (1)$$

Subject to

$$NPV_{n}=\sum_{t=1}^{T}\frac{I_{nt}}{(1+r)^{(tD-U)}}+\sum_{j=1}^{J}\sum_{k=1}^{Kj}\frac{PV_{njkT}x_{jk}}{(1+r)^{DT}},\quad \forall n=1,\ldots,N, \quad (2)$$

$$\mathbb{E}(NPV)=\sum_{n=1}^{N}p_{n}NPV_{n}, \quad (3)$$

$$I_{nt}=\sum_{j=1}^{J}\sum_{k=1}^{Kj}x_{jk}c_{jknt},\quad \forall n=1,\ldots,N\ \ t=1,\ldots,T, \quad (4)$$

$$L_{nt}=[b_{t}-I_{nt}]^{+},\quad \forall n=1,\ldots,N\ \ t=1,\ldots.,T, \quad (5)$$

$$CVaR_{t}=VaR_{t}+\frac{1}{(1-\alpha)N}\sum_{n=1}^{N}[L_{nt}-VaR_{t}]^{+},\quad \forall t=1,\ldots,T, \quad (6)$$

$$\sum_{k=1}^{Kj}x_{jk}=1,\quad \forall j=1,\ldots,J, \quad (7)$$

where the sets, variables, and parameters used in the model are presented in Table 2.

Table 2. Description of used notation.

Symbol	Definition
Sets	
J	The set of stands
K	The set of silvicultural operation prescriptions
N	The set of scenarios that represent uncertainty
T	The set of time periods
Variables	
c_{jknt}	Income generated from stand j by applying prescription k under scenario n for period t
$CVaR_t$	Conditional value at risk for period t
$\mathbb{E}(NPV)$	Expected net present value
I_{nt}	Income for scenario n for period t
L_{nt}	Losses for scenario n for period t
NPV_n	Net present value for scenario n
p_n	Probability of scenario n occurring
PV_{njkT}	Productive value for scenario n at stand j managed according to prescription k for the final period
VaR_t	Value at risk for period t
x_{jk}	Proportion of stand j managed under prescription k
Parameters	
α	Confidence interval for the value at risk
b_t	Target income for period t
$CVaR^*$	Maximum conditional value at risk reached when net present value maximization is considered as the main objective of the model
D	Duration of the period
λ	Risk coefficient
NPV^*	Maximum net present value reached when net present value maximization is considered as the main objective of the model
NPV_*	Minimum net present value reached when conditional value at risk minimization is considered as the main objective of the model
r	Discount rate
U	Timing of a silvicultural operation during the period

To enable the accomplishment of sub-objective 1.3, we introduced the risk coefficient λ to the objective function. The risk coefficient is defined by a decision-maker and allows them to specify their risk aversion preference. In our model, the parameter varied between 0 and 1, bounding the problem between the two extremes of a risk-averse solution with a pure focus on minimizing the CVaR and a risk-neutral solution with a focus on maximizing $\mathbb{E}(NPV)$.

Equation 1 presents the objective function, which aims to maximize $\mathbb{E}(NPV)$ and minimize the CVaR. We used a 3% discount rate (r) and 90% confidence interval (α). Both $\mathbb{E}(NPV)$ and the CVaR are balanced by the maximum and minimum values that could be achieved at extreme points, when the objective function focuses solely on optimizing one of the components (the minimum value of the CVaR will always be zero; therefore, it is not part of the equation). This way, each component of the objective function is scaled between 0 and 1. By changing the risk coefficient λ, the decision-maker is able to express their risk-aversion priority, e.g., with $\lambda = 1$, the focus is only on $\mathbb{E}(NPV)$, with $\lambda = 0$, the focus is only on CVaR, and with $\lambda = 0.5$, both components of the objective function are treated as equally important.

The NPVs for each scenario are computed using Equation 2; they are evaluated as the sum of discounted incomes from silvicultural operations and the productive value of the forest at the end of the planning horizon. The productive value is computed using the models from Pukkala (2005). Equation 3 evaluates $\mathbb{E}(NPV)$. Equation 4 computes the incomes generated by each scenario at each period. Equation 5 calculates the losses for each scenario and each period. The losses are expressed as the difference between the stated periodic income target and achieved incomes. For this study, targets were set by testing and selecting those that remain achievable under most scenarios. Equation 6 computes the CVaR for each period. In Equations 5 and 6, the "+" symbol refers to keeping only positive values, while all negative values are set to 0. Finally, Equation 7 ensures that the entire proportion of the area of each stand is allocated to a specific management prescription. This means that overall, the model formulation follows the Model I formulation of Johnson and Scheurman (1977), where a set of management treatments is developed for each stand.

2.5. Assessment of the value of information

To assess the first research objective, we computed the VoI of the different inventory approaches by following a series of steps. First, the stochastic programming model was run with each of the inventory datasets as input (using an initial Monte Carlo sample set of scenarios generated in the simulation process). A direct comparison between these models' outputs is impossible, as the input inventory data are different. Therefore, to allow the outputs to be compared, we extracted the obtained solution (a list of management prescriptions to be implemented for the forest holding) from each inventory dataset so that they could be evaluated against the reference dataset. This solution was then applied to the simulated reference dataset (obtained with another Monte Carlo scenario set) as input. This is a simple calculation of the expected "real-life" outputs that a decision-maker would have obtained if they designed a management plan based on the data from a certain inventory approach.

To compute the VoI obtained from the four studied approaches, we needed to compare the objective function values for all solutions evaluated using the reference dataset. This required us to run the stochastic programming model with the reference data to obtain the optimal solution with the most accurate data. We then extracted the objective function value obtained with the reference data, as well as the corresponding objective function values obtained after applying the solutions from the inventory approaches' datasets to the reference data. The VoI for each approach was then determined by calculating the difference between the two objective function values (Equation 8), following a methodology similar to that described in Chapter 4 of Birge and Louveaux (2011). We suggest the term value of improved information (VoII):

$$VoII = Obj.function\ value_{reference} - Obj.function\ value_{inventory} \quad (8)$$

Finally, to meet the second research objective, we studied the changes in the VoII for different values of the periodic income target and risk coefficient λ.

3. Results

The obtained VoII varied between the four studied inventory approaches (Fig. 3). These values represent the percentage of the optimal value reduction in the objective function value compared to a case with information of better quality (reference harvester data in this study). From these results, we can infer that semi-ITC and ABA-DD demonstrated very similar results and outperformed the two other inventory approaches across all targets and values of the risk coefficient λ. On the other hand, ABA-MV and ITC showed a clearly weaker performance.

Interestingly, the VoII varied with changes in the periodic income targets (700, 1100, and 1500 € per ha) and the value of the risk coefficient λ. The VoII for semi-ITC and ABA-DD was rather low and stayed under 9% for the low and middle targets, and it was under 6% for the high target. A slight increase in the VoII was observed at intermediate values of λ, where the optimization model aimed to balance the $\mathbb{E}(NPV)$ maximization and CVaR minimization objectives. Similar trends were observed with ITC; however, it performed considerably worse at intermediate values of λ, where the VoII almost reached 30% for the low and high targets. On the other hand, the trends observed in ABA-MV were considerably different. ABA-MV exhibited a substantial increase in the VoII at the low and middle targets when the optimization objectives shifted towards the $\mathbb{E}(NPV)$ maximization problem (λ approaching 1). At the high target, the importance of accurate data remained high and relatively consistent at all values of the risk coefficient. Finally, it is important to note that ABA-DD, ITC, and semi-ITC performed equally well in the case of a pure $\mathbb{E}(NPV)$ maximization problem (λ = 1); in the case of a pure CVaR minimization problem (λ = 0), all four approaches performed equally well at the low and middle targets. This highlights that the VoII highly depends on the decision-maker's targets and risk-aversion preferences.

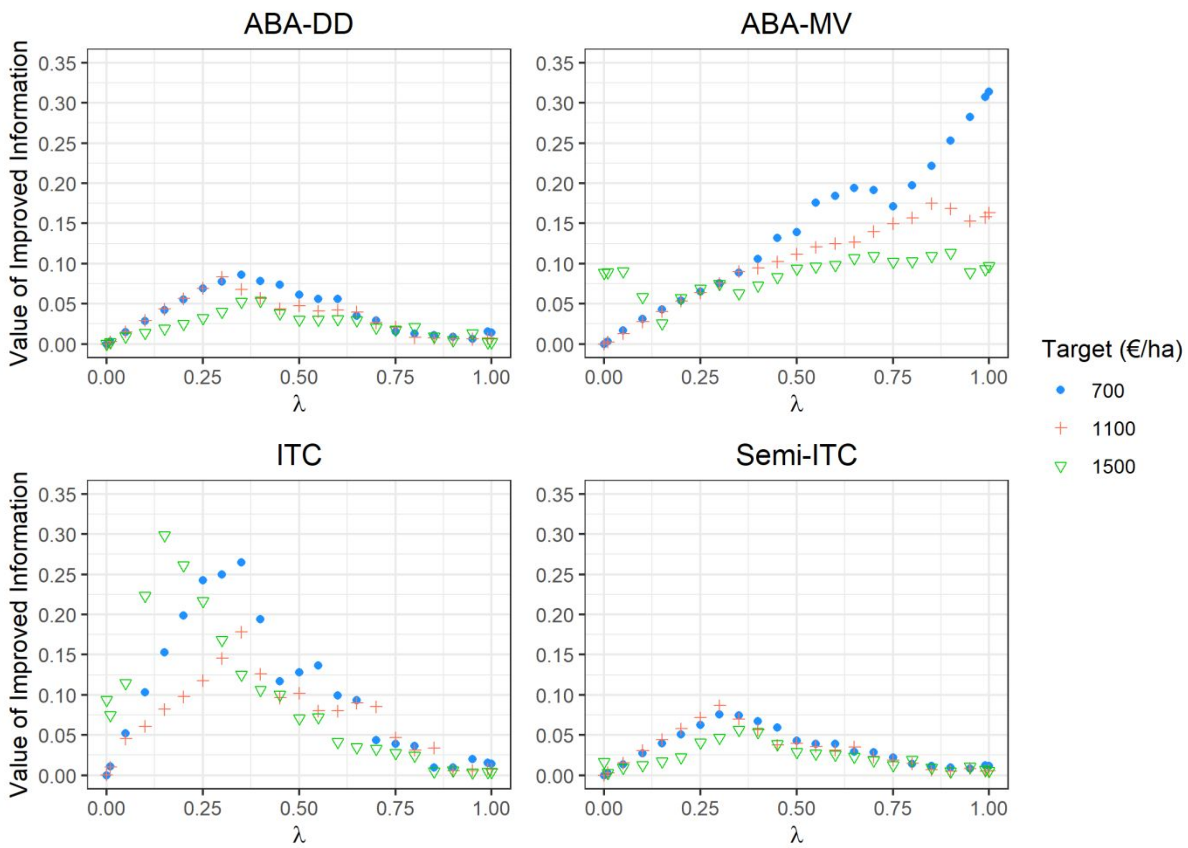

Fig. 3. The value of improved information at different values of the risk coefficient λ obtained for the four inventory approaches, area-based diameter distribution (ABA-DD), area-based mean values (ABA-MV), individual tree crown (ITC), and semi-individual tree crown (semi-ITC), for three periodic income targets.

Although our analysis focused on the outcome of applying management decisions obtained from stochastic programming optimization models with the inventory data compared to the reference dataset (Fig. 4, applied output), the difference between these results and the anticipated outputs from the optimization models (Fig. 4, model output) is worth noting. This difference demonstrates the accuracy of the models' predictions. It shows that semi-ITC and ABA-DD were the closest to the expected outcomes of the use of the generated solution on the reference dataset, followed by ABA-MV and ITC. It is also interesting to highlight certain differences between the anticipated models' outputs. Although a direct comparison between them is not possible, their discrepancies offer valuable insights, highlighting how different the expectations of the decision-maker would be when applying each of the inventory approaches.

When analyzing the obtained values of $\mathbb{E}(NPV)$ and the CVaR after applying the generated management plans to the reference dataset (Fig. 4, applied output), a certain inference can be made. Specifically, the analysis indicates that ABA-MV provided the lowest values of $\mathbb{E}(NPV)$ at high values of the risk coefficient λ and lower CVaR values relative to the other studied inventory approaches in most cases. Semi-ITC and ABA-DD demonstrated the highest values of $\mathbb{E}(NPV)$; however, ABA-DD showed a slightly higher CVaR. Finally, at lower values of the risk coefficient λ, ITC tended to produce solutions with the lowest values of $\mathbb{E}(NPV)$ and slightly higher values of the CVaR. However, as λ approached 1, the ITC approach showed similarly high values of $\mathbb{E}(NPV)$ compared to semi-ITC and ABA-DD, and also the highest values of the CVaR among all approaches.

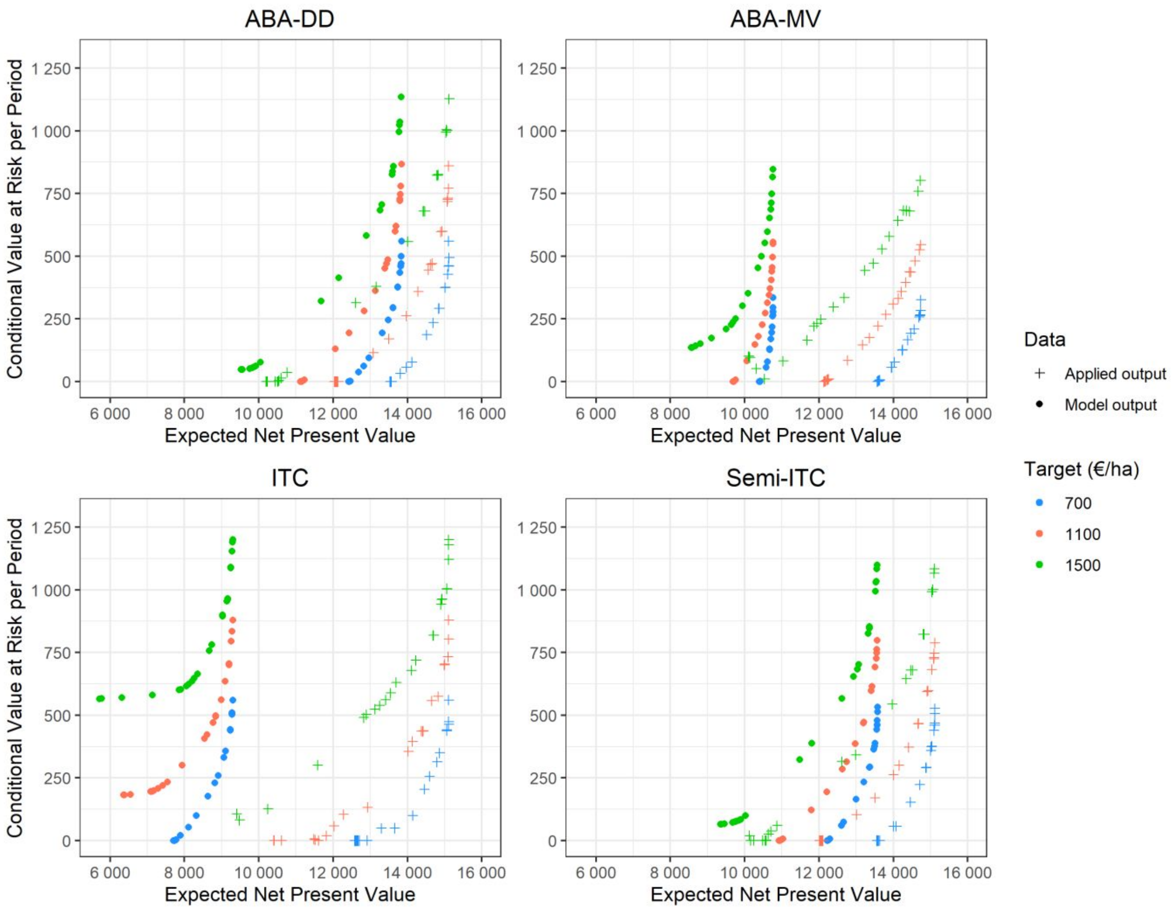

Fig. 4. Anticipated optimization model outputs (model output) for the expected net present value and average periodical conditional value at risk obtained with inventory data from the four inventory approaches (area-based diameter distribution (ABA-DD), area-based mean values (ABA-MV), individual tree crown (ITC), and semi-individual tree crown (semi-ITC)) compared to the outputs of applying the management plans obtained with model outputs to the reference dataset for three periodic income targets (applied output). The risk coefficient λ is lowest (λ = 0) on the left side of each graph and increases towards the right side of each graph (reaching λ = 1).

The incorporation of the CVaR minimization goal as part of the stochastic programming objective function allowed the even flow of periodic incomes to be ensured (Table 3). With CVaR minimization as the main priority, it was possible for all inventory approaches to meet the low- and middle-level income targets. At the high target, semi-ITC and ABA-DD were able to assure periodic incomes in close proximity to the target value, whereas ABA-MV and ITC exhibited more significant reductions in incomes for period 1. This indicates that not all trees were detected when the ITC approach was used, and as a result, timber volumes were underestimated. In the case of ABA-MV, the underestimation was likely due to the application of the diameter distribution model.

At $\lambda = 0.5$, i.e., when the model gave $\mathbb{E}(NPV)$ maximization and CVaR minimization the same importance, the model tended to harvest more in the first period. Additionally, it ensured that target objectives were met until period 5 or 6 for all approaches, except ITC, where the target was not met after period 4. After that, incomes remained below the target value until the end of the planning horizon. This held true for all except the low target, where all four studied approaches reached the target again at period 8 and also at period 9 for ABA-MV and ITC. Finally, when the optimization model focused on $\mathbb{E}(NPV)$ maximization, most of the available volume was harvested at period 1 with all inventory approaches. This was followed by very low or no harvests until period 8, where a new peak was observed for all targets. For the high target, ABA-MV was the only approach for which the peak at period 8 did not reach the income target value.

Table 3. Expected periodic incomes, expected net present value E(NPV) and conditional value at risk (CVaR) per hectare generated with the use of management decisions from the four inventory approaches: area-based diameter distribution (ABA-DD), area-based mean values (ABA-MV), individual tree crown (ITC), and semi-individual tree crown (semi-ITC).

Inventory approach	ABA-DD			ABA-MV			ITC			Semi-ITC		
Periodic income target (€/ha)	700			700			700			700		
Risk coefficient	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99
E(NPV) (€/ha)	13533	14693	15109	13575	14232	14740	12584	14455	15109	13555	14709	15112
CVaR (€/ha)	0	2347	5600	1	1250	3256	0	2049	5600	0	2230	5072
Income Period 1 (€/ha)	6633	9752	12146	6079	7080	8363	4655	8157	12146	6481	9568	11848
Income Period 2 (€/ha)	2527	934	0	3062	2930	2951	2202	1943	0	2754	1159	385
Income Period 3 (€/ha)	862	869	0	1082	1124	1391	1568	1434	0	881	861	0
Income Period 4 (€/ha)	817	793	0	978	974	209	1416	1264	0	780	768	0
Income Period 5 (€/ha)	823	598	0	841	759	0	1270	0	0	813	711	0
Income Period 6 (€/ha)	827	0	0	731	709	0	1172	0	0	814	0	0
Income Period 7 (€/ha)	857	0	0	718	0	0	1132	545	0	873	0	0
Income Period 8 (€/ha)	934	1588	2338	894	1089	1336	773	1231	2338	914	1540	2194
Income Period 9 (€/ha)	700	422	0	700	700	902	722	700	0	700	426	151
Income Period 10 (€/ha)	724	140	0	743	152	38	953	224	0	749	145	0
Periodic income target (€/ha)	1100			1100			1100			1100		
Risk coefficient	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99
E(NPV) (€/ha)	12037	14288	15110	12130	13586	14740	10419	14133	15109	12011	14155	15113
CVaR (€/ha)	0	3595	8603	0	2218	5454	3	3955	8800	0	3005	7886
Income Period 1 (€/ha)	4881	8093	12028	4357	6260	8362	1112	6665	12146	4747	7663	11621
Income Period 2 (€/ha)	1369	1451	150	1823	1784	2951	2303	2377	0	1432	1596	674
Income Period 3 (€/ha)	1281	1335	0	1563	1571	1391	2449	2147	0	1344	1328	0
Income Period 4 (€/ha)	1191	1204	0	1404	1480	210	1776	1801	0	1191	1186	0
Income Period 5 (€/ha)	1225	1259	0	1333	1322	0	1110	0	0	1243	1227	0
Income Period 6 (€/ha)	1261	42	0	1244	1131	0	1347	0	0	1235	658	0
Income Period 7 (€/ha)	1303	0	0	1140	550	0	1664	532	0	1304	0	0
Income Period 8 (€/ha)	1203	1210	2290	1112	892	1336	1600	998	2338	1221	1115	2099
Income Period 9 (€/ha)	1209	556	50	1117	532	902	1365	828	0	1216	549	252
Income Period 10 (€/ha)	1160	208	0	1144	217	38	1575	300	0	1180	202	0
Periodic income target (€/ha)	1500			1500			1500			1500		
Risk coefficient	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99	λ = 0.01	λ = 0.5	λ = 0.99
E(NPV) (€/ha)	10204	14009	15118	10110	12679	14728	9482	14107	15110	10158	13966	15096
CVaR (€/ha)	6	5585	11281	1006	3346	8026	820	6788	11803	6	5439	10664
Income Period 1 (€/ha)	1519	6914	11743	505	3668	8362	784	5679	12028	1517	6629	11346
Income Period 2 (€/ha)	1735	1790	527	2304	2429	2834	1518	3243	150	1576	1968	995
Income Period 3 (€/ha)	1702	1743	0	2086	2027	1516	1516	2910	0	1740	1789	0
Income Period 4 (€/ha)	1608	1637	0	1926	1901	210	2309	1512	0	1621	1633	0
Income Period 5 (€/ha)	1652	1516	0	1755	1837	0	1424	0	0	1686	1520	0
Income Period 6 (€/ha)	1620	0	0	1718	1530	0	1511	0	0	1659	192	0
Income Period 7 (€/ha)	1635	0	0	1557	1326	0	1513	0	0	1655	0	0
Income Period 8 (€/ha)	1651	994	2147	1526	439	1336	1713	885	2290	1676	935	2004
Income Period 9 (€/ha)	1621	640	202	1533	641	896	2053	1110	50	1648	654	353
Income Period 10 (€/ha)	1599	282	0	1533	279	38	2088	217	0	1613	285	0

Note: Solutions obtained within one inventory approach but with different targets should be treated as separate problems, due to the E(NPV) and CVaR normalization parameters being individual for each target.

4. Discussion

To evaluate and compare the VoI of the different inventory approaches, we developed a framework that incorporates stochastic programming and assesses the quality of the solution derived from the inventory data. The inventory approaches contained various errors in the number and dimensions of tree stems. With this information, we were able to determine the cost efficiency of the forest information. The efficiency strongly depends on the management preferences of the decision-maker. In this case, we explored the trade-off between maximizing economic profitability and minimizing the risk of negative deviations from a stated periodic income target, measured using the CVaR.

Through an analysis of the VoI obtained with different inventory approaches, we could assess the usefulness of the given inventory data for decision-making. The VoI studied in the framework of stochastic programming is often expressed as the expected value of perfect information (EVPI), which assumes the availability of accurate information at some time in the future (Chapter 4 of Birge and Louveaux, 2011). For this study, we calculated the difference between less accurate information obtained with inventory approaches and the reference dataset as the VoII. The reference data we used were obtained using a harvester with precise positioning, which produced high-quality (but not perfect) information. To reflect these higher-quality data, we included a much smaller source of uncertainty for this dataset.

The sources of potential errors in the predictions of each inventory approach varied. Regression models used when predicting forest attributes using ABA-MV and ABA-DD could lead to overestimation of small values and underestimation of large values, since models tend to flatten the regression line (Ståhl et al. 2024). For the ITC approach, a single tree is allowed for each crown segment, leading to understory trees to be undetected (Solberg et al. 2006). Semi-ITC compensates for this by accounting for trees otherwise missed. The forest values predicted by studied inventory approaches were used as input for the forest simulator. The simulator consists of several regression models predicting amongst other aspects growth, ingrowth, and mortality. Due to the interactions, it is difficult to assess how the errors from these models propagate for the different inventory approaches. However, for ABA-DD and Semi-ITC having both a similar and close fit to the reference data, the errors should propagate in the same way.

The stochastic programming model applied can be interpreted from a risk preference perspective. In this case, we assessed risk as the negative deviations from a specific periodic income target, allowing the decision-maker to determine their risk-aversion preferences to obtain the specific periodic income. The parameter λ in the objective function (Equation 1) allows the evaluation of a risk-averse solution (minimizing $CVaR$, $\lambda = 0$), a risk-neutral solution (maximizing $\mathbb{E}(NPV)$, $\lambda = 1$), or any permutations between these extreme cases. Our optimization model utilizes the outcomes of these extreme cases as normalization factors. In earlier work, Eyvindson and Cheng (2016) applied a similar risk coefficient without normalization. This created the perception of linearly increasing VoI with the increasing value of the risk coefficient. However, when the CVaR could not be further reduced, the actual outcomes remained unchanged, and the improvement in the VoI resulted solely from an increase in the value of the risk coefficient.

Using the CVaR to meet the stated periodic income targets was demonstrated to be an effective way of including a soft even-flow constraint. The decision-maker could interpret the average periodic value of the CVaR (Fig. 4) as an average periodic loss (i.e., negative deviation from the target) for the worst scenarios. The use of the CVaR can be compared to alternative downside risk measures. For instance, Eyvindson and Kangas (2016) evaluated the total negative deviations from the periodic targets, using shadow prices to weigh the importance of the timeliness of the income. The shadow prices were used to reflect a risk-neutral decision-maker; however, they were estimated from the deterministic equivalent. This could lead to inappropriate suggestions for risk-neutral decision-makers.

The results of this study highlighted that risk-aversion preferences and the stated periodic income targets of the decision-maker would determine the choice of an inventory approach (Fig. 3). When the focus of the decision-maker was to maximize the $\mathbb{E}(NPV)$, ABA-DD, ITC, and semi-ITC performed equally well, while ABA-MV demonstrated much poorer results. When the decision-maker's focus was to minimize the CVaR at low- and middle-level targets, there were no large differences between the approaches. However, for the highest targets, semi-ITC and ABA-DD were preferred. Substantial differences in the VoII were observed between the extreme values of λ, demonstrating that semi-ITC and ABA-DD should provide the best solutions when balancing the objectives.

The results for the extreme case with solely NPV maximization were in accordance with previous work by Bergseng et al. (2015), highlighting that ITC, semi-ITC, and ABA-DD outperform ABA-MV in terms of potential economic losses due to sub-optimal management decisions. As pointed out by Kangas et al. (2018), this must result from additional error that was introduced with the diameter distribution simulation step for ABA-MV. This is because other inventory approaches provide tree lists as input for the forest simulator, whereas ABA-MV provides only mean values per stand. Consequently, the simulator must first simulate the tree list based on the mean values before simulating the development of the forest holding, which likely will not reflect the actual diameter distribution.

The lack of differences between the inventory approaches observed in the extreme case of solely minimizing the CVaR with low targets can be attributed to the size of the specified targets. When the target is achievable according to the specific applied dataset, the model can still achieve the periodic even-flow income. This was the case for ITC, when the data severely underestimated the timber volume in the forest holding. Therefore, to simply assure consistent periodic incomes, less precise inventory approaches could be sufficient. However, such approaches may not guarantee the maximum possible income and optimal NPV, as shown in Fig. 4.

Planning for a single management objective is likely too simplistic for most forest planning cases. Multi-objective forest management could ensure the increased simultaneous production of multiple ecosystem services (Díaz-Yáñez et al. 2021). However, this requires decision-makers to be able to state management goals and risk preferences, which might be very challenging. For cases when decision-makers are unsure of their preferences, there can be substantial value in providing flexibility in the decision-making process. Thus, the availability of more precise inventory data provides decision-makers with increased flexibility, ensuring that even if objectives or risk preferences change, the data will provide high-quality decision support.

The best overall results were observed for semi-ITC and ABA-DD. Both approaches have been reported to provide inventory information with high accuracy, with semi-ITC providing slightly better results (Rahlf et al. 2015; Kandare et al. 2017). However, the inventory costs for the two approaches are substantially different. Single-tree approaches benefit from high-point-density ALS data (10 points m^{-2} in comparison to 1 point m^{-2} used for ABAs), causing the acquisition costs to increase by nearly two times compared to area-based approaches. Therefore, given the difference in the inventory costs and the small difference in the VoII for forest planning, ABA-DD seems to be favorable.

The focus of this study has been to evaluate the improvements various data acquisition approaches provide to the decision-maker. An improved data quality improves the solution quality; however, the specific properties of the data lead to more appropriate applications in specific use cases. Our assessments were not compared to perfect information but rather compared to the best possible data available. This allowed a direct assessment of the inventory approaches while recognizing the potential for further data improvement.

Further avenues of research are possible to assess the required data quality for effective decision-making. This work only explored timber harvesting operations and did not take other ecosystem services into consideration. The incorporation of additional ecosystem services and the quantification of their associated uncertainties are currently of high interest to forest stakeholders (De Pellegrin Llorente et al. 2023). By including those ecosystem services, we could ensure that the proposed methodology effectively identifies the best data collection approach, enabling the design of a management plan that will meet diverse environmental and societal needs. Furthermore, our assessment only accounted for uncertainty in the initial inventory data, without considering other sources of uncertainty, such as growth models and future timber prices. For instance, growth model uncertainty can be incorporated into the simulation process using an auto-regressive function (Pietilä et al. 2010), and timber price uncertainties could be estimated based on approximation from historical datasets with the use of forecasting approaches (Diebold 2001). In our study, due to the acquisition of the reference data by harvester during harvesting operations, the inventory consisted of mature forest stands. This may have been a limiting factor in the choice of the optimal management option. Consequently, future research should focus on forest holdings with a more diversified age structure. Finally, the future development of the proposed VoII assessment method should incorporate the option to transform it into an interactive assessment tool. This tool would assist decision-makers in evaluating whether gathering more precise data is necessary to achieve their specific management goals.

5. Conclusions

The assessment of the value of information of data acquired using different inventory approaches offers valuable insight into the usefulness of the data for effective forest management. A reference dataset used for comparing inventory approaches does not have to provide perfect data; it simply needs to provide data of higher quality in comparison to the existing information. Stochastic programming is an effective tool for VoI analysis that allows the incorporation of different uncertainty scenarios into the decision-making process. By simultaneously considering the two management goals of NPV maximization and CVaR minimization with the inclusion of the decision-maker's risk preference parameter, it is possible to obtain optimal solutions for various risk-aversion levels. Less accurate inventory approaches may offer sufficient data quality for decision-making when the management objective is focused on one extreme goal only. However, more precise methods could guarantee more flexibility for the decision-maker, ensuring optimal management decisions when multiple goals and different risk aversion preferences are considered.

Author statements:

Competing interests statement

The authors declare that there are no competing interests.

Author contribution statement

Conceptualization: ON, TG, KE; Data curation: ON, LN, TG, KE; Formal analysis: ON, KE; Funding acquisition: TG; Investigation: ON, LN, TG, KE; Methodology: ON, LN, KE; Software: ON, KE; Supervision: TG, KE; Visualization: ON; Writing – original draft: ON; Writing – review and editing: ON, LN, TG, KE.

Data availability statement

Data used in this study are available upon reasonable request.

Funding statement

This work was supported by the Research Council of Norway under the project SmartForest [project no. 309671].

References

Bergseng, E., Ørka, H.O., Næsset, E., and Gobakken, T. 2015. Assessing forest inventory information obtained from different inventory approaches and remote sensing data sources. Ann. For. Sci. **72**(1): 33–45. doi:10.1007/s13595-014-0389-x.

Bettinger, P., Boston, K., Siry, J.P., and Grebner, D.L. 2017. Forest management and planning. 2nd ed. Academic Press, London.

Birge, J.R., and Louveaux, F. 2011. Introduction to stochastic programming. 2nd ed. Springer, New York.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., and Solberg, S. 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sens. Environ. **114**(4): 911–924. doi:10.1016/j.rse.2009.12.004.

Burkhart, H.E., Stuck, R.D., Leuschner, W.A., and Reynolds, M.R. 1978. Allocating inventory resources for multiple-use planning. Can. J. For. Res. **8**(1): 100–110. doi:10.1139/X78-017.

Crookston, N.L., and Finley, A.O. 2008. yaImpute: An R package for kNN imputation. J. Stat. Softw. **23**(10): 1–16. doi:10.18637/jss.v023.i10.

De Pellegrin Llorente, I., Eyvindson, K., Mazziotta, A., Lämås, T., Eggers, J., and Öhman, K. 2023. Perceptions of uncertainty in forest planning: contrasting forest professionals' perspectives with the latest research. Can. J. For. Res. **53**(6): 391–406. doi:10.1139/cjfr-2022-0193.

Díaz-Yáñez, O., Pukkala, T., Packalen, P., Lexer, M.J., and Peltola, H. 2021. Multi-objective forestry increases the production of ecosystem services. For. Int. J. For. Res. **94**: 386–394. doi:10.1093/forestry/cpaa041.

Diebold, F.X. 2001. Elements of forecasting. 2nd ed. South-Western, Thomson Learning, Cincinnati, Ohio.

Duffie, D., and Pan, J. 1997. An overview of value at risk. J. Deriv. **4**(3): 7–49. doi:10.3905/jod.1997.407971.

Duvemo, K., Barth, A., and Wallerman, J. 2007. Evaluating sample plot imputation techniques as input in forest management planning. Can. J. For. Res. **37**(11): 2069–2079. doi:10.1139/X07-069.

Eid, T., Gobakken, T., and Næsset, E. 2004. Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. Scand. J. For. Res. **19**(6): 512–523. doi:10.1080/02827580410019463.

Eyvindson, K., and Cheng, Z. 2016. Implementing the conditional value at risk approach for even-flow forest management planning. Can. J. For. Res. **46**(5): 637–644. doi:10.1139/cjfr-2015-0270.

Eyvindson, K., and Kangas, A. 2015. Evaluating the required scenario set size for stochastic programming in forest management planning: incorporating inventory and growth model uncertainty. Can. J. For. Res. **46**(3): 340–347. doi:10.1139/cjfr-2014-0513.

Eyvindson, K., and Kangas, A. 2016. Integrating risk preferences in forest harvest scheduling. Ann. For. Sci. **73**(2): 321–330. doi:10.1007/s13595-015-0517-2.

Eyvindson, K.J., Petty, A.D., and Kangas, A.S. 2017. Determining the appropriate timing of the next forest inventory: incorporating forest owner risk preferences and the uncertainty of forest data quality. Ann. For. Sci. **74**(1). doi:10.1007/s13595-016-0607-9.

Gertner, G.Z., and Dzialowy, P.J. 1984. Effects of measurement errors on an individual tree-based growth projection system. Can. J. For. Res. **14**(3): 311–316. doi:10.1139/x84-057.

Gobakken, T., and Næsset, E. 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. Scand. J. For. Res. **19**(6): 529–542. doi:10.1080/02827580410019454.

Gobakken, T., and Næsset, E. 2005. Weibull and percentile models for lidar-based estimation of basal area distribution. Scand. J. For. Res. **20**(6): 490–502. doi:10.1080/02827580500373186.

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

Hyyppä, J., and Inkinen, M. 1999. Detecting and estimating attributes for single trees using laser scanner. Photogramm. J. Finland, **16**: 27–42.

Johnson, K.N., and Scheurman, H.L. 1977. Techniques for prescribing optimal timber harvest and investment under different objectives—discussion and synthesis. For. Sci. **23**(suppl_1): a0001-z0001. doi:10.1093/forestscience/23.s1.a0001.

Kandare, K., Dalponte, M., Ørka, H.O., Frizzera, L., and Næsset, E. 2017. Prediction of species-specific volume using different inventory approaches by fusing airborne laser scanning and hyperspectral data. Remote. Sens. **9**(5). doi:10.3390/rs9050400.

Kangas, A., Gobakken, T., Puliti, S., Hauglin, M., and Næsset, E. 2018. Value of airborne laser scanning and digital aerial photogrammetry data in forest decision making. Silva Fenn. **52**(1). doi:10.14214/sf.9923.

Kangas, A., and Maltamo, M. 2000. Percentile based basal area diameter distribution models for Scots pine, Norway spruce and birch species. Silva Fenn. **34**(4). doi:10.14214/sf.619.

Kangas, A.S. 2010. Value of forest information. Eur. J. For. Res. **129**(5): 863–874. doi:10.1007/s10342-009-0281-7.

Ketzenberg, M.E., Rosenzweig, E.D., Marucheck, A.E., and Metters, R.D. 2007. A framework for the value of information in inventory replenishment. Eur. J. Oper. Res. **182**(3): 1230–1250. doi:10.1016/j.ejor.2006.09.044.

Kilkki, P., Maltamo, M., Mykkänen, R., and Päivinen, R. 1989. Use of the Weibull function in estimating the basal area dbh-distribution. Silva Fenn. **23**(4). doi:10.14214/sf.a15550.

King, A.J., and Wallace, S.W. 2012. Modeling with stochastic programming. Springer, New York.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. **28**(5): 1–26. doi:10.18637/jss.v028.i05.

Lumley, T. 2004. The leaps package for regression subset selection. R package version 2.9.

Mäkinen, A., Kangas, A., and Mehtätalo, L. 2010. Correlations, distributions, and trends in forest inventory errors and their effects on forest planning. Can. J. For. Res. **40**(7): 1386–1396. doi:10.1139/X10-057.

McRoberts, R.E., Næsset, E., and Gobakken, T. 2015. Optimizing the k-nearest neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. Remote. Sens. Environ. **163**: 13–22. doi:10.1016/j.rse.2015.02.026.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote. Sens. Environ. **80**(1): 88–99. doi:10.1016/S0034-4257(01)00290-5.

Noordermeer, L., Ørka, H.O., and Gobakken, T. 2023. Imputing stem frequency distributions using harvester and airborne laser scanner data: a comparison of inventory approaches. Silva Fenn. **57**(3). doi:10.14214/sf.23023.

Pasalodos-Tato, M., Mäkinen, A., Garcia-Gonzalo, J., Borges, J.G., Lämås, T., and Eriksson, L.O. 2013. Assessing uncertainty and risk in forest planning and decision support systems: review of classical methods and introduction of innovative approaches. For. Syst. **22**(2): 282–303. doi:10.5424/fs/2013222-03063.

Pietilä, I., Kangas, A., Mäkinen, A., and Mehtätalo, L. 2010. Influence of growth prediction errors on the expected losses from forest decisions. Silva Fenn. **44**(5): 829–843. doi:10.14214/SF.111.

Pukkala, T. 2005. Metsikön tuottoarvon ennustemallit kivennäismaan männiköille, kuusikoille ja rauduskoivikoille. Metsätieteen aikakauskirja: 311–322. [In Finnish]

Rahlf, J., Breidenbach, J., Solberg, S., and Astrup, R. 2015. Forest parameter prediction using an image-based point cloud: a comparison of semi-ITC with ABA. Forests, **6**(11): 4059–4071. doi:10.3390/f6114059.

Rasinmäki, J., Mäkinen, A., and Kalliovirta, J. 2009. SIMO: an adaptable simulation framework for multiscale forest resource data. Comput. Electron. Agric. **66**(1): 76–84. doi:10.1016/j.compag.2008.12.007.

Rockafellar, R.T., and Uryasev, S. 2000. Optimization of conditional value-at-risk. J. Risk, **2**: 21–42.

Siipilehto, J. 1999. Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fenn. **33**(4): 281–301. doi:10.14214/sf.650.

Siitonen, M. 1993. Experiences in the use of forest management planning models. Silva Fenn. **27**(2): 167–178. doi:10.14214/sf.a15670.

Solberg, S., Naesset, E., and Bollandsas, O.M. 2006. Single Tree Segmentation Using Airborne Laser Scanner Data in a Structurally Heterogeneous Spruce Forest. Photogramm. Eng. Remote Sens. **72**(12): 1369–1378. doi:10.14358/PERS.72.12.1369.

Ståhl, G., Gobakken, T., Saarela, S., Persson, H.J., Ekström, M., Healey, S.P., Yang, Z., Holmgren, J., Lindberg, E., Nyström, K., Papucci, E., Ulvdal, P., Ørka, H.O., Næsset, E., Hou, Z., Olsson, H., and McRoberts, R.E. 2024. Why ecosystem characteristics predicted from remotely sensed data are unbiased and biased at the same time – and how this affects applications. For. Ecosyst. **11**: 100164. doi:10.1016/j.fecs.2023.100164.

Appendix: Inventory data collection and predictions

A1. Harvester data

The harvester was equipped with a real-time kinematic Global Navigation Satellite System (GNSS) that gave cm-accurate positioning. The GNSS, in conjunction with sensor hardware that measured the crane position, recorded the harvested trees' locations with a planimetric accuracy of approximately 1 meter (for details, see Noordermeer et al. (2021)). Additionally, the harvester data contained data on species, diameter at breast height (dbh), and the stem profile. Because stem profiles only contained data on the processed part of the stem, the total tree height was predicted from the stem profile using taper models (Hansen et al. 2023).

Polygons of harvested areas were generated as unary unions of buffers around the positions of harvested stems. The polygons were then tessellated into regular grid cells of 250 m^2, and stands were defined as clusters of grid cells with a minimum total size of 0.2 ha. The minimum size of 0.2 ha conformed to the typical minimum size of forest stands in commercial Norwegian forest inventories.

A2. ALS surveys

The harvested stands were covered by four ALS surveys which were acquired in 2013, 2016, 2017 and 2019 using different instruments and acquisition parameters. All ALS data were acquired under leaf-on conditions. Flying altitudes ranged from 1150-2900 m, footprint diameters from 0.25-0.73 m, and pulse densities from 5-30 m^{-2}. Different areas were covered by the ALS data, which overlapped in some places. The elapsed time between harvesting and ALS acquisition varied from zero to eight years, with a mean of five years. The raw ALS data were processed, whereby laser echoes were classified as ground or non-ground by the contractors Blom Geomatics AS and Terratec AS. For each harvesting operation, the most recently acquired ALS data were used for tree crown segmentation and computation of ALS metrics (for details, see below). Digital terrain models were constructed as triangulated irregular networks (TIN) from the laser echoes classified as ground, using the lidR package

(Roussel et al. 2020) in R. The ALS data were then normalized by computing the height relative to the terrain height of the TIN for all echoes.

A2.1. Tree crown segmentation

Canopy height models were generated for all stands as rasterized values of maximum ALS height with a spatial resolution of 0.5 m. Trees in the canopy height models were located using a local maximum filter following methods demonstrated by Popescu et al. (2003). Trees were segmented from the canopy height model using a segmentation algorithm proposed by Dalponte and Coomes (2016). The height and coordinates (x,y,z) of the highest laser echo within each crown segment were computed. The tree crown polygons represented the spatial extents of the tree crowns and included the coordinates of the treetops that were detected in the ALS data.

A2.2. Airborne laser scanner metrics

ALS echoes were extracted from within ABA grid cells and tree crown segments and used to compute ALS metrics from all echoes. ALS metrics, including the maximum height (Hmax), mean height (Hmean), standard deviation (Hsd), skewness (Hskew), and kurtosis (Hkurt) were computed. The percentage of echoes above the mean height and 2 m (Habovemean, Habove2, respectively) and the normalized Shanon diversity index (Hentropy, van Ewijk et al. 2011) were also computed. Echo heights at the 10th, 20th, …, and 90th percentiles of the height distributions (H10, …, H90) as well as the 95th percentile (H95) were computed. Lastly, the height range between the lowest canopy point >2 m and H95 was divided into 10 fractions of equal height, and canopy density metrics (D0, D1, …, D9) were computed as the proportion of echoes above each fraction divided by the total number of echoes.

A3. Prediction of stand attributes

Stand attributes were predicted following the methods of Noordermeer et al. (2023). For the ABA-DD, ITC, and semi-ITC approaches, tree lists were imputed for each target observation using nearest neighbor imputation, while for the ABA-MV approach, values of N, Dgm and Hgm were imputed. Three-fold cross-validation was used to obtain predictions for all stands.

Distance matrices were used to identify the k nearest neighbors for each target observation based on selected ALS metrics and the Euclidean distance metric. The similarities between each target observation and the selected k nearest neighbors were computed as the inverse of the distance matrices. For the ITC, semi-ITC and ABA-DD approaches, stem frequencies within diameter classes of 2,4, …, 80 cm were imputed for each target unit and for each tree species separately following the methods described by Packalén & Maltamo (2008). Then for each diameter class, the corresponding tree height was predicted using a height-diameter model with the form of the Näslund function (Näslund 1936). For the ABA-MV approach, species-specific values of N, Dgm and Hgm were imputed for each target observation as averages of corresponding values of nearest neighbors, weighted with the similarity based on the Euclidean distance. Imputed values of N were then summed for each species within each stand. Stand-level mean values of Dgm and Hgm were computed for each species as corresponding mean values imputed for all grid cells, weighted by the proportion of ALS echoes with a height >2 m).

In a preliminary analysis, possible co-location errors between harvester and ALS data were found, where maximum ALS heights within tree crown segments and ABA grid cells differed substantially from the heights of harvested trees located within those tree crown segments and ABA grid cells. Such observations were excluded from the reference data prior to the imputation. Specifically, a simple linear regression model was used, with the maximum tree height from the harvester data as the response variable and the 95th percentile of ALS height as the predictor variable, to label those observations with a Cooks distance > 0.5 of the mean Cooks distance as erroneous. Observations labeled as erroneous were only removed from the reference data, not the target data, before fitting the kNN model in each fold.

References

Dalponte, M., and Coomes, D.A. 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods Ecol. Evol. **7**(10): 1236–1245. doi:10.1111/2041-210X.12575.

van Ewijk, K.Y., Treitz, P.M., and Scott, N.A. 2011. Characterizing Forest Succession in Central Ontario using Lidar-derived Indices. Photogramm. Eng. Remote Sens. **77**(3): 261–269. doi:10.14358/PERS.77.3.261.

Hansen, E., Rahlf, J., Astrup, R., and Gobakken, T. 2023. Taper and volume models for spruce, pine and birch in Norway. Scand. J. For. Res. **38**(6): 413–428. doi:10.1080/02827581.2023.2243821.

Näslund, M. 1936. Thinning experiments in pine forest conducted by the forest experiment station. Meddelanden fran Statens Skogsforsöksanstalt **29**: 1–169. [In Swedish]

Noordermeer, L., Sørngård, E., Astrup, R., Næsset, E., and Gobakken, T. 2021. Coupling a differential global navigation satellite system to a cut-to-length harvester operating system enables precise positioning of harvested trees. Int. J. For. Eng. **32**(2): 119–127. doi:10.1080/14942119.2021.1899686.

Noordermeer, L., Ørka, H.O., and Gobakken, T. 2023. Imputing stem frequency distributions using harvester and airborne laser scanner data: a comparison of inventory approaches. Silva Fenn. **57**(3). doi:10.14214/sf.23023.

Packalén, P., and Maltamo, M. 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. Can. J. For. Res. **38**(7): 1750–1760. doi:10.1139/X08-037.

Popescu, S.C., Wynne, R.H., and Nelson, R.F. 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can. J. Remote Sens. **29**(5): 564–577. doi:10.5589/M03-027.

Roussel, J.R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R.H., Meador, A.S., Bourdon, J.F., de Boissieu, F., and Achim, A. 2020. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. Remote Sens. Environ. **251**: 112061. doi:10.1016/J.RSE.2020.112061.