Biometrics, 2024, 80(3), ujae104
https://doi.org/10.1093/biomtc/ujae104
Biometric Practice

OXFORD

Planning cost-effective operational forest inventories

Santeri Karppinen[1], Liviu Ene[2], Lovisa Engberg Sundström[2], Juha Karvanen[1,*]

[1]Department of Mathematics and Statistics, University of Jyvaskyla, Seminaarinkatu 15, Jyvaskyla 40014, Finland, [2]Skogforsk, Uppsala Science Park, Uppsala SE-751 83, Sweden

[*]Corresponding author: Juha Karvanen, Department of Mathematics and Statistics, University of Jyvaskyla, Seminaarinkatu 15, Jyvaskyla 40014, Finland (juha.t.karvanen@jyu.fi).

ABSTRACT

We address a Bayesian two-stage decision problem in operational forestry where the inner stage considers scheduling the harvesting to fulfill demand targets and the outer stage considers selecting the accuracy of pre-harvest inventories that are used to estimate the timber volumes of the forest tracts. The higher accuracy of the inventory enables better scheduling decisions but also implies higher costs. We focus on the outer stage, which we formulate as a maximization of the posterior value of the inventory decision under a budget constraint. The posterior value depends on the solution to the inner stage problem and its computation is analytically intractable, featuring an NP-hard binary optimization problem within a high-dimensional integral. In particular, the binary optimization problem is a special case of a generalized quadratic assignment problem. We present a practical method that solves the outer stage problem with an approximation which combines Monte Carlo sampling with a greedy, randomized method for the binary optimization problem. We derive inventory decisions for a dataset of 100 Swedish forest tracts across a range of inventory budgets and estimate the value of the information to be obtained.

KEYWORDS: Bayesian modeling; decision making; forestry; quadratic assignment problem; scheduling; value of information.

1 INTRODUCTION

Decisions on harvesting are central in forestry. The scope of decision making varies from the strategic level, where the time scale is decades, to the optimal cutting of a single stem (Kivinen, 2004; Ulvdal et al., 2023). In operational short-term forest planning, the starting point is a set of forest tracts, that is, a set of forest areas that a forestry company has already contracted to be cut in the near future, for instance, in the next 6 months. In this setting, forest growth is irrelevant because the planning horizon is short. Ultimately, the decision of interest is to choose a schedule for the harvesting of the tracts such that the monthly demand for timber can be met as precisely as possible. Both exceeding and subceeding the demand targets cause losses to the industry.

The timber volume available in the tracts is uncertain, which brings the decision problem into the realm of statistics. Before making the harvest scheduling decision, the uncertainty can be reduced by carrying out a forest inventory where the timber volume is estimated with a chosen accuracy (the inventory decision) using field surveys (Nieuwenhuis et al., 1999) and remotely sensed data (Vauhkonen et al., 2014; Siipilehto et al., 2016). A more accurate inventory leads to better scheduling decisions but also costs more because a larger sample size is needed (Gregoire et al., 2016). Although the problem of choosing an inventory method in industrial wood procurement is conceptually described already by Ståhl (1994, p. 27–28), we are not aware of any previous works with practical solutions.

In this work, we develop a Bayesian approach for cost-effective operational (short-term) forest inventory planning. We formulate the inventory decision problem as an optimization problem that involves the maximization of the posterior value of the inventory decision under a budget constraint. The concept of posterior value, that is, the expected value of the new data to be collected, has a crucial role in the inventory decision problem. The difference of the posterior value and the prior value, that is, the expected value before collecting new data, is called value of information (cf. Eidsvik et al., 2015). In forestry, value of information and related concepts have been applied to long-term planning from the perspective of a forest owner (Ståhl, 1994; Duvemo et al., 2014; Kangas et al., 2014; 2015).

Computing the posterior value leads to solving a generalized quadratic assignment problem (GQAP) (Lee and Ma, 2004) which is known to be NP-hard (Cook, 1971). We propose algorithms that employ greedy heuristics and Monte Carlo sampling to find sufficiently good solutions for operational use.

Our approach for inventory decision planning allows for taking into account the prior uncertainties in the timber volume, as well as the uncertainties in the measurements carried out in a forest inventory. The primary output of our method is a plan on how to cost-effectively select tract inventory accuracies under uncertainty. Our practical example illustrates this for 100 Swedish forest tracts.

We begin in the next section by explaining the inventory decision problem of interest in more detail and providing guidance for the rest of the paper. Web Appendix 1 contains a table of all notation used in this paper.

Received: November 20, 2023; Revised: July 6, 2024; Accepted: September 5, 2024

2 OVERVIEW OF THE DECISION PROBLEM

We consider a forestry company that has the right to cut n_S forest tracts that each have n_A timber assortments. The company has some prior (pre-inventory) information on the timber volumes $\mathbf{V}$ of the tracts in the form of an $(n_S \times n_A)$-dimensional prior probability distribution $p_{\mathbf{V}}$. Furthermore, the company has the opportunity to improve the knowledge on $\mathbf{V}$ by making a decision $\mathbf{x}^{(I)}$ on a forest inventory. The result of $\mathbf{x}^{(I)}$ is an inventory dataset $\mathbf{Y} \mid \mathbf{x}^{(I)}$ following an $(n_S \times n_A)$-dimensional probability distribution $p_{\mathbf{Y}|\mathbf{x}^{(I)}}$, which integrates over the prior $p_{\mathbf{V}}$ and a measurement probability distribution $p_{\mathbf{Y}|\mathbf{V},\mathbf{x}^{(I)}}$. The decision vector $\mathbf{x}^{(I)}$ encodes the selection of an inventory method for each tract among n_I predefined inventory methods of varying accuracy. The cost of applying inventory method i to tract s is $C_{s,i}$, a known constant.

Under this setting, we consider the cost-effective maximization of the posterior value of the inventory decision $\mathbf{x}^{(I)}$ which leads us to model a two-stage decision-making problem where an "inner" problem is nested inside an "outer" problem. The inner problem considers scheduling the harvests of the forest tracts in order to meet a known demand $D_{a,t}$ for each assortment a at (future) times $t = 1, 2, \ldots, n_T$, and the outer problem considers the cost-effective selection of $\mathbf{x}^{(I)}$ under a budget constraint, B. The outer and inner problems are linked, since the solution of the inner problem depends on the probabilistic knowledge available on the volumes $\mathbf{V}$, which the company has the opportunity to improve using $\mathbf{Y} \mid \mathbf{x}^{(I)}$ obtained from the outer decision problem.

Even though we model the full two-stage problem, we focus on solving the outer problem. This is because the solution of the inner problem is dependent on which $\mathbf{Y} \mid \mathbf{x}^{(I)}$ is realized. In the planning of $\mathbf{x}^{(I)}$, $\mathbf{Y}$ will necessarily be unknown and random, and therefore a sensible, data-informed solution for the inner problem can only be obtained once (the chosen) $\mathbf{x}^{(I)}$ has been carried out in practice, and a real-world realization $\mathbf{y}$ of the random variable $\mathbf{Y} \mid \mathbf{x}^{(I)}$ collected. In the computational method that selects the inventory decision, however, the inner problem is solved for multiple realizations $\mathbf{Y} \mid \mathbf{x}^{(I)}$ for each candidate decision $\mathbf{x}^{(I)}$. In effect, this evaluates the fitness of a given $\mathbf{x}^{(I)}$ on average, but does not provide a single, concrete solution to the inner problem. We return to this in the discussion.

The rest of the paper is organized as follows. Section 3 introduces the Swedish tract dataset used in our practical illustration. The models for the prior volume distribution $p_{\mathbf{V}}$ and the measurement distribution $p_{\mathbf{Y}|\mathbf{V},\mathbf{x}^{(I)}}$ are described in Section 4. The inventory decision problem is formally presented as an optimization problem in Section 5. In Section 6, we present a practical method for solving the inventory decision problem, and study its performance in simulations. Finally, we apply the method to the Swedish tract dataset in Section 7 and conclude with a discussion in Section 8.

3 DATA

The dataset for the practical illustration is a random sample of a larger dataset and encompasses $n_S = 100$ forest tracts in Southern Sweden. These tracts have been in fact clear-cut but here we consider the pre-harvest situation with three inventory methods ($n_I = 3$), three assortments (pine, spruce, or deciduous trees, $n_A = 3$) and a 6 months' harvest scheduling horizon ($n_T = 6$). Letting $s = 1, 2, \ldots, n_S$, $a = 1, 2, \ldots, n_A$, $i = 1, 2, \ldots, n_I$, and $t = 1, 2, \ldots, n_T$ index tracts, assortments, inventory methods and time points, respectively, the data consist of

1. the prior means $\mu^0_{a,s}$ and variances $(\sigma^0_{a,s})^2$ of timber volumes for each assortment and tract,
2. the measurement variances $\sigma^2_{a,s,i}$ for the inventory methods $i = 1, 2, 3$ for each assortment and tract,
3. the costs $C_{s,i}$ of applying inventory method i in tract s,
4. the inventory budget B that the inventory costs cannot exceed, and
5. the demand targets $D_{a,t}$ for each assortment and month.

The prior means and variances are obtained from a previously developed imputation model, which was fitted using data on clear-cut tracts where both the true timber volumes and the characteristics of the tracts are known. A more detailed description of the data and the process is given in Web Appendix 2.

The measurement variances $\sigma^2_{a,s,i}$ of the three inventory methods ($i = 1, 2, 3$) are also obtained using a previously developed model. The inventory methods are characterized by the number of fixed-area field plots that are used in the estimation of timber volume (5, 10, and 20 plots per forest tract, respectively). The model is based on data on clear-cut tracts where the location and the volume of each tree have been recorded by the harvester. These data were used to simulate inventories in-silico where the timber volume of the trees inside a plot is estimated without cutting the trees. More details are given in Web Appendix 3.

The costs $C_{s,i}$ were assumed to be constant over the set of forest tracts so that $C_{s,i} = C_i$. We assume costs $C_1 = 100$, $C_2 = 150$, and $C_3 = 250$ that consist of a fixed cost of 50 monetary units and the cost of measuring 5, 10, and 20 plots, respectively, at a cost of 10 monetary units per plot. The inventory budget B will be varied from 10 000 to 25 000 in increments of 1000.

The demand targets are set by the industry in the real world but in the practical illustration they are obtained by setting $D_{a,t} = \left(\sum_{s=1}^{n_S} \mu^0_{a,s}\right)/n_T$ for each assortment a and time t. This yields for all 6 months considered the demand targets 1520.11, 7495.42, and 481.26 for pine, spruce, and deciduous trees, respectively.

4 STATISTICAL MODELS FOR ASSORTMENT VOLUMES AND INVENTORY MEASUREMENTS

Next, we detail the prior distribution of the timber volumes, $p_{\mathbf{V}}$, and the measurement model $p_{\mathbf{Y}|\mathbf{V},\mathbf{x}^{(I)}}$ in the description of Section 2 and discuss their estimation.

In general, we use $V_{a,s}$ to refer to the volume of timber assortment (eg, pine, spruce, or deciduous trees) $a \in \{1, 2, \ldots, n_A\}$, in tract $s \in \{1, 2, \ldots, n_S\}$, and denote the corresponding (random) measurement by $Y_{a,s}$. Furthermore, we will use $\mathbf{V}$ and $\mathbf{Y}$, respectively, to denote the volumes and measurements of all assortments in all tracts (stacked to a random vector in some order that plays no particular role).

We model the prior (pre-inventory) knowledge on the (univariate) volume $V_{a,s}$ by the log-normal distribution

$$V_{a,s} \sim \log\text{-}N(\nu_{a,s}^{(p)}, \lambda_{a,s}^{(p)}), \tag{1}$$

where $\nu_{a,s}^{(p)}$ and $\lambda_{a,s}^{(p)}$ are the parameters of the log-normal distribution.

The distribution of the measurement $Y_{a,s}$, given true volume $v_{a,s}$ and inventory method i, is likewise modeled by a log-normal distribution:

$$Y_{a,s} \mid (V_{a,s} = v_{a,s}, x_s^{(I)} = i) \sim \log\text{-}N(\nu_{a,s,i}^{(m)}, \lambda_{a,s,i}^{(m)}), \tag{2}$$

with parameters $\nu_{a,s,i}^{(m)}$ and $\lambda_{a,s,i}^{(m)}$, and with $x_s^{(I)} \in \{1, 2, \ldots, n_I\}$ standing for the inventory decision applied to tract s. We estimate the parameters $\nu_{a,s}^{(p)}$, $\lambda_{a,s}^{(p)}$, $\nu_{a,s,i}^{(m)}$, and $\lambda_{a,s,i}^{(m)}$ in (1) and (2) using the method of moments, by matching the mean of (1) and the variances of (1) and (2) to the data values $\mu_{a,s}^0$, $(\sigma_{a,s}^0)^2$, and $\sigma_{a,s,i}^2$ that were discussed in Section 3. The details are given in Web Appendix 4.

A priori, we assume that each volume $V_{a,s}$ is independent of the volumes of other assortments in the same or other tracts. Similarly, the observations $Y_{a,s}$ are conditionally independent given $V_{a,s} = v_{a,s}$ and the inventory decision. We discuss extensions of this model in Section 8.

Finally, Web Appendix 5 also presents a computationally simpler model where the prior (1) and measurement model (2) are replaced with normal distributions instead. While this model can be unrealistic in the sense that it places probability mass on negative observations, it may still be useful with moderate prior and measurement variances.

5 THE FOREST INVENTORY PLANNING PROBLEM

The planning problem considered in this paper can be formulated mathematically as a constrained optimization problem:

$$\begin{aligned}
&\text{maximize} \quad \text{PoV}(\mathbf{x}^{(I)}) \\
&\qquad \text{(posterior value of inventory decision)} \\
&\text{such that} \quad C_{\mathbf{x}^{(I)}} \le B \\
&\qquad \text{(the inventory cost is less than the budget)} \\
&\text{where} \quad \mathbf{x}^{(I)} \in \mathcal{X}^{(I)} \\
&\qquad \text{(the inventory methods are chosen from 'feasible'} \\
&\qquad \text{solutions)},
\end{aligned} \tag{3}$$

with the notation explained below.

The "inventory decision vector" $\mathbf{x}^{(I)}$ is an integer vector with n_S elements. The sth element of $\mathbf{x}^{(I)}$, $x_s^{(I)} \in \{1, 2, \ldots, n_I\}$, indicates the decision of the inventory method for tract s. We assume that the inventory methods are "cost-accuracy consistent":

Definition 1 *Let $I = \{1, 2, \ldots, n_I\}$ be a set of inventory methods and let $C_{s,i}$ denote the cost of applying inventory method i to tract s. The set I is cost-accuracy consistent if:*

(1) $C_{s,i} > C_{s,j}$, $i, j \in I, i \neq j$, implies that for all $s \in \{1, 2, \ldots, n_S\}$ and for all $a \in \{1, 2, \ldots, n_A\}$ it holds $\sigma_{a,s,i}^2 < \sigma_{a,s,j}^2$,

(2) (without loss of generality) the inventory method indices are ordered so that $i > j$, $i, j \in \{1, 2, \ldots, n_I\}$ implies $C_{s,i} > C_{s,j}$ for all $s = \{1, 2, \ldots, n_S\}$.

In other words, more costly inventories are more accurate and the inventory methods have been ordered such that inventory methods with higher index have higher accuracy. The feasible set $\mathcal{X}^{(I)}$ is formed by all inventory decision vectors in the set $\{1, 2, \ldots, n_I\}^{n_S}$ that satisfy the "prior variance order" constraint:

Definition 2 *An inventory decision vector $\mathbf{x}^{(I)} = (x_1^{(I)}, \ldots, x_{n_S}^{(I)})$ follows prior variance order if for any two tracts $s_1, s_2 \in \{1, 2, \ldots, n_S\}$, $s_1 \neq s_2$, it holds $x_{s_1}^{(I)} \ge x_{s_2}^{(I)}$ when $\sigma_{s_1}^0 \ge \sigma_{s_2}^0$, where $\sigma_s^0 := \sqrt{\sum_{a=1}^{n_A} (\sigma_{a,s}^0)^2}$.*

This constraint is an assumption of "rationality" of the cost-effective forest inventories: for two inventory methods and any two tracts, the inventory that is more accurate can only be carried out for the tract with higher total prior uncertainty. By the independence of the tracts and assortments for (1), the total pre-inventory standard deviation σ_s^0 of the timber volume can be computed as in Definition 2.

We will discuss the posterior value $\text{PoV}(\mathbf{x}^{(I)})$ in Section 5.1. To incorporate the cost-efficiency of the inventory $\mathbf{x}^{(I)}$ to our formulation, we maximize PoV with respect to the fixed (monetary) inventory budget B, which cannot be exceeded by the forest inventory cost:

$$C_{\mathbf{x}^{(I)}} = \sum_{s=1}^{n_S} C_{s,x_s^{(I)}}. \tag{4}$$

5.1 The posterior value of the inventory decision

The posterior value $\text{PoV}(\mathbf{x}^{(I)})$ measures on an average sense the value of the decision $\mathbf{x}^{(I)}$, knowing that the inventory data $\mathbf{Y} \mid \mathbf{x}^{(I)}$ influences the future harvest scheduling decision. Thus, to define $\text{PoV}(\mathbf{x}^{(I)})$, we need to first define the future harvest scheduling problem.

The decision variable for the future harvest scheduling decision is an $n_T \times n_S$ matrix $\mathbf{X}^{(T)}$, with elements $x_{ts}^{(T)} \in \{0, 1\}$, $t = 1, 2, \ldots, n_T$, $s = 1, 2, \ldots, n_S$, denoting whether tract s is clear-cut at time t. The matrix $\mathbf{X}^{(T)}$ is subject to the constraint $\sum_{t=1}^{n_T} x_{ts}^{(T)} \in \{0, 1\}$ for all $s = 1, 2, \ldots, n_S$, since each tract is fully harvested exactly once or not at all. We denote by $\mathcal{X}^{(T)}$ the set of all binary matrices that satisfy this constraint. Furthermore, we use a dot $(\cdot)$ to denote "over all tracts"; the notation $\mathbf{x}_{t\cdot}^{(T)}$, for example, refers to the row t of $\mathbf{X}^{(T)}$.

We model the efficiency of the future harvest scheduling decision in terms of a quadratic utility function:

$$\text{U}(\mathbf{X}^{(T)}, \mathbf{V}) = -\sum_{t=1}^{n_T} \sum_{a=1}^{n_A} \left(\sum_{s=1}^{n_S} x_{ts}^{(T)} V_{a,s} - D_{a,t} \right)^2, \tag{5}$$

which compares the total harvest of assortment a at time t to its industry demand, $D_{a,t}$, and penalizes deviations quadratically. The quadratic form not only facilitates the algorithms presented in Section 6 but also aligns with the notion that the industry can adapt to small deviations from the demand targets but large deviations may cause major difficulties. Note that (5) is a random variable since it depends on the assortment volumes $\mathbf{V}$ which are not known exactly before harvesting.

Consider then, how the inventory decision $\mathbf{x}^{(I)}$ influences the future harvest scheduling decision $\mathbf{X}^{(T)}$ through the data obtained from the inventory decision:

(1) When the inventory decision $\mathbf{x}^{(I)}$ is made, a random inventory dataset $\mathbf{Y} \sim p_{\mathbf{Y}|\mathbf{x}^{(I)}}$ will be obtained. The data $\mathbf{Y}$ are noisy measurements of the uncertain volumes $\mathbf{V}$.
(2) Using the data $\mathbf{Y}$, the knowledge regarding $\mathbf{V}$ can be improved. This amounts to computing the posterior distribution of $\mathbf{V}$ using Bayes' rule:

$$p_{\mathbf{V}|\mathbf{Y},\mathbf{x}^{(I)}}(\mathbf{v} \mid \mathbf{y}, \mathbf{x}^{(I)}) = \frac{p_{\mathbf{Y}|\mathbf{V},\mathbf{x}^{(I)}}(\mathbf{y} \mid \mathbf{v}, \mathbf{x}^{(I)}) p_{\mathbf{V}}(\mathbf{v})}{\int_{\mathbf{V}} p_{\mathbf{Y}|\mathbf{V},\mathbf{x}^{(I)}}(\mathbf{y} \mid \mathbf{v}, \mathbf{x}^{(I)}) p_{\mathbf{V}}(\mathbf{v}) \mathrm{d}\mathbf{v}}.$$

(3) Using the improved knowledge of $\mathbf{V}$, a better future harvest scheduling decision $\mathbf{X}^{(T)}$ can be made. According to Bayesian decision theory (cf. Raiffa and Schlaiffer, 1967; Hirshleifer and Riley, 1979), $\mathbf{X}^{(T)}$ should be chosen to maximize expected utility:

$$\max_{\mathbf{X}^{(T)} \in \mathcal{X}^{(T)}} \mathbb{E}_{\mathbf{V}|\mathbf{Y}=\mathbf{y},\mathbf{x}^{(I)}}[\mathrm{U}(\mathbf{X}^{(T)}, \mathbf{V})] = \max_{\mathbf{X}^{(T)} \in \mathcal{X}^{(T)}} \int_{\mathcal{V}} \mathrm{U}(\mathbf{X}^{(T)}, \mathbf{v}) p_{\mathbf{V}|\mathbf{Y},\mathbf{x}^{(I)}}(\mathbf{v} \mid \mathbf{y}, \mathbf{x}^{(I)}) \mathrm{d}\mathbf{v}.$$

The quantity capturing the logic of the above steps (1)–(3) is the posterior value of the inventory decision:

$$\mathrm{PoV}(\mathbf{x}^{(I)}) = \int_{\mathcal{Y}} \max_{\mathbf{X}^{(T)} \in \mathcal{X}^{(T)}} \left[\int_{\mathcal{V}} U(\mathbf{X}^{(T)}, \mathbf{v}) p_{\mathbf{V}|\mathbf{Y},\mathbf{x}^{(I)}} \times (\mathbf{v} \mid \mathbf{y}, \mathbf{x}^{(I)}) \mathrm{d}\mathbf{v} \right] p_{\mathbf{Y}|\mathbf{x}^{(I)}}(\mathbf{y} \mid \mathbf{x}^{(I)}) \mathrm{d}\mathbf{y}, \quad (6)$$

which contains the outer integral over $p_{\mathbf{Y}|\mathbf{x}^{(I)}}$, since the data $\mathbf{Y}$ obtained in step (1) are random and unknown in the planning of the inventory decision $\mathbf{x}^{(I)}$.

$\mathrm{PoV}(\mathbf{x}^{(I)})$ is related to the value of (imperfect) information (VoI) (cf. Eidsvik et al., 2015) associated with $\mathbf{x}^{(I)}$:

$$\mathrm{VoI}(\mathbf{x}^{(I)}) = \mathrm{PoV}(\mathbf{x}^{(I)}) - \mathrm{PV}. \quad (7)$$

Here, the constant

$$\mathrm{PV} = \max_{\mathbf{X}^{(T)} \in \mathcal{X}^{(T)}} \int_{\mathcal{V}} \mathrm{U}(\mathbf{X}^{(T)}, \mathbf{v}) p_{\mathbf{V}}(\mathbf{v}) \mathrm{d}\mathbf{v} \quad (8)$$

is the prior value, which corresponds to the maximal expected value gained from making the optimal future harvest scheduling decision based on $p_{\mathbf{V}}$ (without considering any inventory data collection). The theoretical upper limit for (7) is the value of perfect information obtained by replacing $\mathrm{PoV}(\mathbf{x}^{(I)})$ by the perfect posterior value PoV^* (see (2) in Web Appendix 7), that is, the value of knowing volume $\mathbf{V}$ exactly before the utility is maximized.

5.2 A computationally attractive representation for the posterior value of inventory decision

Next we will focus on an alternative representation for (6) that allows us to design an algorithm to solve the planning problem (3) and to estimate VoI. To begin, note that the inner integral in (6) can be simplified as follows:

$$\begin{aligned} &\int_{\mathcal{V}} \mathrm{U}(\mathbf{X}^{(T)}, \mathbf{v}) p_{\mathbf{V}|\mathbf{Y},\mathbf{x}^{(I)}}(\mathbf{v} \mid \mathbf{y}, \mathbf{x}^{(I)}) \mathrm{d}\mathbf{v} \\ &= \mathbb{E}\{\mathrm{U}(\mathbf{X}^{(T)}, \mathbf{V}) \mid \mathbf{Y} = \mathbf{y}, \mathbf{x}^{(I)}\} \\ &= -\sum_{t=1}^{n_T} \sum_{a=1}^{n_A} \mathbb{E}\left\{ \left(\sum_{s=1}^{n_S} x_{ts}^{(T)} V_{a,s} - D_{a,t} \right)^2 \middle| \mathbf{Y}_{a\cdot} = \mathbf{y}_{a\cdot}, \mathbf{x}^{(I)} \right\} \\ &= -\sum_{t=1}^{n_T} \sum_{a=1}^{n_A} (\mathbf{x}_{t\cdot}^{(T)})' \boldsymbol{\Sigma}_{a\cdot}^{+} \mathbf{x}_{t\cdot}^{(T)} + \left((\mathbf{x}_{t\cdot}^{(T)})' \boldsymbol{\mu}_{a\cdot}^{+} - D_{a,t} \right)^2, \end{aligned} \quad (9)$$

where $(X)'$ stands for the matrix/vector transpose of X, and we have used the formulas $\mathbb{E}[X^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2$, and $\mathrm{Var}(\mathbf{a}'\mathbf{Z}) = \mathbf{a}'\mathrm{Cov}(\mathbf{Z})\mathbf{a}$ for a random vector $\mathbf{Z}$. Furthermore, we use $\boldsymbol{\mu}_{a\cdot}^{+} := \mathbb{E}[\mathbf{V}_{a\cdot} \mid \mathbf{Y}_{a\cdot} = \mathbf{y}_{a\cdot}, \mathbf{x}^{(I)}]$ and $\boldsymbol{\Sigma}_{a\cdot}^{+} := \mathrm{Cov}(\mathbf{V}_{a\cdot} \mid \mathbf{Y}_{a\cdot} = \mathbf{y}_{a\cdot}, \mathbf{x}^{(I)})$, to denote the posterior mean and covariance of $\mathbf{V}_{a\cdot}$, respectively.

Under the assumptions of prior independence and conditionally independent observations in the models of Section 4, $\boldsymbol{\mu}_{a\cdot}^{+}$ is a vector with elements $\mathbb{E}[V_{a,s} \mid Y_{a,s} = y_{a,s}, x_s^{(I)}]$, $s = 1, 2, \ldots, n_S$, and $\boldsymbol{\Sigma}_{a\cdot}^{+}$ is a diagonal matrix with diagonal elements $\mathrm{Var}(V_{a,s} \mid Y_{a,s} = y_{a,s}, x_s^{(I)})$, $s = 1, 2, \ldots, n_S$. To evaluate $\boldsymbol{\mu}_{a\cdot}^{+}$ and $\boldsymbol{\Sigma}_{a\cdot}^{+}$, it is therefore sufficient to compute only these scalar quantities, which can be done, for example, by numerical integration.

In Web Appendix 6, it is shown that the maximization in (6) can be expressed as:

$$\begin{aligned} &\min \ \frac{1}{2} \sum_{t=1}^{n_T} (\mathbf{x}_{t\cdot}^{(T)})' \mathbf{Q} \mathbf{x}_{t\cdot}^{(T)} + \sum_{t=1}^{n_T} (\mathbf{c}_t)' \mathbf{x}_{t\cdot}^{(T)} + r, \\ &\text{s.t. } 0 \le \sum_{t=1}^{n_T} x_{ts}^{(T)} \le 1 \text{ for all } s = 1, 2, \ldots, n_S, \\ &\text{where } x_{ts}^{(T)} \in \{0, 1\} \text{ for all } t, s, \text{ and } \mathbf{Q}: \\ &= 2 \sum_{a=1}^{n_A} \left(\boldsymbol{\Sigma}_{a\cdot}^{+} + \boldsymbol{\mu}_{a\cdot}^{+} (\boldsymbol{\mu}_{a\cdot}^{+})' \right), \\ &\mathbf{c}_t := -2 \sum_{a=1}^{n_A} D_{a,t} \boldsymbol{\mu}_{a\cdot}^{+}, \text{ and } r := \sum_{t=1}^{n_T} \sum_{a=1}^{n_A} D_{a,t}^2. \end{aligned} \quad (10)$$

This formulation allows us to write the posterior value (6) in the following form, which can be approximated by Monte Carlo

sampling from the distribution $p_{\mathbf{Y}|\mathbf{x}^{(I)}}$:

$$\mathrm{PoV}(\mathbf{x}^{(I)}) = \int_{\mathcal{Y}} \left(-\frac{1}{2} \sum_{t=1}^{n_T} (\hat{\mathbf{x}}_{t\cdot}^{(T)})' \mathbf{Q} \hat{\mathbf{x}}_{t\cdot}^{(T)} - \sum_{t=1}^{n_T} (\mathbf{c}_t)' \hat{\mathbf{x}}_{t\cdot}^{(T)} - r \right) p_{\mathbf{Y}|\mathbf{x}^{(I)}}(\mathbf{y} \mid \mathbf{x}^{(I)}) \mathrm{d}\mathbf{y}. \tag{11}$$

Here, $(\hat{\mathbf{x}}_{t\cdot}^{(T)})_{1 \le t \le n_T}$ denotes the solution of (10), which depends on $\mathbf{Q}$, the $\mathbf{c}_t$'s and r (defined in (10)). Note that $\mathbf{Q}$ and the $\mathbf{c}_t$'s in turn depend on $\mathbf{y}$ through $\boldsymbol{\mu}_{a\cdot}^{+}$ and $\boldsymbol{\Sigma}_{a\cdot}^{+}$.

6 ALGORITHMS FOR APPROXIMATING THE POSTERIOR VALUE AND SOLVING THE PLANNING PROBLEM

The direct solution of the planning problem (3) is difficult, since the evaluation of the posterior value of inventory decision (11) requires the solution of a binary optimization problem (the future harvest scheduling problem) within a multivariate integral. To ease this computational problem, we (1) approximate the solution of the binary optimization problem using a heuristic method and (2) approximate the integral (11) by Monte Carlo sampling where M realizations are simulated from distribution $p_{\mathbf{Y}|\mathbf{x}^{(I)}}$. Next, we will present this kind of approximation for $\mathrm{PoV}(\mathbf{x}^{(I)})$ in Section 6.1 and then use it to construct an algorithm to solve the planning problem (3) in Section 6.2.

6.1 Approximating the posterior value

The optimization problem (10) within (11) can be seen as a special case of the GQAP, first proposed by Lee and Ma (2004). Following the formulation of the GQAP by Hahn et al. (2008), the problem considers the placement of K facilities to L locations with known quadratic and linear coefficients C_{ijkn} and B_{ij}, respectively, and with constraints on the "space limitations" S_j for each location j. In our problem setting, the forest tracts are analogous to the facilities, and the locations to the times at which the tracts should be harvested. However, in our setting, the GQAP is simplified such that there are no space limitations, that is, $S_j = \infty$, and a large number of the coefficients B_{ij} and C_{ijkn} in the objective function are zero.

The GQAP is NP-hard, as it is a generalization of the generalized assignment problem (Koopmans and Beckmann, 1957) shown to be NP-hard by Fisher et al. (1986). Most of the research on the solution of the GQAP has focused on exact solution using branch and bound methods (cf. Hahn et al., 2008; Pessoa et al., 2010). In our context, these kinds of methods are too slow, since the dimension of the problem is relatively high, and we need to be able to solve the problem M times in order to evaluate one inventory decision, using the Monte Carlo approximation for (11).

Thus, we rely instead on a fast and greedy heuristic method, which we call the "random sweep method". The method considers multiple random initializations (starting values) for the decision variable $\mathbf{X}^{(T)}$. For each random initialization, the method applies a sequence of "local moves" where each move improves the objective function value. Once the solution cannot be improved further, the final objective function value is recorded and the process repeats for the next random initialization. Finally, the method returns the smallest objective function value found among all random initializations considered. In our context, this strategy is computationally attractive for two reasons. First, the change in the objective function value for each local move is cheap to evaluate in comparison to the full objective function value of (10). Second, the local moves can explore the full feasible region and always preserve the feasibility of the improving candidate solutions.

The local moves correspond to changes in the harvest scheduling of a single tract while the harvest scheduling of the other tracts remains fixed. Invoking each local move changes the objective function value of (10) by an "adjustment". For each tract s, there are $n_T + 1$ possible adjustments: the adjustment corresponding to the local move that does not change the timing (0) and the adjustments for moves that alter the harvest timing (given current timing):

$$\text{From time } u \text{ to 'no harvest':} \quad -c_{us} - (\mathbf{Q}_{\cdot s})' \mathbf{x}_{u\cdot}^{(T)} + \frac{1}{2} Q_{ss}, \tag{12}$$

$$\text{From 'no harvest' to time } u\text{:} \quad c_{us} + (\mathbf{Q}_{\cdot s})' \mathbf{x}_{u\cdot}^{(T)} + \frac{1}{2} Q_{ss}, \tag{13}$$

$$\text{From time } u \text{ to time } v\text{:} \quad -c_{us} + c_{vs} + (\mathbf{Q}_{\cdot s})' \mathbf{x}_{v\cdot}^{(T)} - (\mathbf{Q}_{\cdot s})' \mathbf{x}_{u\cdot}^{(T)} + Q_{ss}, \tag{14}$$

where c_{us} stands for the sth element of the vector $\mathbf{c}_u$, Q_{ss} is the sth diagonal element of $\mathbf{Q}$, and $\mathbf{Q}_{\cdot s}$ is the sth column of $\mathbf{Q}$. The adjustments in (12)–(14) have been derived by inspecting the changing terms in the objective function when the local moves are applied.

Using adjustments (12)–(14), Algorithm 1 gives pseudo-code for the random sweep method that can be used to approximately solve the optimization problem (10) given the specification $(\mathbf{Q}, (\mathbf{c}_t)_{1 \le t \le n_T}, r)$ and n_{ri} random initializations. The name RandomSweep for the algorithm comes from line 7, where a random processing order (a "sweep" over the tracts) is drawn. Lines 6–15 carry out the core computation, where local moves and adjustments are chosen for the tracts, and the solution and the current objective function value are updated, until the current solution cannot be improved further.

On line 10, there may be multiple local moves with negative adjustments, and the choice of the local move in such a situation affects how greedy the resulting algorithm will be. The greediest choice is to always select the local move that improves (decreases) the objective function value the most. In this paper, we refer to this strategy as the "greedy descent" strategy. Instead, it is also possible to choose a less greedy descent strategy that we call "conservative descent", which always chooses the local move that improves the objective function value the least (but improves nonetheless). The conservative descent strategy increases the

Algorithm 1 RandomSweep$(\mathbf{Q}, (\mathbf{c}_t)_{t=1,2,\ldots,n_T}, r; n_{ri})$

1: Initialise the return value $m = \infty$.
2: **for** $i \in 1, 2, \ldots, n_{ri}$ **do**
3: Sample a random feasible decision matrix $\mathbf{X}_{cur}$.
4: Set m_{cur} to the full objective function value associated with $\mathbf{X}_{cur}$.
5: If $m_{cur} < m$, set $m = m_{cur}, \mathbf{X}_{best} = \mathbf{X}_{cur}$.
6: **while** true **do**
7: Sample a random processing order p (a permutation of tract indices $1, 2, \ldots, n_S$).
8: **for** tract index $s \in p$ **do**
9: With $\mathbf{X}_{cur}$ and $(\mathbf{Q}, \mathbf{c}_t)$, compute all adjustments for tract s using (12)-(14).
10: Choose a local move with adjustment $c \leq 0$. Denote by l the chosen local move.
11: Apply local move l to the sth column of $\mathbf{X}_{cur}$.
12: Set $m_{cur} = m_{cur} + c$. If $m_{cur} < m$, set $m = m_{cur}, \mathbf{X}_{best} = \mathbf{X}_{cur}$.
13: **end for**
14: If $\mathbf{X}_{cur}$ did not change in the above loop, break out of the while loop.
15: **end while**
16: **end for**
17: Return $m, \mathbf{X}_{best}$.

convergence time of the method. However, even with this strategy, we have observed that the while loop in Algorithm 1 very often still terminates in at most a few dozen iterations. We will investigate the effects of the choice of the descent strategy in Section 6.3.

We remark that Algorithm 1 is similar to the Hero method of Pukkala and Kangas (1993), which has been found to strike an efficient tradeoff in terms of computational complexity vs. quality of solution (Pukkala and Kurttila, 2005) in forest planning problems. Algorithm 1 extends Hero by introducing the random processing order of the tracts on each iteration; the Hero method would instead process the tracts without varying the order. This randomization ensures that no particular tract has "priority" over the other tracts and introduces variation to the solutions explored by the method. Furthermore, in contrast to Hero, Algorithm 1 provides "greedyness control" with the two descent strategies and has been tailored to solve general optimization problems of the form (10) by taking advantage of the adjustment formulas (12)–(14). Note that it is not strictly necessary to use Algorithm 1 to solve the harvest scheduling problems arising in this work; we discuss refinements and altenatives in Section 8.

Algorithm 2 $\widehat{\text{PoV}}(\mathbf{x}^{(I)}; M, n_{ri})$

1: **for** $k = 1, 2, \ldots, M$ **do**
2: Simulate $\mathbf{y}^{(k)}$ from $p_{\mathbf{Y}|x^{(I)}}$.
3: Given $\mathbf{y}^{(k)}$, compute $\boldsymbol{\mu}_{a\cdot}^{+}$ and $\boldsymbol{\Sigma}_{a\cdot}^{+}$ for each $a = 1, 2, \ldots, n_A$.
4: Using $(\boldsymbol{\mu}_{a\cdot}^{+}, \boldsymbol{\Sigma}_{a\cdot}^{+})_{1\leq a\leq n_A}$ and $(D_{a,t})_{1\leq a\leq n_a, 1\leq t\leq n_T}$, construct $\mathbf{Q}$, $(\mathbf{c}_t)_{1\leq t\leq n_T}$ and r in (10).
5: $m^{(k)}, \mathbf{X}_{best}^{(k)} \leftarrow$ RandomSweep$(\mathbf{Q}, (\mathbf{c}_t)_{1\leq t\leq n_T}, r; n_{ri})$
6: **end for**
7: **return** $-\frac{1}{M}\sum_{k=1}^{M} m^{(k)}$

With the help of Algorithm 1, we present Algorithm 2 which is a Monte Carlo approximation of (11) given the (candidate) inventory decision $\mathbf{x}^{(I)}$, M Monte Carlo samples and n_{ri} random initializations for Algorithm 1. For each iteration k, Algorithm 2 proceeds by simulating a measurement vector $y^{(k)}$ from the marginal distribution $p_{\mathbf{Y}|\mathbf{x}^{(I)}}$ (line 2). This can be carried out using the models of Section 4 by first simulating $\mathbf{v}$ from the prior distribution $p_{\mathbf{V}}$, and then simulating $\mathbf{y}^{(k)} \sim p_{\mathbf{Y}|\mathbf{V},\mathbf{x}^{(I)}}(\cdot \mid \mathbf{V} = \mathbf{v}, \mathbf{x}^{(I)})$. The measurement $\mathbf{y}^{(k)}$ is subsequently used to compute the posterior means and covariances $(\boldsymbol{\mu}_{a\cdot}^{+}, \boldsymbol{\Sigma}_{a\cdot}^{+})_{1\leq a\leq n_A}$ (line 3). Finally, these quantities (together with the demands $D_{a,t}$) are used to construct the minimization problem (10), which is then approximately solved using Algorithm 1 and the minimum $m^{(k)}$ recorded (line 5). The use of Algorithm 1 on line 5 can be replaced with another optimization method targeting the problem (10). The output of Algorithm 2 is the mean of the negative $m^{(k)}$'s, where the change of sign is applied since we are interested in maximizing rather than minimizing.

Finally, we remark that with minor modifications to the minimization problem (10) and Algorithms 1 and 2, the methods in this section may also be leveraged to approximate PV and PoV*. The details are given in Web Appendix 7.

6.2 The practical solution method

Algorithm 3 presents pseudo-code for our practical method that approximately solves the planning problem (3) for an inventory budget B, using M Monte Carlo samples and n_{ri} random initializations.

Algorithm 3 SolvePlanningProblem$(B; M, n_{ri})$

1: Initialise an empty container O for results.
2: Construct the set $\mathcal{X}^{(I)}$.
3: **for** $x \in \mathcal{X}^{(I)}$ **do**
4: Evaluate $p = \widehat{\text{PoV}}(x; M, n_{ri})$ (Algorithm 2).
5: Evaluate $c = C_x$ using (4).
6: Save (x, p, c) to the container O.
7: **end for**
8: Set $O_B = \{(x, p) : (x, p, c) \in O \text{ and } c \leq B\}$
9: Find (x^*, p^*) from O_B such that $p^* \geq p$ for all elements $(x, p) \in O_B$.
10: **return** x^*, p^*

The algorithm first constructs the feasible set $\mathcal{X}^{(I)}$, and then, for each $\mathbf{x} \in \mathcal{X}^{(I)}$, computes the inventory cost using (4) and the approximate posterior value using Algorithm 2. The output of Algorithm 3 is the highest approximate posterior value and the associated inventory decision with cost less than B (lines 8–9). The inventory budget B is only required by Algorithm 3 after processing the set $\mathcal{X}^{(I)}$. Therefore, Algorithm 3 can solve (3) approximately for multiple budgets with a single run of the processing loop (lines 3–7).

Under the assumption of "cost-accuracy consistency" (Definition 1) and distinct total prior volume variances, each inventory decision vector $\mathbf{x} \in \mathcal{X}^{(I)}$ can be constructed as follows. First, define

$$\kappa(n_I, n_S) := \{(z_1, z_2, \ldots, z_{n_S}) : z_i \in \{1, 2, \ldots, n_I\}, \\ z_i \leq z_j \, \forall i < j\},$$

that is, $\kappa(n_I, n_S)$ is the set of ordered vectors of length n_S with elements in $\{1, 2, \ldots, n_I\}$. For each $\mathbf{z} \in \kappa(n_I, n_S)$, we can construct a corresponding $\mathbf{x} \in \mathcal{X}^{(I)}$ by setting $x_{o_j} = z_j$ for $j \in 1, 2, \ldots, n_S$, where o is the vector of tract indices sorted in the order of increasing total prior standard deviation σ_s^0 (see Definition 2). It is well known that $|\kappa(n_I, n_S)| = \binom{n_S+n_I-1}{n_S}$, implying $|\mathcal{X}^{(I)}| = \binom{n_S+n_I-1}{n_S}$, as well.

6.3 Investigating the performance of Algorithm 1

Our approximation of the posterior value uses Algorithm 1 to solve the minimization problem (10). The performance of the algorithm is therefore important for the overall performance of

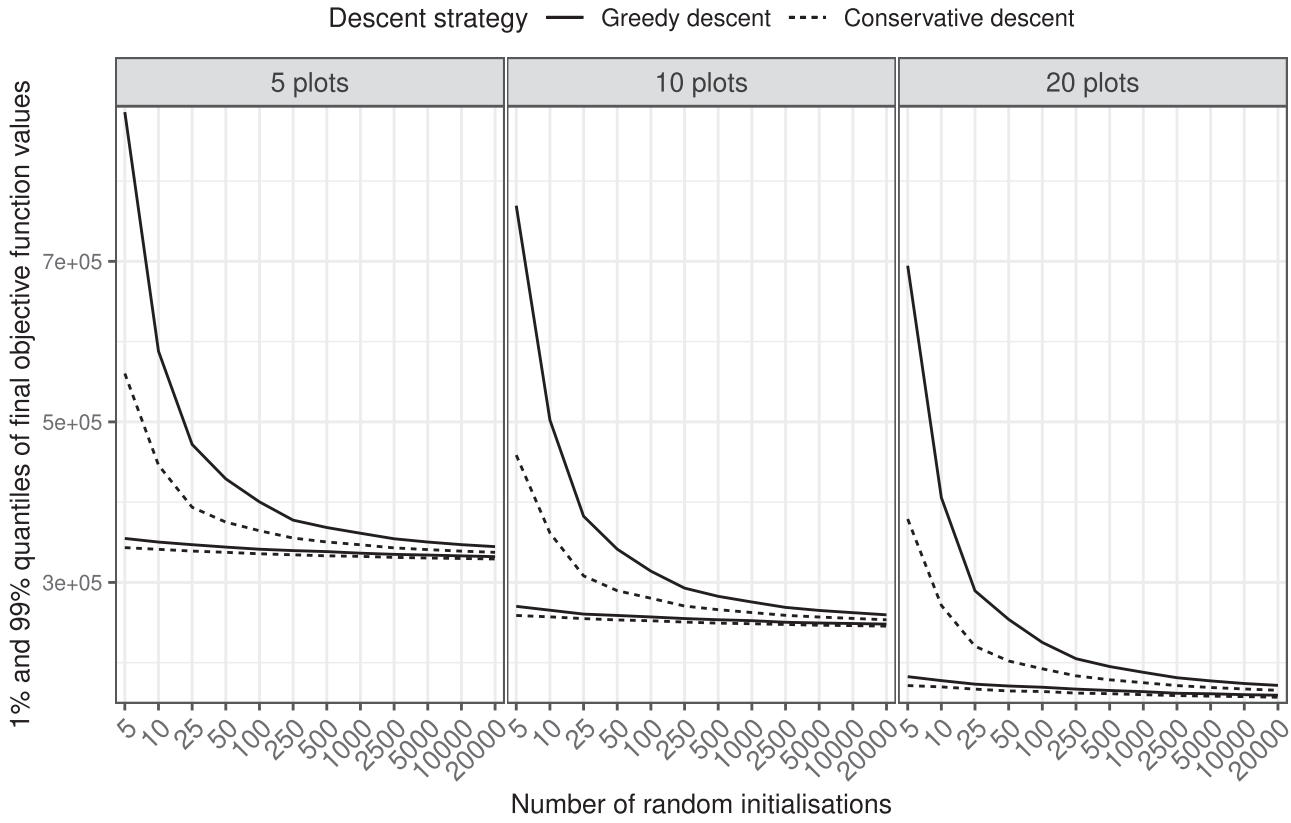

FIGURE 1 The variability in the final objective function values of Algorithm 1 with respect to the number of random initializations and the descent strategy discussed in Section 6.1. For both descent strategies, the depicted lines correspond to the 1% and 99% quantiles of final objective function values obtained, when Algorithm 1 was run 10 000 times with the number of initializations shown in the horizontal axis. Note that the horizontal axis does not increase linearly. The panes indicate the inventory decision for all 100 tracts that was used to construct the three minimization problems of the form (10) studied here.

the approximation and thus the solution of the full planning problem (3).

As Algorithm 1 features randomness, we studied the variation in the found optima. We constructed three test minimization problems of the form (10), as follows. First, using the tract data of Section 3, we simulated three inventory datasets with 5, 10, or 20 plots for every tract from the normal model of Web Appendix 5. For each test problem, we ran Algorithm 1 10 000 times using both descent strategies from Section 6.1) and varied the number of random initializations n_{ri} from 5 up to 20 000. After each run, we recorded the final objective function value.

Figure 1 summarizes the results of this experiment by displaying the empirical 1% and 99% quantiles of the final 10 000 objective function values given the descent strategy and the number of random initializations n_{ri} of Algorithm 1. As expected, the variability of the final objective function values quickly diminishes as n_{ri} increases.

As the optimization problem is NP-hard, it is possible that the true global minimum is even lower than the achieved solutions. To study whether there exist better solutions, we further ran the Gurobi optimiser (Gurobi Optimization, LLC, 2023) for 14 h in the case of the minimization problem constructed when the inventory decision was 20 plots for each tract. Our implementation in Gurobi used Algorithm 1 as a heuristic for the mixed-integer solver. The results for this experiment are shown in Web Figure 4 in Web Appendix 8 and indicate that the minima found by Algorithm 1 are typically within 5%–15% of the best minimum found using Gurobi when a sufficient number of random initializations are used. Note that compared to the 14 h runtime of Gurobi, Algorithm 1 finishes in seconds.

Interestingly, in Figure 1, the conservative descent strategy exhibits less variability for all n_{ri} studied, and also appears to find solutions with slightly better objective function values. Our intuition is that this occurs since the conservative descent explores the solution space more efficiently by avoiding moves that improve the objective function value too eagerly, when the randomly sampled solution candidates will still be of poor quality.

Finally, the panes from left to right in Figure 1 show decreasing objective function values as the inventory decision is made more accurate and expensive. This makes sense, since more accurate inventory data will provide better information on the unknown volumes $\mathbf{V}$, improving the found objective function value in the harvest scheduling problem (10).

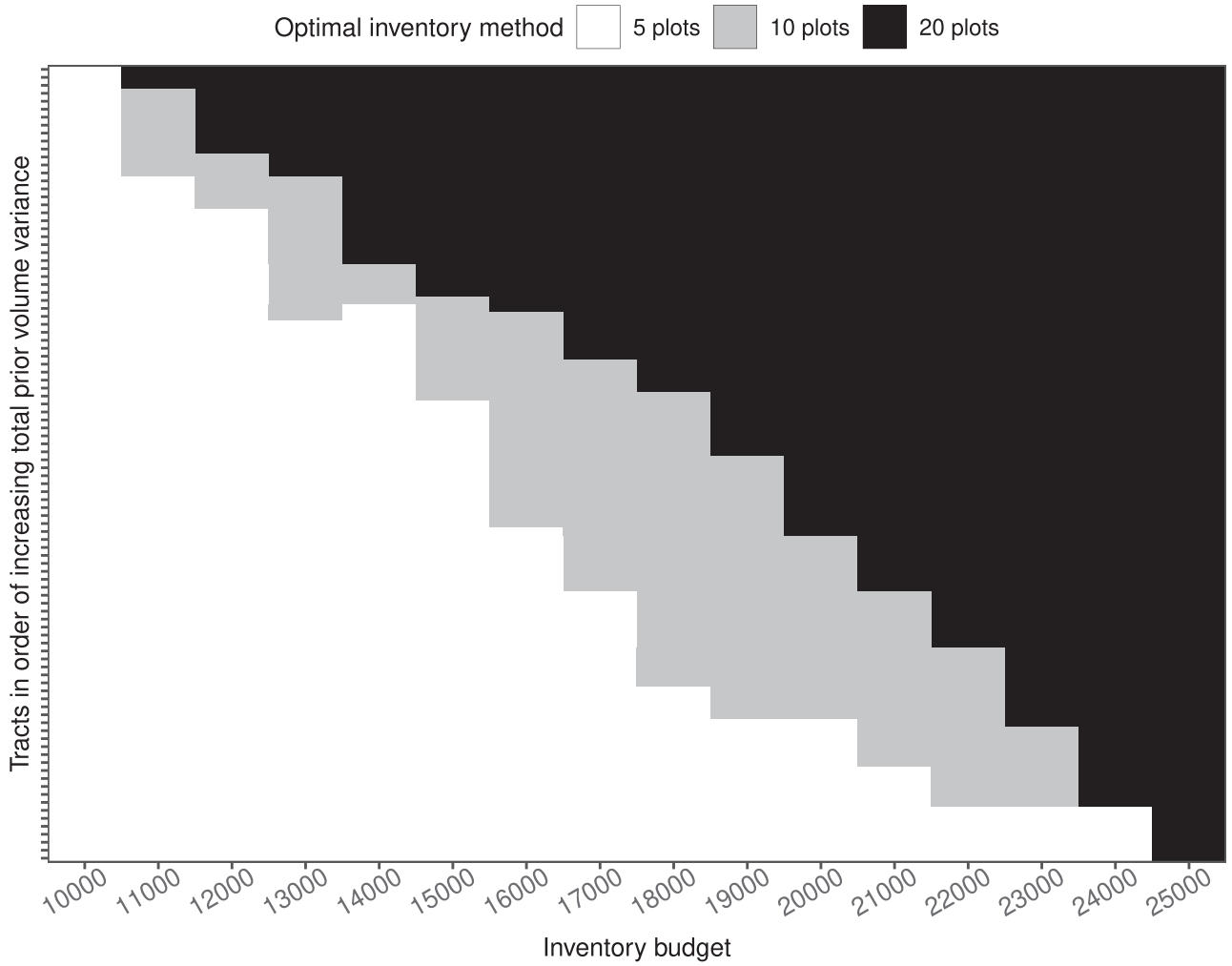

FIGURE 2 The inventory decisions returned by Algorithm 3 for the Swedish tract dataset with budget constraints from 10 000 to 25 000 by increments of 1000. The shade of a cell shows the preferred inventory decision for a tract (vertical axis) under the given budget (horizontal axis).

7 APPLICATION TO THE SWEDISH TRACT DATASET

Next, we apply the methods of Section 6 to the Swedish tract dataset discussed in Section 3. For each budget, we apply Algorithm 3 with the model discussed in Section 4. This yields the preferred inventory method (sampling 5, 10, or 20 plots) for each tract in the dataset. Motivated by our findings in Section 6.3, we use 2500 Monte Carlo samples, 2500 random initializations and the conservative descent strategy for Algorithm 1 within Algorithm 3.

Figure 2 summarizes the results. The inventory decisions follow an organized structure due to the ordering of the tracts by the prior volume variance (on the vertical axis). For the extreme budgets 10 000 and 25 000, the cheapest (5 plots inventory for every tract) and the most expensive inventory (20 plots inventory for every tract) are chosen, as the costs of these inventories coincide exactly with the extreme budgets. In contrast, all three inventory methods are used for some tracts when the budget is between 11 000 and 23 000. For these budgets, the proportions of the inventory methods selected for each tract vary, which occurs because of the difference between the assortment composition in the tracts and the different costs of the inventories. For budgets 16 000–19 000, it appears that the proportions of tracts with 10 or 20 plots inventory are roughly the same, whereas for budgets 20 000–25 000, the 20 plots' inventories begin to "dominate", as there is enough money to spend in the budget. For budgets less than 15 000, especially, the proportions of tracts to measure with 10 and 20 plots vary without a clear pattern.

Figure 3 displays the maximal posterior value associated with the decisions computed by Algorithm 3 that costed less than each inventory budget considered. In addition, the prior value PV and the perfect posterior value PoV* discussed in Section 5.1 are shown with horizontal lines. The increasing posterior values depict the increasing information gain from larger budgets. In comparison to PV, the figure shows that a significant information gain is obtained even by carrying out the cheapest possible forest inventory costing 10 000. The discrepancy between PoV* and the maximal posterior value obtained for budget 25 000 depicts the information gain that could still be obtained by even more expensive and accurate inventory methods.

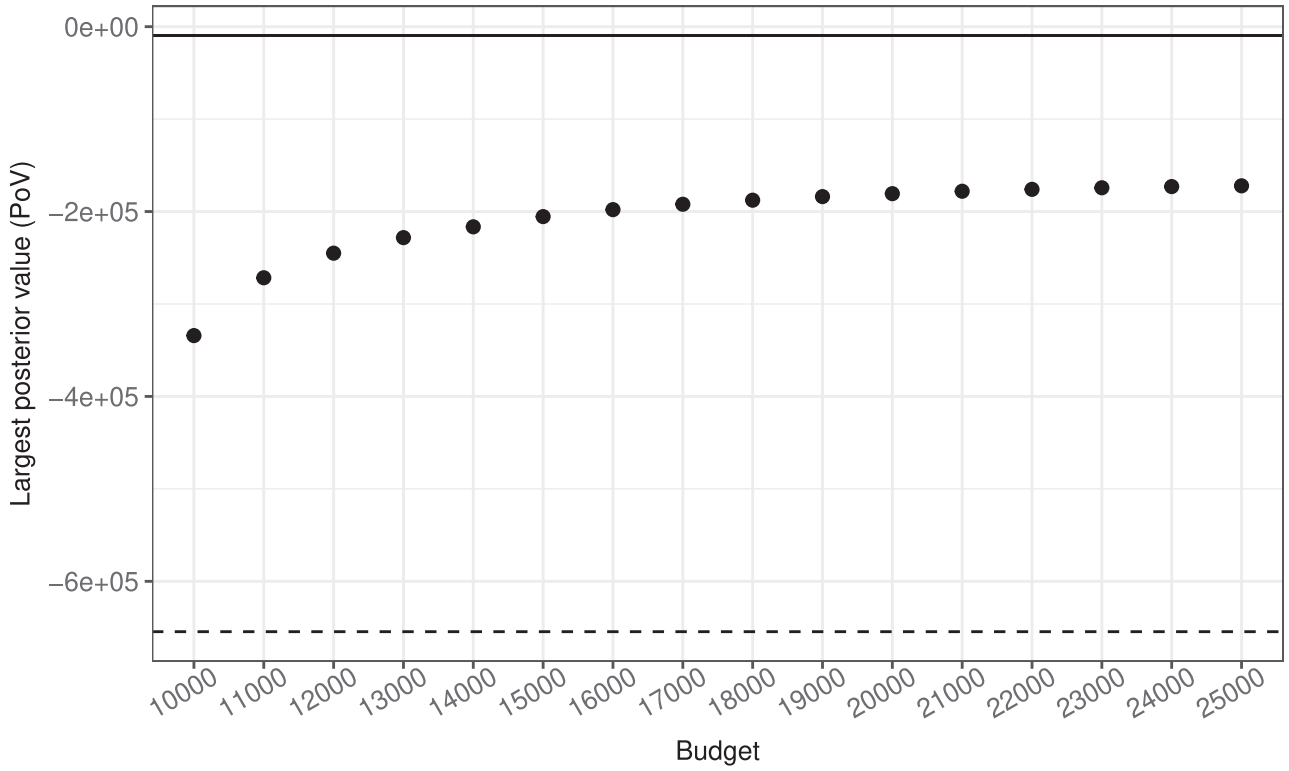

FIGURE 3 The largest posterior value obtained for each budget constraint (points) considered with the Swedish tract dataset. The horizontal dashed and solid lines correspond to approximations of the prior value PV (8) and the perfect posterior value, respectively. These quantities were approximated as discussed in Web Appendix 7. The difference between the solid and the dashed line and the difference between the points and the dashed line correspond to approximate values of perfect and imperfect information, respectively.

Inspecting the rate of increase of the posterior values in Figure 3 can be used to help decision-making: spending more than 16 000–18 000 in the forest inventory appears to bring diminishing returns from a Bayesian decision-theoretic perspective.

8 DISCUSSION

We developed a computational method for selecting a cost-effective operational forest inventory under uncertainty. Our approach is based on maximizing the posterior value of the data collection associated with an inventory decision. This leads us to model a Bayesian two-stage decision problem where the inner stage considers scheduling the harvesting to fulfill the demand targets and the outer stage considers the problem of interest, selecting the accuracy of pre-harvest inventories. As the output of our method, we obtain a plan that indicates how the inventory budget should be divided between the forest tracts in order to provide maximally beneficial information for scheduling the harvests of the tracts.

The steps to use the method for also solving the harvest decision problem require collecting inventory data in practice, and proceed as follows. First, the inventory budget is fixed and the solution to the inventory decision problem is obtained from one of the columns of Figure 2 (or equivalent). Then, the selected inventory is carried out (in practice), which yields a realization $\mathbf{y}$. The data $\mathbf{y}$ are then used to construct a final harvest scheduling problem of the form (10); solving it yields the solution to the harvest scheduling problem. Here, the problem is solved only once, and thus it is feasible to use computationally intensive branch and bound mixed-integer solvers (similar to the Gurobi solver in Section 6.3) to obtain as efficient a solution as possible.

The developed approach scales well with respect to the number of assortments n_A and the number of tracts n_S (assuming that n_I is small compared to n_S) but increasing the number of inventory methods n_I may multiply the computational cost by $n_S + n_I - 1$. In practical settings, the scalability with respect to n_S is, however, arguably more important. The approach could also be applied in a dynamic setting, where new tracts are added every month and the decision-making is done repeatedly.

Since the optimization problems arising in this paper are NP-hard, the proposed algorithms rely on heuristics and the obtained solutions may not be globally optimal. When we studied the performance of the random sweep method of Section 6.3, we found that in problems constructed from the Swedish tract data, it could reach minimums that were within 5%–15% of minimums computed by a vastly more computationally intensive mixed-integer solver. Even though this does not constitute perfect performance, we believe that the overall developed method could already be a useful tool in operational forestry. Note that when considering the full two-stage problem, the noted 5%–15% deviation only applies in the harvest scheduling problems solved as a part of the Monte Carlo approximation when selecting the inventory decision. This is because when the found

inventory decision is operationalized, the final harvest scheduling problem can be solved with more computationally intensive methods as discussed above.

Within the presented method, it would be possible to use other methods than Algorithm 1 for solving the harvest scheduling problems (10). Potential improvements include simulated annealing (SA) (Kirkpatrick et al., 1983) and multimoves that move the harvest timing for many tracts simultaneously. Heinonen and Pukkala (2004) studied these methods and found that they may improve harvest optimization results over the Hero method, which is similar to Algorithm 1. However, these results do not necessarily mean that using more intensive harvest optimization methods will lead to better estimates of the posterior value PoV (6). This is because the choice of the harvest scheduling optimizer acts as a tradeoff in the bias-variance decomposition of the PoV Monte Carlo approximation (Algorithm 2). In particular, for a fixed computational budget, faster heuristic methods allow for more Monte Carlo samples (less variance at the cost of bias) whereas computationally intensive methods decrease bias–at the cost of more variance.

There are many ways to broaden and refine the problem studied in this paper. Instead of log-normal models, alternative statistical models (such as the Gamma distribution) under which the posterior means and variances for assortment volumes can be computed, (to a sufficient numerical approximation) could be used for the assortment volumes and volume measurements. It might also be possible to extend the developed model with a spatial correlation structure between the tracts and/or with a correlation structure for the assortments within the tracts. Modeling these correlations would likely make the inventory returned by our method cheaper as each inventory observation would yield more information at the same cost. The spatial correlation structure could be formulated, for example, using the distance between tracts, which could also be useful for formulating further constraints to the inventory decision or harvest scheduling problems, similar to the ideas of Eyvindson et al. (2018).

'The prior variance order' constraint (Definition 2) can be seen as a simplification that reduces the large solution space of the unconstrained forest inventory problem to a smaller solution space that omits many "redundant" solutions with cost greatly larger or smaller than the budget B. In scenarios where some of the assortments are more important or valuable than others, the prior variance ordering is not justified, and it may be beneficial to order the tracts based on some other quantity. An example is an ordering based on (monetary) value uncertainty computed using crude estimates of value per unit of timber volume. In general, the ordering should reflect the amount of uncertainty (of some quantity), because it only makes sense to pay for measuring what is unknown.

To further take into account differences between the assortments, the utility function (5) could be modified to include (constant) assortment and time-specific weights. This would only lead to minor changes to the parameters of (10), and thus could be carried out with the presented method. Furthermore, formulating (5) in monetary terms would allow for direct VoI-based optimization without budget constraints.

Finally, the full problem setup could be developed into the direction of multi-objective decision-making, for example, by considering also the quality of the harvested material in the inner stage. Such extensions might better account for complexities in operational forestry but might also create new computational challenges for Bayesian decision-making.

ACKNOWLEDGMENTS

The authors thank Lars Eliasson, Oskar Gustavsson, Salme Kärkkäinen, Kaisa Miettinen, and Babooshka Shavazipour for useful discussions.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices are available with this paper at the Biometrics website on Oxford Academic. The source code used to draw the conclusions of the paper is available on Oxford Academic and at https://github.com/skarppinen/cost-eff-forest-inv.

FUNDING

The work has been funded by Peter Wallenberg Foundation (grant number 2022.0008). The research is related to the thematic research area DEMO (Decision Analytics utilizing Causal Models and Multiobjective Optimization) of the University of Jyväskylä.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The data on the 100 Swedish forest tracts and the numerical results needed to create the figures of the paper are available at https://github.com/skarppinen/cost-eff-forest-inv.

REFERENCES

Cook, S. A. (1971). The complexity of theorem-proving procedures, In: *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, 151–158.

Duvemo, K., Lämås, T., Eriksson, L. O. and Wikström, P. (2014). Introducing cost-plus-loss analysis into a hierarchical forestry planning environment. *Annals of Operations Research*, 219, 415–431.

Eidsvik, J., Mukerji, T. and Bhattacharjya, D. (2015). *Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis*. United Kingdom: Cambridge University Press.

Eyvindson, K., Rasinmäki, J. and Kangas, A. (2018). Evaluating a hierarchical approach to landscape-level harvest scheduling. *Canadian Journal of Forest Research*, 48, 208–215.

Fisher, M. L., Jaikumar, R. and Van Wassenhove, L. N. (1986). A multiplier adjustment method for the generalized assignment problem. *Management Science*, 32, 1095–1103.

Gregoire, T., Næsset, E., Mcroberts, R., Stahl, G., Andersen, H.-E., Gobakken, T. et al. (2016). Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. *Remote Sensing of Environment*, 173, 98–108.

Gurobi Optimizer. (2023). *Gurobi Optimization, LLC*.https://www.gurobi.com/documentation/current/refman/index.html [Accessed 12 September 2024].

Hahn, P. M., Kim, B.-J., Guignard, M., Smith, J. M. and Zhu, Y.-R. (2008). An algorithm for the generalized quadratic assignment problem. *Computational Optimization and Applications*, 40, 351–372.

Heinonen, T. and Pukkala, T. (2004). A comparison of one-and two-compartment neighbourhoods in heuristic search with spatial forest management goals. *Silva Fennica*, 38, 319–332.

Hirshleifer, J. and Riley, J. G. (1979). The analytics of uncertainty and information—an expository survey. *Journal of Economic Literature*, 17, 1375–1421.

Kangas, A., Hartikainen, M. and Miettinen, K. (2014). Simultaneous optimization of harvest schedule and measurement strategy. *Scandinavian Journal of Forest Research*, 29, 224–233.

Kangas, A., Hartikainen, M. and Miettinen, K. (2015). Simultaneous optimization of harvest schedule and data quality. *Canadian Journal of Forest Research*, 45, 1034–1044.

Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220, 671–680.

Kivinen, V.-P. (2004). A genetic algorithm approach to tree bucking optimization. *Forest Science*, 50, 696–710.

Koopmans, T. C. and Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica: Journal of the Econometric Society*, 25, 53–76.

Lee, C.-G. and Ma, Z. (2004). The generalized quadratic assignment problem, Technical report, *Department of Mechanical and Industrial Engineering*, University of Toronto, Canada.

Nieuwenhuis, M., Malone, L., Mchugh, F. and Layton, T. (1999). Development and evaluation of a pre-harvest inventory and cross-cutting simulation procedure to maximise value recovery. *Irish Forestry*, 56, 12–28.

Pessoa, A. A., Hahn, P. M., Guignard, M. and Zhu, Y.-R. (2010). Algorithms for the generalized quadratic assignment problem combining Lagrangean decomposition and the Reformulation-Linearization Technique. *European Journal of Operational Research*, 206, 54–63.

Pukkala, T. and Kangas, J. (1993). A heuristic optimization method for forest planning and decision making. *Scandinavian Journal of Forest Research*, 8, 560–570.

Pukkala, T. and Kurttila, M. (2005). Examining the performance of six heuristic optimisation techniques in different forest planning problems. *Silva Fennica*, 39, 319–332.

Raiffa, H. and Schlaiffer, R. (1967). *Applied Statistical Decision Theory*. New York, NY: Wiley.

Siipilehto, J., Lindeman, H., Mikko, V., Yu, X. and Uusitalo, J. (2016). Reliability of the predicted stand structure for clear-cut stands using optional methods: Airborne laser scanning-based methods, smartphone-based forest inventory application Trestima and pre-harvest measurement tool EMO. *Silva Fennica*, 50, 1–24.

Ståhl, G. (1994). *Optimizing the Utility of Forest Inventory Activities*, PhD thesis, Swedish University of Agricultural Sciences, Department of Biometry and Forest Management.

Ulvdal, P., Öhman, K., Eriksson, L. O., Wästerlund, D. S. and Lämås, T. (2023). Handling uncertainties in forest information: the hierarchical forest planning process and its use of information at large forest companies. *Forestry*, 96, 62–75.

Vauhkonen, J., Maltamo, M., McRoberts, R. E. and Næsset, E. (2014). *Introduction to Forestry Applications of Airborne Laser Scanning*, 1–16. Dordrecht: Springer.

Received: November 20, 2023; Revised: July 6, 2024; Accepted: September 5, 2024