Optimizing high-dimensional stochastic forestry via reinforcement learning

Tahvonen, Olli

2022-12

Tahvonen , O , Suominen , A , Malo , P , Viitasaari , L & Parkatti , V-P 2022 , ' Optimizing high-dimensional stochastic forestry via reinforcement learning ' , Journal of Economic Dynamics & Control , vol. 145 , 104553 . https://doi.org/10.1016/j.jedc.2022.104553

http://hdl.handle.net/10138/351513
https://doi.org/10.1016/j.jedc.2022.104553

Journal of Economic Dynamics & Control 145 (2022) 104553

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

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

Optimizing high-dimensional stochastic forestry *via* reinforcement learning

Olli Tahvonen[a], Antti Suominen[b], Pekka Malo[b], Lauri Viitasaari[c], Vesa-Pekka Parkatti[a]

[a] University of Helsinki, Dept. of Economics, Finland
[b] Aalto University School of Business, Dept. of Information and Service Management, Finland
[c] Department of Mathematics, Uppsala University, Sweden

ARTICLE INFO

Article history:
Received 22 June 2022
Revised 8 September 2022
Accepted 17 October 2022
Available online 20 October 2022

JEL classification:
C61
Q23

Keywords:
Artificial intelligence
Reinforcement learning
Forestry
Stochasticity
Curse of dimensionality
Optimal rotation
Natural resources

ABSTRACT

In proceeding beyond the generic optimal rotation model, forest economic research has applied various specifications that aim to circumvent the problems of high dimensionality. We specify an age- and size-structured mixed-species optimal harvesting model with binary variables for harvest timing, stochastic stand growth, and stochastic prices. Reinforcement learning allows solving this high-dimensional model without simplifications. In addition to presenting new features in reservation price schedules and effects of stochasticity, our setup allows evaluating the simplifications in the existing research. We find that one- or two-dimensional models lose a high fraction of attainable economic output while the commonly applied size-structured matrix model overestimates economic profitability, yields deviations in harvest timing, including optimal rotation, and dilutes the effects of stochasticity. Reinforcement learning is found to be an efficient and promising method for detailed age- and size-structured optimization models in resource economics.

1. Introduction

The classic economic approach to forestry aims to maximize stand value by the choice of rotation periods. This Samuelson (1976) interpretation of the Faustmann (1849) forest valuation equation is generic in revealing the economics of clear-cut timing. Research has proceeded from this basis, but the development has reached a specific limit; optimal solutions become hard or impossible to find after expanding the model details toward a level that appears necessary for emerging extensions. Our aim is to show that recent achievements in reinforcement learning (RL) algorithms[1] can be applied when extending important model details and realism without forgoing optimization and economically justifiable model structure.

1.1. Related literature

The problems of high dimensionality are already present in deterministic models but are most pressing when specifications include stochastic factors. This is more or less the reason why a major part of stochastic forest economic literature

E-mail addresses: olli.tahvonen@helsinki.fi (O. Tahvonen), antti.suominen@aalto.fi (A. Suominen), pekka.malo@aalto.fi (P. Malo), lauri.viitasaari@math.uu.se (L. Viitasaari), vesa-pekka.parkatti@helsinki.fi (V.-P. Parkatti).

[1] For economic applications, see Charpentier et al. (2021); Maliar and Maliar (2022); Maliar et al. (2021); Scheidegger and Bilionis (2019).

https://doi.org/10.1016/j.jedc.2022.104553

relies on the unidimensional Samuelson-Faustmann (S-F) setup. Within the "tree paradigm", Miller and Voltaire (1983) view the rotation choice as a problem of single- or multistage optimal stopping, describe stand cutting value as a linear Brownian process, and show the optimality of a barrier rule on the stand cutting value. Willassen (1998), Sødal (2002) and Chang (2005) extend this by assuming a linear, geometric, or logistic Brownian process and obtain explicit barrier rules where stochasticity increases the stand cutting value. Their assumptions on Brownian process and a fixed initial level for the system state after each cut imply that the origin of stochasticity is in physical growth instead of price. This allows analytical solutions but leaves open the effects of stochastic wood price.

In Brazee and Mendelsohn (1988), empirical data on wood price suggest an independent and identically distributed (i.i.d.) price process. The application of numerical dynamic programming (and ignoring stochasticity in growth) produces a result where optimal rotation is determined by a reservation price schedule. A similar result is obtained by the implicit finite difference method by Insley and Rollins (2005). Clarke and Reed (1989) and Reed and Clarke (1990) include stochasticity in both physical growth and price and make a sharp distinction between age- and size-dependent growth. By the former they refer to cultivated tree stands and by the latter to untended forests where growth is determined by various density factors. Their assumption of specifying the price process as the geometric Brownian process and ignoring costs lead to barrier rules for age or size depending on the growth specification. Observing that age- and size-dependent growth are extremes of a continuum, Reed and Clarke (1990) suggest to simultaneously include these features albeit warn that difficulties will likely preclude the extension.

Ignoring the partial harvest or thinning, which are inherent in almost any actual forestry, is a crucial tractability simplification in the tree paradigm and closely related models. Helmes and Stockbridge (2011) include one thinning specified as a switch from one diffusion process to another and optimize its timing with the rotation period. Saphores (2003) and Alvarez and Koskela (2007) are earlier extensions with partial harvest, and they apply pure size-dependent growth, which, with thinning, brings their models close to the general model of biomass harvesting by Clark (1976).

Including simultaneous age- and size-dependencies and both partial harvests and rotation becomes natural after describing a forest through stage-, age-, or size-structured specifications[2]. This extension is typically carried out by Markovian transition matrix models with predetermined size classes (Getz and Haight, 1989). However, stochastic optimization with the resulting 10–50 continuous state and control variables is nontrivial. A well-known approach is to optimize parameters for a feedback thinning function that determines the number of harvested trees as a function of stand cutting value (Haight, 1990; Lu and Gong, 2003). This reduces the number of optimized variables to 3–10 and is explained to be in line with the results on reservation price schedule (Brazee and Mendelsohn, 1988). However, how optimizing the parameters for the assumed function deviates from optimizing the genuine decision variables, i.e. the number of trees in various size classes, remains open.

Another approach, initiated in Kaya and Buongiorno (1987), applies a stochastic (transition) matrix model for computing a reduced model with a lower number of discrete states. The reduced model describes the probabilities of entering the next state given the present state and a (discrete) harvesting choice. Stochastic prices are specified as discrete realizations with associated probabilities leading to a tractable number of stand-market states. In extensive later works, this framework is extended by including mixed stands, biodiversity, financial risks, risk aversion, and several other factors (Buongiorno and Zhou, 2020; Zhou and Buongiorno, 2006; 2019). As the authors emphasize, their approach aims to reduce model dimensionality. This leads to asking how well the reduced model with a sparse state structure represents the original problem under study.

Malo et al. (2021) apply a size-structured matrix model with twelve pre-specified size classes, four tree species, and a continuous number of trees in each size class. Their matrix model is a direct simplification from a more detailed individual-tree model obtained by a well-known transformation (Getz and Haight, 1989). Solutions for deterministic price are obtained by an RL algorithm without further simplifications. The solutions are shown to coincide with those computed by nonlinear programming and interior point methods but in only 0.06% of the original computing time.

All these three branches of research, aiming to proceed beyond the S-F and the tree paradigm approaches, apply the matrix model either as such or under simplifications. Although this model has a long history and is among the most common tools in specifications of forest growth, it is challenged due to its predefined size classes (Easterling et al., 2000). This feature may yield "fast and slow pathways", i.e. a fraction of trees transits unrealistically rapidly or slowly between the classes. This is problematic in ecology, and research has proceeded to integral projection models (Easterling et al., 2000) with continuous size structure. In forest models, a similar step is the "individual-tree model" (Liang and Picard, 2013). In this dominant model type in present forest ecology, the survival and growth of individual trees (or groups of identical trees) depend on tree characteristics (e.g. diameter) and interactions with other trees. This setup avoids the problems of the matrix model, includes both age and size dependencies[3], and allows for both thinning and clear-cuts but with a heavy cost of expanding the model dimensions.

[2] Forest economic models with these extensions have close relationship with age-structured models in vintage capital and economic demography literature (Boucekkine et al., 2011).

[3] Age dependencies are implicitly in the growth and mortality processes of individual trees. Both these processes and natural regeneration are density dependent, implying size dependencies as defined in Clarke and Reed (1989)

1.2. Our setup

We apply, for the first time, a mixed-species individual-tree model for deterministic and stochastic problems and with smooth optimization between regimes relying entirely on partial harvesting or thinning, i.e. continuous cover forestry (CCF) and regimes with both thinning and finite rotation, i.e. rotation forestry (RF). Earlier economic studies (Haight and Monserud, 1990a; Tahvonen, 2011) based on individual-tree structure are deterministic, restrict the CCF/RF regime choice, include fixed harvest timing, and reduce optimized variables by grouping harvested trees. In our model the revenue formation is based on tree diameters and a detailed model for variable and fixed harvesting costs separately on thinning and clear-cuts. Stochasticity is included in (log-normally distributed) wood prices, in individual-tree diameter growth and in natural regeneration. Individual-tree growth and regeneration are based on growth models estimated using empirical data with conditionally Gaussian innovations (Pukkala et al. 2013).

Our Theorem 1 proves that a deterministic stationary harvesting policy can be found that maximizes the objective functional. The result follows from the general theory on dynamic decision processes by Schäl (1983). Additionally, our model can be extended to cover more general price dynamics than the log-normally distributed prices used here. By the existence of a deterministic stationary Markov policy, we are able to extend general reinforcement learning algorithms, such as proximal policy optimization (Schulman et al., 2017), to solve the stochastic harvesting problem.

First, we show how the model performs if reduced to the unidimensional tree paradigm/S-F setup. This serves as a benchmark for the gains obtained with expanded detailed structure but additionally produces results that are most easily comparable with existing research. Our i.i.d. price stochasticity produces one- and two-dimensional reservation price schedules, but pure growth stochasticity ála the tree paradigm does not cause any effects on optimal rotation.

Secondly, we compute solutions where the optimal management regime consists of optimized partial harvesting followed by clear-cuts. The results reveal the losses and limitations of the S-F setup and that the reservation price schedules become dependent on both price and multidimensional forest state. The lower-dimensional matrix model in existing literature is shown to overestimate growth, the size of felled trees, and the value of the forest, in addition to deviations in harvest frequency and rotation period length.

Third, we obtain solutions relying solely on partial harvesting (the CCF regime) and show their properties under stochastic growth and price and compute reservation price schedules, none of which have been presented in existing studies. We find that including a wider range of harvesting alternatives decreases the overall level of economic risk. The inclusion of thinning and the CCF regime turns out to be essential for profitability, implying that optimizing rotation timing and clear-cuts solely, as in the S-F and tree paradigm models, serves for theoretical purposes but in any wider contexts calls for expanded dimensions.

The model with four tree species includes 1900 state and 950 control variables and binary harvest timing choices. Extending harvesting alternatives from clear-cuts to thinning and from a single-species setup to mixed species suggest that simpler specifications dilute the effects of stochasticity. The lower-dimensional matrix model has a similar effect. Additionally, it leads to overestimation of stand growth and of the economic outcome compared with the detailed individual-tree model.

Obtaining high-dimensional results based on individual-tree structure is possible by RL despite being beyond reach by earlier methods. This allows us to evaluate the accuracy of the feedback thinning function applied in the existing literature. This idea to circumvent the curse of dimensionality turns out to be inaccurate but based on our results we suggest a modification that may more closely mimic optimal harvests.

The next section specifies the optimization model. This is followed by an existence theorem for the optimal solutions as stationary Markov policy. Next, we present the computation results and finally a discussion and conclusions. Proofs, details on empirical model properties and the reinforcement algorithm are provided in the appendix.

2. Stochastic Optimization based on a Mixed Individual-Tree Model

Let $\tilde{x}_{j,q,w,t}$ denote the number of trees (per ha) of species $j \in \{1,\dots,l\}$ in size classes $q \in \{1,\dots,m\}$ and age cohorts $w \in \{1,\dots,n\}$ at the beginning of period $t = 0,1,2,\dots$ and $h_{j,q,w,t}$ the number of trees harvested at the end of the periods, respectively. The variables $\tilde{d}_{j,q,w,t}$ denote the diameter (centimeters) of size class q of age w trees at the beginning of period t. To allow any initial stand state and utilization of per-period information on growth stochasticity, we introduce variables $x_{j,q,w,t}$ and $d_{j,q,w,t}$ to represent the state variables at the end of the periods before harvesting. The stand state showing the number of trees in various species, size, and age classes at the beginning and end of the periods but before harvesting are then given by $\tilde{x}_t \in \mathbb{R}^{l\times m\times n}$ and $x_t \in \mathbb{R}^{l\times m\times n}$, respectively. Similarly, we use d_t, $\tilde{d}_t$, and h_t to denote the corresponding diameter data and the number of harvested trees by species, size, and age class. During each period t, the fraction of trees of species j that survive from age class w to the next age class $w+1$ is denoted by $0 \le \alpha_j(\tilde{x}_t, \tilde{d}_t) \le 1$. The per-period diameter growth is $I_{j,q,w}(\tilde{x}_t, \tilde{d}_t, \varepsilon_t)$, where the development of ε_t is stochastic. Natural regeneration of species j and size class q is given by the ingrowth function $\phi_{j,q}(\tilde{x}_t, \tilde{d}_t, \omega_t) \ge 0$, where ω_t is the stochastic variation. The diameters of the new naturally regenerated trees are $\widehat{d}_{j,q,0}$.

Let r be the rate of interest p.a., and $\gamma = 1/(1+r)^5$ be the discount factor for a 5-year period. Let $(p_t)_{t=0}^{\infty}$, $p_t \in \mathbb{R}^{2l}$ denote a species-specific stochastic price process for saw timber and pulp, and let $\delta_t^{th} \in \{0,1\}$ and $\delta_t^{cc} \in \{0,1\}$ denote boolean vari-

ables that determine periods when a thinning or a clear-cut take place, respectively. The gross revenues from clear-cutting and thinning are $R_{cc}(h_t, d_t, p_t)$ and $R_{th}(h_t, d_t, p_t)$ and the clear-cut and thinning costs $C_{cc}(h_t, d_t)$ and $C_{th}(h_t, d_t)$, respectively. When either a thinning or a clear-cut takes place, we include a fixed harvesting cost C_f (planning and transporting the harvester to the site), implying that harvesting may not be carried out at each period. In the case of a clear-cut, the stand state is always reset to bare land that is immediately artificially regenerated to the state $\bar{x}_{j,q,0}, \bar{d}_{j,q,0}$, where the former is the number of seedlings and the latter their initial size. Given that the regeneration cost is C_r, the per-period net revenues are

$$\pi(x_t, d_t, p_t, h_t, \delta_t^{th}, \delta_t^{cc}) = \begin{cases} R_{th}(h_t, d_t, p_t) - C_{th}(h_t, d_t) - C_f & \text{if } \delta_t^{th} = 1 \\ R_{cc}(x_t, d_t, p_t) - C_{cc}(x_t, d_t) - C_f - C_r & \text{if } \delta_t^{cc} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Let the expected present value of net revenues be denoted as $J(x_0, d_0, p_0)$, where an initial stand state is x_0, diameter data d_0, and market prices p_0. The stochastic problem is written as

$$J(x_0, d_0, p_0) = \max_{h_t, \delta_t^{th}, \delta_t^{cc}} \mathbb{E}\left[\sum_{t=0}^{\infty} \pi(x_t, d_t, p_t, h_t, \delta_t^{th}, \delta_t^{cc})\gamma^t\right], \tag{2}$$

subject to

$$x_{j,q,1,t} = \phi_{j,q}(\tilde{x}_t, \tilde{d}_t, \omega_t)(1 - \delta_{t-1}^{cc}) + \delta_{t-1}^{cc}\bar{x}_{j,q,0}, \quad j \in \{1, \ldots, l\}, \quad q \in \{1, \ldots, m\}, \tag{3}$$

$$\tilde{x}_{j,q,w+1,t+1} = \left(x_{j,q,w,t} - h_{j,q,w,t}\delta_t^{th}\right)(1 - \delta_t^{cc}), \quad j \in \{1, \ldots, l\}, \quad q \in \{1, \ldots, m\}, \quad w \in \{1, \ldots, n-1\}, \tag{4}$$

$$x_{j,q,w,t} = \alpha_{j,q,w}(\tilde{x}_t, \tilde{d}_t)\tilde{x}_{j,q,w,t}, \quad j \in \{1, \ldots, l\}, \quad q \in \{1, \ldots, m\}, \quad w \in \{2, \ldots, n\}, \tag{5}$$

$$\tilde{d}_{j,q,1,t} = \widehat{d}_{j,q,0}(1 - \delta_{t-1}^{cc}) + \delta_{t-1}^{cc}\bar{d}_{j,q,0}, \quad j \in \{1, \ldots, l\}, \quad q \in \{1, \ldots, m\}, \tag{6}$$

$$d_{j,q,w,t} = \tilde{d}_{j,q,w,t} + I_{j,q,w}(\tilde{x}_t, \tilde{d}_t, \varepsilon_t), \quad j \in \{1, \ldots, l\}, \quad q \in \{1, \ldots, m\}, \quad w \in \{1, \ldots, n\}, \tag{7}$$

$$\tilde{d}_{j,q,w+1,t+1} = d_{j,q,w,t}, \quad j \in \{1, \ldots, l\}, \quad q \in \{1, \ldots, m\}, \quad w \in \{1, \ldots, n-1\}, \tag{8}$$

$$\delta_t^{th}\delta_t^{cc} = 0, \tag{9}$$

$$x_t, \tilde{x}_t \geq 0, \quad 0 \leq h_t \leq x_t, \tag{10}$$

where $t = 0, 1, 2, \ldots$. In addition, our existence proof applies a technical assumption $x_t, d_t \leq K$ element-wise for some finite constant K. Thus, the number of trees or the diameter of each size, age, and species class cannot exceed some arbitrarily large value that never becomes binding and exceeds levels that are biologically attainable in a one-hectare forest.

In (1)–(10), the problem is to choose the number of harvested trees h_t and the timings for thinning and clear-cuts, δ_t^{th} and δ_t^{cc}, over the time periods $t = 0, 1, 2, \ldots$. Equation (3) defines the number of trees in new cohorts at the end of each period before harvesting in the cases with and without clear-cuts. Equation (4) gives the number of trees in various cohorts at the beginning of the periods. Equation (5) includes the survivability of trees over periods. If any cohort reaches age n, it will die naturally. Equation (6) determines the initial tree diameters depending on whether regeneration is natural or artificial, and equations (7)-(8) determine their development thereafter.

One possible solution regime includes only thinning ($\delta_t^{cc} = 0, t = 0, 1, 2, \ldots$), i.e. it is the CCF regime. Solutions with a chain of clear-cuts and possibly thinnings are RF regime solutions. In intermediate cases, clear-cuts only occur occasionally (e.g. at the initial state) although thinning is the dominant harvesting method. Empirical details for the stand growth models are given in Appendix A including the interaction between tree species in growth, mortality and regeneration as well as the connection between the matrix and the individual-tree model. The economic parameter values, including the estimated model for stochastic prices, are in Appendix B.

3. Existence of an optimal solution as a stationary Markov policy

The forest management problem (1)–(10) can be described as a dynamic discrete-time decision process $(S, \{D(s) : s \in S\}, q, q_0, \pi, \gamma)$, where the components of the tuple are defined as follows:

(i) S stands for the state space, where $s_t = (x_t, d_t, p_t) \in S$ represents the number of trees x_t, tree diameters d_t, and market prices p_t.
(ii) A is the action space, where each action $a_t = (h_t, \delta_t^{th}, \delta_t^{cc}) \in A$ represents a harvesting decision. We note that δ_t^{th} and δ_t^{cc} are independent variables and not functions of h_t and x_t. This ensures that when studying continuity with respect to δ_t^{cc} and δ_t^{th}, taking the limit ensures that both of these variables remain 1 or 0 throughout the limiting procedure. We define $D(s)$ as a non-empty subset of A, representing the set of actions feasible at state S (i.e., the set of implementable harvesting decisions). Requirement $a_t \in D(s_t)$ ensures that the number of trees harvested cannot exceed the number of trees available. We also denote $\operatorname{gph} D = \{(s, a) \in S \times A : a \in D(s)\}$.
(iii) The transition probability is $q : \operatorname{gph} D \to \mathcal{P}(S)$, where $\mathcal{P}(S)$ denotes the set of all probability measures on S. In this notation, the transition probability q is directly defined by the ecological growth model equations presented in Appendix A and the price dynamics given in Appendix B. That is, q describes the probability measure of variables $x_{j,k,t+1}$ and p_t given $(s_t, a_t) \in \operatorname{gph} D$.
(iv) $q_0 \in \mathcal{P}(S)$ is the so-called initial state distribution.
(v) $\pi : \operatorname{gph} D \to \mathbb{R}$ is a measurable reward function.
(vi) $\gamma \in (0, 1)$ is a discount factor.

A policy is understood as a mapping from past observations to a distribution over the actions (harvesting decisions). Formally, a policy is defined as a sequence $\mu = (\mu_t)$ of transition probabilities, where $\mu_t(\cdot|\tau)$ provide the conditional probabilities of various actions given a history of forest management decisions $\tau = (s_0, a_0, \ldots, s_t)$.

We write Δ for the set of all policies and $\mathbf{F}$ for the subset representing all deterministic policies (decision functions), i.e. $f \in \mathbf{F}$ iff $f : S \to A$ is a deterministic function such that $a = f(s) \in D(s)$ for $s \in S$. A policy is called Markov if the distribution depends only on the last state. We write $\Delta_M \subset \Delta$ for the set of all deterministic Markov policies. Finally, if the policy does not change over time, we define deterministic stationary policy as a Markov policy (f_t), where $f_t = f$ is independent of t. We identify $\mathbf{F}$ with the set of all deterministic stationary policies.

The value of a policy μ is defined as the total expected discounted reward that is encountered while the policy is executed starting from initial state s:

$$V_\mu(s) = \mathbb{E}_\mu\left[\sum_{t=0}^{\infty} \gamma^t \pi(s_t, a_t) | s_0 = s\right], \tag{11}$$

where the expectation is taken with respect to the probability measure defined by policy μ. The goal is then to find a policy μ that maximizes the expected reward,

$$V^\star(s) = \sup_{\mu \in \Delta} V_\mu(s)$$

for any initial state $s \in S$. The optimal policy $\mu^\star$ is related to $V^\star$ through $\mu^\star \in \operatorname*{argmax}_{\mu \in \Delta} T^\mu V$, where

$$(T^\mu V)(s) = \mathbb{E}_{a \sim \mu(\cdot|s)}\left[\pi(s, a) + \gamma \int_S V(y) q(dy|s, a)\right] \tag{12}$$

is the Bellman operator. While $\mu^\star \in \Delta$ can be a randomized policy, the existence of a corresponding optimal deterministic stationary policy means that for each $s \in S$ the supremum in

$$\sup_{a \in D(s)} \left\{\pi(s, a) + \gamma \int_S V^\star(y) q(dy|s, a)\right\}$$

can be achieved by selecting the actions according to a deterministic function $f^\star : S \to A$, $f^\star \in \mathbf{F}$. The existence of a solution is guaranteed by the following result:

Theorem 1. *The stochastic optimization problem (1)–(10) corresponds to a dynamic discrete-time decision process $(S, \{D(s) : s \in S\}, q, q_0, \pi, \gamma)$, which has an optimal solution that can be represented as a deterministic stationary Markov policy. That is, there exists a function $f \in \mathbf{F}$ such that $\max_{\mu \in \Delta} V_\mu(s) = \max_{f \in \mathbf{F}} V_f(s)$ for each $s \in S$.*

The proof of the proposition is given in Appendix C along with a discussion of technical details. The result has a number of important implications for the solvability of the problem. In particular, this allows us to use efficient reinforcement learning (RL) algorithms, such as proximal policy optimization (PPO) (Schulman et al., 2017), that have been popularized in machine learning literature to solve stochastic optimization problems that can be represented as Markov decision processes.

In reinforcement learning, the process of finding the optimal policy parameters is conceptualized as an agent that learns by interacting with its environment and by gathering experiences that will help the agent evaluate what was successful and explore what potentially optimal actions are available in various situations (Sutton and Barto, 2018). The interaction

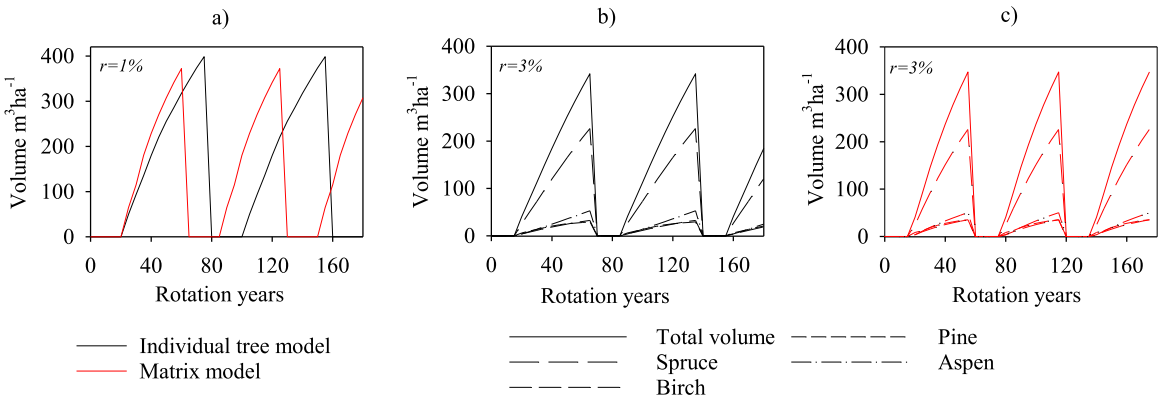

Fig. 1. Deterministic Samuelson-Faustmann model for pine and a mixed stand. a) Interest rate 1%, BLV and rotation € 14064, 80 yrs in the individual-tree model and 17623€, 60 yrs in the matrix model. b) Mixed stand, individual-tree model, interest rate 3%, BLV and rotation 65 65 yrs. c) Mixed stand, matrix model, interest rate 3%, BLV and rotation 1099 55 yrs. (BLV=bare land value, initial state, Appendix A) .

between a learning agent and its environment is defined using the formal framework of Markov decision processes, but in contrast to dynamic programming, the transition probabilities do not need to be known in exact form when learning the policy parameters. The RL methods also avoid an exhaustive search over the entire state and action spaces, which makes them computationally more efficient than classical methods, such as policy and value iteration, for large state and action spaces.

The environment is commonly defined as a large simulation model representing how the actual environment would respond to the actions taken by the agent. In our case, the environment is essentially represented by the forest growth model. The agent and the environment interact at discrete time steps $t = 0, 1, 2, \ldots$. At each time step t, the agent receives a description of the forest stand state s_t and, based on the state, selects an action $a_t = (h_t, \delta_t^{th}, \delta_t^{cc})$, where the agent chooses between thinning, clear-cutting, and doing nothing, and how much is thinned. As a consequence of its action, the agent receives a reward, i.e. a per-period profit, $\pi(s_t, a_t)$, and observes a new stand state s_{t+1} one time step later. The Markov decision process underlying the environment and agent together thereby give rise to a trajectory of states, forest management decisions, and gross profits: $s_0, a_0, \pi_0, s_1, a_1, \pi_1, \ldots$. In RL, each of these trajectories begins independently of how the previous one ended. As the objective of the agent is to maximize the expected net present value (NPV) over each trajectory, the agent can learn from the rewards it has received by pursuing various forest management policies, as represented by the sequence of actions it has taken.

To find a policy that maximizes the expected discounted reward, we apply a modified version of proximal policy optimization (PPO) to sequentially maximize a surrogate objective function that serves as an approximate lower bound for the expected reward (Schulman et al., 2017). While the original PPO algorithm treats all actions as continuous, the algorithm can easily be modified to handle a mixture of discrete and continuous actions, where the continuous actions are viewed as parameters of the discrete actions (Fan et al., 2019). Therefore, similar to Malo et al. (2021), we approach the problem of continuous action and state spaces using the notion of parameterized action spaces, where the choice between thinning, clear-cutting, and doing nothing is modeled as discrete actions and the actual harvesting quantities are modeled as their continuous parameters. Further details on the algorithm are discussed in Appendix D.

4. Results

Model (1)–(10) can be solved under a various number of details. To reveal their role, we proceed from a simple setup toward more detailed ones. For this purpose, we first remove thinning (but maintain all other model properties) and solve the rotation as a single optimized variable in spirit of the Samuelson (1976) interpretation of the Faustmann (1849) land value analysis and the closely related "tree paradigm" literature (Miller and Voltaire, 1983). Next, we include optimized thinning and finally both thinning and multiple tree species.

4.1. Samuelson-Faustmann specification

In purely deterministic solutions for pine (Fig. 1a), the maximum sustainable yield rotation is 70 years (not shown), while the economic rotation is 80 (60) years when interest rate is 1% (3%, not shown). In Fig. 1a, the matrix model yields a 20-year shorter rotation and overestimates bare land value (BLV). In Figs. 1b and c, the interest rate is 3%, the stand consists of four tree species (which in the S-F setup that ignore partial harvests, are all clear-cut simultaneously), and the individual-tree based optimal rotation is 65 years, while it is 10 years shorter in the matrix model and again BLV is overestimated. These are the first signs of the matrix model tendencies to overestimate volume growth leading to differences in both optimal harvests and profitability.

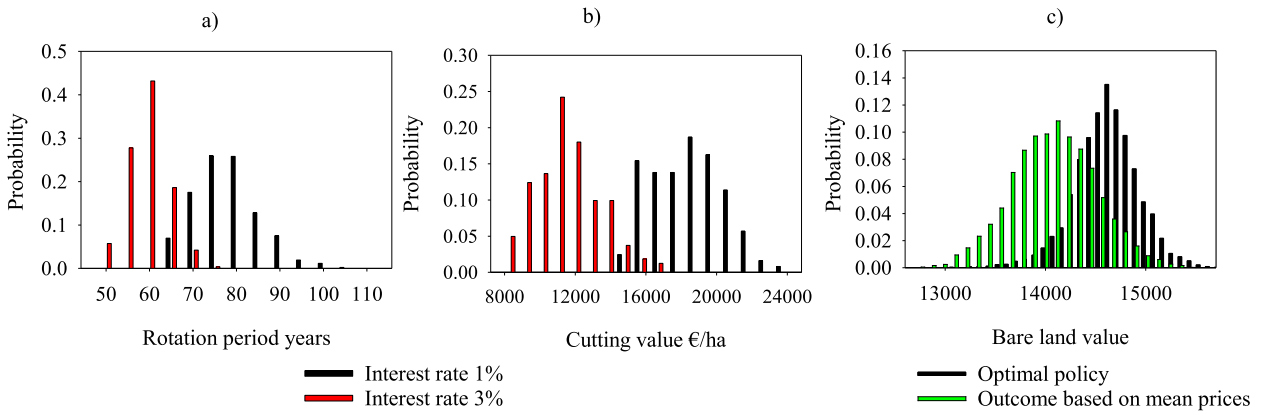

Fig. 2. Samuelson-Faustmann model with stochastic price for pine. a) Expected rotations 77.9 and 59.9 years. b) Interest rate 1%, expected cutting value € 18188, SD € 1968, 3% expected cutting value € 11704, SD € 1843. c) Interest rate 1%, optimal policy expected BLV 14597, SD € 326, policy trained in the deterministic environment, expected BLV 14089, SD € 424 (BLV=bare land value, SD=standard deviation).

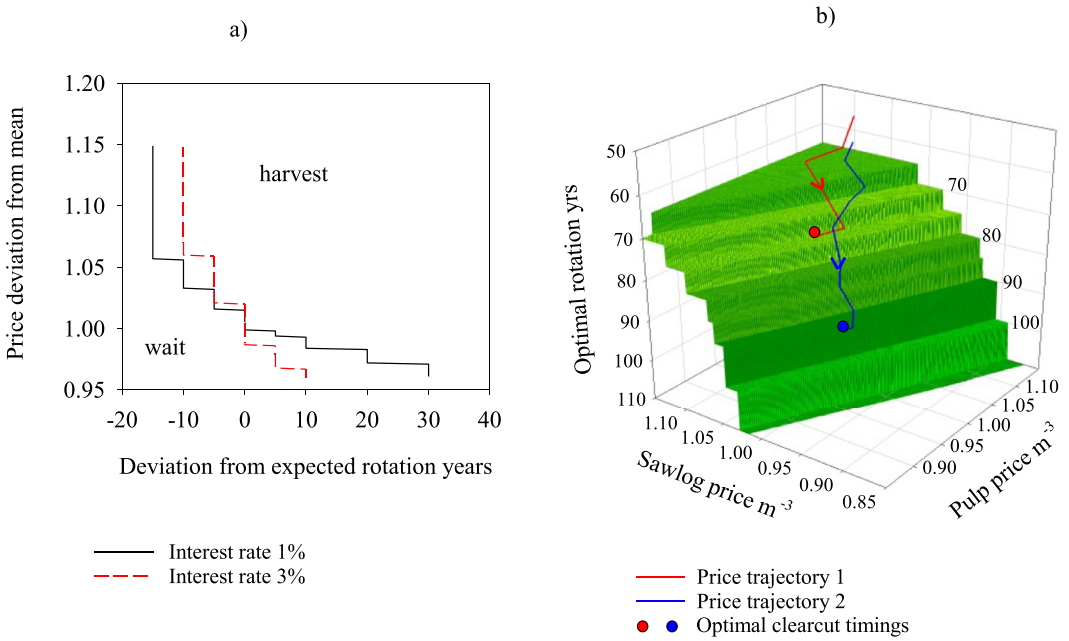

Fig. 3. Reservation price schedules (pine). a) With equal price deviations, b) Independent saw timber and pulp price deviations, interest rate 3%.

Following the tree paradigm, we next include stochastic growth into the S-F setup. This yields expected clear-cut harvests and BLVs within 0.1% of their deterministic levels with very low standard deviations (SD) (∼ 1%). Additionally, growth stochasticity is not high enough to produce any deviation from the optimal deterministic rotations. These outcomes follow from the low stochasticity in tree diameter growth and from the minor role of stochasticity in natural regeneration in the S-F solutions, where rotation is rather short.

In contrast to stochastic stand growth, i.i.d. stochastic pulp and saw timber prices (Appendix B) imply a deviation of up to 30 years compared with the expected rotation (Fig 2a). Given 1% and 3% interest rates and 5000 stochastic realizations, the expected rotations are slightly below their deterministic levels (77.9 and 59.9 years vs. 80 and 60 years). Figs. 2a,b show that the optimal cutting is not determined by barrier rule for age or cutting value. The BLV distributions in Fig. 2c show that the optimal policies yield higher expected BLVs and lower SDs compared with the policy computed under deterministic mean prices.

The optimal reservation price schedule in Fig. 3a shows that with a 1% interest rate, a price deviation of ca. +6% is enough to advance clear-cutting by 15 years and a deviation of ca. -3% causes a postponement of up to 30 years. However, clear-cut timing is less flexible with a higher interest rate. In Fig. 3a, the saw timber and pulp price deviations from the mean are of equal percentage, reflecting the positive covariance of these variables in our price model estimation (Table 4, Appendix B). Fig. 3b relaxes this coupling, implying that the reservation price schedule forms a surface with its diagonal as

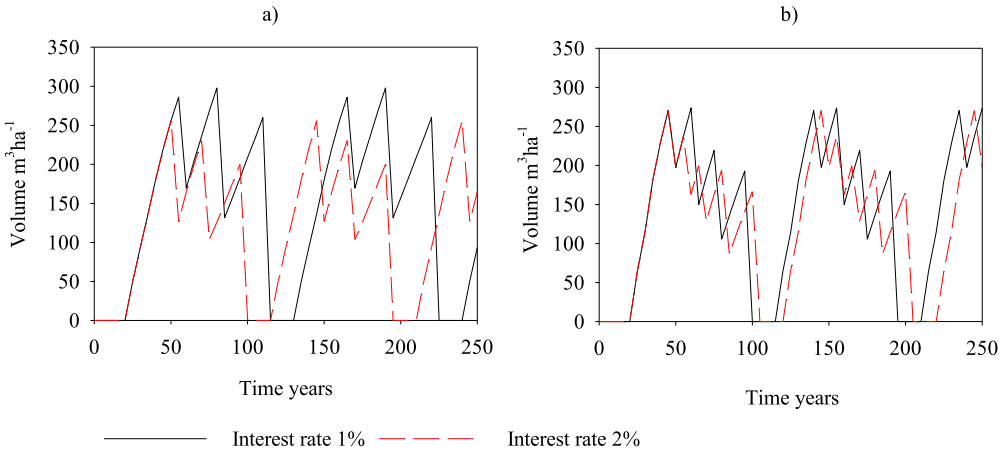

Fig. 4. Deterministic rotation forestry solutions (pine). a) Individual-tree model: optimal rotation 110 (95) years, BLV € 16676 (€ 5437), interest rate 1% (2%). b) Matrix model: optimal rotation 95 (100) years, BLV € 21393 (€ 7732), interest rate 1% (2%).

the two-dimensional schedule in Fig. 3a. A closer examination reveals that optimal timing reacts more sensitively to saw timber compared with pulp price.

In the stochastic realizations of Figs. 2b and c, the optimal stochastic policy produces a higher expected BLV compared with the policy computed under deterministic mean prices, but the differences are 3.6% (3.2%) with an interest rate of 1% (3%), i.e. the benefit from a flexible stochastic policy is rather low. Closely related to this, the realized prices at clear-cuts are € 61 and € 31.7 for saw and pulp logs, respectively (interest rate 1%), while the expected prices are € 58.6 and € 30.05. The explanation for the rather low benefit from this flexible policy is the low SD ($\sim$ 4%) of the saw and pulp log prices compared with expected prices in our estimated price model. To demonstrate these effects, we increased the SD in our price model to 30%. Given this change, the variation in optimal rotation realizations becomes larger, the benefit from optimizing under stochastic prices increases to ca. 40%, and reservation price levels always exceeded the expected price levels.

4.2. Optimized rotation, management regime choice, and thinning

Optimized thinning, deterministic setups, and interest rates of 1-2% imply that the RF regime with two or three thinnings is optimal for pine (Figs. 4a,b). Compared with the S-F model (Figs. 1a,b), thinning lengthens the rotation by 30–45 years and increases BLV by 19–28%. The matrix model overestimates the BLV by 28–42%, yields different numbers and timings of thinning, and different rotations compared with the individual-tree model. Volume developments in the models are additionally rather different. Similarly as in the S-F setup, stochasticity in stand growth alone does not cause any variations in harvest timing and only minor changes in cutting value.

In Fig. 5a, stochasticity in prices changes both the timing (5–10 years) and intensity of thinnings along with the timing of clear-cuts. This differs sharply from deterministic growth and no thinning (Fig. 3a), where stand state is a deterministic function of time and reservation price schedule is a priori determined in the time–price plane. In Fig. 5a, thinning reacts to stochastic prices, implying that the stand state varies at the moments of clear-cuts. Thus, harvest timing becomes dependent on varying stand state in addition to price. To obtain a comparable figure with schedules without thinning, the reservation price schedules in Fig. 5b assume the state from the deterministic solution after the second thinning (Fig. 5a) as the initial state. Comparing with the S-F model (Fig. 3a), clear-cut timing is more sensitive to timber price variations with optimized thinning. With a 1% interest rate, ca. 38% of clear-cuts are expected to deviate by 10 years or more from the deterministic 110-year rotation (Fig. 5c). 5000 realizations with simultaneous price and growth stochasticity show an increase from the deterministic BLV of € 5437 to an expected BLV of € 5569 (SD € 142). Fig. 5d compares the policy computed under deterministic environment and the policy computed under stochastic prices and growth. Comparing with BLVs in the S-F model shows that the share of SD on BLV is lower under thinning: 1.7% vs. 2.2%, with an interest rate of 1% and 2.5% vs. 3.7% with an interest rate of 3%. Thus, flexibly optimized thinning acts as risk-spreading.

Including optimized thinning for deterministic Norway spruce implies an infinitely long rotation, i.e. optimality of the CCF regime and convergence toward a cyclical steady state (Fig. 6). The differences with solutions without thinning are remarkable, as the rotations in the S-F solutions are 90 and 70 years and BLVs are 25% and 96% lower (interest rates 1% and 3%).

Fig. 6a shows how the matrix model produces faster initial growth, earlier and more frequent thinning (20 vs. 15 years), and smoother volume development compared with the individual-tree model. In Fig. 6b (upper panel), the tree classes in the individual-tree model are grouped into 5-cm size classes. The matrix model overestimates the number of trees at both ends and particularly the size of the largest trees. Together, these differences imply that the matrix model overestimates the BLV by a factor of three, although the overestimation is lower for the steady-state mean annual net revenues (26%).

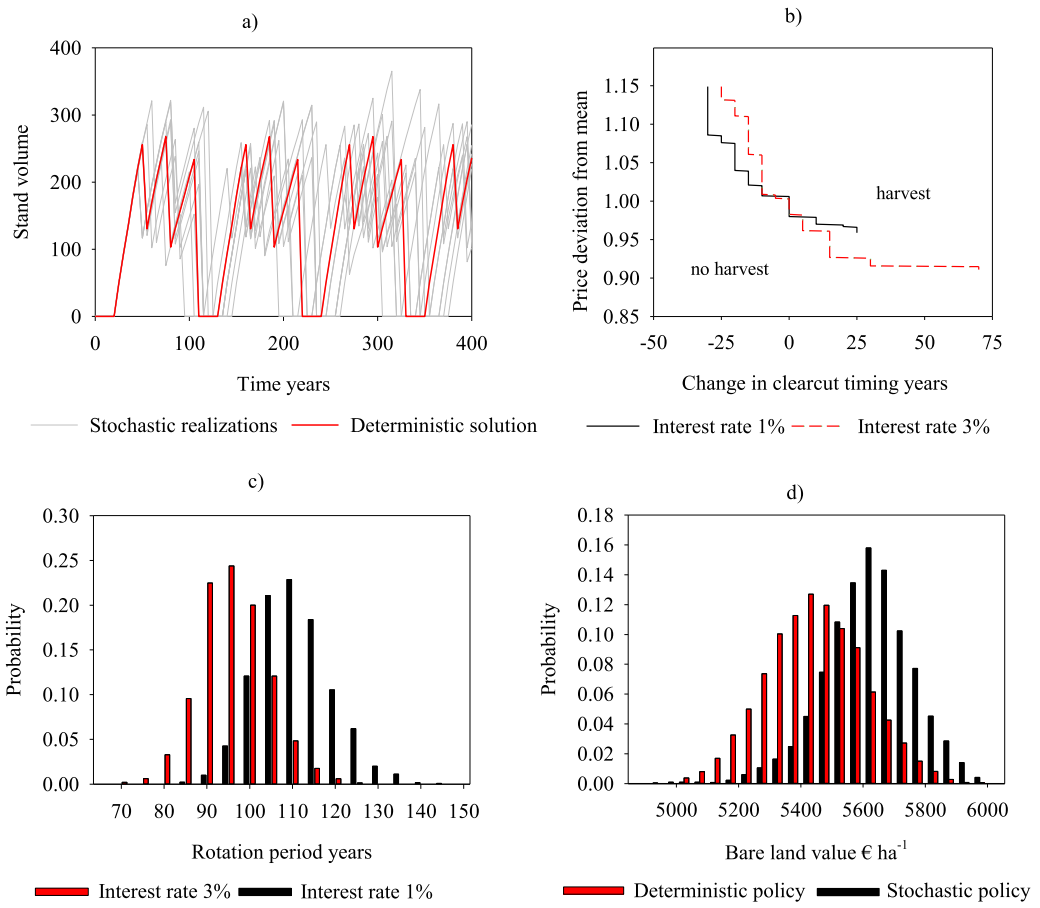

Fig. 5. Optimal stochastic rotation solutions (pine). a) Growth and price stochasticity, interest rate 1%, b) Reservation price schedules, c) Rotation period realizations d) Bare land value (BLV) realizations, stochastic price and growth, interest rate 2%, expected BLV under optimal policy € 5569 (SD € 142) and € 5542 (SD € 158) under policy trained in the deterministic environment.

In Fig. 6b (lower panel), the individual-tree size structure before the steady-state harvest is shown by 416 (nonzero) state variables (tree size classes and diameters). Accordingly, the number of (nonzero) control variables is 208 in addition to the binary harvest timing variables.

Pure growth stochasticity (not shown) causes 10-year deviations from the deterministic 25-year harvesting interval and decreases the expected harvest interval to 23 years (interest rate 1%). The SD in harvest levels is 35% (20%) with a 1% (3%) interest rate. Compared with the S-F model, where growth stochasticity has negligible effects, these variations are high but still low compared with stochasticity and SD in natural regeneration, which is ca. 2.8 times the expected value.

To study price stochasticity in the CCF regime, we first assume an initial state from the deterministic steady-state solution five years after the optimal harvest. In Figs. 7a and b, the mean price would imply a harvesting rate of 25 (20) years with an interest rate of 1% (3%) but a 15% higher price advances the harvest by 10 years. Similarly as in the RF regime (Fig. 5b), the change in harvest timing is more sensitive to low price realizations (and low interest rate), and prices that remain below 10% of the mean postpone the harvest by 20 years or more. Fig. 7c is based on 57 500 price realizations and the stand state after previous harvest varies between realizations, implying that at each date the minimum price with harvest realization varies as a consequence of varying stand state. This allows computing both the mean reservation price and SD. In Fig. 7d, both price and stand growth are stochastic, implying that the SD from the mean timing is larger (575 000 realizations). Thus, thinning and/or including both growth and price stochasticity imply that the reservation prices are dependent on stochastically evolving stand growth in contrast to the simplified schedules in Figs. 7a and 5b.

Despite the differences in timing, the size variation in the harvested trees is negligible, i.e. 19.0 cm with an SD of 4 mm (interest rate 3%) and 19.1 cm with an SD of 1 mm (interest rate of 1%). Instead, the variation occurs in the number of trees harvested, implying an SD of 31 trees when the expected number of harvested trees per harvest is 273.

A run with 5000 realizations produces an expected BLV equal to € 622 with an SD of € 41. Mean BLV decreases to € 601 and SD increases to € 43 if the harvesting policy is trained based on deterministic mean prices (Fig. 7e). In Fig. 7f, the deterministic solution is contrasted to 10 realizations under stochastic price and stand growth. In stochastic solutions, the mean steady-state volume $112m^3$ is above the deterministic average steady-state volume of $107m^3$.

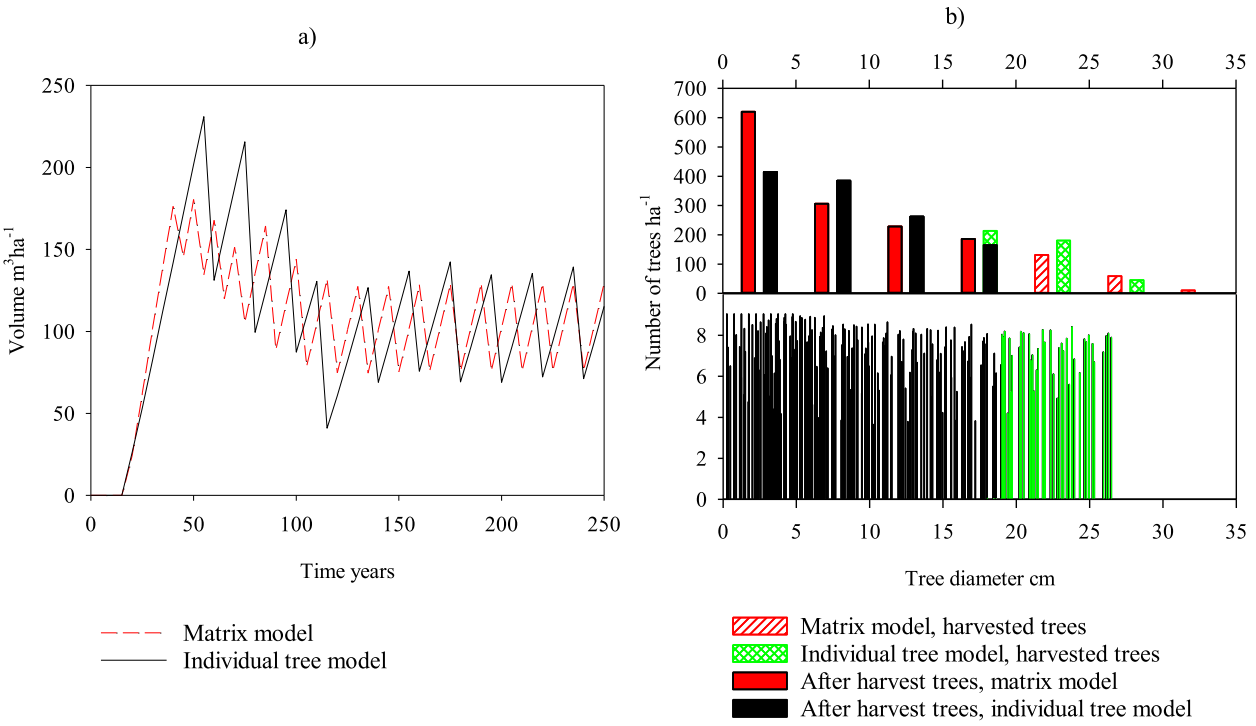

Fig. 6. Deterministic continuous cover solution (spruce). a) Bare land value (BLV) in the individual-tree (matrix) model € 602 ha-1 (€ 1900 ha-1). b) Steady-state stand structures in the matrix and individual-tree models. Interest rate 3%.

Table 1
Comparison of matrix and individual-tree deterministic models.

Species composition	Bare land value €		Harvest interval years		Average revenues net € a^{-1}		Harvest m^3a^{-1}		Average volume m^3		Diameters of felled trees cm	
	M	I	M	I	M	I	M	I	M	I	M	I
spruce	1900	602	15	20	215	171	5.3	4.4	103	107	20–35	19–26
spruce, birch	2039	595	15	20	236	168	6.1	4.9	104	105	20–35	12–25
spruce, birch, pine, aspen	2339	653	15	20	225	159	6.3	5.3	105	110	15–35	14–25

Note: M=matrix model, I=individual-tree model, interest rate 3%. All figures per hectare. Excluding BLV, figures are for a CCF steady-state cycle.

Similarly as with RF solutions for pine, optimized thinning decreases the relative SD of BLVs compared with solutions without thinning, i.e. 2.3%, 4.7%, and 7.4% vs. 1.6%, 2.5%, and 6.6% (interest rates 1%, 2%, and 3%). Intuitively, more frequent fellings and a lower number of trees per felling decrease the overall stand-level risk.

Price stochasticity increases the BLV in the individual-tree model by 3.3% and by 1.4% in the matrix model. Accordingly, the steady-state net revenues increase by 7.3% in the individual-tree model and by 3.5% in the matrix model. In both models, price stochasticity increases the mean steady-state volume, but the increase is 5% in the individual-tree model and 1% in the matrix model. The simpler matrix model thus tends to dilute the effects of stochasticity.

4.3. Continuous cover forestry for mixed-species stands

With four tree species, the number of nonzero tree classes in the individual-tree model is 950, implying, along with the variables for tree diameter, that the number of (nonzero) state variables is 1900 and the number of control variables is 950 in addition to the binary harvest timing variables. Assuming interest rates of 1–3%, the optimal deterministic solution in both the matrix and individual-tree models is a spruce-dominated CCF solution (Fig. 8a). However, the matrix model solution is more even over time, and harvesting interval is shorter. Additionally, the matrix model overestimates the number of trees in both the small and large size classes (Fig. 8b).

A major difference between the models' outputs are in their objective values, and the matrix model produces more than three times higher BLV compared with the individual-tree model (Table 1). This is caused by differences in harvest timing and the levels of yield and net revenues. The first harvest is 15 years sooner in the matrix model solution, and the largest harvested tree is 32.5 cm in the steady-state harvests and 25.6 cm in the individual-tree model. Additionally, the matrix model suggests that a mixed-species stand clearly outperforms single-species stands, while a similar outcome

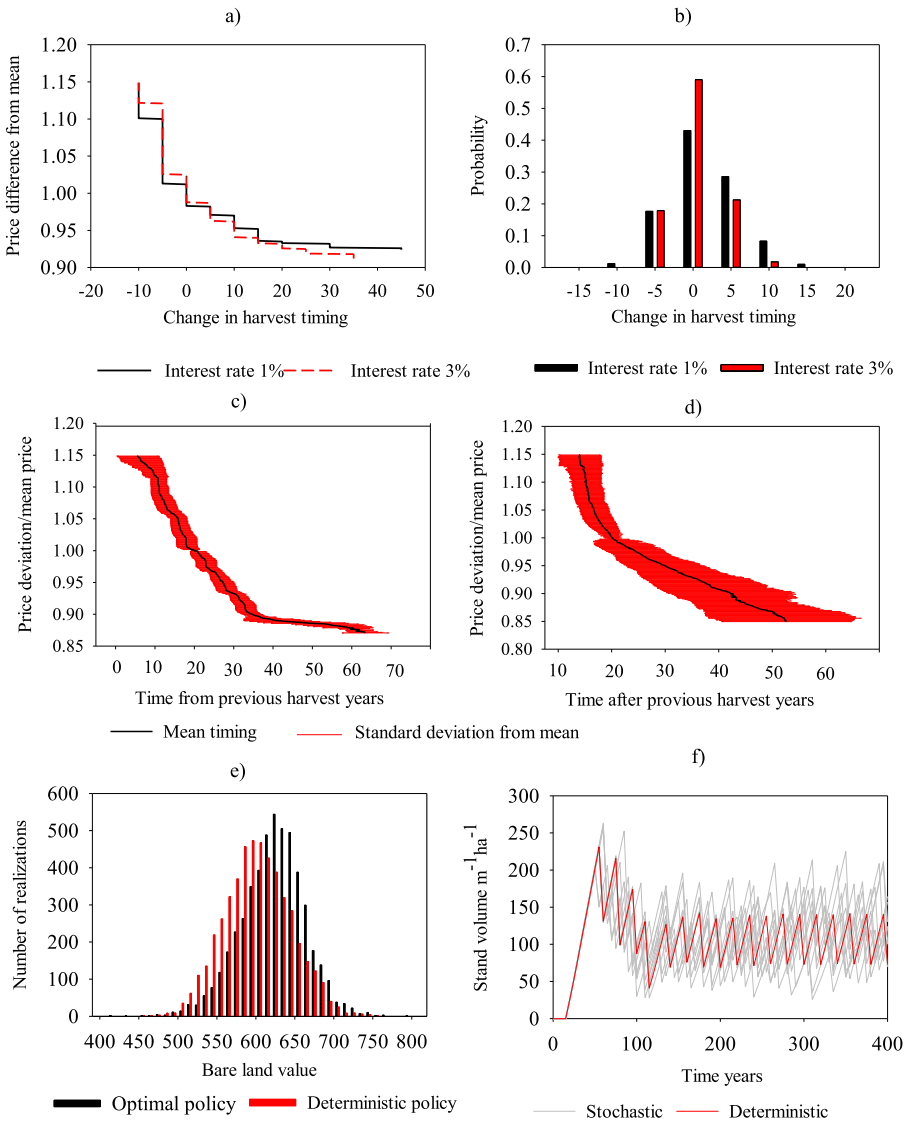

Fig. 7. Stochastic continuous cover solutions (spruce). a) A steady-state reservation price schedule under fixed previous harvest state (stochastic price), b) Harvest timing distributions as deviations from deterministic timing (stochastic price), c) A steady-state reservation price schedule under stochastic previous harvest state (stochastic price), d) A steady-state reservation price schedule under stochastic previous harvest state (stochastic price and growth), e) Bare land value (BLV) distributions (stochastic price), f) Deterministic solution and stochastic realizations (stochastic price and growth), interest rate 3% in c-f.

is less clear in the individual-tree output (Table 1). The models' differences at the steady-state cycle are less dramatic but still remarkable: the matrix model net revenues are 25-41% and the timber output is 18-24% higher compared with the individual-tree model.

The main effect of growth stochasticity occurs in natural regeneration, and the SD averages ca. 2.5 times the mean number of ingrowth seedlings. This causes variability in steady-state net revenues per harvest with an SD of 22% (Fig. 9b). To a large extent, this high variability is cancelled out in the BLV distribution, where the fraction of SD from the expected BLV is 6.3% (Fig. 9c). Growth stochasticity lengthens the expected harvesting interval from its deterministic 20 years to 21.3 years (Fig. 9a). As a consequence, growth stochasticity increases the average stand volume and average annual steady-state harvest yield (Figs 10). Given the S-F setup, growth stochasticity does not have any effects on clearcut timing implying that Fig. 1b shows the optimal S-F solutions for mixed species forest with and without stochasticity. Thus, the comparison with Fig. 10 reveals the effects of fully utilizing the model dimensions.

By comparison, the stochastic matrix model results show lower SDs in BLV (2.2%) while the SD with the individual-tree model is 6.3%. In the matrix model, stochasticity shortens the expected harvesting interval from 15 to 12.8 years (Fig. 9a). Additionally, the SDs of the steady-state harvest, stand volume, and net revenues are lower in the matrix model, suggesting that it dilutes the effects of growth stochasticity, as in the case with single-species spruce CCF solutions.

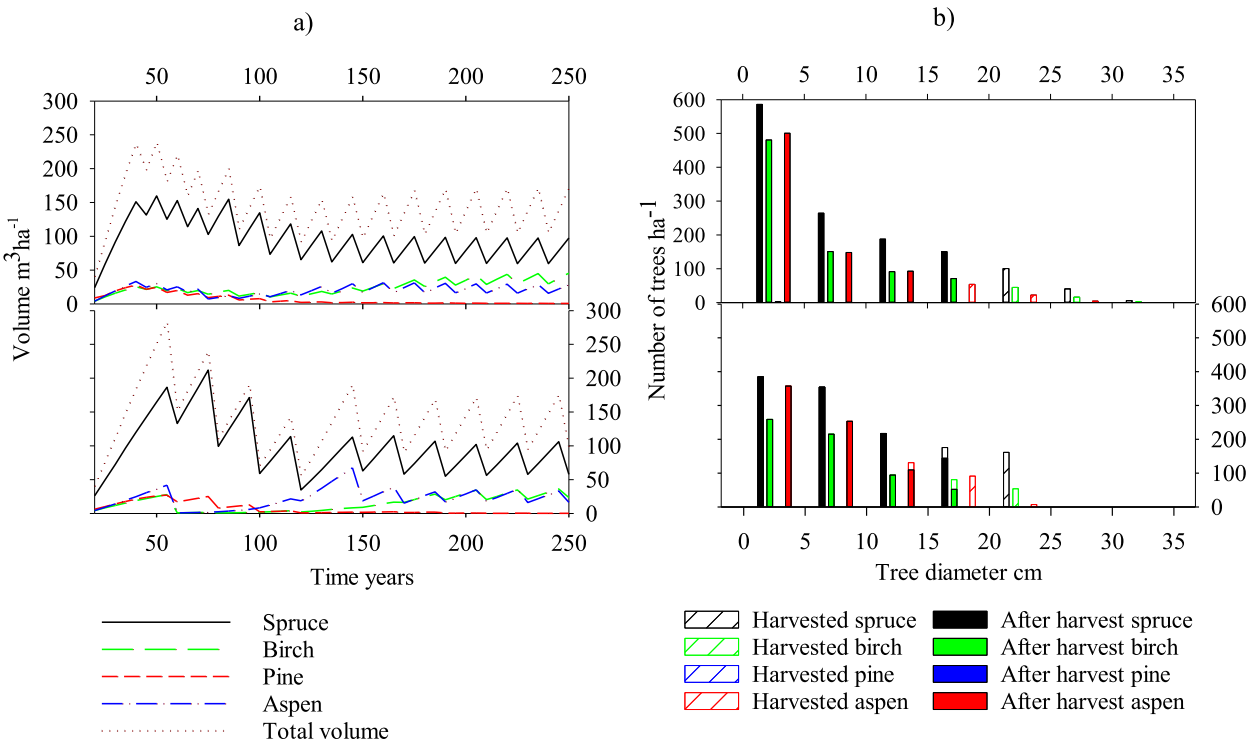

Fig. 8. Mixed-species deterministic solutions. a) and b) Upper (lower) panel matrix (individual-tree) model. Interest rate 3%.

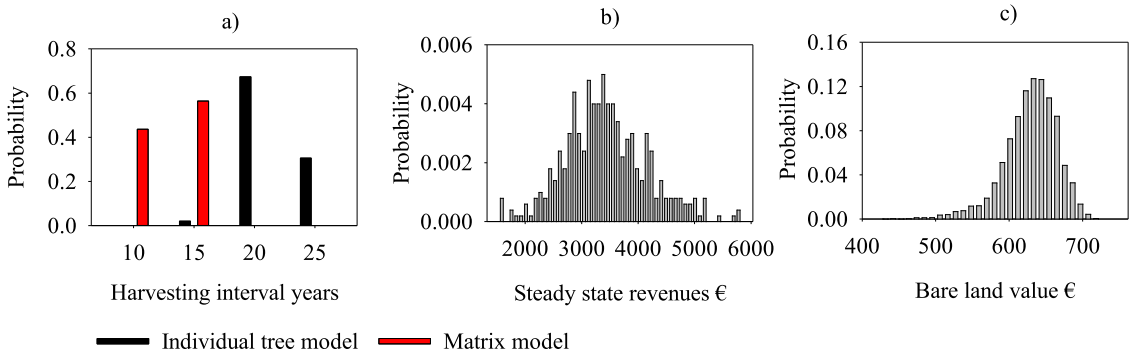

Fig. 9. Mixed-species model with stochastic stand growth. a) Individual-tree (matrix) model expected harvesting interval 21.3 years (12.8). b) Expected revenues € 3425, SD € 737. c) Expected BLV € 628, SD € 41. Interest rate 3% (b,c).

5. Discussion and conclusions

The most commonly applied stochastic approach for forest resources, i.e. the "tree paradigm", only optimizes rotation length and specifies stand clear-cut value as geometric Brownian motion, with the result that stochasticity increases both expected rotation and BLV (Chang, 2005; Miller and Voltaire, 1983; Willassen, 1998). Brownian motion with a fixed initial state after each harvest implies that stochasticity originates from stand growth. In our empirically detailed model and data, including only growth stochasticity into this Samuelsson-Faustmann model (i.e. no partial harvests) does not lead to any departures from the deterministic optimal rotation.

Given the specification with no partial harvests, our results on i.i.d. price stochasticity are qualitatively similar to Brazee and Mendelsohn (1988) and Insley and Rollins (2005), but we find that clear-cut timing reacts more sensitively to saw timber compared with pulp price, and that timing is more flexible under a low interest rate. In our results, the stochastic policy increases the expected BLV ca. 3.5% above the outcome of the deterministic policy, while the gain is 42–88% in Brazee and Mendelsohn (1988) and as high as 550% in Insley and Rollins (2005). In our setup, the low increase in expected BLV follows from the low SD in our price data. When the SD in prices is increased to similar levels as in Brazee and Mendelsohn (1988), the gain from optimal policy in the expected BLV (~ 40%) is closer to their results.

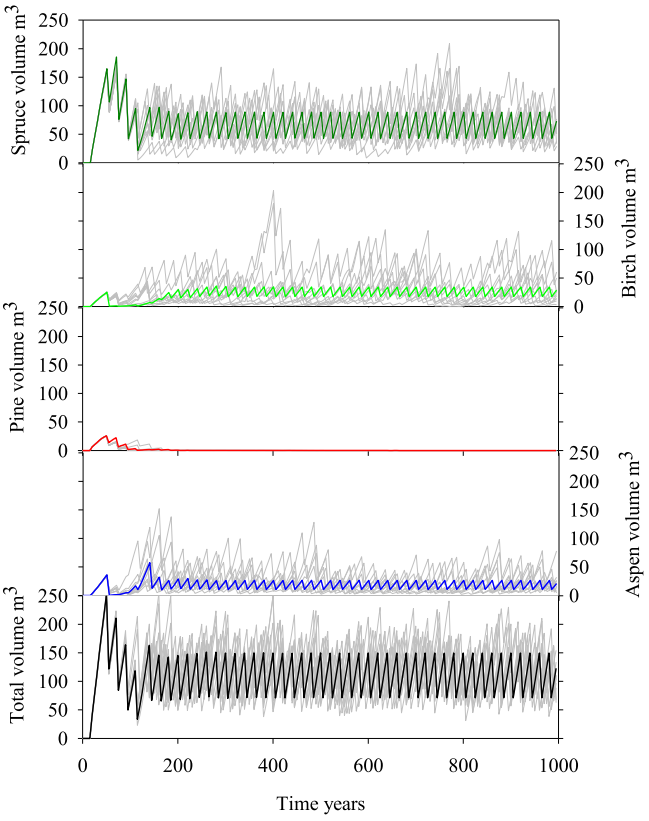

Fig. 10. The optimal deterministic and 10 stochastic realizations for a mixed-species stand.
Notes: Interest rate 3%, Expected BLV € 628.

While Clarke and Reed (1989) and Reed and Clarke (1990) make a sharp distinction between age-dependent and size-dependent models, our setup includes these features simultaneously. Steps toward this direction are taken in Helmes and Stockbridge (2011), Sødal (2002), and Alvarez and Koskela (2007) by the inclusion of thinning (partial harvests) but without expanding the model dimensions. In our size- and age-structured model, the inclusion of thinning lengthens the rotation period, which, depending on tree species, may become infinitely long. In these cases, the solution represents continuous cover forestry, i.e. harvesting based entirely on thinning instead of clear-cuts. Additionally, after including thinning, the maximized stand value increases (19–900%) and optimal solutions react more sensitively to both stochastic forest growth and stochastic wood prices. However, with thinning, SD becomes lower, suggesting that more frequent smaller harvests act as risk-spreading. Together these results reveal the clear reason for proceeding from the classic S-F setup and the tree paradigm to models with extended dimensions.

The optimization of intermediate harvests or thinning brings out the model dimensions and the curse of dimensionality. The individual-tree structure applied here (with 200–2000 state and control variables) is an exception even without stochasticity. Haight and Monserud (1990a,b) apply an individual-tree model, solve both RF and CCF solutions but group harvested trees into seven size classes, and do not apply flexible optimization of harvest timing. Tahvonen (2011) uses a fully continuous diameter variable but assumes only one ingrowth tree class per cohort and circumvents the mixed-integer control variables by assuming a fixed harvesting interval. Fixed harvest timing is a clear departure from the optimal timing/stopping spirit and distorts the solutions when the initial state is far from the CCF steady state and when the question is CCF/RF optimal choice.

In previous research with i.i.d. price stochasticity (Brazee and Mendelsohn, 1988; Insley and Rollins, 2005), the reservation price schedules are two-dimensional, decreasing the relations between price and clear-cut age. We show that including partial harvests with price stochasticity or both price and growth stochasticity changes this established result and the price/timing combination becomes dependent on the forest state that evolves stochastically. Computation realizations suggest that the effects of a stochastically evolving stand state make the guidance of the mean timing schedule inaccurate. Optimal timing must be based on observing both stand state and prices.

Within stochastic Markov decision models, Haight (1990, 1991) include i.i.d. stochastic timber price within the matrix model approach and assume a feedback thinning function that maps the stand cutting value with the number of harvested trees. The function parameters are optimized by Monte Carlo simulation, and the results suggest no harvest when the stand

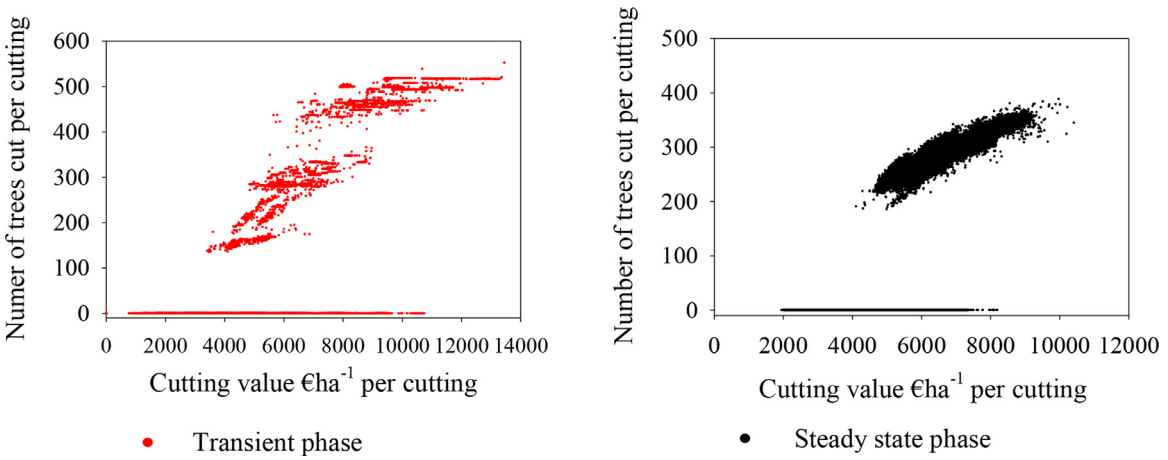

Fig. 11. Relationship of cutting value - number of harvested trees. Note: Spruce, interest rate 3%.

cutting value is low and a positive increasing harvest with a higher value of standing trees. Later, similar methods motivated by the aim to circumvent the curse of dimensionality are applied in Lu and Gong (2003, 2005) and Pukkala (2015), among others.[4]

We can analyze the accuracy of the thinning function approximation by running the optimal policy (that depends on both stand state and price) with a sufficiently high number of stochastic price realizations and by plotting the stand cutting value with the number of trees harvested (Fig. 11). The outcomes show that neither the steady-state solutions nor the solutions for the transient period can be accurately represented in our model by a function starting horizontally from the origin and continuing as an increasing concave function after some cutting value of standing trees, as in Haight (1990, 1991). This outcome is surely case-specific, but at least in our setup a single function cannot include the feature where waiting may be optimal, even with a high stand cutting value, if the price realization is (still) low.

However, in our results, price stochasticity changes the number of harvested trees but the minimum diameter of harvested trees is rather stable (albeit depends on the species and interest rate). Based on this, a feedback function that maps the cutting value of trees exceeding the typical minimum diameter and the number of trees harvested may produce a reasonable approximation.

A very different approach to circumventing the curse of dimensionality is developed in Kaya and Buongiorno (1987), where a matrix model is used to compute transition probabilities for a reduced (linear) system with a finite number of discrete stand states. Given 41–64 states and three discrete prices, dynamic programming is used to obtain a feedback harvest policy for stochastic problems. This approach is applied widely, and e.g. Buongiorno and Zhou (2015) study mixed-species CCF forestry with price and growth stochasticity and under various objectives. The question remains how well the reduced model is able to represent the original higher-dimensional nonlinear specification. Compared with the feedback thinning function or Buongiorno and his coauthors, we apply a generalization of the matrix model and maintain the relationship between the genuine state variables and harvesting control variables. Additionally, our state variables and stochastic prices are continuous and the results suggest that reducing the model dimensions is not harmless.

Getz and Haight (1989) compare a two-species matrix model with an individual-tree model and observe that the matrix model provides a method for simplifying individual-tree models with minimum loss of biological details. This is in line with the still prevailing view (Liang and Picard, 2013) that individual-tree models are not suited for optimization because of the excessively large number of state and control variables. In contrast, matrix models are written to contain fewer variables and are computationally suitable for optimization studies. We challenge this view based on optimization output, while these earlier comparisons observe deterministic stand development without harvests (Picard and Liang, 2014) or outcomes from pre-specified non-optimized harvests and do not compare the maximized objective values produced by the two model variants.

One possibility may be to decrease the size-class width in the matrix model to decrease the differences with the individual-tree model. In our model this, however, implies that the fraction of trees moving to the next size class may exceed one. Preventing this generates non-differentiability in the transition and mortality functions; an unfortunate feature for the optimization algorithms typically used with the matrix models. Similar problems occur in a specification without the "Usher property" (Picard and Liang, 2014), i.e. when some fraction of individuals are allowed to move up more than one class. These problems could perhaps be smaller if the period length in forest growth models were less than the typical 5 or 10 years.

[4] For a similar discussion in another context, see Maliar and Maliar (2022).

The curse of dimensionality is a barrier to model development in many fields of dynamic economics (Scheidegger and Bilionis, 2019), and optimizing the use of forest resources is not an exception. Specifying our model under a varying degree of dimensionality reveals that increasing model structure and details changes the optimization solutions both qualitatively and quantitatively. Methods developed in reinforcement learning are promising in proceeding beyond the existing boundaries and should have high potential in all sub-fields of resource economics, including the economics of climate change (Cai and Lontzek, 2019; Rolnick et al., 2022).

Data availability

Optimizing high-dimensional stochastic forestry *via* reinforcement learning

Appendix A. Forest Growth

Following Pukkala et al. (2013), the survivability functions in (5) are given as

$$\begin{aligned}\alpha_{j,q,w}(\tilde{x}_t,\tilde{d}_t) = [1+\exp(&-b_{0,j}-b_{1,j}\sqrt{\tilde{d}_{j,q,w,t}}-b_{2,j}\tilde{d}_{j,q,w,t}-b_{3,j}\sqrt{B_{p,q,w}(\tilde{x}_t,\tilde{d}_t)}\\ &-b_{4,j}\sqrt{B_{sp,q,w}(\tilde{x}_t,\tilde{d}_t)}-b_{5,j}\sqrt{B_{b,q,w}(\tilde{x}_t,\tilde{d}_t)+B_{a,q,w}(\tilde{x}_t,\tilde{d}_t)}\\ &-b_{6,j}\sqrt{B_{b,q,w}(\tilde{x}_t,\tilde{d}_t)+B_{a,q,w}(\tilde{x}_t,\tilde{d}_t)+B_{p,q,w}(\tilde{x}_t,\tilde{d}_t)}-b_{7,j}\varpi)]^{-1},\end{aligned} \tag{A.1}$$

where functions $B_{j,q,w}$, $j=p,sp,b,a$, represent the basal area (m^2ha^{-1}) of trees larger than $\tilde{d}_{j,q,w,t}$ diameter of Scots pine (Pinus sylvestris L.), Norway spruce (Picea abies (L.) Karst.), silver birch (Betula pendula Roth), and European aspen (Populus tremula L.) trees respectively. Functions B are continuous and bounded. Parameter ϖ denotes the length of period, which is 5 years. The diameter growth in (7) is given as

$$\begin{aligned}I_{j,q,w}(\tilde{x}_t,\tilde{d}_t,\varepsilon_t) = \exp\Big(&b_{8,j}+b_{9,j}\sqrt{\tilde{d}_{j,q,w,t}}+b_{10,j}\tilde{d}_{j,q,w,t}+b_{11,j}\ln(TS)+b_{12,j}SX+\ b_{13,j}\ln(B(\tilde{x}_t,\tilde{d}_t))\\ &+b_{14,j}\frac{B_{p,q,w}(\tilde{x}_t,\tilde{d}_t)}{\sqrt{d_{j,q,w,t}+1}}+b_{15,j}\frac{B_{sp,q,w}(\tilde{x}_t,\tilde{d}_t)}{\sqrt{d_{j,q,w,t}+1}}+b_{16,j}\frac{B_{b,q,w}(\tilde{x}_t,\tilde{d}_t)+B_{a,q,w}(\tilde{x}_t,\tilde{d}_t)}{\sqrt{d_{j,q,w,t}+1}}\\ &+b_{17,j}SD+b_{18,j}d_{j,t}SD+b_{19,j}a\Big)+\varepsilon_{j,w,t},\end{aligned} \tag{A.2}$$

where B denotes the total stand basal area (m^2ha^{-1}), TS the temperature sum (degree days, dd), SX an indicator variable if site fertility is "sub-xeric" instead of the more fertile "mesic", and SD is the standard deviation of the diameter (cm). The last term $\varepsilon_{j,w,t}$ denotes the stochastic variation around the expected diameter increment and is obtained by aggregating the stochastic tree-level model defined by Pukkala et al. (2013). For pine, we assume the site to be 1350dd/sub-xeric, and for spruce and mixed-species stands the site is 1100dd/mesic.

When the individual-tree model is applied as the matrix model, the right-hand side of (A.2) is divided by the size class width, implying that the modified (A.2) determines the transition rate (Liang and Picard, 2013). The matrix model we use in comparison is based on a 50-mm size class depth suggested by Bollandsås et al. (2008) and applied in many studies such as Malo et al. (2021).

Let $N_{in,j}$ and $p_{in,j}$ denote the number of ingrowth trees and the probability of ingrowth, respectively. The stochastic ingrowth function in Pukkala et al. (2013) is given by

$$\phi_j(\tilde{x}_t,\tilde{d}_t,\omega_t) = N_{in,j}\times p_{in,j}, \tag{A.3}$$

$$\begin{aligned}N_{in,j} &= \exp(b_{20,j}+b_{21,j}\sqrt{B(\tilde{x}_t,\tilde{d}_t)}+b_{22,j}\ln(B_b(\tilde{x}_t,\tilde{d}_t))+b_{23,j}B(\tilde{x}_t,\tilde{d}_t)+\omega_{j,t}),\\ p_{in,j} &= \frac{1}{1+\exp\Big(-b_{24,j}-b_{25,j}\ln(B_j(\tilde{x}_t,\tilde{d}_t))-b_{26,j}\sqrt{B_p(\tilde{x}_t,\tilde{d}_t)}-b_{27,j}B_p(\tilde{x}_t,\tilde{d}_t)}\\ &\quad -b_{28,j}B(\tilde{x}_t,\tilde{d}_t)+b_{29,j}SX\Big),\\ \omega_{j,t} &= \tilde{\rho}_j\omega_{j,t-1}+u_{j,t},\quad u_t=(u_{1,t},\ldots,u_{l,t})\sim N(0,\Sigma_u)\end{aligned} \tag{A.4}$$

where B_j denotes the total basal area of trees of species j, and $\omega_{j,t}$ denotes the residual variation in ingrowth for species j and a 5-year period t. The relative growth variation is especially large among small trees and in the ingrowth estimates. The ingrowths of consecutive 5-year periods and the residuals of predicted ingrowths are positively correlated in the Pukkala et al. (2013) model. The species-specific temporal autocorrelation coefficient is given by $\tilde{\rho}_j$. This is largely explained by the fact that one good regeneration year tends to generate ingrowth for several years. In addition to autocorrelation, the

Table A1
Parameter values for the growth model (Pukkala et al. 2013).

	Spruce	Birch	Pine	Aspen		Spruce	Birch	Pine	Aspen
$b_{0,j}$	5.871	0.433	2.333	0.433	$b_{15,j}$	-0.1473	-0.1044	-0.1399	-0.1044
$b_{1,j}$	1.536	2.284	1.518	2.284	$b_{16,j}$	-0.08277	-0.1554	-0.1797	-0.1554
$b_{2,j}$	0.122	-0.217	-0.083	-0.217	$b_{17,j}$	0	-0.1137	0	-0.1137
$b_{3,j}$	0.106	0	-0.602	0	$b_{18,j}$	0	0.004857	0	0.004857
$b_{4,j}$	0.69	-0.483	-0.686	-0.483	$b_{19,j}$	0	0	0	0.2017
$b_{5,j}$	0.226	0	-0.332	0	$b_{20,j}$	4.378	6.368	6.109	4.829
$b_{6,j}$	0	-0.266	0	-0.266	$b_{21,j}$	-0.0265	0	-0.844	-0.0236
$b_{7,j}$	0.465	0	0	0	$b_{22,j}$	0	0.496	0	0
$b_{8,j}$	-9.6448	-6.0405	-5.9901	-6.0405	$b_{23,j}$	0	-0.161	0	0
$b_{9,j}$	0.455	0.9309	0.5057	0.9309	$b_{24,j}$	1.001	1000	-0.375	1.202
$b_{10,j}$	-0.05741	-0.1441	-0.07699	-0.1441	$b_{25,j}$	0.641	0	1.045	0
$b_{11,j}$	1.4551	0.812	0.987	0.812	$b_{26,j}$	0	0	-0.556	0
$b_{12,j}$	-0.04910	-0.09846	-0.07558	0	$b_{27,j}$	0.046	0	0	0
$b_{13,j}$	-0.3081	-0.1424	-0.3593	-0.1424	$b_{28,j}$	-0.0658	0	0	-0.027
$b_{14,j}$	-0.02915	-0.03275	-0.141	-0.03275	$b_{29,j}$	0	-0.301	0.277	-1.07

residuals of predicted ingrowths are cross-correlated across species, which is captured by the random factors u_t that are assumed to follow a multivariate normal distribution and be independent within time periods t with a constant covariance matrix Σ_u. Parameter values are based on values presented in Pukkala et al. (2013), and the deterministic ingrowth model is obtained by averaging. The numerical values for regression coefficients $b_{\cdot,j}$ in (A.1)–(A.3) can be found in Table A1. When the initial state is bare land, it takes 20 years in mesic sites (25 years in sub-xeric sites) until the stand is artificially regenerated and the number of trees is 1750 (spruce) and 250 (other species in mixed-stands cases) or 2100 (pure pine stands). At these states, trees are equally distributed into ten classes with diameters 5.25, 5.75,..., 9.75 (in cm). The number of ingrowth trees are equally distributed into ten classes with diameters 0.25, 0.75,..., 4.75 (in cm), meaning that the ingrowth for individual size classes is $\phi_{j,q}(\tilde{x}_t, \tilde{d}_t, \omega_t) = \phi_j(\tilde{x}_t, \tilde{d}_t, \omega_t)/10$. In the matrix model, the artificially generated state is 1750 spruce trees and 250 other species in the 7.5-cm size class. Naturally regenerated trees enter the 2.5-cm size class.

Appendix B. Harvesting revenues and costs under stochastic prices

Harvesting revenues are defined separately for saw logs and pulpwood using species-specific market prices. The gross revenue per period is given by

$$R(h_t, d_t, p_t) = \sum_{j=1}^{l}\sum_{q=1}^{m}\sum_{w=1}^{n}(p_{1,j,t}v_{1,j,w}(d_{j,q,w,t}) + p_{2,j,t}v_{2,j,w}(d_{j,q,w,t}))h_{j,q,w,t}, \tag{B.1}$$

where $p_{1,j,t}$ and $p_{2,j,t}$ are the inflation-adjusted real prices of saw timber and pulpwood calculated using 2019 as a base year, while $v_{1,j,w}(d_{j,q,w,t})$ and $v_{2,j,w}(d_{j,q,w,t})$ are saw timber and pulpwood volumes per tree (Table B.3). Throughout, functions v are continuous and bounded functions of d.

The per-period species-specific saw timber and pulpwood prices are modeled as iid. random variables with a multivariate log-normal distribution,

$$\ln(p_t) \sim N(\mu_p, \Sigma_p), \tag{B.2}$$

where $\mu_p \in \mathbb{R}^{2l}$ represents long-term equilibrium prices and $\Sigma_p \in \mathbb{R}^{2l\times 2l}$ is the covariance matrix for normally distributed random innovations. While the long-term equilibrium prices can be specified as long-term means, we estimate the covariance structure using historical price data from the last 30 years obtained from the statistics of Natural Resources Institute, Finland. Even though the observed annual prices have short-term dependencies, we find them significant only within each 5-year period. Therefore, when modeling prices in steps of 5-year periods, it is justified to assume that the per-period prices are independent and identically distributed variables with a finite covariance matrix. To account for short-term effects when estimating Σ_p from annual data, we apply a Vector Error Correction Model (VECM) suggested by Engle and Granger (1987), which is essentially a vector autoregression that contains an equilibrium correction term. Recently, similar modeling approaches have been applied e.g. by Kuuluvainen et al. (2018), who studied price integration between domestic and imported saw logs in Finland. The annual price model is then given by

$$\ln(p_\tau) - \ln(p_{\tau-1}) = \Pi p_{\tau-1} + \sum_{i=1}^{k}\Gamma_i(\ln(p_{\tau-1}) - \ln(p_{\tau-2})) + \text{const} + \xi_\tau, \quad \xi_\tau \sim N(0, \tilde{\Sigma}_p), \tag{B.3}$$

where $\Pi \in \mathbb{R}^{2l\times 2l}$, $\Gamma_i \in \mathbb{R}^{2l\times 2l}$, for all $i = 1, \ldots, k$, and $\tilde{\Sigma}_p \in \mathbb{R}^{2l\times 2l}$ is the covariance matrix for normally distributed random innovations. The number of lags k is specified based on information criteria. The remaining parameters can be estimated using Johansen's maximum likelihood procedure (Johansen, 1995). The first term $\Pi p_{\tau-1}$ on the right-hand side of (B.3) is known as the error correction term. The term can be decomposed as $\Pi p_{\tau-1} = \eta\beta' p_{\tau-1}$, where $\eta \in \mathbb{R}^{2l\times 2l}$ and $\beta \in \mathbb{R}^{2l\times 2l}$

Table B1
Saw timber and pulpwood volumes for spruce, pine, birch, and aspen (Parkatti and Tahvonen 2020).

	Mesic forest type							
	Spruce		Pine		Birch		Aspen	
d	$v_{1,1}(d)$	$v_{2,1}(d)$	$v_{1,2}(d)$	$v_{2,2}(d)$	$v_{1,3}(d)$	$v_{2,3}(d)$	$v_{1,4}(d)$	$v_{2,4}(d)$
6.5	0	0	0	0	0	0	0	0
7.5	0.01374	0	0.03458	0	0.01591	0	0.01591	0
12.5	0.06664	0	0.06659	0	0.07464	0	0.07464	0
17.5	0.1669	0	0.10166	0.09764	0.18005	0	0.18005	0
22.5	0.0808	0.23419	0.03905	0.27034	0.07854	0.25137	0.07854	0.25137
27.5	0.06482	0.44578	0.03001	0.48515	0.06655	0.45137	0.06655	0.45137
32.5	0.05975	0.68392	0.0275	0.74205	0.05827	0.69732	0.05827	0.69732
37.5	0.04978	0.96304	0.02647	1.04106	0.04978	0.96304	0.04978	0.96304
42.5	0.05039	1.25313	0.02596	1.38216	0.04865	1.24859	0.04865	1.24859
47.5	0.04324	1.57421	0.02567	1.76537	0.04463	1.55035	0.04463	1.55035
52.5	0.03925	1.89981	0.02549	2.29067	0.03891	1.86531	0.03891	1.86531
	Sub-xeric forest type							
	Spruce		Pine		Birch		Aspen	
d	$v_{1,1}(d)$	$v_{2,1}(d)$	$v_{1,2}(d)$	$v_{2,2}(d)$	$v_{1,3}(d)$	$v_{2,3}(d)$	$v_{1,4}(d)$	$v_{2,4}(d)$
6.5	0	0	0	0	0	0	0	0
7.5	0.01285	0	0.03342	0	0.01445	0	0.01445	0
12.5	0.06061	0	0.06370	0	0.06552	0	0.06552	0
17.5	0.15062	0	0.09685	0.09244	0.15522	0	0.15522	0
22.5	0.06857	0.21435	0.03738	0.25616	0.07000	0.21483	0.07000	0.21483
27.5	0.06052	0.39553	0.02917	0.46094	0.05743	0.39299	0.05743	0.39299
32.5	0.04872	0.61681	0.02690	0.70979	0.04731	0.59908	0.04731	0.59908
37.5	0.04593	0.85638	0.02597	0.99370	0.04769	0.82020	0.04769	0.82020
42.5	0.04370	1.11749	0.02551	1.32168	0.04179	1.06492	0.04179	1.06492
47.5	0.03787	1.40218	0.02525	1.69072	0.03290	1.32770	0.03290	1.32770
52.5	0.03573	1.68841	0.02509	2.10083	0.03096	1.59244	0.03096	1.59244

Parameter d is tree diameter at breast height, $v_{1,j}(d)$ is the pulpwood volume (m^3) per tree, and $v_{2,j}(d)$ is the saw timber volume (m^3) per tree. Aspen and birch volumes are assumed to be equal. We use linear interpolation to obtain volumes for arbitrary tree diameters.

Table B2
The covariance matrix for the logarithms of the wood prices.

	saw log, pine	saw log, spruce	saw log, birch	pulp, pine	pulp, spruce	pulp, birch
saw log, pine	0.00145	0.00108	0.00081	0.00126	0.00106	0.00096
saw log, spruce		0.00094	0.00058	0.00096	0.00080	0.00073
saw log, birch			0.00062	0.00084	0.00076	0.00062
pulp, pine				0.00154	0.00125	0.00124
pulp, spruce					0.00110	0.00099
pulp, birch						0.00114

Table B3
Species-specific mean roadside saw timber and pulpwood prices in euros per m^3.

	Spruce	Pine	Birch	Aspen[a]
saw timber	58.44	58.64	49.73	30.16
pulpwood	34.07	30.51	30.50	19.74

[a] Prices for spruce, pine, and birch are based on data from the Natural Resources Institute Finland, while prices for aspen are based on personal communication with the Central Union of Agricultural Producers and Forest Owners.

denote the cointegration matrix and the loading matrix, respectively. The cointegration matrix β contains information on the equilibrium relationships between the variables in levels. Vector $\beta' p_{\tau-1}$ can be interpreted as the distance of the price variables from their equilibrium values. The loading matrix β describes the speed at which the wood prices converge back to their equilibrium values. Given that the effects of price shocks vanish already during a 5-year period, we can also use the error covariance structure, estimated using (B.3), in the simplified period-level model (B.2) without loss of rigor. The parameter values used for the stochastic price model are given in Table B.4. The mean prices are according to Table B.5.

Harvesting costs depend on species, tree diameter, and harvested wood quantity. The variable harvesting and hauling costs are derived from the detailed empirical model by Nurminen et al. (2006) for both clear-cut $z = cl$ and thinning $z = th$:

Table B4
Species-specific parameters for harvesting-cost functions.

Species	q	$c_{j,q,0}$	$c_{j,q,1}$	$c_{j,q,2}$	$c_{j,q,3}$	$c_{j,4}$	$c_{j,5}$
Spruce	th	2.415	0.412	0.758	-0.180	2.272	0.535
	cl	2.100	0.412	0.758	-0.180	1.376	0.393
Pine	th	2.415	0.547	0.196	0.308	2.272	0.535
	cl	2.100	0.532	0.196	0.308	1.376	0.393
Birch and Aspen	th	2.415	0.420	0.797	0.174	2.272	0.535
	cl	2.100	0.430	0.756	0.174	1.376	0.393

th=thinning, cl=clear-cut.

$$C_z(h_t, d_t) = \sum_{j=1}^{l} c_{j,z,0} \sum_{q=1}^{m} \sum_{s=1}^{n} h_{j,z,s,t}(c_{j,z,1} + c_{j,z,2} v_{j,s}(d_{j,z,s,t}) + c_{j,z,3} v_{j,s}(d_{j,z,s,t})^2) \tag{B.4}$$

$$+ c_{z4} \sum_{j=1}^{l} \sum_{s=1}^{n} h_{j,s,t} v_{j,s}(d_{j,z,s,t}) + c_{z,5} \sum_{j=1}^{l} \sum_{s=1}^{n} h_{j,z,s,t} v_{j,s}(d_{j,z,s,t})^{0.7}, \quad z = th, cl, \tag{B.5}$$

where $v_{j,s}(d_{j,z,s,t})$ is the total tree volume, again assumed continuous and bounded, and $c_{j,z,s}$ are parameters given in Table B.6. The model defines cutting (B.4) and hauling (B.5) costs separately. According to this model, the variable harvesting costs increase with total harvested volume but decrease with tree volume. Only cutting costs have species-specific parameters, while hauling and felling costs are determined without separating between tree species.

Appendix C. Existence of an optimal deterministic Markov policy

To prove Theorem 1, we use the general result, Theorem 2, below, that requires the following assumptions.

Assumption 1 (Compactness)**.** The state space S is a non-empty Borel subset of $\mathbb{R}^{dim(s)}$, where each state is defined as a triplet $s = (x, d, p)$, and the action space A is a non-empty Borel subset of $\mathbb{R}^{\dim(a)}$, where each action is defined as a triplet $a = (h, \delta^{th}, \delta^{cc})$. For each $s \in S$, $D(s) \in 2^A$, where 2^A denotes the family of all nonempty compact subsets of A.

Assumption 2 (Continuity)**.** We assume that the following conditions are satisfied:

(W1) The set valued mapping $s \mapsto D(s)$ is upper semicontinuous.
(W2) The reward function $\pi : \operatorname{gph} D \to \overline{\mathbb{R}}$ is upper semicontinuous.
(W3) The transition law $q : \operatorname{gph} D \to \mathcal{P}(S)$ is weakly continuous; that is, $\int_S \theta(z) q(dz|s,a)$ is a continuous function on $\operatorname{gph} D$ for each bounded continuous function $\theta : S \to \mathbb{R}$.

While the continuity and compactness assumptions are sufficient for the existence of a stationary optimal policy in the negative case where $\pi \leq 0$, they are not enough to guarantee the existence of a stationary optimal policy in the positive case where $\pi \geq 0$. To guarantee that the total expected rewards are well defined, we will make use of the following additional convergence assumption:

Assumption 3 (General convergence condition)**.** For all $s \in S$, $\mu \in \Delta$, at least one of the numbers $V_\mu^+(s) := \mathbb{E}_\mu\left[\sum_{t=0}^{\infty} \pi_t^+ | s\right]$ and $V_\mu^-(s) := \mathbb{E}_\mu\left[\sum_{t=0}^{\infty} \pi_t^- | s\right]$ is finite, where $\pi_t = \pi(s_t, a_t)\gamma^t$ is a short-hand for the discounted reward observed at time t when action a_t is taken in state s_t. For any $\pi_t \in \overline{\mathbb{R}}$, here $\pi_t^+ = \max\{\pi_t, 0\}$ and $\pi_t^- = \max\{-\pi_t, 0\}$ so that $\pi_t = \pi_t^+ - \pi_t^-$.

Theorem 2 (Schäl (1983))**.** *. Under the compactness and continuity conditions in Assumptions (1)-(2) and the general convergence condition in Assumption (3), a deterministic stationary optimal policy $f \in \mathbf{F}$ exists such that $\sup_{\pi \in \Delta} V_\pi(s) = \sup_{f \in \mathbf{F}} V_f(s)$ for each $s \in S$.*

Using the above theorem by Schäl (1983), we can verify that the stochastic problem defined in Section 2 can be viewed as a dynamic decision model, where the optimal value of the expected total rewards under all stationary policies is equal to the optimal value under all (possibly randomized and non-Markovian) policies.

Proof of Theorem 1

Theorem 1 follows directly from Theorem 2 once we verify that Assumptions 1–3 hold. For this, we first note that Assumption 1 is valid. Indeed, the state and action spaces S and A are obviously non-empty Borel subsets of $\mathbb{R}^{dim(s)}$ and $\mathbb{R}^{dim(a)}$. Moreover, for given $s \in S$, the set $D(s)$ of possible actions $a = (h, \delta^{th}, \delta^{cc})$ satisfies $h_{j,q,w} \in [0, x_{j,q,w}]$ (i.e. more trees cannot be harvested than actually exist). As all intervals $[0, x_{j,q,w}]$ are closed and bounded, so is $D(s)$ as a direct product of such intervals. Let us next verify Assumption 2, i.e. conditions $(W1) - (W3)$.

Condition $(W1)$**:** Recall that, by the very definition, a set valued mapping $s \mapsto D(S) \in 2^A$ is upper semicontinuous if the set

$$\{s : D(s) \cap M \neq \emptyset\}$$

is closed in the state space S for every closed $M \subset A$. For a given state $s = (x, d, p) \in S$, we have $x_{j,q,w} \geq 0$ and for the possible harvesting decision $h \in \mathbb{R}^{l\times m\times n}$ we have $h_{j,q,w} \in [0, x_{j,q,w}]$, for all $1 \leq j \leq l$, $1 \leq q \leq m$ and $1 \leq w \leq n$. We also have $A = [0,\infty)^{l\times m\times n} \times \{0,1\} \times \{0,1\}$. Let M be a closed set in A. Then it can be represented as

$$M = M_1 \times \ldots M_{l\times m\times n} \times \{0,1\} \times \{0,1\},$$

where each $M_k \subset \mathbb{R}_+$ is closed. Consequently, the claim follows directly by considering the one-dimensional case and observing that the set $\{x \in \mathbb{R} : [0,x] \cap M \neq \emptyset\}$ is closed for every closed $M \subset [0,\infty)$.

Condition $(W2)$**:** As the cost function $C_z(h_t, d_t)$ and the reward function $R_z(h_t, d_t, p_t)$ are continuous functions of states s_t and h_t, and the reward function can be expressed as the linear combination of these continuous functions given δ_t^{th} and δ_t^{cc}, the upper semicontinuity follows.

Condition $(W3)$**:** Let $\theta : X \mapsto \mathbb{R}$ be bounded and continuous. We have to show that the mapping

$$(s,a) \mapsto \mathbb{E}[\theta(s_{t+1})|(s_t,a_t) = (s,a)]$$

is continuous. Note that now the transition law q is the probability distribution of s_{t+1} given (s_t, a_t). Moreover, using defining equations (1)–(10) we observe that s_{t+1} contains two random innovations ω_t and ε_t, but otherwise is a deterministic continuous transformation of (s_t, a_t). As ω_t and ε_t are independent of (s_t, a_t), ε_t is an independent sequence and ω_t is an AR(1) Gaussian sequence, condition $(W3)$ is obviously satisfied.

It remains to verify Assumption 3. We will show that the negative part $V_\pi^-(s) := \mathbb{E}_\mu\left[\sum_{t=0}^\infty \pi^-(s_t,a_t)\gamma^t|s\right]$ is finite, allowing to drop the price dynamics p from the considerations.

Let $x_t, d_t \leq K$ elementwise for some finite constant K. That is, the number of trees x_t or the diameter d_t of each size, age, and species classes cannot exceed a certain number determined by the area of the forest. Using elementary fact $(a-b)^- \leq b$ and the definition (1) of π, we see that

$$\pi^-(s_t,a_t)\gamma^t \leq \big(C_z(x_t,d_t) + C_f + C_r\big)\gamma^t \leq (K + C_f + C_r)\gamma^t$$

for a constant K as, by assumption, $C_z(x_t, d_t)$ is a bounded function for $z \in \{th, cc\}$. Thus

$$\mathbb{E}_\mu\left[\sum_{t=0}^\infty \pi^-(s_t,a_t)\gamma^t|s\right] \leq (K + C_f + C_r)\sum_{t=0}^\infty \gamma^t = \frac{K + C_f + C_r}{1-\gamma} < \infty.$$

This verifies Assumption 3. Thus, since Assumptions 1–3 hold, the statement of Theorem 1 follows from Theorem 2. This concludes the proof.

Appendix D. Finding the Markov policy using proximal policy optimization

Reinforcement learning algorithms, such as proximal policy optimization, are closely related to dynamic programming. Similar to learning methods, many dynamic programming algorithms are iterative and incremental procedures that find the correct solution through successive approximations (Sutton and Barto, 2018). Reinforcement learning methods can basically be seen as attempts to achieve the same objective as dynamic programming but with less computation and without assuming a perfect model of the environment. Though the forest environment considered in this study is known, it is still too complicated because of the large number of states and actions (i.e., the curse of dimensionality) to be solved using classical dynamic programming techniques. To deal with large state and action spaces, we apply policy gradient methods that learn a parameterized policy directly via gradient ascent.

D1. Proximal policy optimization as a policy gradient algorithm

Assume that a policy is parameterized by a vector θ, and $J(\theta) = V_{\mu_\theta}(s)$ is a measure for the performance of the parameterized policy μ_θ. The idea of policy gradient methods is to maximize performance by updating the policy parameters using approximate gradient ascent in J:

$$\theta \leftarrow \theta_{old} + \beta\nabla J(\theta_{old}),$$

where β denotes the learning rate parameter for the current update.

Most of the policy gradient algorithms would approximate the gradient by

$$\nabla_\theta \mathbb{E}_{\mu_\theta}\left[\sum_{t=0}^\infty \gamma^t \pi(s_t,a_t)\right] \approx \hat{\mathbb{E}}_t\left[\nabla_\theta \log \mu_\theta(a_t|s_t)\hat{A}_t\right], \tag{D.1}$$

where the expectation $\hat{\mathbb{E}}_t$ denotes the empirical average over a finite batch of sample trajectories, and $\hat{A}_t$ is an estimate for the advantage function at time t. The trajectories are obtained by interacting with the environment. These trajectories are also known as policy rollouts, as they are produced by repeatedly applying the current policy. The advantage for a policy μ is defined as $A_\mu(s,a) = Q_\mu(s,a) - V(s)$, where $Q_\mu(s,a) = \pi(s,a) + \gamma\int_S V_\mu(y)q(dy|s,a)$ is the state-action value function for

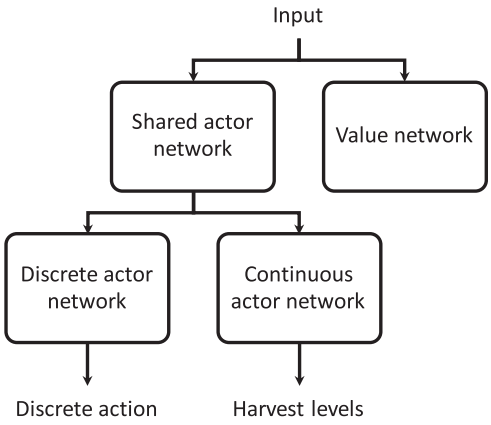

Fig. D1. The neural network structure. The stand structure is given as an input. The actor networks output the discrete harvesting decision and the continuous harvest levels. The value network is used in the advantage estimation.

policy μ. This class of algorithms that use the advantage function while approximating the gradient are known as actor–critic methods.

To prevent destructively large parameter updates, the policy gradient algorithms commonly use surrogate objectives instead of $J(\theta)$ directly. For instance, trust region algorithms have applied a surrogate objective,

$$\max_{\theta} J_{TR}(\theta) = \hat{\mathbb{E}}_t\left[\frac{\mu_\theta(a_t|s_t)}{\mu_{\theta_{old}}(a_t|s_t)}\hat{A}_t\right] \tag{D.2}$$

subject to

$$\hat{\mathbb{E}}_t\left[KL\left(\mu_{\theta_{old}}(\cdot|s_t), \mu_\theta(\cdot|s_t)\right)\right] \leq \kappa, \tag{D.3}$$

where KL denotes Kullback-Leibler divergence and θ_{old} is the parameter vector before the update, and κ is a constant that limits the size of the policy updates.

The proximal policy optimization algorithm (PPO) can be seen as a policy gradient algorithm that is based on the well-known actor–critic architecture (Schulman et al., 2017). Similar to trust region methods, PPO maximizes a surrogate objective function that forms a lower bound on the performance of the policy. This corresponds to performing approximate gradient ascent in

$$J_{KL}(\theta) = \hat{\mathbb{E}}_t\left[\frac{\mu_\theta(a_t|s_t)}{\mu_{\theta_{old}}(a_t|s_t)}\hat{A}_t - \beta KL\left(\mu_{\theta_{old}}(\cdot|s_t), \mu_\theta(\cdot|s_t)\right)\right], \tag{D.4}$$

where the penalty parameter β is assumed to be updated using a suitable adaptive strategy. Hence, instead of attempting to constrain the size of the policy updates using a strict constraint, PPO applies clipped surrogate functions or a soft penalty that limits the KL divergence of the policy update. Most empirical studies suggest that clipping seems to work at least as well as PPO with KL penalty, but it is considerably easier to implement. The clipped surrogate objective that is used in most PPO applications is given by

$$J_{PPO}(\theta) = \hat{\mathbb{E}}_t\left[\min\left(\frac{\mu_\theta(a_t|s_t)}{\mu_{\theta_{old}}(a_t|s_t)}\hat{A}_t,\ g(\epsilon, \hat{A}_t)\right)\right], \tag{D.5}$$

where the clipping function g is defined as

$$g(\epsilon, A) = \begin{cases}(1+\epsilon)A & \text{if } A \geq 0,\\ (1-\epsilon)A & \text{if } A < 0.\end{cases} \tag{D.6}$$

D2. Proximal policy optimization with parameterized action spaces

Following Malo et al. (2021), we use a modified version of PPO known as hybrid proximal policy optimization algorithm (Fan et al., 2019), where the surrogate objective function is extended to include both continuous and discrete actions. The algorithm is essentially similar to the original PPO, except that the policy is represented by two parallel actor networks, one for the discrete actions and one for the continuous actions (see Fig. D.12). The implementation leverages the special hierarchical structure of the action space, where the continuous actions (i.e. harvesting amounts) can be interpreted as parameters of the discrete actions (i.e. choice between thinning, clear-cut, and waiting).

To avoid the computationally expensive use of KL divergence directly, the policy parameters are updated via a clipped PPO objective function

$$J_{H-PPO}(\theta) = \hat{\mathbb{E}}_t[L(s_t, a_t, \theta_{old}, \theta)], \tag{D.7}$$

where the policy is parameterized by vector $\theta = (\theta_c, \theta_d)$, where θ_c and θ_d are parameters that correspond to continuous and discrete decisions, and the surrogate objective L is defined as the sum of losses corresponding to the discrete and continuous decisions $a = (h, \delta^{th}, \delta^{cc})$, i.e.

$$L(s, a, \theta_{old}, \theta) = L_c(s, h, \theta_{old,c}, \theta_c) + L_d(s, \delta^{th}, \delta^{cc}, \theta_{old,d}, \theta_d) \tag{D.8}$$

where

$$L_c(s, h, \theta_{old,c}, \theta_c) = \min\left(\frac{\mu_{\theta_c}(h|s)}{\mu_{\theta_{old,c}}(h|s)}\hat{A}^{GAE(\gamma,\lambda)},\ g(\epsilon, \hat{A}^{GAE(\gamma,\lambda)})\right) \tag{D.9}$$

$$L_d(s, \delta^{th}, \delta^{cc}, \theta_{old,d}, \theta_d) = \min\left(\frac{\mu_{\theta_d}(\delta^{th}, \delta^{cc}|s)}{\mu_{\theta_{old,d}}(\delta^{th}, \delta^{cc}|s)}\hat{A}^{GAE(\gamma,\lambda)},\ g(\epsilon, \hat{A}^{GAE(\gamma,\lambda)})\right) \tag{D.10}$$

The advantages are approximated using the generalized advantage estimation (GAE) approach proposed by Schulman et al. (2015):

$$\hat{A}_t^{GAE(\gamma,\lambda)} = (1-\lambda)\left(\hat{A}_t^{(1)} + \lambda\hat{A}_t^{(2)} + \lambda^2\hat{A}_t^{(3)} + \cdots\right), \tag{D.11}$$

where the overall advantage is expressed as the sum of 1 to k-step look-ahead functions

$$\begin{aligned}
\hat{A}_t^{(1)} &= -V(s_t) + \pi(s_t, a_t) + \gamma V(s_{t+1}) \\
\hat{A}_t^{(2)} &= -V(s_t) + \pi(s_t, a_t) + \gamma V(s_{t+1}) + \gamma^2 V(s_{t+2}) \\
&\vdots \\
\hat{A}_t^{(k)} &= -V(s_t) + \pi(s_t, a_t) + \gamma V(s_{t+1}) + \cdots + \gamma^{k-1} V(s_{t+k-1}) + \gamma^k V(s_{t+k}),
\end{aligned}$$

where $\lambda \in [0, 1]$ is a hyper-parameter. In practice, the state value function V (i.e. the "critic") is approximated using a parameterized function V_ϕ, which is typically assumed to be a neural network. The pseudo-code of the algorithm is presented in Algorithm 1.

Algorithm 1 PPO with hybrid actions and a clipped objective function.

procedure PPO-CLIP
Input: initial policy parameters θ_0, initial value function parameters ϕ_0
for k=0,1,2,... **do**
Sample a set of trajectories $D_k = \{\tau_i\}$ each with T time steps by running policy μ_{θ_k} in the environment.
Compute rewards to go $\hat{U}_t = \sum_{t=t}^{T} \pi(s_t, a_t)\gamma^t$.
Compute advantage estimates $\hat{A}_t^{GAE(\gamma,\lambda)}$ based on the current value function V_{ϕ_k}.
Update the policy by maximizing the clipped surrogate objective:

$$\theta_{k+1} = argmax_\theta \frac{1}{|D_k|T}\sum_{\tau\in D_k}\sum_{t=0}^{T} L(s_t, a_t, \theta_k, \theta)$$

using (stochastic) gradient ascent.
Estimate value function by minimizing mean-squared error:

$$\phi_{k+1} = argmin_\phi \frac{1}{|D_k|T}\sum_{\tau\in D_k}\sum_{t=0}^{T} (V_\phi(s_t) - \hat{U}_t)^2$$

using gradient descent.
end for
end procedure

The policy function μ_θ (i.e. the "actor") is implemented as two neural networks with two heads, one for discrete actions and one for continuous actions, i.e. harvest levels. A softmax-layer is used to compute the action probabilities for each stochastic policy. The continuous part is implemented as a Gaussian policy, which outputs the harvest levels that are transformed to relative proportions of trees belonging to predefined bins. The expected values of the Gaussian are the outputs of the continuous head of the neural network. The standard deviations are treated as additional parameters that are optimized

using gradient descent together with the neural network parameters. In our implementation, we used 1-cm bins, with the exception of combining the bins with diameters less than 5cm and bins above 35 cm. An accuracy of 1cm is considered sufficient to approximate the optimal solutions without significant sacrifice in the objective function value. Moreover, harvesting instructions with higher precision in diameter values would hardly be feasible in real-world harvesting.

The performance of RL algorithms is often sensitive to the state representation. Thus, the actor and critic apply preprocessing to the original state variables before feeding the observations to the neural network inputs. As a preprocessing step, we use the tree counts and diameters for each size and age class to compute a histogram with predefined fixed bins for the diameter distribution of each species. The tree frequencies for each bin are fed to the neural network input nodes. Learning policy and value functions is considerably more efficient using this alternative state representation compared with feeding raw tree count and diameter pairs to the network. As the PPO algorithm is prone to local optima, we run the training multiple times. The policy with best performance was chosen in the evaluation. Calculations were performed using computer resources within the Aalto Science-IT project. For further details considering the PPO algorithm used, see Fan et al. (2019); Schulman et al. (2015). Our H-PPO implementation follows the details presented in Malo et al. (2021).

References

Alvarez, L.H., Koskela, E., 2007. Optimal harvesting under resource stock and price uncertainty. Journal of Economic Dynamics and Control 31 (7), 2461–2485.

Bollandsås, M., Buongiorno, J., Gobakken, T., 2008. Predicting the growth of stands of trees of mixed species and size: A matrix model for norway. Scandinavian Journal of Forest Research 23 (2), 167–178.

Boucekkine, R., Hritonenko, N., Yatsenko, Y., 2011. Optimal control of age-structured populations in economy, demography, and the environment. Routledge.

Brazee, R., Mendelsohn, R., 1988. Timber harvesting with fluctuating prices. Forest science 34 (2), 359–372.

Buongiorno, J., Zhou, M., 2015. Adaptive economic and ecological forest management under risk. Forest Ecosystems 2 (4), 1–15.

Buongiorno, J., Zhou, M., 2020. Consequences of discount rate selection for financial and ecological expectation and risk in forest management. Journal of Forest Economics 35 (1), 1–17.

Cai, Y., Lontzek, T.S., 2019. The social cost of carbon with economic and climate risks. Journal of Political Economy 127 (6), 2684–2734.

Chang, F.-R., 2005. On the elasticities of harvesting rules. Journal of Economic Dynamics and Control 29 (3), 469–485.

Charpentier, A., Elie, R., Remlinger, C., 2021. Reinforcement learning in economics and finance. Computational Economics 1–38.

Clark, C.W., 1976. Mathematical bioeconomics: the optimal management of renewable resources. Wiley.

Clarke, H.R., Reed, W.J., 1989. The tree-cutting problem in a stochastic environment: The case of age-dependent growth. Journal of Economic Dynamics and Control 13 (4), 569–595.

Easterling, M.R., Ellner, S.P., Dixon, P.M., 2000. Size-specific sensitivity: applying a new structured population model. Ecology 81 (3), 694–708.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: Representation, estimation, and testing. Econometrica 55 (2), 251–276.

Fan, Z., Su, R., Zhang, W., Yu, Y., 2019. Hybrid actor-critic reinforcement learning in parameterized action space. arXiv:1903.01344 [cs.LG]. 1903.01344.

Faustmann, M., 1849. Berechnung des wertes, welchen waldboden, sowie noch nicht haubare holzbestände für die waldwirtschaft besitzen [calculation of the value which forest land and immature stands possess for forestry]. Allgemeine Forst- und Jagdzeitung 25, 441–455.

Getz, W., Haight, R., 1989. Population harvesting: Demographic models of fish. Forest and Animal Resources, Princeton University, Princeton, NJ.

Haight, R.G., 1990. Feedback thinning policies for uneven-aged stand management with stochastic prices. Forest science 36 (4), 1015–1031.

Haight, R.G., 1991. Stochastic log price, land value, and adaptive stand management: Numerical results for california white fir. Forest science 37 (5), 1224–1238.

Haight, R.G., Monserud, R.A., 1990a. Optimizing any-aged management of mixed-species stands: *ii.* effects of decision criteria. Forest science 36 (1), 125–144.

Haight, R.G., Monserud, R.A., 1990b. Optimizing any-aged management of mixed-species stands: *i.* performance of a coordinate-search process. Canadian Journal of Forest Research 20 (1), 15–25.

Helmes, K.L., Stockbridge, R.H., 2011. Thinning and harvesting in stochastic forest models. Journal of Economic Dynamics and Control 35 (1), 25–39.

Insley, M., Rollins, K., 2005. On solving the multirotational timber harvesting problem with stochastic prices: a linear complementarity formulation. American Journal of Agricultural Economics 87 (3), 735–755.

Johansen, S., 1995. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press.

Kaya, I., Buongiorno, J., 1987. Economic harvesting of uneven-aged northern hardwood stands under risk: A markovian decision model. Forest Science 33 (4), 889–907.

Kuuluvainen, J., Korhonen, J., Xu, D., Toppinen, A., 2018. Price integration for domestic and imported sawlogs and pulpwood in finland: an update. Scandinavian Journal of Forest Research 33 (1), 71–80. doi:10.1080/02827581.2017.1327614.

Liang, J., Picard, N., 2013. Matrix model of forest dynamics: An overview and outlook. Forest Science 59 (3), 359–378.

Lu, F., Gong, P., 2003. Optimal stocking level and final harvest age with stochastic prices. Journal of forest economics 9 (2), 119–136.

Lu, F., Gong, P., 2005. Adaptive thinning strategies for mixed-species stand management with stochastic prices. Journal of forest economics 11 (1), 53–71.

Maliar, L., Maliar, S., 2022. Deep learning classification: Modeling discrete labor choice. Journal of Economic Dynamics and Control 135, 104295.

Maliar, L., Maliar, S., Winant, P., 2021. Deep learning for solving dynamic economic models. Journal of Monetary Economics 122, 76–101.

Malo, P., Tahvonen, O., Suominen, A., Back, P., Viitasaari, L., 2021. Reinforcement learning in optimizing forest management. Canadian Journal of Forest Research 51 (10), 1393–1409.

Miller, R.A., Voltaire, K., 1983. A stochastic analysis of the tree paradigm. Journal of Economic Dynamics and Control 6, 371–386.

Nurminen, T., Korpunen, H., Uusitalo, J., 2006. Time consuming analysis of the mechanized cut-to-length harvesting system. Silva Fennica 40, 335–363.

Picard, N., Liang, J., 2014. Matrix models for size-structured populations: unrealistic fast growth or simply diffusion? PloS one 9 (6), e98254.

Pukkala, T., 2015. Optimizing continuous cover management of boreal forest when timber prices and tree growth are stochastic. Forest Ecosystems 2 (1), 1–13.

Pukkala, T., Lähde, E., Laiho, O., 2013. Species interactions in the dynamics of even- and uneven-aged boreal forests. Journal of Sustainable Forestry 32, 371–403.

Reed, W.J., Clarke, H.R., 1990. Harvest decisions and asset valuation for biological resources exhibiting size-dependent stochastic growth. International Economic Review 147–169.

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al., 2022. Tackling climate change with machine learning. ACM Computing Surveys (CSUR) 55 (2), 1–96.

Samuelson, P.A., 1976. Economics of forestry in an evolving society. Economic inquiry 14 (4), 466–492.

Saphores, J.-D., 2003. Harvesting a renewable resource under uncertainty. Journal of Economic Dynamics and Control 28 (3), 509–529.

Schäl, M., 1983. Stationary policies in dynamic programming models under compactness assumptions. Mathematics of Operations Research 8 (3), 366–372.

Scheidegger, S., Bilionis, I., 2019. Machine learning for high-dimensional dynamic stochastic economies. Journal of Computational Science 33, 68–82.

Schulman, J., Moritz, P., Levine, S., Jordan, M., Abbeel, P., 2015. High-dimensional continuous control using generalized advantage estimation. arXiv:1506.02438 [cs.LG]. 1506.02438.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. arXiv:1707.06347 [cs.LG]. 1707.06347.

Sødal, S., 2002. The stochastic rotation problem: A comment. Journal of Economic Dynamics and Control 26 (3), 509–515.
Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.
Tahvonen, O., 2011. Optimal structure and development of uneven-aged norway spruce forests. Canadian Journal of Forest Research 41 (12), 2389–2402.
Willassen, Y., 1998. The stochastic rotation problem: a generalization of faustmann's formula to stochastic forest growth. Journal of Economic Dynamics and Control 22 (4), 573–596.
Zhou, M., Buongiorno, J., 2006. Forest landscape management in a stochastic environment, with an application to mixed loblolly pine–hardwood forests. Forest Ecology and Management 223 (1-3), 170–182.
Zhou, M., Buongiorno, J., 2019. Optimal forest management under financial risk aversion with discounted markov decision process models. Canadian Journal of Forest Research 49 (7), 802–809.