



Spatial interactions and optimal forest management on a fire-threatened landscape



Christopher J. Lauer^{a,*}, Claire A. Montgomery^b, Thomas G. Dietterich^c

^a Oregon State University, Applied Economics, 213 Ballard Extension Hall, Corvallis, OR 97331, United States

^b Oregon State University, Forest Engineering, Resources & Management, 218 Snell Hall, Corvallis, OR 97331, United States

^c Oregon State University, Electrical Engineering and Computer Science, 2067 Kelly Engineering Center, Corvallis, OR 97331, United States

ARTICLE INFO

Keywords:

Wildland fire

Spatial

Ecological disturbance

Risk

Approximate dynamic programming

Reinforcement learning

Forestry

ABSTRACT

Forest management in the face of fire risk is a challenging problem because fire spreads across a landscape and because its occurrence is unpredictable. Accounting for the existence of stochastic events that generate spatial interactions in the context of a dynamic decision process is crucial for determining optimal management. This paper demonstrates a method for incorporating spatial information and interactions into management decisions made over time. A machine learning technique called approximate dynamic programming is applied to determine the optimal timing and location of fuel treatments and timber harvests for a fire-threatened landscape. Larger net present values can be achieved using policies that explicitly consider evolving spatial interactions created by fire spread, compared to policies that ignore the spatial dimension of the inter-temporal optimization problem.

1. Introduction

Forest management is a dynamic problem; actions taken today have important consequences for the future value of forest land. Determining optimal management is further complicated by unpredictable ecological disturbances; one important disturbance on many forested landscapes is wildfire. While fire is a natural process that can be vital to the health of forest ecosystems, it can threaten values on the landscape such as timber, homes in the wildland urban interface, watershed health, air quality, and wildlife habitat. Significant amounts of money and resources have been devoted to the task of eliminating damaging wildfire. According to the 2015 Fire Budget Report, produced by the USDA Forest Service, > 50% of the Forest Service's 2015 operating budget was devoted to fire-related activities, compared to just 16% in 1995. The USDA Forest Service spent more than \$1.7 billion on fire suppression costs alone in 2015 and has spent more than \$1 billion in eight of the last ten years on fire suppression activity (USDA, 2015).

Value on a forest landscape can be diminished, or destroyed, by unpredictable events like fire, and the extent of the damage is at least partially outside of manager control. Routledge (1980) and Reed (1984) demonstrated that the optimal rotation age for timber harvest determined by Faustmann (1968) can be adjusted to account for the possibility of an unpredictable natural disaster damaging or destroying a stand. They showed that the optimal rotation age decreases as the

probability of stand destruction increases. Reed built upon his model in subsequent papers to demonstrate how optimal management will be affected if fire arrival rate is a function of stand characteristics. He also explored how to determine the optimal schedule of investment in fire protection, such as fuel treatment or fire-fighting infrastructure (Reed, 1989, 1993). Many authors build upon Reed's insights; examples include Amacher et al. (2005) and Garcia-Gonzalo et al. (2014) who look at how fuel treatment and silvicultural interventions affect optimal rotation age for fire threatened forest stands. Daigneault et al. (2010) examine how carbon sequestration is affected by manager response to fire risk.

An important shortcoming of these models is that they focus on management at the stand level and do not account for fire's ability to spread between stands. Because fire can travel large distances across a landscape, fire risk on an individual stand is a function of the condition of the entire landscape. Several authors have created models to account for fire behavior that includes spatial interactions. Wei et al. (2008) separated fire arrival probability into ignition and spread probabilities. These probabilities account for varying ecological factors such as slope and wind direction. They used the conditional spread probabilities to find locations of fuel treatments in the current time period to minimize expected loss to fire on the modeled landscape. Ager et al. (2010) used repeated simulation to compare the damage probabilities for structures in the wildland-urban interface under different fuel management

* Corresponding author.

E-mail addresses: lauer@oregonstate.edu (C.J. Lauer), claire.montgomery@oregonstate.edu (C.A. Montgomery), tgd@cs.orst.edu (T.G. Dietterich).

strategies when fire spreads. These models are static because they consider only treatment in the current period and the expected value of the treated landscape. Chung et al. (2013) developed a model that places fuel treatments over two time periods to minimize expected loss. The vegetation evolves over time, but no fire actually occurs. While this model is intertemporal, it is not dynamic because it does not account for how optimal management will adjust on a post-fire landscape, should fire actually occur.

Stochastic dynamic programming is a natural method for determining the optimal management of a fire-threatened forest landscape. Actions are selected to maximize the current period value plus the expected future value of the treated landscape given that optimal choices are made in future periods as new knowledge becomes available (e.g., stochastic events are realized). This optimization strategy depends on the principle of optimality (Bellman, 1957) which is the assumption that no matter what the current action is, actions in all future states will be optimally chosen. Solving a stochastic dynamic programming problem requires complete enumeration of all possible outcomes in each time period; this can make even very simple problems intractably large – the so-called “curse of dimensionality.” Konoshima et al. (2008, 2010) was able to formulate and solve the optimal fuel treatment and harvest problem as a stochastic dynamic problem by considering only 2-periods, 7 management units, 4 stand age classes, 4 treatment options, and 2 weather conditions. They demonstrated that landowners will try to protect on-site timber values by shortening rotations, as suggested by Reed (1984), but will try to protect adjacent timber values by postponing harvest to avoid high spread rates associated with young stands. This study demonstrated the importance of accounting for stochasticity and spatial interactions in a dynamic decision framework, but its practical usefulness for decision-making and policy analysis is limited.

Both the spatial and intertemporal aspects of solving for the optimal management of a fire-threatened landscape contribute to the curse of dimensionality. Heuristic solution methods, such as simulated annealing or tabu search, are widely used to identify management that is approximately optimal in a given time period for large landscapes with spatial interactions. Chung et al. (2013) is one example. See also Bettinger et al. (2003). But these methods alone are not sufficient for solving dynamic problems in which managers respond to new information as it becomes available. For that, we need an approximate approach to solving dynamic programming problems.

In this study, we develop a method for incorporating spatial interactions into a dynamic decision process that accounts for the stochastic nature of fire by using a machine learning technique known as approximate dynamic programming (Sutton and Barto, 1998; Powell, 2007, 2009). We applied our method to model optimal timing and placement of timber harvest and fuel treatment on a 64-unit landscape that we parameterized to represent the ecological conditions of southwest Oregon. The value functions we estimate provide a way to model the expected benefits, costs, and externalities associated with different management actions which have uncertain consequences in multiple locations on the landscape. The value function implies an optimal policy for management actions. We evaluate the effectiveness of our method by comparing simulated outcomes over 150-year time horizons to outcomes generated using Reed and Faustmann timber harvest rotations and to a rule-of-thumb approach to placing fuel treatments. The policies generated by our model lead to higher average net present values (NPV) on the landscape compared to the benchmark policies. In the concluding section, we highlight the relevance of our key findings, describe model limitations, and discuss plans to extend the basic model to address relevant questions in forest management.

2. Model

A bio-economic model can provide insight into factors that drive landowner behavior and can be used to determine an optimal set of actions for a decision-making agent when landscape conditions give rise

to fire risk that affects stand value. Our model accounts for the financial incentives faced by the agent. It could also be specified to include non-market incentives. The ecological processes that determine the evolution of the landscape can be thought of as a constraint on the agent's actions. Vegetation evolves and stochastic disturbances (in this case, fire) occur over time so that future costs and revenues depend on actions taken today and, likewise, the present value depends on those future rewards. Finally, because fire spreads across the landscape creating interactions that affect stand value, our model is explicitly spatial.

2.1. Markov decision process and Bellman's equation

We represent the economic and ecological components of this problem as a Markov Decision Process (MDP) (Puterman, 1994) that has five different components.

1. A set of possible *States*, which depends on the attributes of the individual stands contained by the landscape. The state S_t describes the conditions of the landscape at time t .
2. A set of *Actions* that describes what a land manager can do. In our setting, the overall action at time t is a vector x_t of the management activities applied to each stand in the landscape. For each stand, there are four possible management activities: harvest timber (clear-cut), treat fuel to reduce fire risk, implement both activities, or do nothing; in each time-step one of these four options must be chosen for each stand.
3. A *Reward Function*, $C(S_b, x_t)$ that describes the immediate (i.e. current period) costs or revenues associated with a particular action for a particular state.
4. A *State Transition Model*, $S_{t+1} = S'(S_b, x_t, W_t)$, that describes how the state evolves over time. The transition from S_t to S_{t+1} , is a function of the current state, S_b , the current action, x_t , and a vector of stochastic events, W_t , which includes fire arrival and weather.
5. A *Discount Factor*, δ , that determines how current rewards are valued relative to future rewards.

To analyze this MDP, we employ Bellman's equation (Bellman, 1957), a recursive equation that assigns a value to a particular state Eqs. (1a), (1b) by computing an expectation over the values of future states. This equation represents a landowner's decision-making process. It can be broken into two parts – immediate cost or benefit of a management action, which is captured by the reward function, and the expectation of the maximized value of the next period's state, $V(S'(S_b, x_t, W_t))$. We employ the so-called action-value representation, also known as the “Q-value” (Watkins and Dayan, 1992). The quantity $Q(S_b, x_t)$ is the expected return of taking action x_t in state S_t and then behaving optimally thereafter. In approximate dynamic programming, it is often useful to first compute $Q(S_b, x_t)$ for each possible action x_t and then choose the action that has the highest Q value. This is the optimal action in state S_b , and it defines the value $V(S_b)$.

$$V(S_t) = \max_{x_t} Q(S_t, x_t) \quad (1a)$$

where $Q(S_b, x_t)$ is the value of performing action x_t on the landscape in state S_t and behaving optimally thereafter:

$$Q(S_t, x_t) = C(S_t, x_t) + \mathbb{E}_{S_{t+1}}[\delta V(S'(S_t, x_t, W_t))] \quad (1b)$$

2.2. Spatial interactions

To incorporate spatial interactions, we decompose $Q(S_b, x_t)$ into separate action-value functions for each stand in the landscape. Let $Q_i(S_b, x_t)$ denote the contribution of stand $i = 1, 2, \dots, I$ to the value of the overall landscape at time t . Note that the state S_b , action x_b , and stochastic event W_t are not indexed by i because they represent the

entire landscape:

$$Q_i(S_t, x_t) = C_i(S_t, x_t) + \mathbb{E}_{S_{t+1}}[\delta V_i(S'(S_t, x_t, W_t))] \quad (2)$$

$V_i(S'_t, x_t, W_t)$ is stand i 's contribution to the value of the landscape in state $S(S_t, x_t, W_t)$ given that the landscape is managed optimally in future periods, i.e. $\sum_i V_i(S'_t, x_t, W_t) = V(S'_t, x_t, W_t)$. The landowner chooses x_t to maximize the overall Q -value for the landscape which is the sum of the contributions from each stand:

$$V(S_t) = \max_{x_t} Q(S_t, x_t) = \max_{x_t} \sum_{i=1}^I Q_i(S_t, x_t) \quad (3a)$$

This value function implies an optimal policy for the management of the landscape:

$$x^*(S_t) = \operatorname{argmax}_{x_t} Q(S_t, x_t) \quad (3b)$$

The individual stand value functions depend not only on that stand's vector of attributes, but also on the attributes of the entire landscape (as indicated by S_t) to account for potential fire spread. Therefore, the actions that maximize the value of the landscape may not maximize the value of each individual stand.

3. Solution method

The optimal management policy is found by solving Eqs. (3a), (3b), a difficult problem because of the unknown value function and the curse of dimensionality created by spatial interactions and unpredictable ecological processes. We resolve this difficulty by modifying Eqs. (3a), (3b) using two techniques described by Powell (2007): post-decision states, and approximate dynamic programming (ADP) using value function iteration. Either value iteration or policy iteration could be used to estimate the value function and determine an optimal policy. Policy iteration works by calculating the value for a policy and uses that value to find a better policy. Value iteration works by finding the best action based on the current estimate of the value function and uses the realized value for this action to improve the estimate of the value function. For this problem, value iteration is a better strategy than policy iteration because of the size of the action space. Additionally, computing the value for a particular policy is difficult because of the need to simulate many time steps into the future. It would be intractable to calculate the value for even a small portion of the possible policies.

3.1. Post-decision state

The post-decision state technique is a method for simplifying the transition model and the action-value functions. It was first described by Van Roy et al. (1997) and can be defined as the state observed by the agent after actions have been taken, but before any stochastic process (in this case fire arrival) occurs. The state-transition model, $S_{t+1} = S'(S_t, x_t, W_t)$, is divided into two steps. First, let $g(S_t, x_t)$ be a deterministic process that applies management action vector x_t to the landscape to produce the post-decision state $S_t^x = g(S_t, x_t)$. The important property of the post-decision state is that it can be anticipated accurately by the decision-making agent because it depends only on the current state, which is observed by the agent, and the actions of the agent. The transition to the next period state S_{t+1} is a function of the post-decision state and the stochastic fire event $S_{t+1} = S'(S_t^x, W_t)$. With this change, we can rewrite the Q and V functions in terms of the post-decision state S_t^x :

$$Q_i(S_t, x_t) = C_i(S_t, x_t) + V_i(g(S_t, x_t)) = C_i(S_t, x_t) + V_i(S_t^x) \quad (4)$$

We can then rewrite the Bellman equation in terms of the post-decision state as:

$$V(S_t^x) = \mathbb{E}_{S_{t+1}}[\delta \{\max_{x_{t+1}} \{C(S_{t+1}, x_{t+1}) + V(g(S_{t+1}, x_{t+1}))\}\}] \quad (5)$$

where $C(S_{t+1}, x_{t+1}) = \sum_i C_i(S_{t+1}, x_{t+1})$ and $V(S_{t+1}^x) = \sum_i V_i(S_{t+1}^x)$. This version of the Bellman equation relates the value of the post-decision state S_t^x to the expected value of the next period reward and the value of the next period post-decision state. Hence, this form of the Bellman equation goes from post-decision state to post-decision state, whereas the standard Bellman equation goes from (pre-decision) state to (pre-decision) state. The advantage of this is that the expectation in Eq. (5) is outside of the max operation, whereas in Eqs. (1a), (1b) we must compute a separate expectation for each possible action before we can take the max.

3.2. Value function approximation

The reformulated Bellman equation, using the post-decision state, can be solved by applying a form of approximate dynamic programming (ADP) known as value function approximation. The main idea is to estimate the coefficients for a linear approximation of the true, but unobservable, value of the post-decision state as a function of observable attributes of that state. Each stand's contribution to the value of the post-decision state $V_i(S_t^x)$ is represented by a linear model defined with respect to a set of basis functions $\phi = (\phi_1, \dots, \phi_J)$, and coefficients $\theta = (\theta_1, \dots, \theta_J)$ as shown in Eq. (6):

$$V_i(S_t^x) \approx \bar{V}_i(S_t^x) = \sum_{j=1}^J \theta_j \times \phi_j(S_t^x) \quad (6)$$

Basis functions $\phi = (\phi_1, \dots, \phi_J)$ are a numeric representation of observable attributes for the stand and surrounding landscape that influence the net present value of the stand.¹ Knowing values for θ allows us to compute the value of any state by computing the basis function values in that state, multiplying by the coefficients, and summing the results. The Q -values for each stand are modified to include an approximation of the stand's contribution to the value of the post-decision state:

$$Q_i(S_t, x_t) = C_i(S_t, x_t) + \bar{V}_i(S_t^x | \theta) \quad (7)$$

The landscape value function and the optimal policy implied by this value function are now defined using the modified Q -values shown in Eq. (7).

The value function approximation, $\bar{V}_i(S_t^x | \theta)$, is conceptually similar to a hedonic model for determining the value of a forest stand. Hedonic models estimate how attributes of a composite good contribute to its value. They are used for appraising forest land or real estate market values. In this case, the composite good is the stand, which depends on the attributes of the landscape.

3.3. Value iteration algorithm

Here we give a brief description of the value iteration algorithm employed to estimate the unknown coefficients θ of $\bar{V}_i(S_t^x | \theta)$; we provide details in the Appendix A. To compute the optimal values of the coefficients, we initialize $\theta = \theta^0$ for $\bar{V}_i^0(S_t^x | \theta^0)$ and then update the coefficients through a series of iterations $n = 1 \dots N$. The iterative updates to the value function coefficients happen during a number of simulation cycles. In each cycle, we start in a chosen starting state, S_0 , and simulate management for several years into the future. In each time-step simulation outcomes are used to update the estimate of the value function coefficients. Finding the optimal action, $x_t^* = \operatorname{argmax}_{x_t} \sum_{i=1}^I [C_i(S_t, x_t) + \bar{V}_i^n(S_t^x | \theta^n)]$, in each time-step, given the current state S_t and the most recent estimate of the value function $\bar{V}_i^n$, is a large combinatorial problem. We employ a heuristic optimization method that combines simulated annealing (Kirkpatrick, 1984) with tabu search (Reeves, 1993) to quickly find a near optimal solution

¹ Examples of basis functions include a single variable, a product of multiple variables, power functions, and so on.

for each time step. This vector of actions x_t^* is applied to the landscape to find the post-decision state. We then simulate state transitions (which could include a fire) on the post-decision landscape by drawing random ignition and weather events from the distributions described in Section 4. These simulations result in a set of next period states $S_{t+1}^m, m = 1, \dots, M$. The optimal action and the value for each of the states resulting from the each of the M simulations is found to calculate the realized contribution each stand makes to the value of the post-decision state: $V(S_{t+1}^m) = \max_{x_{t+1}} \sum_{i=1}^I [C_i(S_{t+1}^m, x_{t+1}) + \bar{V}_i^n(S_{t+1}^m)]$. In the next step, the coefficients θ are updated from θ^n to θ^{n+1} to reduce the difference between our current estimate of stand value and the average of the realized values of the states resulting from the simulations. This completes one “Bellman backup”² step for value iteration. Then we choose one of the M simulated resulting states S_{t+1} as the next starting state. If there are no time-steps left in the current cycle, a new cycle is started with the starting state S_0 . Additional steps are necessary to adequately explore the state space and avoid local optima, these are also described in the Appendix A. When the algorithm converges, we have an approximation of the value function $\bar{V}_i^*(S_i^x | \theta^*)$ that implies the stationary policy $x_t^* = \operatorname{argmax}_{x_t} \sum_{i=1}^I [C_i(S_t, x_t) + \bar{V}_i^*(S_t^x | \theta^*)]$.

4. Data and parameters

The focus of this article is on agent behavior and developing an optimization framework for managing fire-threatened landscapes. While our study is not rooted in a specific location, the underlying ecology is a crucial driver of agent behavior. To demonstrate this optimization framework, a representative landscape for SW Oregon was created using pre-existing ecological models and parameters which are described in this section. While there are many possible objectives for forest management, in this application we assumed that the landowner's goal is to maximize the expected net present value of harvested timber on the landscape. This objective is easy to define and is likely to characterize many forest landowners.

4.1. Landscape parameters

We modeled a representative forest landscape as an 8×8 grid consisting of 64 40-acre square stands that are flat (no elevation change) with identical soil conditions, climate, and weather. The defining feature of a forest stand is that its vegetation is relatively homogeneous and can be treated in a uniform manner (Tappeiner et al., 2007). Smaller stand sizes increase the resolution and landscape heterogeneity, especially for determining the effect of fire on the landscape; larger stand sizes decrease modeling complexity. We selected 40 acres as a reasonable minimum size for a timber harvest unit. In order to account for edge effects, and because our model landscape is relatively small, we model the landscape as a torus that wraps on itself. This ensures that all effects of the fire are captured in the model, because a fire that burns to the Eastern boundary (for example) wraps around and continues spreading inward from the Western boundary. It is mathematically possible for a fire to wrap around and return to the point of ignition, but this was a rare event in our simulations occurring only in the worst fuel and weather conditions. This construction is an advantage for our model landscape because it eliminates the need to model the costs associated with fire that spreads from outside the landscape, or fire that spreads off the edge of the landscape. The effect of these events is captured when the fire wraps around an edge. Ignoring the edge effects would mean that stands on the edge of our model landscape face a lower fire risk.

The initial landscape is created by randomly assigning an age class, with an associated vector of attributes, to each stand in the landscape. Each stand evolves over time, independently of the other stands. We

tracked the evolution of stand characteristics over time using a transition table—attributes we track are stand age, total cubic feet of biomass per acre, merchantable cubic feet per acre, merchantable board feet per acre, quadratic mean diameter, crown base height, tree height, and fuel model. These characteristics were used to drive simulations of fire event, and to compute the reward function resulting from landowner actions.

Stands transition into the next state as a result of vegetation growth, fire, harvest, and fuel treatment. Vegetation growth was simulated using the Inland CA/Southern Cascades variant of Forest Vegetation Simulator (FVS) (Dixon et al., 2002). In the vegetation simulations, bare ground is prepared for planting by piling and burning surface fuel and planting 500 Douglas-fir seedlings (*Pseudotsuga menziesii*) per acre. At age 15, the stands are thinned from below to a density of 300 trees per acre. After this time, trees are allowed to grow until they are harvested or destroyed in a fire. This approximates typical even-age stand management that would occur in this type of forest (Hobbs et al., 1992; Tappeiner et al., 2007). We used a real discount rate of 4% ($\delta = 0.96$; Row et al., 1981). Log prices were obtained from the Oregon Department of Forestry (2016), and harvest/haul costs were based on a harvest cost model (Fight et al., 1984).

We developed a simple model to simulate fire. Ignition probability was determined using statistics from the SW District of the Oregon Department of Forestry (Thorpe, 2011). All stands on the landscape had equal probability of ignition. To model the ability of fire to spread from one stand to another when an ignition occurs, we used the BEHAVE fire modeling system (Andrews et al., 2003) to determine the fire spread rates associated with each fuel model/weather combination. A surface fuel model represents surface fuels in order to classify a wide number of ecosystems for the purpose of modeling wildfire spread (Anderson, 1982). The Fire and Fuel Extension to the Forest Vegetation Simulator (FFE-FVS; Reinhardt and Crookston, 2003) generates fuel models for each stand in each time step as it grows, receives silvicultural treatments, is harvested, and has fuel treatments applied (simulated in FVS as piling and burning surface fuel). In fact, fire damage and fire spread are not necessarily the same. However, for this application, we assumed that fire arrival leads to stand destruction. Therefore, we only modeled fire spread and not also fire intensity. We used FireFamily Plus (Bradshaw and McCormick, 2000), a software tool that analyzes weather observations and computes fire danger indices, to analyze Remote Automatic Weather Station (RAWS) data for several weather stations in SW Oregon in order to determine average wind speed and fuel moisture conditions for 4 different fire danger categories: low fire danger was the average conditions of the 0th–15th percentile of the fire danger index, moderate was the average conditions of the 16th–89th percentile, high was the average of the 90th–97th percentile, and extreme was the average of the 98th–100th percentile. Weather was drawn according to the following discrete distribution: lower fire danger, probability 0.15; moderate fire danger, 0.65; high fire danger; 0.07; extreme danger 0.03. Wind was assumed to be equally probable from all directions in all conditions.

The extent of fire spread for each ignition is controlled by fire duration. Duration was a randomly drawn number between 24 and 96 h, with longer durations more likely under more dangerous weather conditions. Fire duration is not based on empirical data; instead, repeated simulations on randomly generated landscapes using the parameters described above were performed to find a distribution of durations that led to a fire size distribution similar to that which has historically occurred in SW Oregon (Thorpe, 2011). Fuel treatment costs were determined based on a study by Calkin and Gebert (2006).

The representative landscape outlined above is a simplification of the real world, but it approximates the ecological processes that managers in the real world face. The underlying ecological models are complex enough that it is intractable to enumerate all the states, actions, or stochastic events that could occur on the landscape – this is the curse of dimensionality which precludes the use of exact dynamic

² A Bellman backup maps the current estimate of the value function to a new estimate.

programming methods.

4.2. Value function specification

The choice of basis functions in Eq. (6), $\phi = (\phi_1, \dots, \phi_J)$, is crucial for accurately approximating the value function for a stand. Several different specifications were tried in preliminary experiments for this model landscape. The specification that performed best is described here:

$$\begin{aligned} \bar{V}_i(S_i^x | \theta) = & \theta_1 + \theta_2 LTV_i + \theta_3 LTV_i^2 + \theta_4 SR_i + \theta_5 SR_i^2 + \theta_6 LTV_i SR_i \\ & + \theta_7 \sum_{e \in E_i} SR_e + \theta_8 \sum_{c \in C_i} SR_c + \theta_9 \sum_{e \in E_i} SR_e^2 + \theta_{10} \sum_{c \in C_i} SR_c^2 \\ & + \theta_{11} SR_i \sum_{e \in E_i} SR_e + \theta_{12} SR_i \sum_{c \in C_i} SR_c + \theta_{13} A_i \sum_{e \in E_i} SR_e \\ & + \theta_{14} A_i \sum_{c \in C_i} SR_c \end{aligned} \quad (8)$$

In this formula, the following quantities are included:

- LTV_i is the land and timber value according to the Faustmann rotation without fire. It is calculated using the vegetation growth simulations and cost equations described in the previous section.
- A_i is the age of stand i .
- SR_i is the predicted down-wind spread rate under extreme weather conditions. It is difficult to directly compare the likely effect of one fuel condition to another. This helps capture the risk created by fuel conditions on the landscape that facilitate fire spread.
- E_i is the set of neighboring stands that share an edge with stand i (Fig. 1).
- C_i is the set of stands that share a corner with stand i .

In several of the basis functions, we first sum the spread risk over all stands in E_i or in C_i and then multiply this by other quantities. This works because in our model the wind is equally probable in every direction, so all of the neighbors in E_i or in C_i pose the same risk to stand i and can be handled interchangeable. If the wind had a prevailing direction, we would specify a separate basis function and θ coefficient for each neighbor. The basis function values that determine the value function approximation for the initial landscape are depicted in Fig. 2. Fig. 2a shows the land and timber value (LTV_i) for each stand, Fig. 2b shows the spread rate (SR_i) associated with the stand's fuel conditions. Note that the basis function values are correlated with one another because SR_i and LTV_i both depend heavily on stand age, A_i .

This specification of the value function approximation includes fire risk from the first ring of adjacent stands only. It does not incorporate all of the spatial information about potential fire risk; it is possible for a

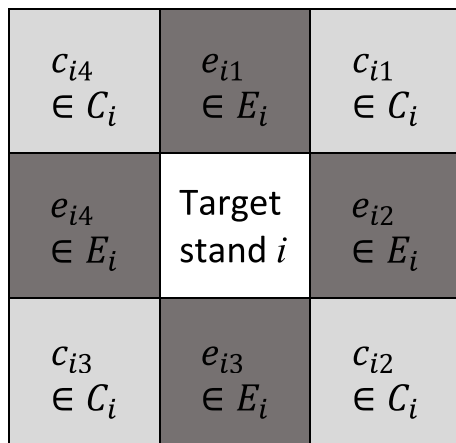


Fig. 1. Depicts adjacent stands whose attribute are included in the value function approximation.

fire to travel from any stand on the landscape to any other stand if the fuel and weather conditions are right. However, nearby stands have a greater impact on fire risk than distant stands. On the representative landscape we created, preliminary experiments showed that the inclusion of more distant stand attributes did not significantly improve the performance of the value function approximation.

The coefficients for this value function were estimated using the algorithm described in Section 3 and the appendix; the resulting approximation for the value function, $\bar{V}_i^*(S_i^x | \theta^*)$, should accurately approximate the expected value of being in state S_i^x . To initialize the algorithm we set $\theta_2^0 = 1$ and all other coefficients to zero for our initial guess θ^0 . This seemed to be a good first approximation because the Faustmann rotation has long been recognized as optimal in the absence of complicating factors such as fire risk (Samuelson, 1976). A reasonable initial estimation of the value function is important for ensuring the convergence of the value function iteration algorithm to a good optimum. This is especially important in the context of timber harvest because most stand value depends on anticipated future value and not the current reward; there is a long interval between positive rewards which occur when trees are harvested.

5. Results

5.1. Evaluating the value function approximation

The value function iteration algorithm used to generate the value function coefficients does not provably converge to the true coefficient values except in very specific conditions, which are not met for this case (Powell, 2007). Hence, it is important to assess whether the chosen specification of the value function accurately estimates the expected landscape value. We evaluated the value function approximation, $\bar{V}_i^*(S_i^x | \theta^*)$, by addressing two questions: (1) Does each coefficient estimate converge to a stable value? and (2) Are the landscape values returned by the value function approximation close to the landscape values computed by Monte Carlo simulations?

To answer the first question, we tracked the evolution of the value function coefficients. Coefficients should move rapidly from their initial (inaccurate) value toward a stable value and then make small oscillations around this stable value. The coefficients will never completely converge because of stochastic ignition and weather events. The coefficients evolved as we expected (Fig. 3): in the early iterations, they make wide oscillations that trend toward their stable value; in the later iterations, the coefficients make smaller oscillations around a fixed value with no apparent trend in at least the last 2000 iterations.

Second, the value function is a prediction of the discounted stream of rewards that will be generated over an infinite time horizon for that starting state. To determine if the value function approximation generates an accurate prediction, we generated 500 150-year simulations of fire events for the same starting landscape, optimally choosing timber harvests and fuel treatments in each time step (i.e. $x_t^* = \operatorname{argmax}_{x_t} \sum_{i=1}^I [C_i(S_i, x_t) + \bar{V}_i^*(S_i^x | \theta^*)]$). Each simulation followed a different, randomly generated stream of fire events drawn from distributions of ignition location, fire duration, and weather as described in Section 4.1. Each simulation led to different states and different reward streams, because each simulation faced different weather events. We tracked the rewards earned in each period on each stand for every simulation and compared the discounted sum of these rewards for the entire landscape, (Eq. (9) - V_k is the realized value for simulation $k = 1 \dots 500$), to the value predicted by the value function approximation, ($\bar{V}$) Eq. (10).

$$V_k = \sum_{t=0}^{T=150} \left[\delta^t \sum_{i=1}^{I=64} C_i(S_i, x_t^*) \right] + \delta^{150} \sum_i \bar{V}_i^*(S_{i150}^x) \quad (9)$$

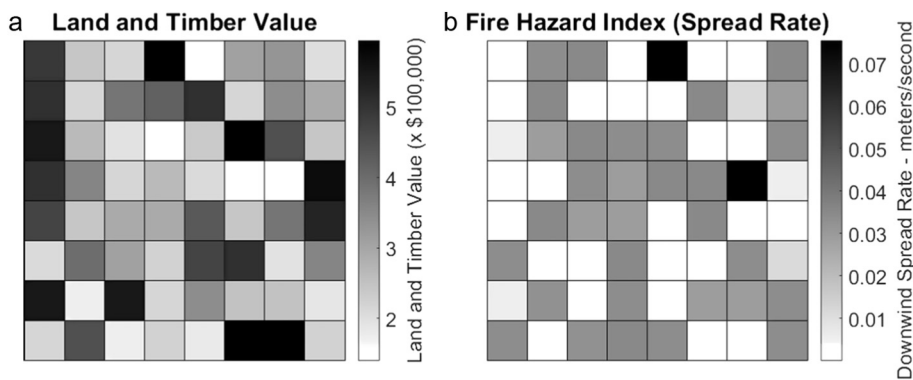


Fig. 2. Variables that represent the initial landscape, determined based on underlying landscape attributes. 2a shows the land and timber value calculated based on the Faustmann rotation. 2b shows the downwind spread rate for extreme weather conditions.

$$\bar{V} = \max_x \sum_{i=1}^{I=64} C_i(S_0, x_0^*) + \bar{V}_i^*(S_0^*) \quad (10)$$

The ending value, $\bar{V}_i^*(S_{t50}^*)$, was computed using the value function approximation and the state of the ending landscape. Although this is a finite sum approximation to the infinite horizon return, the discount factor δ^{150} is 0.0022, so the difference is negligible.

The value function approximation $\bar{V}(S|\theta)$ cannot perfectly predict the value of a specific stream of rewards because of stochasticity, but it should be very close to the average V_k generated by the simulations. For the study landscape, $\bar{V}(S|\theta) = \$15.2719$ million and the average $V_k = \$15.2939$ million, a difference of only 0.14%.

5.2. Policy evaluation

In order to evaluate the policy determined by the approximate dynamic programming algorithm, we compared the ADP policy to 3 benchmark policies:

1. Harvest at Faustmann rotation age. This policy ignores fire risk entirely, but is the standard rule in a risk-free setting. The Faustmann rotation provides no guidance for fuel treatment, and no fuel treatments occur with this policy.
2. Harvest at Reed rotation age. This policy incorporates fire risk at the stand level, but omits spatial interactions with other stands in the landscape. No fuel treatments occur with this policy either.
3. Harvest at Reed rotation age and treat fuels whenever fuel treatment leads to an improvement in fuel conditions. We chose this as a reasonable rule-of-thumb policy for harvest and fuel treatment, and it is the standard by which we judge the ability of the ADP policy to improve management decisions.

We simulated outcomes under each benchmark policy using the same 500 ignition and weather trajectories that we employed in the ADP policy simulations. Fig. 4 is a series of histograms of the realized outcomes (V_k , $k = 1 \dots 500$) for each benchmark policy and for the ADP policy. The means and variances are shown in Table 1. We conducted a variety of pair-wise statistical tests to confirm that the differences in outcomes between each of the policies were significant. The results of these tests are shown in Table 2. The ADP policy outperforms the benchmark policies on average, and the distribution of outcomes under the ADP policy is statistically greater than it is under the alternative policies. However, this does not imply that the ADP policy will outperform the benchmarks in every simulation. If there is a particularly favorable stream of fire and weather events, the Faustmann rotation obtains a better outcome. Conversely, in particularly unfavorable weather streams, the Reed policy leads to a better outcome for the landscape. The Faustmann policy is the only strategy that can achieve the maximum attainable landscape value. However, this only occurs if there are no fires over the simulation period. In fact, it generated worse outcomes on average and had a much higher variance.

In Fig. 5a–c, scatterplots comparing the outcomes for the ADP policy to the benchmark policies are shown. Each point on the plot represents one of the 500 ignition/weather trajectories. Points below the diagonal line show weather/ignition trajectories where the ADP policy was better; points above the diagonal line show trajectories where the benchmark policy was better. The Faustmann policy outperformed the ADP policy in only 11 (2%) of the 500 simulations (see Fig. 5a). The Reed policy increases the value of the average outcome compared to the Faustmann policy by harvesting the trees at an earlier age, which reduces the probability that they will be damaged by fire, but also reduces the profits generated by harvest. The Reed policy outperformed the ADP policy in 35 (7%) of the 500 simulations (Fig. 5b). The final benchmark is the Reed policy with fuel treatment. It outperforms the ADP policy on 69 (13.8%) of the 500 simulations and mitigates losses in the worst ignition/weather trajectories, as can be seen in Fig. 5c. The ADP policy achieves larger average outcomes compared to this policy; however, the ADP policy also has a larger variance, and additional probability of loss in the worst weather/ignition trajectories.

5.3. Implications for forest management

The actions taken using the benchmark policies are straightforward and easy to understand. If a stand reaches the target rotation age, (or in one case, target fuel conditions) it should be harvested (or treated). The Reed harvest age of 35 years represents a “harvest young to avoid loss” strategy, while the Faustmann harvest age of 54 years ignores fire risk altogether. The average ADP harvest age of 39 years falls between these two extremes. There is a distribution, shown in Fig. 6, of harvest ages found by the ADP policy, because this policy takes into account how an action will affect the entire neighborhood value and not just the individual stand value. Hence, harvest age differs depending on the condition of the surrounding stands. Under higher risk landscape conditions, it may be optimal to harvest early; in other safer configurations, it may be optimal to delay harvest.

Land managers can exert some control over fire arrival probability by using fuel treatments to slow spread. However, determining when and where fuel treatments should be placed on the landscape is a difficult problem. The cost of the fuel treatment is only justified if it is exceeded by the increase in expected value created by changing the fire arrival probabilities across the landscape. In the third benchmark policy, fuel treatment was applied any time it would create a state with less dangerous fuel conditions, regardless of economic considerations. With the ADP policy, fuel treatment is applied only when expected benefits exceed costs. In our simulations, that occurs at 3 stages in the life of the stand. The ADP policy led to fuel treatment for all stands at age 20. This reduces moderate-hazard fuels to low-hazard fuels at a time when the stand is beginning to produce merchantable volume. Fuel hazard is highest in very young stands that have no merchantable volume. There are two reasons to treat these stands—one is to avoid having to replant and start a new rotation, thereby postponing receipt of future harvest revenue; the other is to protect merchantable volume

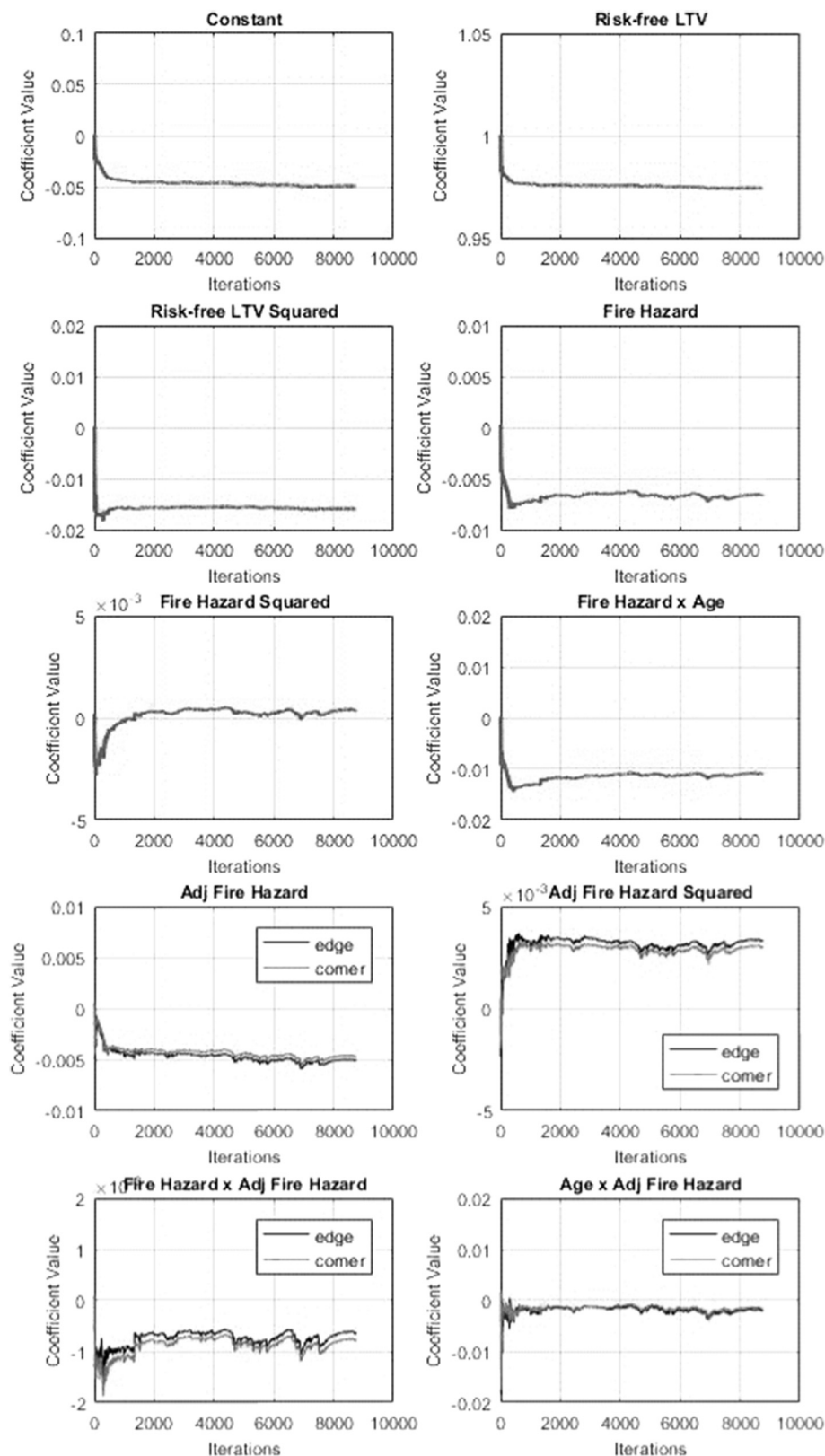


Fig. 3. Evolution of value function approximation coefficients during the value iteration process.

on adjacent stands. For about half of the stands in the earliest age classes and highest hazard fuel conditions, the ADP policy finds that the benefits of fuel treatment exceed the costs, and treatments are applied. A few of the oldest 45-to-47-year-old strategically-placed stands receive fuel treatment that reduces fire hazard only slightly but protects standing volume in these and adjacent stands.

While, in our model, wealth maximization through timber management is the only objective, we recognize that changing the management policy will also change ecological outcomes. As an example, we present histograms of fire size under the ADP policy compared to benchmark policies for the 500 simulations in Fig. 7a–c. Forest land managers may prefer policies that make large fires less likely. While this

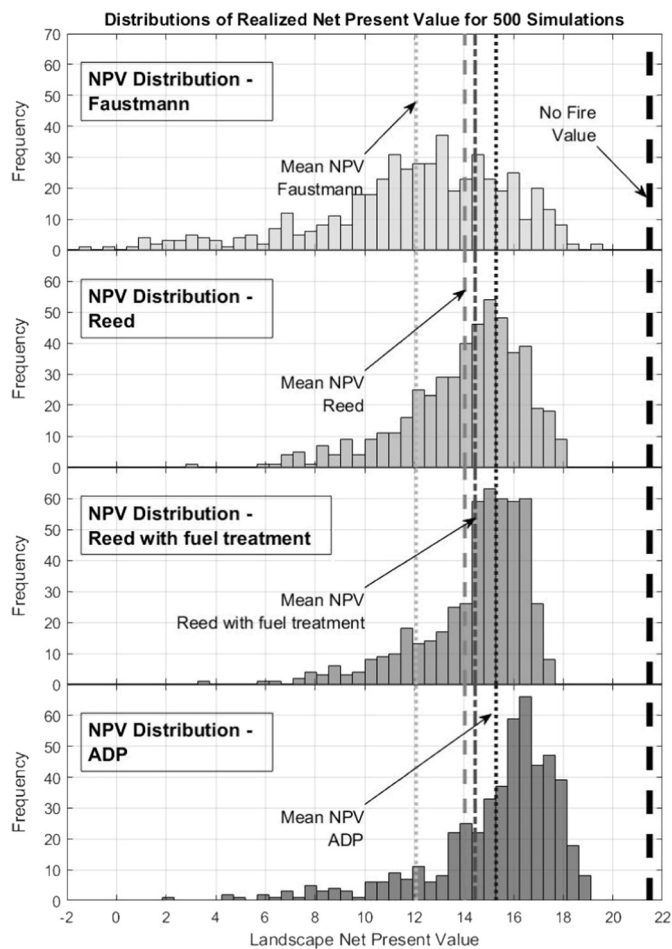


Fig. 4. Histograms depicting the distribution of net present values generated by the ADP policy and the benchmark comparison policies for 500 trajectories of ignition/weather events.

Table 1

Mean and variance of the net present value (NPV) generated by benchmarks and ADP policy.

Policy	Mean V_k (\$ million)	Variance
Faustmann	12.07	14.40
Reed	14.03	5.97
Reed with fuel treatment	14.45	4.81
ADP	15.29	7.18

Table 2

Pairwise comparisons for ADP policy outcomes and benchmark policy outcomes.

	Average difference (\$ million)	p-Value of one-tail Welch's t -test	p-Value of one-tail Kolmogorov-Smirnov test	Effect size-Cohen's d
ADP v. Faustmann	3.2222	0.0000	0.0000	0.9809
ADP v. Reed	1.2642	0.0000	0.0000	0.4719
ADP v. Reed with fuel treatment	0.8423	0.0000	0.0000	0.3144

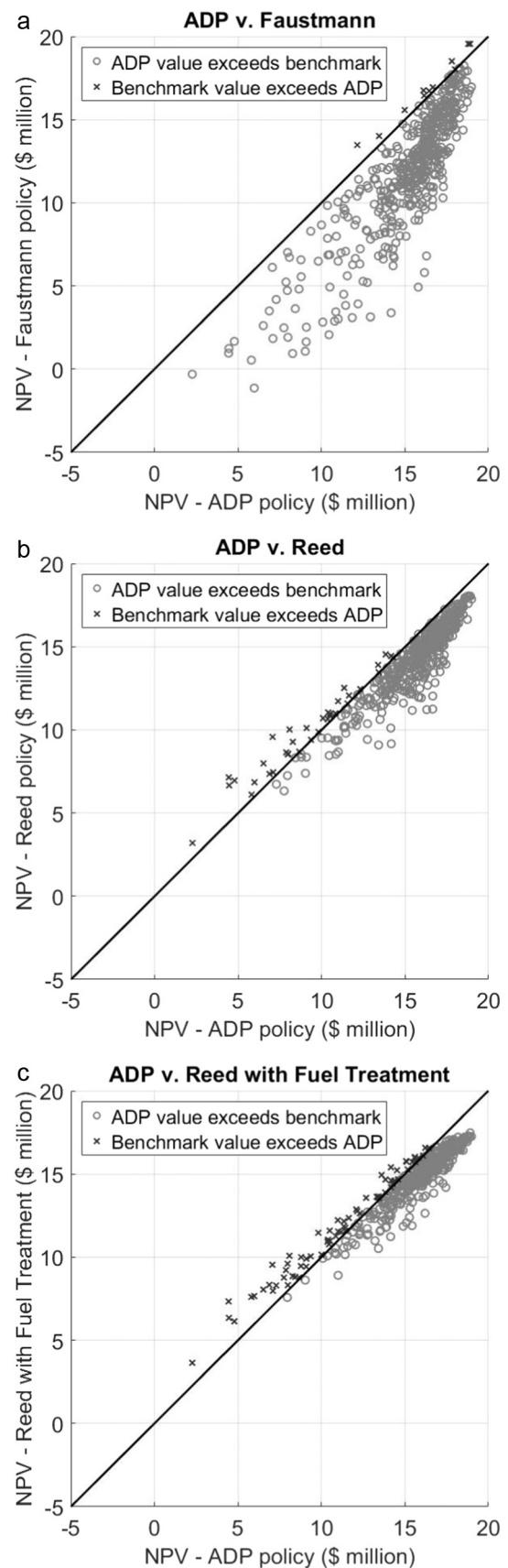


Fig. 5. Compares two different policies for the same trajectory of fire and weather events. Each of the 500 points represents 1 trajectory. Points below the diagonal line (circles) represent trajectories in which the ADP policy leads to a higher NPV than the benchmark. a-Faustmann, b-Reed, c-Reed with fuel treatment.

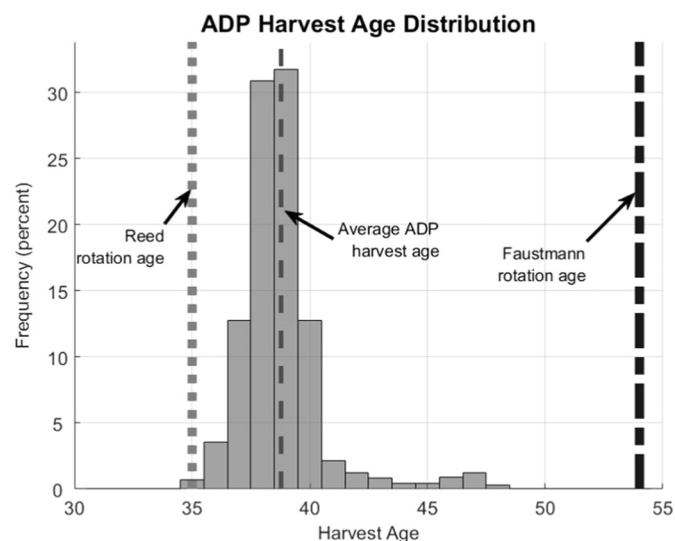


Fig. 6. Distribution of harvest ages that result from following the ADP policy. Faustmann and Reed harvest ages are included for comparison.

was not explicitly considered in our objective function, the choice of policy will affect this outcome. The ADP policy leads to fewer large fires and more small fires than any of the alternative policies because spatial interactions that facilitate fire spread are explicitly considered. In a second example, we show histograms for volume of merchantable timber destroyed by each fire in Fig. 8a–c. Of all the policies we considered, fires destroy the most merchantable biomass under the Faustmann policy and the least amount of merchantable biomass under the Reed policies. This is true even though fires under the Reed policy are more likely to burn a larger area. Despite the fact that fires under the ADP policy destroy more merchantable timber than they do under the Reed policy, the ADP policy leads to higher average value, because the stands harvested under the ADP policy are more mature and generate more revenue.

6. Conclusions

In this study, we applied an approximate optimization method to the problem of forest management in the face of fire on a 64-unit landscape of 4 mile² that represents conditions in southwestern Oregon. This method uses approximate dynamic programming (also known as reinforcement learning) to estimate a value function, and is combined with heuristic optimization techniques using the construct of the post-decision state to find optimal actions in each period. To our knowledge, this is the first application of approximate dynamic programming to the problem of optimal management of a fire-prone landscape. Previous studies either ignored spatial interactions created by fire spread or ignored the ability a manager has to adjust his or her actions optimally in future periods once stochastic outcomes have been realized. Applications of stochastic dynamic programming that accounted for both issues have been limited to problems with very simple ecological models (e.g. Konoshima et al., 2008, 2010).

We believe our analysis establishes a platform that will provide the means to explore many vexing wildland fire policy problems. In this study, we considered only the objective of net present value of timber harvest, building upon established ideas for determining optimal timber management in the face of possible stand destruction. We demonstrated that optimal harvest age and other forest management choices vary depending on the surrounding landscape conditions that contribute to fire risk. With respect to harvest age, we confirmed the finding of Konoshima et al. (2008, 2010) that there exists a trade-off between harvesting early to protect on-site timber value and holding standing timber longer to reduce fire spread risk to adjacent stands. The

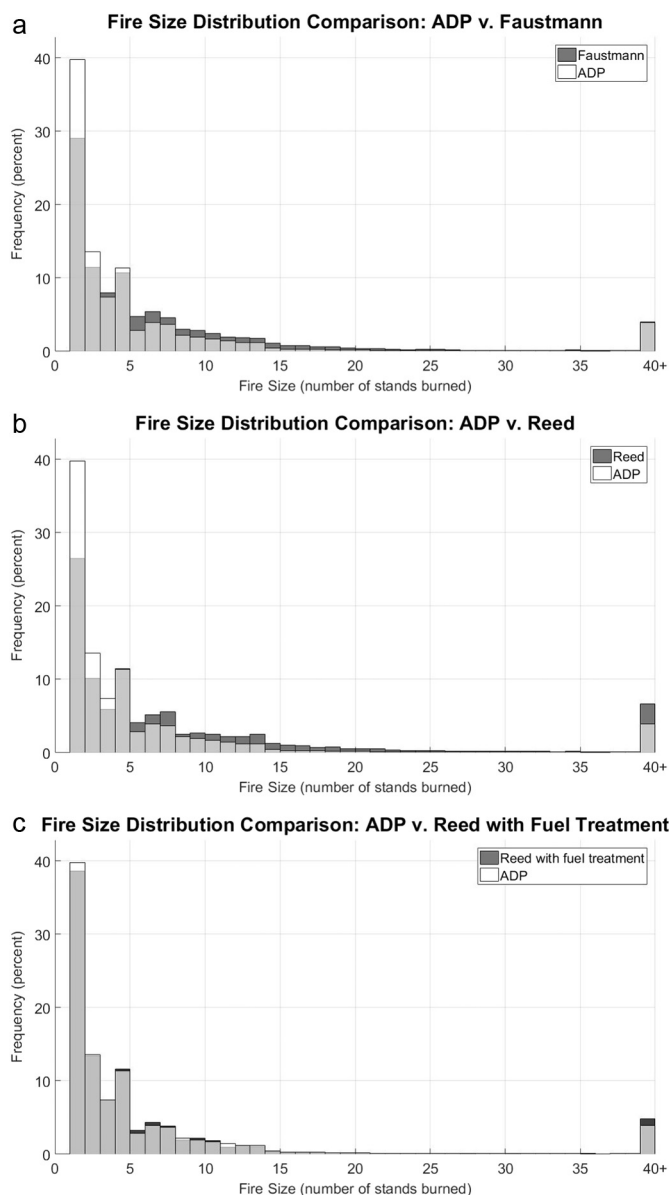


Fig. 7. Distribution of fire sizes that arise as a result of following the ADP policy compared the distribution of fire sizes that result from the benchmark policies for the same set of ignition/weather trajectories. a-Faustmann, b-Reed, c-Reed with fuel treatment.

distribution of optimal harvest ages falls between those predicted by the Reed and Faustmann models, which are optimal in non-spatial settings. We found that the placement of fuel treatments also depends on conditions of the surrounding landscape—with stands more likely to receive treatment when there are timber values at risk on adjacent stands.

Some possible extensions include:

Nontimber objectives – We modeled optimal management for timber as a financial asset under risk of stand-destroying fire. Fire is one example of a stochastic ecological process that creates spatial interactions—windthrow, the spread of invasive species, and the movement of wildlife are others. Likewise, maximizing the financial asset value of the forest is just one example of a forest management objective. A landowner might instead want to manage fire to minimize threat to habitat connectivity, protect structures in the wildland urban interface or meet a restoration objective. The optimization framework demonstrated here is flexible enough to be adapted for these and other problems related to the management of complex forested landscapes.

Complexity of the landscape and/or ecological models – While the ecological models that underlie this work are complex enough to

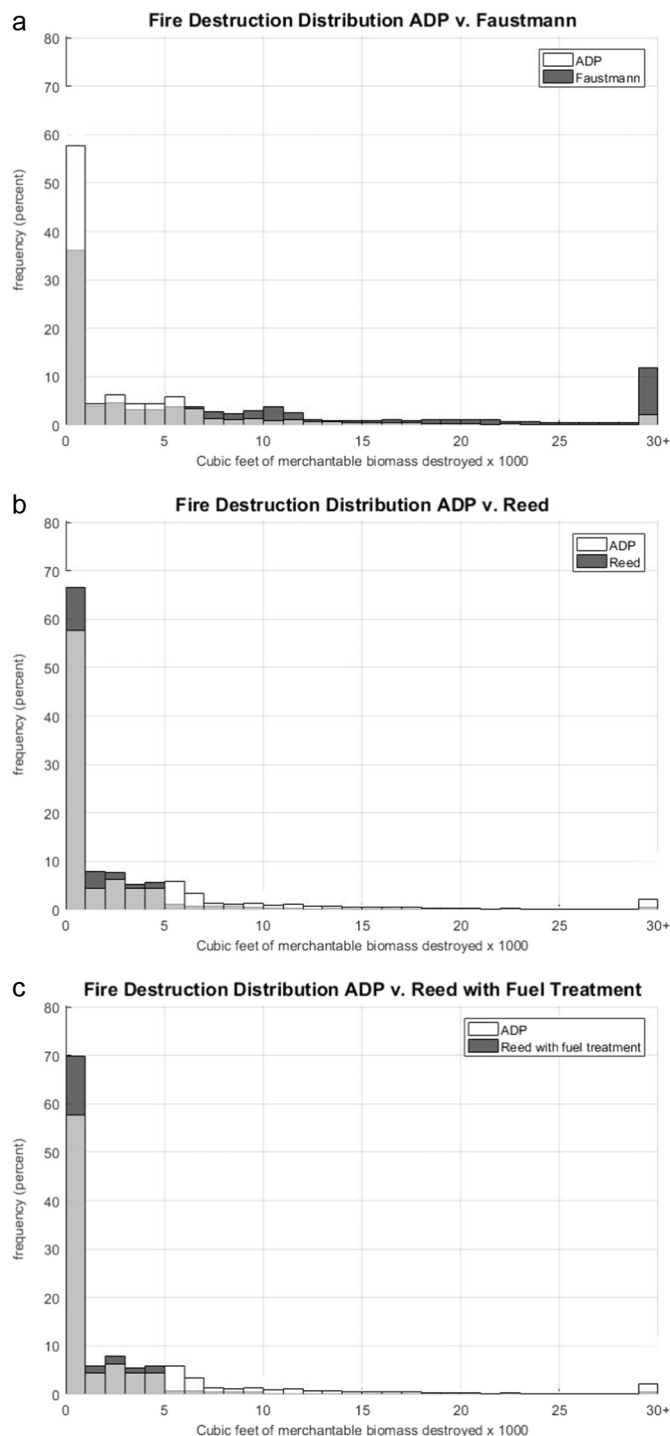


Fig. 8. Distribution of merchantable biomass destroyed by fire as a result of following the ADP policy compared the distribution that result from the benchmark policies for the same set of ignition/weather trajectories. a-Faustmann, b-Reed, c-Reed with fuel treatment.

preclude the use of exact solution methods, we made several

Appendix A

Value iteration algorithm

We used a value function iteration algorithm to estimate the coefficients (θ) for the value function approximation—this process is described by Powell (2007) and was modified by us to account for challenges specific to this problem. The procedure used in this paper for approximating the value function is depicted in Fig. A.1 and a short, written description of how each step is implemented is included below.

simplifications that we hope to revisit as we extend this work. While it is possible to suggest general ideas from our analysis, it is difficult to draw inferences for a particular site without accurately modeling attributes specific to that site. In a planning context, the model landscape could be aligned to the real landscape – e.g. the number of units could be increased and the spatial configuration does not need to be specified as a grid. Topography and site-specific weather patterns that affect fire behavior, such as prevailing wind direction and intensity, could be added to the model by tailoring the spread rates associated with each stand across the landscape. By incorporating more sophisticated ecological models, we could address questions about fire as a management tool. For example, we assumed all fire is stand-destroying. By also modeling non-catastrophic surface fire, we could differentiate between “good” fire that improves forest health and “bad” fire that destroys things we value.

Risk preferences – Our model of a profit-maximizing agent assumes that the agent is risk neutral, which implies that the agent will try to maximize the expected value of the landscape. In some cases, this might be a reasonable assumption. Large landowners, or investors who hold timber as part of a large portfolio of assets where risk is spread over time and space may prefer a risk-neutral management strategy. For other forest landowners, especially smaller landholders, a risk-averse strategy may be preferred instead. Their forest may represent their retirement savings or their children's college fund. A risk-averse agent might prefer the Reed with fuel treatment policy over the ADP policy, while this policy has a lower average outcome it also has a smaller variance in outcomes.

Multiple landowners – Many of the most difficult policy challenges arise when fire crosses property boundaries. We hope to extend this model to incorporate multiple landowners so that spatial externalities created by fire spread can be analyzed.

In this paper, we show the importance of accounting for spatial interactions created by ecological processes and we demonstrate a method for analyzing those interactions in a dynamic decision-making framework. We showed how Bellman's equation can be modified to account for spatial interactions and how machine learning techniques provide a tractable method for incorporating spatial information into an already complex stochastic dynamic optimization problem. The optimization framework we developed provides a platform for policy analysis in the context of complex ecological systems where disturbance is a factor. The value function we estimate provides a means to model the cost of risk created by altering landscape conditions without damage needing to be explicitly realized. It allows the marginal effect of management actions that change the state of the landscape to be characterized, and it provides a way to compare the known cost of an action to the expected benefits of that action on the landscape.

Acknowledgements

This material is based upon work supported by the National Science Foundation under CyberSEES Computing and Visualizing Optimal Policies for Ecosystem Management (grant no. 1331932). The authors would like to thank Rachel Houtmann for her insights about parameterizing models of vegetation growth and fire spread. We would also like to thank two anonymous reviewers for their helpful comments.

Step 1

Define the landscape parameters, the initial state, and the initial estimation of the value function. These are described in detail in the data and parameters section.

The following series of steps is repeated until the maximum number of cycles (indexed by $k = 1 \dots K$) is reached, or the stopping criteria is met.

Step 2

The current period is set to $t = 0$. The starting state for the cycle is determined; with some probability, it is the initial state defined in step one, otherwise it is a randomly generated starting state. Choosing a different starting state allows the algorithm to learn from states it may not otherwise see, and is one strategy for ensuring adequate exploration of the state space. The number of periods T (years) in the cycle is defined. For early cycles, the number of periods is relatively small. This is because the landscape evolves over time based in part on actions chosen by the agent. These actions are determined based on the current estimation of the value function. We know that the initial approximation of the value function is likely to be wrong and we expect it to improve as the approximation process progresses. It is important to expose the algorithm to states that are similar to those that are likely to occur under the optimal policy in order to accurately approximate the value function. In the early stage of the value function estimation process, going too many time steps into the future may lead to states that are unlikely to occur under the optimal policy. In later cycles, the number of periods is increased (i.e. $T(k)$, $\frac{dT}{dk} > 0$).

Step 3

In this step, the optimal action for the current state, given our current estimation of the value function, is found. The optimal action is implied by the value function estimate, and can be written as $x_t^*(S_t) = \operatorname{argmax}_{x_t} \sum_{i=1}^I [C_i(S_t, x_t) + \bar{V}_i^n(S_t^x | \theta^n)]$. This is the argmax for the sum of Eq. (7) for all stands. Because of the spatial interactions on the landscape, determining the optimal action is a complex combinatorial problem. Actions taken on a particular stand not only affect the value of that stand, but also the value of other stands on the landscape. For each stand in the landscape, the agent can harvest timber, treat fuel, do both, or do neither. On our 64-stand landscape, 4^{64} different action combinations are possible. It is not computationally practical to evaluate each one, or use exact integer programming techniques; instead, we employ heuristic methods (hybrid simulated annealing/tabu search algorithm) (Kirkpatrick, 1984; Reeves, 1993) to rapidly find a near-optimal solution for the current state, given our current estimate of the value function.

Step 4

The current approximation of the value function, $\bar{V}_i^n(S_t^x | \theta^n)$ determines the chosen actions. These actions determine which states are visited. It is important to ensure that we explore states that might not otherwise be found by an agent to learn about the value of these other states. This is especially important in early iterations because the value function approximation is not likely to be accurate in these iterations, and may incentivize actions that would not be chosen if the true value function were known. Because of the size of the state space, it is impossible to see every potential state; for states that are explored, it is impossible to see every potential stochastic outcome that could occur. Therefore, it is important to focus exploration in likely regions of the state space. This can be accomplished by modifying the action (x_t^*) found in step 3 by making a small change to this action with probability (P). The probability of deviating from the optimal action (x_t^*) decreases as the number of iterations increases and the approximation of the value function improves, $P(n)$, $\frac{dP}{dn} < 0$. In the case of this problem, small modifications to the optimal action might be postponing harvest one period, or choosing one extra fuel treatment on the landscape. The action for period t , either the optimal action (x_t^*) or modified action ($x_t' = f(x_t^*)$), is used to calculate the post-decision state. The post-decision state is used in fire simulations and in updating the value function approximation.

Step 5

In this step, we advance to the next time period (i.e. $t = t + 1$) and fire events are simulated. Multiple simulations ($m = 1 \dots M$) are created. A vector of random ignition and weather variables (W^m) is drawn to determine if and where fire occurs. Next the post-decision state S_{t-1}^x , the possible fire occurrence, and vegetation-growth models are used to generate a set of new states $S_t^m = S'(S_{t-1}^x, W^m)$ for ($m = 1 \dots M$) and each resulting state, is saved. The average value of the resulting states will be used to update the value function coefficients in the next steps. To prevent the value function update from over-reacting to low probability events that are drastically different from the expected outcome, the number of simulations is higher in early cycles. After a number of cycles have been completed, the value function approximation should move toward a more accurate estimate, so the number of simulations per time step is decreased to get more variance in the observed outcomes and fine-tune the value approximation $M(k)$, $\frac{dM}{dk} < 0$.

Step 6

In this step, the optimal action for each of the M states resulting from the fire simulations is determined. The optimal action for state S_t^m given the current approximation of the value function is $x_t^{m*} = \operatorname{argmax}_{x_t^m} \sum_{i=1}^I [C_i(S_t^m, x_t^m) + \bar{V}_i^n(S_t^{x,m} | \theta^n)]$. This action is used to calculate the value of being in state S_t^m , and determine each stand's contribution to the value of that state. This contribution is found using Eq. (7), i.e. $Q_i(S_t^m, x_t^{m*} | \bar{V}_i^n)$. The discounted value of being in the current state is a realized value of being in last period's post-decision state.

Step 7

In this step ($n = n + 1$), and the value function approximation $\bar{V}_i^n$ is updated. The updating process is written as $\bar{V}_i^n \leftarrow U(S_{t-1}^x, \bar{V}_i^{n-1}, Q_i^{avg})$. This

process uses the value for the post-decision state for the previous period calculated using the current value function approximation $\bar{V}_i^{n-1}(S_{t-1}^x | \theta^{n-1})$ and the average of the realized values of the current period states S_t^m , ($m = 1 \dots M$). The average value for each stand is $Q_i^{avg} = \sum_M Q_i(S_t^m, x_t^{m*})/M$. The method used for updating the coefficients for the value function estimate is a stochastic gradient which works as follows:

$$\theta^n = \theta^{n-1} - \alpha_{n-1} (\bar{V}_i^{n-1}(S_{t-1}^x) - \delta Q_i^{avg}(S_t^m, x_t^{m*})) \frac{\partial \bar{V}_i^{n-1}}{\partial \theta} \quad (\text{A.1})$$

θ^n is the vector of new parameter estimates for the value function. These new parameters are determined by adjusting the current parameter values using the difference between the predicted value of the stand in the post decision state $\bar{V}_i^{n-1}(S_{t-1}^x)$, and the average of the realized values of the stand after it has transitioned Q_i^{avg} , multiplied by the change in the value function for a change in the parameter. A learning rate α_{n-1} , a number between zero and one, also referred to as a step size, determines how much weight is placed on the most recent observation versus how much weight is placed on the current parameter value.

The value function we estimate, $\bar{V}_i(S_t^x)$, is an approximation of the contribution a stand makes to the overall landscape value. Since all stands on the landscape are homogeneous and use the same value function, it is possible to use every stand in the landscape to update the estimate of the coefficients. Since there are 64 stands in our landscape, each stand's new coefficients are given a weight of 1/64 to determine the coefficients for the next iteration's value function. If this method were to be applied on a landscape with heterogeneous stands, the attributes that create heterogeneity would need to be included in the value function or a separate value function approximation would need to be estimated for each stand.

Since the value function estimate $\bar{V}_i^{n-1}$ is linear in θ , the derivative of the value function estimate with respect to the coefficient parameters is the vector of basis functions.

$$\theta^n = \theta^{n-1} - \alpha_{n-1} (\bar{V}_i^{n-1}(S_{t-1}^x) - \delta Q_i^{avg}(S_t^m, x_t^{m*})) \begin{pmatrix} \phi_1(S_{t-1}^x) \\ \phi_2(S_{t-1}^x) \\ \dots \\ \phi_J(S_{t-1}^x) \end{pmatrix} \quad (\text{A.2})$$

The learning rate or step size is a very important factor in determining the ability of this algorithm to accurately approximate the value function. The step size can take many different forms, but it must have certain properties to guarantee convergence; these properties are met with a step size of $\alpha_n = 1/n$ (Powell, 2007). However, the step size $1/n$ goes to zero too fast to get convergence in practice. This is because updates to the value function coefficients are initially made using incorrect estimates of the true value of the state. Because of this, it is important to weight later observations more heavily than early observations. Another important factor that leads to slower convergence is variance in the outcomes caused by the stochastic process. An event that deviates significantly from the true expected value may send the wrong signal if too much weight is placed on that observation. Therefore, it is important to carefully choose a step size rule. There are several different ways for choosing a step size, including constant step sizes or step size rules that decrease over a number of iterations to a target step size.

For this problem, we used the bias-adjusted Kalman filter rule outlined by Powell (2007). This rule chooses a step-size by estimating the bias and variance of the value function approximation after n iterations. It increases the step-size if bias is large and decreases the step size if variance is large. By increasing the step-size when bias is large, the parameters move more quickly toward their true values. By decreasing the step-size when the variance is large, the parameter updates are prevented from overreacting to any particular stochastic realization. Accounting for bias and variance, and adjusting the step-size accordingly leads to faster convergence. The step sizes created by this process are depicted in Fig. A.2.

Step 8

If there are more time periods in the current cycle, one of the states M created in step 5 is randomly chosen as the current state and the algorithm returns to step 3. If no time steps are left in the current cycle and the stopping criteria have not been achieved, the algorithm returns to step 2. If the stopping criteria have been met, or if the specified number of cycle has been completed, the algorithm is stopped.

The algorithm is stopped early in two cases. First, we stopped the algorithm if the change in the coefficients (θ) between cycles was less than some tolerance (i.e. $|\theta_j^n - \theta_j^{n-1}| < \text{tolerance}$, for all $j = 1 \dots J$); coefficients will never completely converge because their value depends on stochastic realizations. The other situation where the algorithm is stopped early is when something has gone wrong. If one or more of the coefficients in θ is outside of a 'reasonable' range at the end of an updating cycle, we assume the algorithm will not converge and stop the algorithm early. This can happen if the basis functions (ϕ) are poorly specified, or a bad initial guess θ^0 for the coefficients is chosen. Mathematically, the initial guess shouldn't matter, but we found that it did matter in practice.

We confirm that convergence is achieved by comparing the value for the landscape predicted by our approximation of the value function to the average value of the landscape for simulations under the optimal policy as described in Section 5.1. We also confirm that coefficients (θ) have likely converged by tracking the evolution of the coefficient values and looking at a linear trend line over the last iterations. The slope of the trend line should be near zero.

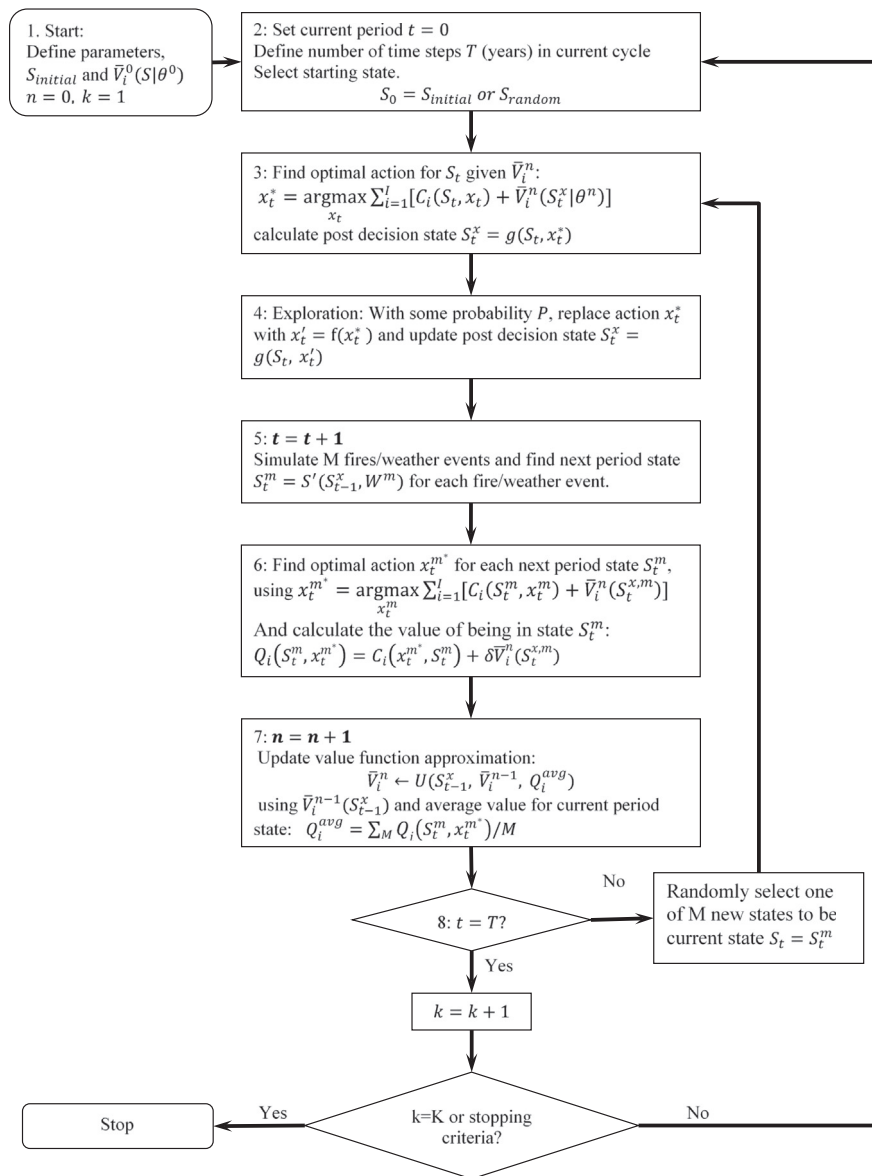


Fig. A.1. Value function iteration algorithm. Index definitions - n indexes number of value function updates, k indexes number of updating cycles, t indexes number of time-steps (years) in a cycle, m indexes number of fire/weather simulations per time-step.

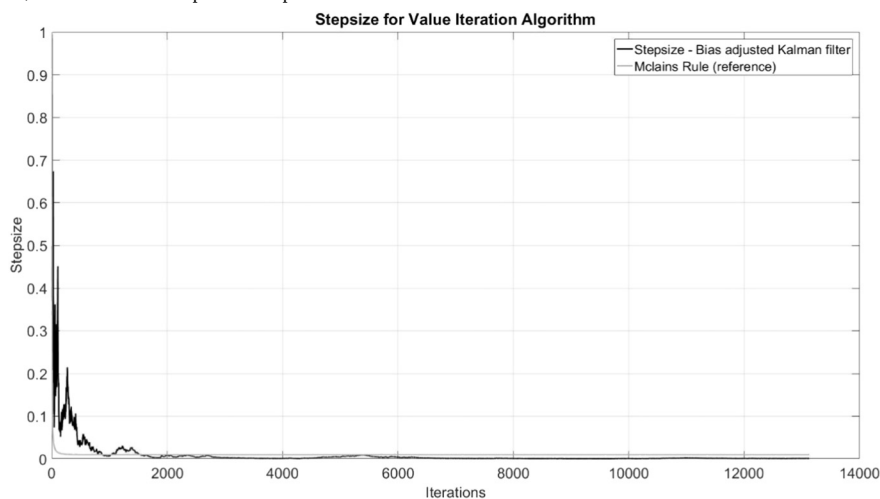


Fig. A.2. Step size/learning rate.

References

- Ager, Alan A., Vaillant, Nicole M., Finney, Mark A., 2010. A comparison of landscape fuel treatment strategies to mitigate wildland fire risk in the urban Interface and preserve old forest structure. *For. Ecol. Manag.* 259 (8), 1556–1570.
- Amacher, Gregory S., Malik, Arun S., Haight, Robert G., 2005. Not getting burned: the importance of fire prevention in forest management. *Land Econ.* 81 (2), 284–302.
- Anderson, Hal E., 1982. Aids to determining fuel models for estimating fire behavior. In: *Gen. Tech. Rep. INT-122*. U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, Utah 22pp.
- Andrews, Patricia L., Bevins, Collin D., Seli, Robert C., et al., 2003. BehavePlus Fire Modeling System: Version 2.0: User's Guide. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.
- Bellman, Richard E., 1957. *Dynamic Programming*. Princeton University Press pp. 151.
- Bettinger, Pete, Sessions, John, Chung, Woodam, Graetz, David, Boston, Kevin, 2003. Eight heuristic planning techniques applied to three increasingly difficult wildlife planning problems: a summary. In: *Systems Analysis in Forest Resources*. Springer, Netherlands, pp. 249–257.
- Bradshaw, Larry S., McCormick, Erin, 2000. FireFamily Plus User's Guide. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.
- Calkin, David, Gebert, Krista, 2006. Modeling fuel treatment costs on forest service lands in the Western United States. *West. J. Appl. For.* 21 (4), 217–221.
- Chung, Woodam, Jones, Greg, Krueger, Kurt, Bramel, Jody, Contreras, Marco, 2013. Optimizing fuel treatments over time and space. *Int. J. Wildland Fire* 22 (8).
- Daigneault, A., Miranda, M., Sohngen, B., 2010. Optimal forest management with carbon sequestration credits and endogenous fire risk. *Land Econ.* 86 (1), 155–172.
- Dixon, Gary E., et al., 2002. *Essential FVS: A User's Guide to the Forest Vegetation Simulator*. USDA-Forest Service, Forest Management Service Center, Fort Collins, CO.
- Faustmann, M., 1968. In: Gane, M. (Ed.), "Calculation of the Value Which Forest Land and Immature Stands Possess for Forestry." Translated by Linnard, W. in Martin Faustmann and the Evolution of Discounted Cash Flow. University of Oxford, pp. 1849.
- Fight, Roger D., LeDoux, Chris B., Ortman, Tom L., 1984. Logging costs for management planning for young-growth coast Douglas-fir. In: *Gen. Tech. Rep. PNW-GTR-176*. U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station, Portland, OR (10 pp.).
- Garcia-Gonzalo, J., Pukkala, T., Borges, J.G., 2014. Integrating fire risk in stand management scheduling. An application to maritime pine stands in Portugal. *Ann. Oper. Res.* 219 (1), 379–395.
- Hobbs, S.D., Tesch, S.D., Owston, P.W., Stewart, R.E., Tappeiner II, J.C., Wells, G.E., 1992. Reforestation Practices in Southwestern Oregon and Northern California. Forest Research Laboratory, Oregon State University, Corvallis.
- Kirkpatrick, Scott, 1984. Optimization by simulated annealing: quantitative studies. *J. Stat. Phys.* 34 (5–6), 975–986.
- Konoshima, Masashi, Montgomery, Claire A., Albers, Heidi J., Arthur, Jeffrey L., 2008. Spatial-endogenous fire risk and efficient fuel management and timber harvest. *Land Econ.* 84 (3), 449–468.
- Konoshima, Masashi, Albers, Heidi J., Montgomery, Claire A., Arthur, Jeffrey L., 2010. Optimal spatial patterns of fuel management and timber harvest with fire risk. *Can. J. For. Res.* 40 (1), 95–108.
- Oregon Department of Forestry, 2016. *Timber Sales – Log Prices*. <https://www.oregon.gov/ODF/Working/pages/TimberSales.aspx>.
- Powell, Warren B., 2007. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. vol. 703 John Wiley & Sons.
- Powell, Warren B., 2009. What you should know about approximate dynamic programming. *Naval Research Logistics (NRL)* 56 (3), 239–249.
- Puterman, Martin L., 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.
- Reed, William J., 1984. The effects of the risk of fire on the optimal rotation of a forest. *J. Environ. Econ. Manag.* 11 (2), 180–190.
- Reed, W.J., 1989. Optimal investment in the protection of a vulnerable biological resource. *Nat. Resour. Model.* 3 (4), 463–480.
- Reed, William J., 1993. The decision to conserve or harvest old-growth forest. *Ecol. Econ.* 8 (1), 45–69.
- Reeves, Colin R., 1993. *Modern Heuristic Techniques for Combinatorial Problems*. John Wiley Sons, Inc.
- Reinhardt, E., Crookston, N.L., 2003. The fire and fuels extension to the forest vegetation simulator. In: *General Technical Report, RMRS-GTR-116*. USDA Forest Service, Rocky Mountain Research Station, Ogden, UT (209 pp.).
- Routledge, R.D., 1980. The effect of potential catastrophic mortality and others unpredictable events on optimal forest rotation policy. *For. Sci.* 26 (3), 389–399.
- Row, Clark, Fred Kaiser, H., Sessions, John, 1981. Discount Rate for long-term forest service investments. *J. For.* 79 (6), 367–376.
- Samuelson, P.A., 1976. Economics of forestry in an evolving society. *Econ. Inq.* 14, 466–492.
- Sutton, Richard S., Barto, Andrew G., 1998. *Introduction to Reinforcement Learning*, 1st ed. MIT Press, Cambridge, MA, USA.
- Tappeiner, J.C., Maguire II, D.A., Harrington, T.B., 2007. *Silviculture and Ecology of Western US Forests*. Oregon State University Press.
- Thorpe, Dan, 2011. *Boot Prints: A Centennial Summary of Activities and Events of Oregon Department of Forestry in Jackson and Josephine Counties*. Southwest Oregon District, Oregon Department of Forestry, Central Point, Or.
- USDA, 2015. Forest Service "The Rising Cost of Wildfire Operations: Effects on the Forest Service's Non-Fire Work".
- Van Roy, B., Bertsekas, D.P., Lee, Y., Tsitsiklis, J.N., 1997. A neuro-dynamic programming approach to retailer inventory management. In: *Proceedings of the 36th IEEE Conference on Decision and Control*, vol. 4. pp. 4052–4057.
- Watkins, C., Dayan, P., 1992. Q learning: technical note. *Mach. Learn.* 8, 279–292.
- Wei, Yu, Rideout, Douglas, Kirsch, Andy, 2008. An optimization model for locating fuel treatments across a landscape to reduce expected fire losses. *Can. J. For. Res.* 38 (4), 868–877.