Jul 1st, 12:00 AM

Decision Support for Environmental Monitoring and Restoration: Application of the Partially Observable Markov Decision Process

David Tomberlin

International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Decision Support for Environmental Monitoring and Restoration: Application of the Partially Observable Markov Decision Process[1]

David Tomberlin
National Marine Fisheries Service[2]
1315 East-West Highway, Silver Spring, MD 20901 USA
david.tomberlin@noaa.gov

Abstract: Environmental monitoring programs can improve management over time, but generally require that correspondingly less time and money be put into direct restoration efforts such as revegetation or dam removal. Thus, budget constraints compel environmental managers to make difficult decisions regarding the allocation of scarce funds and personnel between environmental monitoring and environmental restoration. Among other factors, the best allocation of resources between monitoring and restoration—or, more generally, learning and doing—will depend on the quality of information available from a monitoring program. This paper demonstrates the application of the partially observable Markov decision process (POMDP) as a framework for investigating the optimal intensity of monitoring given stochastic state dynamics and imperfect observations on state variables. Specifically, the paper addresses the problem of choosing among a set of available monitoring protocols that differ in their costs and the type of information they provide. An empirical application of the model to erosion control in California watersheds demonstrates the utility of the resulting decision policy as well as limitations to the approach.
.

Keywords*:* monitoring; optimal learning; value of information; adaptive management.

1. INTRODUCTION

Environmental management budgets often seem small relative to the protection and restoration work managers would like to pursue. A natural temptation is to skimp on monitoring because it diverts funds from projects that appear to provide more direct conservation benefit, such as cleaning up polluted sites or staffing nature reserves. While the information derived from environmental monitoring programs has obvious potential to support better long-term management decisions, the opportunity cost of dedicating scarce funds to monitoring is often enough to induce managers to forego monitoring. And while there is a general appreciation that monitoring can contribute to management, weighing the dynamic and uncertain benefits and costs of monitoring presents a difficult decision problem. This paper develops a framework for analyzing the role of monitoring as part of a more general environmental management program, specifically for assessing the choice of monitoring protocol when managers have access to more than one protocol and when they may also choose to forego monitoring entirely.

While the issue of choosing among monitoring protocols that differ in cost and information quality arises in many areas of environmental management—and the rest of life, for that

[1] An earlier version of this paper was given at the HydroPredict International Conference on Predictions for Hydrology, Ecology and Water Resource Management, 15-18 Sept. 2008, Prague, Czech Republic.
[2] This paper represents the opinions of the author and not necessarily those of the National Marine Fisheries Service.

matter—the application considered here is the choice among surface water quality monitoring protocols within broader program to control sediment loading. In northern coastal California, excess sediment is considered one of the primary threats to the freshwater habitat of endangered salmonids and is the leading cause for streams to be listed as impaired under the US Clean Water Act. Erosion on roads with stream connectivity is thought to be an especially significant problem. Different approaches to monitoring sediment production on road networks are possible, with the more expensive protocols generally yielding better information (in the sense of providing more precise estimates of the true rate of sediment loading). The question that forest managers face, then, is this: is it better to opt for an expensive monitoring protocol that yields high-quality information, a less expensive protocol that yields poorer quality information, or to skip monitoring entirely and spend the resultant savings on engineering projects to reduce erosion rates?

Below, the challenge of weighing trade-offs between the cost and quality of candidate monitoring protocols is treated within a partially observable Markov decision process (POMDP). The POMDP is an extension of the Markov decision process to handle imperfectly observed state variables (such as erosion rates). That is, in the POMDP not only are the state transitions uncertain, but the state at any given point in time is observed with error. In the case of erosion, it is simply not possible to measure the volume of surface material removed from a site with great exactness, and even rough estimates require staff time and equipment. The modeling approach taken here may be thought of as a particular formalization of adaptive management, in that decisions are made in a dynamic and stochastic environment based on beliefs that change as new information becomes available. Given the pervasiveness of errors in observations on many variables that are key to environmental management (e.g., animal populations), a principled framework for incorporating observation error into decision-making is a valuable tool.

2. BACKGROUND ON ROAD EROSION IN THE REDWOOD REGION

The modeling approach to monitoring design demonstrated below, while widely applicable in environmental management, was inspired by the difficulty of developing erosion control strategies in the Middle Fork Caspar Creek watershed on the Jackson Demonstration State Forest Mendocino County, California, part of the coastal redwood region. This watershed covers approximately 1300 acres, with elevation ranging from 100 to 1000 feet and slopes from 0-100%. The climate is Mediterranean, with average annual precipitation between 40 and 70 inches. Second- and third-growth redwood and Douglas-fir are the dominant vegetative community. The watershed is underlain by the Coastal Belt of the Franciscan Assemblage, consisting of marine sedimentary and volcanic rocks. Much of the watershed has been logged two or more times, and numerous logging roads, mostly dating from the 1960s and 1970s, run through the watershed.

Erosion on forest roads can impair road function, increase transportation costs, reduce access, and necessitate substantial expenditures on repair. It may also cause undesired changes in habitat and hydrologic function, which can in turn lead to higher water treatment costs, increased flood risk, reduced recreational opportunity, and lower aquatic habitat quality. In many northern California coastal streams, the contribution of logging roads to in-stream sediment loads is a particular concern, as excess sediment can adversely affect endangered salmon and steelhead trout.

Beginning in 2005, a field study of erosion rates at ten sites in the watershed was begun, with settling basins used to measure coarse sediment production, laboratory analysis of runoff to estimate fine sediment production, and tipping buckets to estimate runoff (Barrett and Tomberlin 2007 provide preliminary analysis). The cost of instrumentation, staff time, and laboratory analysis confronted agency staff directly with questions about the role of monitoring in management and in particular what level of monitoring best balanced the costs and benefits of the program, given the information being generated. Realizing that significant time and money could be saved by scaling back the monitoring effort to measure only coarse sediment, managers asked for a comparison of the desirability of

continuing the extensive monitoring program vs. pursuing the less intensive monitoring protocol.

In developing a road erosion control strategy, forest managers must consider both natural stochasticity in the erosion processes and uncertainty about the efficacy of various management treatments in reducing erosion risk. Some treatments, such as road removal, may be both quite costly and substantially irreversible, creating an incentive to defer the decision. Maintenance, in contrast, can be continued indefinitely while the landowner retains the option to upgrade or remove the road later. Adding to the challenge of developing an erosion control strategy is the difficulty of knowing which roads or road segments are most in need of treatment. Because budgets are generally nowhere near what would be required to treat entire watersheds, prioritization among possible sites and treatments is key, and monitoring (both before and after treatments) can potentially shed important light on this prioritization. The analysis here is motivated by an important practical question in water quality management, namely, how much time and money should be put into monitoring sediment loading when those same resources could be spent instead on remedial measures to reduce sediment production at its source?

3. A DYNAMIC MODEL OF EROSION CONTROL

This section presents a modeling approach to choosing among monitoring protocols, or more generally among a candidate set of actions that differ in their costs and in the information they yield. The POMDP is a collection of sets $\{S, P, A, W, \Theta, R\}$ (Cassandra 1994), where S is the system's state variables, P represents state dynamics as transition probabilities, A is the actions available to an agent, W is the rewards to taking particular actions in particular states, Θ is a set of possible observations on the state variables, and R is a set of observation probabilities. Observations $\theta \in \Theta$ are the only information the agent has on the unobservable true state, S. The observation model R describes the probabilistic relationship between observations θ and the true state S. The decision-maker (here, a land manager) uses observations θ and the observation model R to estimate the state S.

The model assumes the manager's goal is to minimize long-run discounted total costs of sediment control, which includes the cost of monitoring sediment production. The actions that achieve this goal are identified with dynamic programming (Bertsekas 2000) through a recursively defined value function V:

$$V_t(\pi) = \max_a \left[\sum_i \pi_i q_i^a + \beta \sum_{i,j,\theta} \pi_i p_{ij}^a r_{j\theta}^a V_{t+1}[T(\pi \mid a, \theta)] \right] \tag{1}$$

where

π_i = subjective probability of being in state $i \in S$ at time t

q_i^a = immediate reward for taking action $a \in A$ in state $i \in S$ at time t

β = discount factor

p_{ij}^a = probability of moving from state $i \in S$ at time t to state $j \in S$ at time $t+1$ after taking action $a \in A$

$r_{j\theta}^a$ = probability of observing $\theta \in \Theta$ after taking action $a \in A$ and moving to state $j \in S$

T = function updating beliefs based on prior beliefs and observed θ

V is the greatest expected net benefit that the agent can achieve over time, taking into account that as conditions change in the future, different actions may be warranted. The solution of V yields an optimal policy, which is a mapping from beliefs about the current state, π, into the optimal action.

In our setting, the state variable *S* is a forest road segment's potential to deliver sediment to the stream system, which for expository purposes takes only two possible values, *High Erosion* and *Low Erosion.* The action set *A* consists of *Maintain* (i.e., neither monitor sediment loading nor take remedial measures to reduce loading), *Monitor Low* (i.e., monitor with settling basins only), *Monitor High* (i.e., monitor with settling basins augmented by laboratory analysis of suspended sediment), and *Treat* (i.e., take measures to reduce the potential for sediment production). The observation set consists of the same two possible values as *S*, *High Erosion* and *Low Erosion*, but an observation of θ = *High Erosion* does not necessarily mean that the true state *S* =*High Erosion*. Instead, we define an observation model *R* as follows:

$$R^1_{j\theta} = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} \quad R^2_{j\theta} = \begin{bmatrix} 0.63 & 0.37 \\ 0.25 & 0.75 \end{bmatrix} \quad R^3_{j\theta} = \begin{bmatrix} 0.83 & 0.17 \\ 0.12 & 0.88 \end{bmatrix} \quad R^4_{j\theta} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

Each matrix, with the state $j \in S$ defined by row and each observation θ defined by column, defines the probabilistic relationship of observation to true state under a different action. $R^1_{j\theta}$, for example, tells us that after taking action *a*=1 (*Maintain*) and moving to the unobservable state *j*=*Low Erosion*, we would observe θ=*Low Erosion* with 60% probability and θ=*High Erosion* with 40% probability. That is, maintaining the *status quo* provides some weak information, presumably through casual observation of the road. $R^2_{j\theta}$, in contrast, tells us that implementing a monitoring plan with settling basins (*a*=2, *Monitor Low*), yields a stronger basis for inference on *S*, and $R^3_{j\theta}$ that the more sophisticated scheme *Monitor High* yields still more information. Finally, $R^4_{j\theta}$ indicates that immediately after treating the road, observations tell us nothing about the true state of erosion, a reflection of how treatments often cause transient changes in erosion rates that tell us little about the true state of the road.

The stochastic dynamics of the erosion level *S* are given by transition probability matrices defined as follows:

$$P^1_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad P^2_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad P^3_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad P^4_{ij} = \begin{bmatrix} 0.95 & 0.05 \\ 0.90 & 0.10 \end{bmatrix}$$

The first two matrices embody the assumption that under *status quo* maintenance or either monitoring program, the state remains unchanged. The final matrix tells us that under *a*=4 (*Treat*), a *Low Erosion* road stays in that same state with 95% probability, but there is a 5% chance that the treatment will backfire and create a *High Erosion* road. Similarly, treating a *High Erosion* road has an 90% chance of successfully creating a *Low Erosion* road and a 10% chance of failure (meaning the *High Erosion* road stays that way). These values are not derived from field data, but are chosen to reflect a plausible scenario for analysis.

Finally, the reward structure (actually, cost structure) is as follows:

$$W^1_j = \begin{bmatrix} 0 & -15 \end{bmatrix} \quad W^2_j = \begin{bmatrix} -0.5 & -15.5 \end{bmatrix} \quad W^3_j = \begin{bmatrix} -2.5 & -17.5 \end{bmatrix} \quad W^4_j = \begin{bmatrix} -8 & -8 \end{bmatrix}$$

Here the columns of each vector represent the rewards (in USD x 10^3) of taking a particular action in a particular state. W^1 tells us that maintaining the road in *Low Erosion* state will cost nothing, while the cost of maintaining the road in *High Erosion* state is \$15,000 per year, due to the need for cleanup and repair. W^2 and W^3, the payoffs to monitoring, are the same as W^1 less the periodic cost of the monitoring program (\$500 for *Monitor Low* or \$2500 for *Monitor High*, assuming the equipment is already on hand). W^4 tells us that treating the road (by adding rock and mechanically treating likely problems) will cost us the same \$8000 regardless of whether the road is in *Low Erosion* or *High Erosion* state. Comparing all these costs, it's obvious that if the manager knew the true state to be *Low Erosion*, the best choice would be to *Maintain* (*a*=1), and if the manager knew the true

state to be *High Erosion*, the best thing to do would be to *Treat* (a=4). However, the premise of our model, and the reality that managers generally face, is that the true state is unknown.

4. MODEL SOLUTION AND RESULTS

Model solution consists of identifying the optimal value V in (1) and the decision rule, or optimal policy, associated with it. This optimal policy is a mapping from the decision-maker's beliefs about the true state of the system into an optimal action set. That is, it is a state-dependent rule in which the state that is the argument of the rule is the decision-maker's beliefs. Because these beliefs evolve according to Bayes' rule and can in principle take on any value on the unit interval, they are uncountably infinite. This is in contrast to most applications in stochastic dynamic programming, in which numerical solution is attained by some form of backward recursion over a discrete set of possible states.

While allowing for uncountably infinite states is a significant complication, it has been shown that the value function associated with (1) can be represented as a convex and finite set of piecewise linear segments (Monohan 1982). This result enables solution of (1) by backward recursion with a nested linear programming routine that identifies optimal candidate segments over the entire belief space (details are available in Monohan 1982 and Cassandra 1994). The backward recursion from the planning horizon T identifies the set of linear segments that comprise the value function V, which gives the expected value of the optimal policy for every belief at any given number of periods prior to T.

Figure 1 shows the value function, V, for the model parameters given above, as it evolves over a 5-period decision horizon (the highest solid line is V at T-1, the next down is V at T-2, etc.). On the x-axis is π_1, the subjective probability that the road is in the Low Erosion state. The solid lines are the segments that constitute the value function. The dashed lines show the division of the state space into policy regions, i.e., the beliefs for which the actions *Treat*, *Monitor High*, *Monitor Low*, or *Maintain* are optimal. The most salient features of the solution are that 1) as the decision horizon lengthens, maintaining the status quo occupies progressively less of the belief space, and 2) *Monitor High* enters the optimal policy at T-3 and *Monitor Low* at T-5.

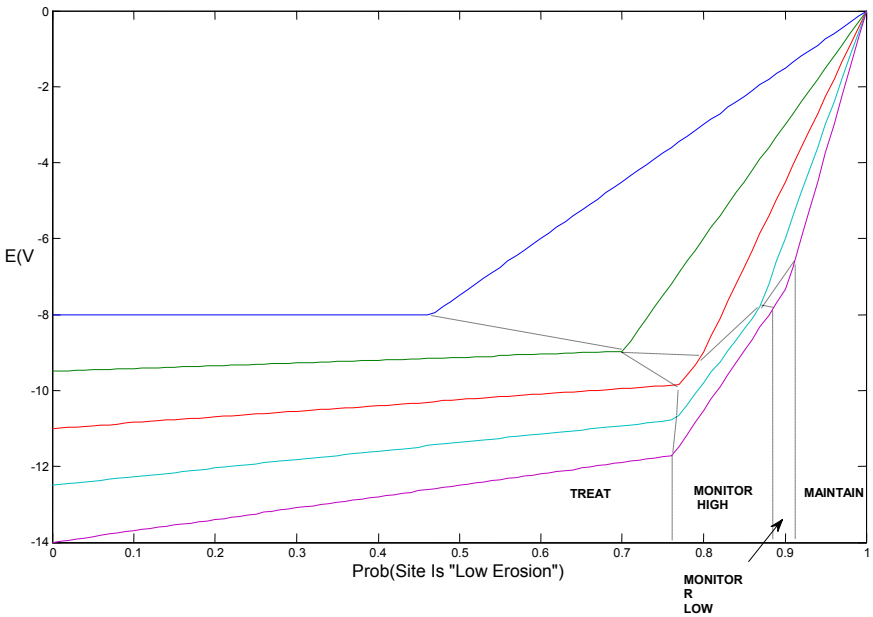

Fig. 1. The value function, V, and optimal policy as a function of beliefs about sediment production, for five different time horizons. The dotted black lines show the division of the belief space into regions associated with each action for a given decision horizon.

These features derive from the cost of the monitoring strategies and the value of the information they provide: only after the decision horizon has lengthened sufficiently to merit the increased expenditure on information gathering. However, *Monitor Low* provides relatively little extra information given its cost, hence it is the optimal action for a small set of beliefs. The choice of *Treat*, by contrast, occupies more than half the belief space: because the cost of maintaining and repairing a high-erosion road is substantial, treating the road without bothering to monitor first is optimal for beliefs (in the 5-period decision horizon) up to π_1 around 0.76. Of course, the results depicted in Figure 1 are specific to the model structure and parameters described above, and should not in any way be construed as general.

Continued iteration on the value function yields an approximately stationary optimal policy in belief space. In this case, the policy showed approximate stationarity by a decision horizon of 10 periods, with a policy in which *Treat* was optimal on $\pi_1 \in [0, 0.76)$, *Monitor High* on $\pi_1 \in [0.76, 0.93)$, *Monitor Low* on $\pi_1 \in [0.93, 0.97)$, and *Maintain* on $\pi_1 \in [0.97, 1.0]$.

The performance of a policy can be examined by simulating the policy's application to the system as it evolves stochastically and provides observations that are used by the decision-maker to estimate the true state. Figure 2 shows the distribution of cumulative performance over a 50-year horizon for 1000 simulations of the approximately stationary optimal policy applied to the model (i.e., 1000 sequences of the policy being applied for 50 years to a sequence of states and observations following the Markov processes given above). The distribution of outcomes is strongly bimodal, due to the particular parameterization adopted: because *Treat* as an action generates no new information, it does not change beliefs, so for many initial beliefs the optimal policy gets stuck in long series of repeatedly choosing *Treat* (note that -400 is the cumulative value of choosing *Treat* every year over a 50-year horizon, regardless of the road's true state). A similar effect, though not as strong applies to the choice of *Maintain*; *i.e.*, because little new information is generated when *Maintain* is the action chosen, it tends to be chosen many times in a row. The relatively few cumulative rewards that lie in the middle of the distribution are associated with initial beliefs such that monitoring is chosen in more periods, i.e., where there is more ambiguity about the current state and where the information signal is strong enough to change beliefs more significantly.

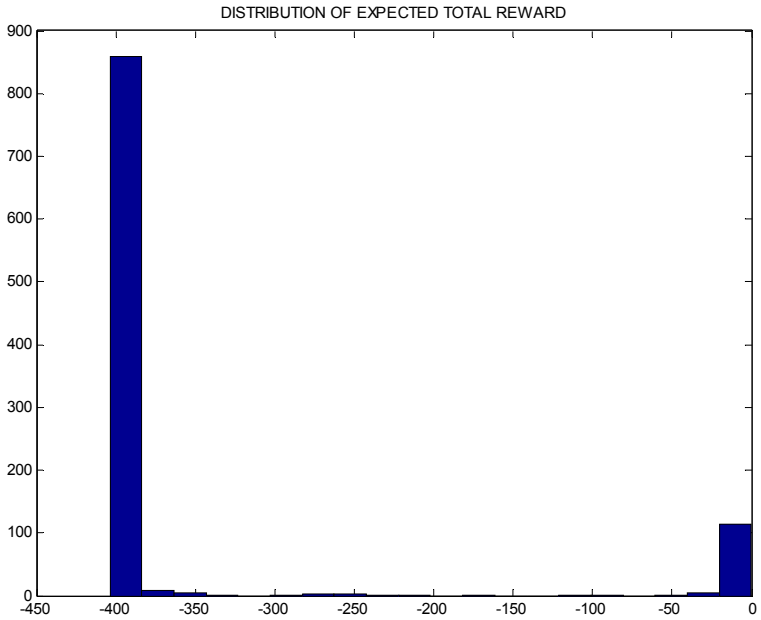

Figure 2: Histogram of cumulative reward over a 50-year horizon, based on 1000 stochastic simulations of the optimal policy.

Sensitivity analysis of the assumed signal strengths of monitoring relative to the *Treat* and *Maintain* actions, not shown here due to space constraints, revealed that the bimodal distribution of outcomes was still present as more information was gleaned from *Treat* and *Maintain*, though in less pronounced form. Similarly, changing the state transitions *P* to allow for greater randomness in the underlying state variable yielded optimal policies that changed more often over time, leading to a more even distribution of cumulative rewards.

5. CONCLUSIONS

The uncertainty inherent in natural resource management requires that we think carefully about the allocation of scarce resources between monitoring and restoration. The POMDP provides a tool for analyzing the conditions under which more or less monitoring is desirable. This approach accounts for the costs (and benefits) of different actions, including different monitoring schemes, in a formulation that captures stochastic changes in state variables and errors in observations taken on these state variables.

In the application presented here, for a 10-period decision horizon, both of the monitoring schemes considered (more and less intensive) were part of the optimal policy, but only for about 21% of possible beliefs about the true state, while in the other 79% of the belief space the preferred actions were to treat without monitoring or to maintain the *status quo* (*i.e.*, to do nothing). That is, in this example, for most beliefs the costs of monitoring exceed the expected benefit—monitoring was preferred only when the manager had a fairly strong *a priori* belief that the road was a low-erosion site, with monitoring serving essentially to rule out the need for more aggressive and expensive treatment.

The optimal policy computed for this problem was then applied to the stochastic environment (*i.e.*, to random sequences of state changes and observations generated from the assumed model), which yielded a strikingly bimodal distribution of cumulative rewards over a 50-year time horizon. The optimal policy of a POMDP is by definition the one that maximizes the *expected value* of cumulative reward, and is not influenced by the variability of that reward. Many decision-makers may be uncomfortable with a policy that generates a very wide range of possible outcomes. Much work is currently being done in operations research to address risk-sensitive and robust approaches to control problems, but the canonical form of the POMDP as presented here is not risk-sensitive, though changing model parameters, such as increasing the designated costs of certain outcomes, can aid in the search for more conservative policies.

Our example here has been stylized both for ease of presentation and because POMDPs are known to be computationally intractable. Due to the larger (and possibly continuous) state space that will apply in many environmental management settings, model solution will have to draw on heuristic techniques for the POMDP, an active field of research in applied mathematics and computer science. The issues of state space aside and risk sensitivity aside, the POMDP as presented here has several important limitations. The reward structure, transition probabilities, and observation model are all assumed known and fixed. Establishing reasonable parameter sets, especially for the observation model, will be a significant challenge in many applied settings. Applying more general dynamic optimization techniques, such as Bayesian reinforcement learning, may provide a way to address some of these concerns, though allowing more general model formulations will generally come at the cost of requiring more data.

Lastly, a word on the subjective beliefs that form the basis for the model presented here is in order. These beliefs may come from personal experience, field experiments, studies from other areas, or other sources, and evolve over time as new information becomes available. Because the beliefs evolve from some prior, different agents looking at the same sequence of observations may still disagree about the 'facts' of a decision setting. Some parties may object that subjective beliefs are not a valid basis for environmental management decisions, but it's hard to see what alternative we have, given that in most settings relatively little is known with certainty. Indeed, subjective probabilities are the *de*

facto basis of most management decisions, and we're probably better off acknowledging this reality rather than igoring it.

ACKNOWLEDGMENTS

Thanks to Brian Barrett of the California Department of Forestry and Fire Protection and Nikos Vlassis of the Technical University of Crete for helpful discussions, and to Teresa Ish of Kuulakai Consulting for work on an earlier application of the model to erosion control. This work was partially supported by funding from the Southwest Fisheries Science Center of the National Marine Fisheries Service.

REFERENCES

Barrett, B., and D. Tomberlin. 2007. Sediment Production on Forest Road Surfaces in California's Redwood Region: Results for Hydrologic Year 2005-2006. In: *Proceedings of the 2007 Society of American Foresters Convention*. Washington, DC: Society of American Foresters.

Bertsekas, D. 2000. *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific.

Cassandra, A. 1994. *Optimal Policies for Partially Observable Markov Decision Processes*. Brown University Department of Computer Science Technical Report CS-94-14.

Monohan, G. 1982. A Survey of Partially Observable Markov Decision Processes: Theory, Models, and Algorithms. *Management Science* 28(1):1-16.