LEARNING AND THE ADAPTIVE MANAGEMENT OF FISHERIES RESOURCES

David Tomberlin
National Marine Fisheries Service, Santa Cruz

Teresa Ish
University of California, Santa Cruz

MARCH 2006

ABSTRACT

Natural resource management under uncertainty frequently raises questions of this form: is current information sufficient to support a decision, or is it preferable to defer a decision while gathering costly new information? This paper presents a framework for systematically addressing questions of this type. The partially observable Markov decision process (POMDP) is a dynamic optimization modeling approach that accounts for uncertainty regarding the system state and for the opportunity to obtain costly information on this state. The main idea of the POMDP is that information has costs and benefits—both of which may be uncertain—and that these must be considered along with other costs and benefits relevant to the decision problem. Here, we demonstrate the use of the POMDP in the context of Pacific salmon recovery efforts. Specifically, we consider the problem of a planner who must choose among three possible habitat management actions: maintaining the *status quo* regime, which is inexpensive but relatively high risk; implementing a monitoring program that will have no immediate habitat benefit but may help produce better decisions in the future; or launching a habitat rehabilitation program without waiting for further information. In our example, we find that it is often preferable to proceed with habitat rehabilitation projects rather than to implement habitat monitoring programs or to maintain the *status quo*. We stress that this result is entirely due to this particular model's parameterization and is in no way generalizable to other similar questions that arise in habitat or general fisheries management. While we believe that the POMDP is the best available modeling framework for rigorously studying information-gathering strategies in fisheries management, it confronts the researcher with serious computational challenges even for relatively simple problems such as that presented here.

INTRODUCTION

While there seems to be no consensus on a definition of 'adaptive management,' clearly a necessary condition for management to be adaptive is that it account for the arrival of new information. Within the natural resource management literature, most work has focused on 'passive adaptive management,' in which new information is incorporated into decisionmaking as it becomes available. A more difficult approach is that of 'active adaptive management,' in which new information is sought optimally: the manager considers the short-term cost of information gathering vs. the potential long-term benefits, and decides whether the costly information is worth having[1].

Markov decision processes (MDPs), which, when solved with the techniques of stochastic dynamic programming, yield a mapping from system state into an optimal policy, may be thought of as a formal representation of adaptive management. However, MDPs assume that state variables are observed perfectly, an assumption that clearly does not hold in many natural resource management problems: animal populations, mineral reserves, and water quality, at least in many situations, cannot be known with certainty, and even developing good estimates is generally very expensive and time-consuming.

The theory of partially observable Markov decisions processes (POMDPs) was developed in response to this shortcoming of MDPs, but no numerical algorithms existed for POMDP solution until Sondik (1971). Despite a steady stream of improvements in both exact and heuristic solution techniques since then, most applied work in dynamic optimization (including control engineering, economics, and behavioral ecology) has continued to rely on MDPs built around certainty-equivalent measures, rather than face the numerical difficulties inherent in POMDP. These difficulties are two-fold. First, POMDPs inherit from MDPs the well-known 'curse of dimensionality,' by which is meant that solution times explode as the number of admissible states and the length of the time horizon increase. Second, POMDPs are fundamentally Bayesian decision processes, in the sense that an agent's *beliefs* about state variables become the basis for the optimal decision rule. The agent may change these beliefs, via Bayes' Theorem, when new information becomes available. While this is conceptually appealing, the practical result is that we move from a world of finitely countable MDP states to one of infinitely uncountable belief states, since an agent may come to have any set of beliefs, depending on how their prior beliefs and new information combine to yield updated beliefs. Thus, for POMDPs we can no longer use the standard techniques of stochastic dynamic programming as presented in, for example, Bertsekas (2000).

The difficulty in implementing POMDP solutions is indeed a strong incentive to assume certainty-equivalence and stay within the relatively comfortable confines of MDPs. We have done this ourselves (Tomberlin et al., in prep). However, the substantial—and sometimes paralyzing—uncertainty we face in thinking about the most important variables in fisheries management (population levels and trends, habitat metrics, fishing effort) really demands a better response. In our own work on salmon habitat

[1] Though these ideas are actually lifted whole-cloth from the control engineering literature, the most thorough treatment in a natural resource management context seems to be Walters (1986).

management, we have found, for example, that the amount of sediment loading from logging roads is essentially a mystery, and only a few consultants are even willing to hazard a guess (for the correctness of which they will never be held accountable). How can managers reasonably approach sediment control decisions when they don't even know whether it's a serious problem?

Below, we use this sediment loading example to structure our discussion of the POMDP, mostly for the purpose of keeping the discussion from becoming too abstract. However, questions of exactly the same form arise anytime we consider fisheries management under uncertainty with an opportunity to invest in learning, which will generally mean incurring some short-term cost to achieve greater overall long-term net benefit. For example, should stock assessment efforts be tripled this year in the hope that better management will result in future years? (The answer is not necessarily yes, even if you are a stock assessment biologist!) Or, moving beyond managerial to behavioral applications, we can think about applying the POMDP to a fisherman's dynamic choice of fishing location, given that he combines prior beliefs and current catch to assess the current state of his current location, and decides whether to move or not based on a comparison of his current beliefs for all possible fishing locations. Indeed, Lane (1989) does just this.

We believe the POMDP is the best existing tool for exploring a variety of questions related to learning and the adaptive management of fisheries. Rather than make that argument directly, however, here we offer an expository example that we hope shows how the POMDP is precisely the tool needed to address a very important question in Pacific salmon recovery planning: when are habitat monitoring programs are justified? We ask the question because various federal, state, and local governments, as well NGOs and community groups, are either monitoring or developing plans to monitor freshwater salmon habitat conditions such as water temperature, turbidity, sediment loads, stream complexity, and so on. All this monitoring seems unobjectionable from a conservation point of view, and a lot of it is actually quite fun, so it might seem ungenerous to ask whether it's justified. However, even given the high level of enthusiasm and public funding for habitat monitoring work, only a small fraction of streams can be monitored, and those only for a few habitat indicators and for a limited period of time. Since monitoring one stream may contribute far less to management than monitoring another, our question is an important one if we're serious about seeing that our conservation efforts result in as much conservation benefit as possible.

Below, we focus on the question of whether logging road erosion monitoring is worth the time and expense, given that we could decline to monitor in favor of i) applying rehabilitative treatments immediately, without bothering to collect data first, or ii) ignoring the problem entirely and hoping it goes away—by far the most common practice. The logging road example works well for our purposes for at least two reasons. First, surface erosion is by its nature difficult to assess without special equipment, making the partial observability approach very apt. Second, logging road erosion control can quite realistically be represented in terms of a few states and actions, and we have good estimates of the costs of these actions. We take the point of view of a land manager who wants to minimize long-run total cost, and will engage in monitoring only if it's expected to help with long-term decision performance. Our model's purpose is to help the land manager decide when monitoring is worth the trouble.

MODEL

We begin with some general notes on POMDP models and then construct a simple model that illustrates the use of POMDP for analyzing the desirability of information-gathering programs (in our example, habitat monitoring programs).

The traditional MDP is a collection of sets $\{S, P, A, W\}$, where S represents state variables, P represents state dynamics as transition probabilities, A represents the actions available to an agent, and W represents the rewards to taking particular actions under particular conditions. A POMDP is an MDP with two additional components, a set of observations Θ and an observation model R. Observations $\theta \in \Theta$ are the only information the agent has on the true state S, which is unobservable. The observation model R describes the probabilistic relationship between observations θ and the true state S. In other words, the agent uses the observation model R to make inferences about the true state S based on noisy observations θ.

Solution of MDPs and POMDPs proceeds through a recursively defined value function V. In the case of POMDPs, this value function is:

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

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

The value function V is simply 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. From the solution of V we can also derive an optimal policy

$$\delta_t^*(\pi_i) = \arg\max_{a \in A} \left[\sum_i \pi_i q_i^a + \beta \sum_{i,j,\theta} \pi_i p_{ij}^a r_{j\theta}^a q_j^a \right]$$

which is a mapping from beliefs about the current state, π, into the optimal action. In other words, for any possible set of beliefs about the true state *S* at any time in the decision problem, the optimal policy identifies the action that will have the great long-term expected net benefit.

While the above formulation may seem abstract, the concept it represents is very intuitive. We live in a world that we understand almost exclusively through limited and imperfect observations. To take an example from fisheries management, we do not set Total Allowable Catch based on the number of fish in the sea, rather on the basis of how many fish we *think* are in the sea, and even the best efforts of a stock assessment team often yield a broad range of estimates of the true fish population.

To make the POMDP formulation more concrete and its significance clearer, we now proceed to construct a particular POMDP that addresses an important question in Pacific salmon habitat management. Sediment loads in salmon-bearing streams significantly impair habitat quality in many Pacific coastal rivers and streams, but identifying the sources of sediment loading is difficult. On forested lands with logging roads, managers may often suspect that particular roads are providing excessive sediment loading, but visual inspection is a very unreliable means of gauging surface erosion levels. We consider the problem faced by a manager who has three actions available to address erosion on a suspected problem road: to maintain the road as it is, to monitor the road's erosion level (by installing field instruments), or to treat the road (e.g., by putting down new gravel). The first of these is quite inexpensive but does nothing to reduce current erosion rates or generate better estimates of these rates; the second is less expensive and does nothing to reduce erosion, but does provide information for subsequent decision-making; the last is quite expensive but has a good chance of effectively curbing the problem (if there is a problem, which the manager can't know with certainty).

In terms of the POMDP formulation, the action set *A* thus consists of {*maintain, monitor, treat*} and the state variable *S* is surface erosion. To keep the model tractable, we restrict this state variable to only two possible values, *High Erosion* and *Low Erosion*. The observation set consists of the same two possible values, *High Erosion* and *Low Erosion*, but of course an observation of *θ = High Erosion* does not necessarily mean that the true state *S =High Erosion*! Instead, we have to define the observation model *R*:

$$R_{j\theta}^1 = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} \quad R_{j\theta}^2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} \quad R_{j\theta}^3 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

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. $R_{j\theta}^1$, for example, tells us 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=*Low Erosion* with 40% probability. That is, maintaining the *status quo* provides some information, presumably through casual observation of the true current erosion level, but it's very weak information, not too much better than a coin toss. $R^2_{j\theta}$, in contrast, tells us that implementing a monitoring plan (a=2), yields a much stronger basis for inference based on observations: in this case, taking an observation when the true state is j=*Low Erosion* yields θ=*Low Erosion* with 90% probability and θ=*Low Erosion* with 10% probability. Finally, $R^3_{j\theta}$ indicates that immediately after treating the road, observations tell us nothing about the true state of erosion, because the treatment itself causes a short-term spike in erosion that is confounded with background erosion processes.

The stochastic dynamics of the state S are given by transition probability matrices:

$$P^1_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad P^2_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad P^3_{ij} = \begin{bmatrix} 0.95 & 0.05 \\ 0.80 & 0.20 \end{bmatrix}$$

The first two matrices indicate that under actions a=1 and a=2 (*maintain* and *monitor*, respectively), a *Low Erosion* road will stay a *Low Erosion* road and a *High Erosion* road will stay a *High Erosion* road. P^3_{ij} tells us that under a=3 (*treat*), a *Low Erosion* road stays in that same state with 95% probability, but allows a 5% chance that the treatment will actually backfire and create a *High Erosion* road. Similarly, treating a *High Erosion* road has an 80% chance of successfully creating a *Low Erosion* road and a 20% chance of failure.

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

$$W^1_{ij\theta} = \begin{bmatrix} -1 & -20 \\ -1 & -20 \end{bmatrix} \quad W^2_{ij\theta} = \begin{bmatrix} -3 & -22 \\ -3 & -22 \end{bmatrix} \quad W^3_{ij\theta} = \begin{bmatrix} -6 & -6 \\ -6 & -6 \end{bmatrix}$$

Here the columns of each matrix represent the possible states j and the rows represent possible observations θ. [We have suppressed the i-dimension of the cost structure since we assume the cost depends only on the state and not how the transition to the state occurred, as in the general POMDP formulation.] In each submatrix of W, the rows are the same because in our case the observation per se does not affect costs, though of course observations do affect beliefs through the observation model R. The costs here are in \$000s/mile of logging road. The matrix W^1 tells us that maintaining a road in *Low Erosion* state will cost \$1000, which is very cheap compared to \$20,000, the cost of maintaining a road in *High Erosion* state. W^2, the payoffs to monitoring, are the same as W^1 plus the cost of the monitoring program itself (\$2000). That is, monitoring does nothing to change the costs associated with the erosion *per se*, it is a pure additional cost. W^3 tells us that treating the road will cost us the same \$6000 regardless of whether the road is in *Low Erosion* or *High Erosion* state. Comparing all these costs, it's obvious that if the decisionmaker knew the true state to be *Low Erosion*, the best choice would be to maintain the current situation (a=1), and if the decisionmaker knew the true state to be

High Erosion, the best thing to do would be to treat the road right away (a=3). However, the premise of our paper, and the reality that habitat managers face, is that the true state is unknowable.

Finally, we assume a discount factor of β=0.95, which completes our model specification.

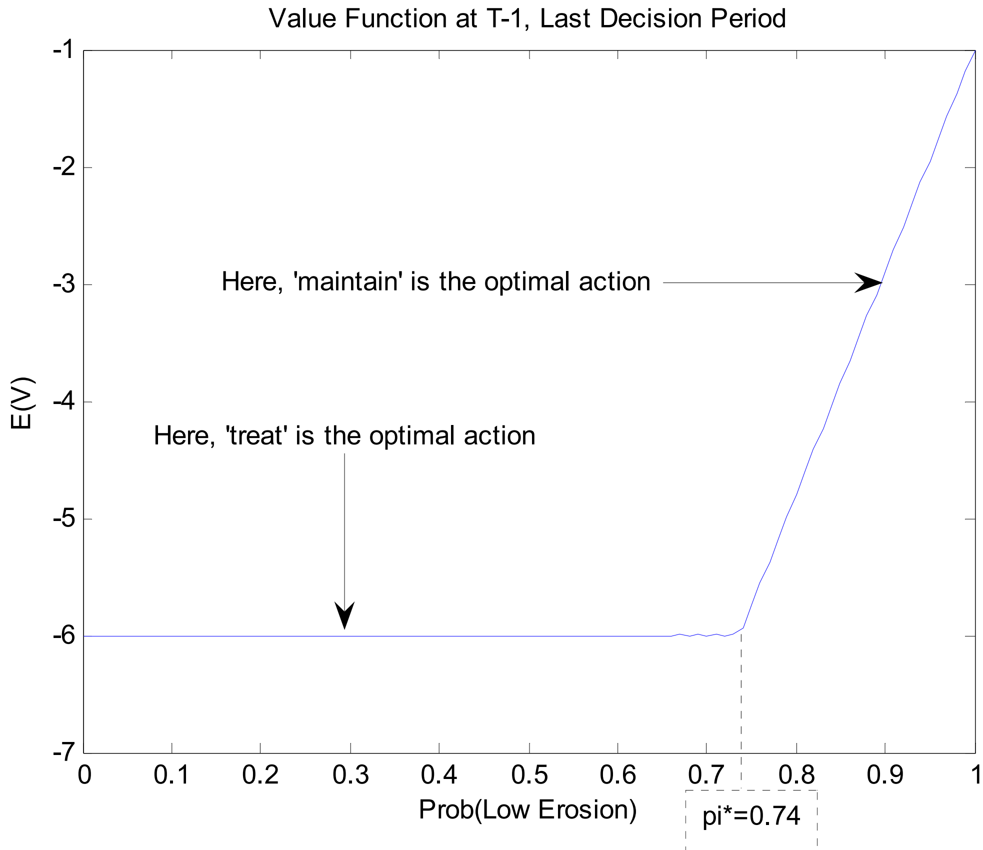
RESULTS

A POMDP solution consists of the recursively defined value function $V_t(\pi)$ and the associated optimal policy $\delta_t(\pi)$. Because the solution technique is rather complicated, we will not describe it here; interested readers can consult Cassandra (1994, pp. 45-54) for a good discussion of the Monohan/Eagle algorithm. POMDP algorithms share with MDP algorithms the basic notion of backward recursion from an arbitarily defined end of time, *T*. In our model, time *T* comes after all decisions have been made and after uncertainty about the true state has been resolved (which is important because it allows unambiguous values to be assigned to each of the possible final states).

Figure 1 (next page) shows the value function at the final time period in which a decision is to be made, *T-1*. The x-axis is the belief simplex for the two possible states in *S*: p(*Low Erosion*) runs from left to right, and since p(*High Erosion*) must be 1-p(*Low Erosion*), p(*High Erosion*) runs from right to left. The y-axis is the expected value of taking particular actions. The blue line is the value function, giving the expected value at *T-1* of taking whichever action has the lowest expected cost. Here, since there is only one decision period before the end of the problem, these values have very straightforward interpretations. If the manager's current belief is that p(*Low Erosion*) is anything less than 74%, the optimal action is to treat the road, which has a payoff of –6 regardless of the current true state. If, however, the manager's current belief is that p(*Low Erosion*) > 74%, then the optimal action is to maintain the *status quo* road, which has an expected value of [–1*p(*Low Erosion*) + –20*p(*High Erosion*)]. In other words, the more certain the manager is that the true state is in fact *Low Erosion*, the greater the expected payoff to doing simple maintenance work. Thus, Figure 1 not only shows the value function but also partitions the belief space into regions on which each possible action is optimal, i.e., visually lays out the optimal policy.

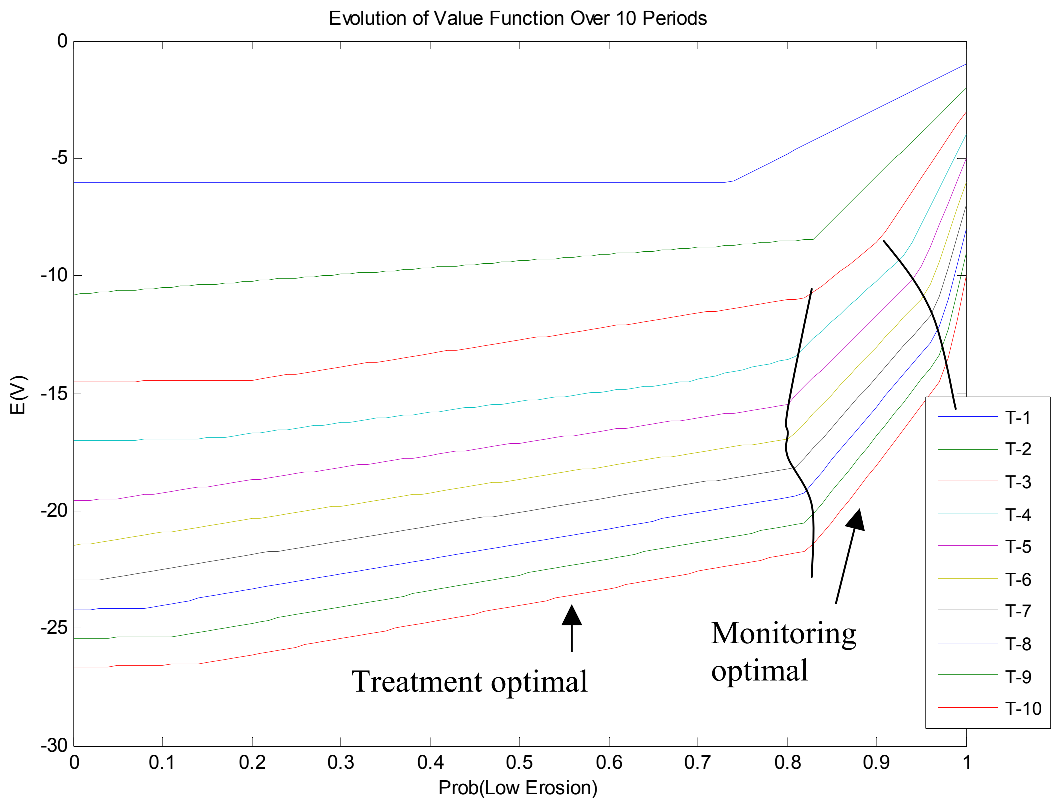
Notice that the action *monitor* does not appear as part of the optimal strategy in Figure 1. The reason is simply that, with no further actions to be taken after *T-1*, there is no justification for paying to gather information that can't provide future benefits in the form of improved decisionmaking.

Fig. 1: The value function in the last decision period, showing the partition of the belief space into regions associated with different optimal actions. Here, *treat* is optimal for beliefs p(*Low Erosion*) <= 74%, and *maintain* is optimal for beliefs p(*Low Erosion*) > 74%.



Of course, we are almost always interested in problems that have at least several decision periods (and sometimes even infinitely many). Figure 2 (next page) shows the evolution of the value function over a 10-period time horizon. The most obvious effect of lengthening the time horizon is that the value function moves steadily downward, due to the expectation of increased future costs (a direct result of our model setup). However, the shape of the value function also changes, as do the actions that form the optimal policy. Specifically, for T-3 and all earlier periods, *monitor* becomes part of the optimal strategy. The belief ranges for which *monitor* is optimal are between the two black curves; the beliefs to the left of the left-most black curve are those for which the optimal action is *treat*, and those to the right of the right-most black curve are those for which the optimal action is *maintain*. There is a bit of back and forth in the left-most curve, because the optimal policy has not fully converged to a stable mapping, but the general picture is clear. Monitoring enters the optimal policy at T-3, once the time horizon has become long enough that information generated by monitoring can yield sufficient benefits (in expectation) to offset the cost of the monitoring program. As the time horizon deepens, the range of beliefs over which monitoring is part of the optimal strategy increases, from about [0.81 0.90] at T-3 to about [0.81 0.97] at T-10. Due to computational limitations, we have not been able to explore longer time horizons, but the belief range seems to be on course to converge at about the T-10 level. Immediate treatment is still the optimal action for beliefs up to around p(*Low Erosion*)=80%, and simply maintaining the status quo is preferred only for beliefs such that p(*Low Erosion*) is well over 90%.

Fig. 2: The evolution of the value function over 10 decision periods. The value function moves monotically downward as the time horizon increases, which is an artifact of our cost-only model. For T-3 and earlier periods, monitoring becomes part of the optimal strategy for those beliefs between the two black curves.



The above discussion and figures may seem abstract, but in fact they correspond quite nicely to the way most of us get through life. We routinely make decisions that are based not on directly observable facts, but on our beliefs about those underlying facts, which for a variety of reasons we either can't or don't want to know with certainty. Both the facts and our beliefs about them may change over time, but at any point in time we make decisions based on our beliefs at that time. Of course, people don't apply Bayes' theorem with the kind of ruthless efficiency that a computer does, but the general notion of combining old and new information seems to reflect a lot of human decisionmaking. Perhaps more importantly for our present purposes, the partition of the belief space into regions corresponding to different optimal actions is very intuitive and also useful. As Figure 2 shows, one person may believe p(*Low Erosion*)=10%, another that p(*Low Erosion*)=40%, and a third that p(*Low Erosion*)=70%, but the POMDP makes clear that they should all still be able to agree on treating the erosion problem immediately.

CONCLUSIONS

The partially observable Markov decision process (POMDP) provides a formal framework for exploring when information-gathering is likely to be worth the cost and when not. Given the expense of monitoring programs in fisheries management (stock assessments, studies of ecosystem indicators, etc.), budgetary realities ensure that managers have to choose among candidate monitoring programs. POMDP provides a tool for thinking carefully about such choices.

Here, we have presented an application of POMDP to a simplified problem in salmon habitat management. Because POMDPs are even more subject than traditional MDPs to the 'curse of dimensionality,' research on numerical techniques for POMDP solution is a very active field in engineering and artificial intelligence, and more sophisticated applications in fisheries management will have to drawn on recent advances in interior-point methods and witness algorithms. However, even our simple example has shown that, under reasonable assumptions, the costs of monitoring may exceed the benefits. In our example, we found that, for problems with a time horizon of at least 5 decision periods, implementing a habitat quality improvement project *without first monitoring* was optimal as long as the subjective probability of existing habitat conditions being good was less than about 80%. Monitoring was optimal over a narrower range of beliefs, specifically when the belief that existing habitat conditions were good was between about 80% and about 95%. That is, monitoring in our example was preferred only when the manager had a pretty strong hunch that current conditions were good, in which case the monitoring served essentially to rule out the need for more aggressive and expensive treatment.

In developing our case for the POMDP as a useful decision-making tool, we deliberately touched lightly on the nature of the subjective probabilities π, which are really the heart of the POMDP. While from a technical point of view there's not much to say about π, which is simply a vector of conditional probabilities, we should address a concern that might arise from a philosophical point of view. Some may object that subjective probabilities have no place in policy making, which should strive at all times to be as objective and scientific as possible. Without rehearsing the centuries-long Bayesian-vs-frequentist struggle, we note that many Bayesians consider subjective probability the only sensible notion of probability, and so would dismiss this criticism as invalid on principle. For our purposes, it doesn't seem necessary to take that rigorous Bayesian position. We are satisfied with the more mundane argument that subjective probabilities are so manifestly the basis for current policymaking that it would impossible to imagine any policy getting made without them. In short, we think subjective probabilities in fisheries decision-making are perfectly sensible and almost perfectly unavoidable.

The deep uncertainty we face in many aspects of fisheries managment requires that we think carefully about when to invest in learning about the systems we manage. The POMDP provides a coherent (and beautiful) framework for such thinking.

LITERATURE CITED

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.

Lane, D. 1989. A Partially Observable Model of Decision Making by Fishermen. *Operations Research* 37(2):240-254.

Sondik, E.J. 1971. The Optimal Control of Partially Observable Markov Decision Processes. Ph.D. dissertation, Stanford University.

Tomberlin, D., W. Baxter, M. Thompson, and T. Ish. In prep. A Decision Model for Cost-Minizing Erosion Control on Logging Roads.

Walters, C. 1986. Adaptive Management of Renewable Resources. New York: Macmillan Publishing.