A Comparison of Optimal Decisions Computed Using Bayesian Updating and Stochastic Dynamic Programming

Robert M. Dorazio
U. S. Geological Survey
Florida Integrated Science Center
7920 NW 71 Street
Gainesville, Florida 32653
email: bdorazio@usgs.gov

Fred A. Johnson
U.S. Fish and Wildlife Service
Division of Migratory Bird Management
7920 NW 71 Street
Gainesville, Florida 32653
email: fred_a_johnson@fws.gov

May 7, 2003

1 Introduction

The last meeting of the Adaptive Management Conference Series (April 2003, Laurel, Maryland) ended with considerable discussion and confusion about similarities and differences in the optimal decisions obtained using Bayesian updating and stochastic dynamic programming. Although both of these approaches are capable of solving Markov decision problems, there are subtle differences in the nature of the solutions, which we believe may be safely ignored when decision making is practiced in an adaptive manner (either passively or actively). Here, we use a simple Markov decision problem to illustrate theoretical and empirical differences in the optimal solutions obtained with Bayesian updating and stochastic dynamic programming.

2 A Markov Decision Problem

2.1 Model of state dynamics

Let $\mathcal{Y} = \{0, 1\}$ denote a discrete state space, and consider an observable sequence of states, $\boldsymbol{y} = (y_0, y_1, \ldots, y_T)$, where $y_t \in \mathcal{Y}$. Assume that the temporal dynamics are governed by the following Markov process (Bonney, 1987):

$$
\begin{aligned}
f(y_t \mid x_{t-1}, y_{t-1}) &= \frac{\exp(y_t \eta_{t-1})}{1 + \exp(\eta_{t-1})} \qquad (1) \\
\eta_{t-1} &= (1 - x_{t-1})\alpha_1 + x_{t-1}\alpha_2 + (2y_{t-1} - 1)\beta
\end{aligned}
$$

where $x_{t-1} \in \{0, 1\}$ denotes a management action taken at time $t - 1$, and α_1, α_2, and β are known parameters. Thus, (1) specifies a set of state- and action-dependent transition probabilities:

Action		$y_t = 1$	$y_t = 0$
"do nothing"	$y_{t-1} = 1$	$\frac{\exp(\alpha_1+\beta)}{1+\exp(\alpha_1+\beta)}$	$\frac{1}{1+\exp(\alpha_1+\beta)}$
$\Updownarrow$			
$x_{t-1} = 0$	$y_{t-1} = 0$	$\frac{\exp(\alpha_1-\beta)}{1+\exp(\alpha_1-\beta)}$	$\frac{1}{1+\exp(\alpha_1-\beta)}$
"manage"	$y_{t-1} = 1$	$\frac{\exp(\alpha_2+\beta)}{1+\exp(\alpha_2+\beta)}$	$\frac{1}{1+\exp(\alpha_2+\beta)}$
$\Updownarrow$			
$x_{t-1} = 1$	$y_{t-1} = 0$	$\frac{\exp(\alpha_2-\beta)}{1+\exp(\alpha_2-\beta)}$	$\frac{1}{1+\exp(\alpha_2-\beta)}$

where $x = 0$ denotes an absence of management (i.e., "do nothing") and $x = 1$ denotes the alternative action (i.e., "manage"). Note that the model parameters have simple interpretations. α_1 represents the effect of management action $x = 0$ on system state, and α_2 represents the effect of management action $x = 1$ on system state. In contrast, β represents serial dependence in system state. The odds that $y_t = 1$ increase by e^{β} if $y_{t-1} = 1$; the odds that $y_t = 1$ decrease by $e^{-\beta}$ if $y_{t-1} = 0$.

2.2 Valuing the consequences and costs of management

Suppose a sequence of T management actions $\boldsymbol{x} = (x_0, x_1, \ldots, x_{T-1})$ may be selected and a corresponding sequence of system states $\boldsymbol{y} = (y_0, y_1, \ldots, y_T)$ is observable. Let $U(\boldsymbol{x}, \boldsymbol{y})$ denote a scalar-valued utility function that provides the basis for comparing an observed sequence of system states, which, in part, reveal the consequences of a selected sequence of management actions (as in (1)). In many decision problems $U(\boldsymbol{x}, \boldsymbol{y})$ is simply a sum of time-dependent utilities

$$U(\boldsymbol{x}, \boldsymbol{y}) = \sum_{t=1}^{T} U(x_{t-1}, y_t) \tag{2}$$

where $U(x_{t-1}, y_t)$ denotes the value of selecting management action x_{t-1} and observing its consequences (state y_t). As an example, assume the following utilities for $U(x_{t-1}, y_t)$:

x_{t-1}	$y_t = 1$	$y_t = 0$
0	1.0	0.0
1	0.8	-0.2

Given our definition of management actions, these utilities are consistent with the notion that $y = 1$ corresponds to a "favorable" state and $y = 0$ corresponds to an "unfavorable" state. For example, the highest utility is obtained if a favorable state is observed at time t ($y_t = 1$) after doing nothing at time $t - 1$ ($x_{t-1} = 0$). Likewise, there is zero utility when failing to manage at time $t - 1$ ($x_{t-1} = 0$) results in an unfavorable state at time t ($y_t = 0$). Management usually involves some cost; therefore, the lowest utility is obtained if an unfavorable state is observed at time t ($y_t = 0$) after managing at time $t - 1$ ($x_{t-1} = 1$).

2.3 Management objective

Suppose an initial system state y_0 is observed, and we seek a procedure for sequentially selecting T management actions in a manner that maximizes the total utility (summing over time) expected to result from those actions. Clearly, this procedure must depend on the model-based predictions of state dynamics (described in Section 2.1) and on the utility function $U(\boldsymbol{x}, \boldsymbol{y})$ (described in Section 2.2) for valuing the costs and predicted consequences (states) of the proposed actions. Given the time- and state-dependence inherent in Markov decision problems, procedures that satisfy our management objective are necessarily adaptive, allowing different actions to be selected based on the observed trajectory of system states. Such procedures are known as *closed-loop* control strategies (Intriligator, 1971) and differ from *open-loop* control strategies, which specify an optimal sequence of management actions using only the initial (observed) state of the system.

3 Optimal solutions

3.1 Stochastic dynamic programming

Stochastic dynamic programming (SDP) is a popular method for solving Markov decision problems (Puterman, 1994). SDP relies on a backward-induction algorithm to calculate an optimal control strategy or policy, which includes both time- and state-dependent optimal actions for the entire sequence of decisions. The backward-induction algorithm begins at the end of the sequence and for each state determines that action which produces the highest expected utility given the model of state dynamics and the utility function. The next stage of the algorithm considers the previous time step and determines for each state that action which produces the highest expected sum of current and future utilities, assuming that optimal state-dependent actions are always selected in future time steps. Backward-induction continues this process by considering each of the earlier decisions in succession and ends by computing the optimal state-dependent actions for the initial decision at time 0. The result is an optimal policy of time- and state-dependent actions.

Given the model of state dynamics specified in (1) and the additive utility function specified in (2), the first step of the backward-induction algorithm may be expressed as follows:

$$\begin{aligned} \overline{U}(x^*_{T-1} \mid y_{T-1}) &= \max_{x_{T-1}} \quad \mathrm{E}_{(y_T \mid x_{T-1}, y_{T-1})} \left[U(x_{T-1}, y_T) \right] \quad (3) \\ &= \max_{x_{T-1}} \sum_{y_T=0}^{1} U(x_{T-1}, y_T) \; f(y_T \mid x_{T-1}, y_{T-1}) \quad (4) \end{aligned}$$

where the subscript $(y_T \mid x_{T-1}, y_{T-1})$ indicates that expectation occurs over the distribution of states predicted for time T conditioned on the previous state and on the management action proposed for time $T-1$. Thus, $x^*_{T-1} \mid y_{T-1}$ denotes the optimal state-specific management action at time $T-1$, the end of the decision sequence. The remaining steps of the backward-induction algorithm are computed recursively

$$\overline{U}(x^*_{t-1} \mid y_{t-1}) = \max_{x_{t-1}} \; \mathrm{E}_{(y_t \mid x_{t-1}, y_{t-1})} \left[U(x_{t-1}, y_t) + \overline{U}(x^*_t \mid y_t) \right] \quad (5)$$

and produce a time-ordered array of optimal, state-dependent management actions at times $T-2$, $T-3$, $\ldots$, 0.

3.2 Bayesian updating

Bayesian updating is an inherently adaptive procedure that provides a coherent framework for decision making in problems of natural resource management (Dorazio and Johnson, 2003). In principle, Markov decision problems may be solved using a combination of Bayesian updating and backward induction (Carlin, Kadane, and Gelfand, 1998; Müller, 1999; Müller, Berry, Grieve, Smith, and Krams, 2000); however, in practice, even simple problems demand an enormous number of calculations, often exceeding the capabilities of modern computers. Here, we describe a procedure for solving Markov decision problems that uses Bayes Theorem to adaptively update an open-loop control strategy. In Section 4 we demonstrate that this procedure yields a closed-loop control strategy that approximates the optimal solution provided by SDP.

Consider the expected utility of an entire sequence of proposed management actions. This expectation, which we denote by $\overline{U}(\boldsymbol{x} \mid y_0)$, averages over the distribution of predicted system states given the model of state dynamics and the initial (observed) state y_0. By definition, this utility is:

$$\overline{U}(\boldsymbol{x} \mid y_0) = \mathrm{E}_{(y_1,\ldots,y_T \mid x_0,\ldots,x_{T-1},y_0)}\left[U(\boldsymbol{x},\boldsymbol{y})\right] \tag{6}$$

where the subscript $(y_1, \ldots, y_T \mid x_0, \ldots, x_{T-1}, y_0)$ indicates the joint distribution of predicted states conditioned on the initial state and the proposed sequence of management actions.

If $U(\boldsymbol{x}, \boldsymbol{y})$ is a sum of time-dependent utilities (as in (2)), considerable simplification of (6) is possible in Markov decision problems, where the joint density of predicted states may be expressed as a product of conditional densities. For example, consider a problem involving a sequence of 2 management actions where the state dynamics are governed by the Markov process defined in (1). In this problem the joint density of states predicted at times 1 and 2 is

$$f(y_1, y_2 \mid x_0, x_1, y_0) = f(y_1 \mid x_0, y_0)\ f(y_2 \mid x_1, y_1). \tag{7}$$

Combining this joint density with the utility function in (2) yields the expected utility (defined in (6)) of management actions (x_0, x_1):

$$\begin{aligned} \overline{U}(x_0, x_1 \mid y_0) &= \mathrm{E}_{(y_1,y_2 \mid x_0,x_1,y_0)}\left[U(x_0, y_1) + U(x_1, y_2)\right] & (8) \\ &= \sum_{y_1=0}^{1}\sum_{y_2=0}^{1}\left[U(x_0, y_1) + U(x_1, y_2)\right]\ f(y_1, y_2 \mid x_0, x_1, y_0) & (9) \end{aligned}$$

Substituting the right-hand-side of (7) for the joint density function in (9) and simplifying yields

$$\begin{aligned} \overline{U}(x_0, x_1 \mid y_0) &= \sum_{y_1=0}^{1}\left[U(x_0, y_1) + \sum_{y_2=0}^{1} U(x_1, y_2)\, f(y_2 \mid x_1, y_1)\right]\ f(y_1 \mid x_0, y_0) & (10) \\ &= \mathrm{E}_{(y_1 \mid x_0,y_0)}\Big[U(x_0, y_1) + \mathrm{E}_{(y_2 \mid x_1,y_1)}[U(x_1, y_2)]\Big] & (11) \end{aligned}$$

A similar derivation provides the expected utility of a sequence of 3 management actions

$$\overline{U}(x_0, x_1, x_2 \mid y_0) = \mathrm{E}_{(y_1 \mid x_0,y_0)}\left[U(x_0, y_1) + \mathrm{E}_{(y_2 \mid x_1,y_1)}\Big[U(x_1, y_2) + \mathrm{E}_{(y_3 \mid x_2,y_2)}[U(x_2, y_3)]\Big]\right] \tag{12}$$

It is obvious from inspection of (11) and (12) that evaluation of the expected utility of a sequence of management actions simplifies to a sequence of nested calculations because the expected utility of a management action at time t is nested within the expected utility of a management action at time $t - 1$. This is a general characteristic of Markov decision problems whose utility function is a sum of time-dependent utilities.

Initially (at $t = 0$), the optimal sequence of management actions maximizes the expected utility of the entire sequence of actions. For example, in a problem involving only $T = 2$ decisions, the expectation in (11) may be used to express the optimal solution

$$\overline{U}(x_0^*, x_1^* \mid y_0) = \max_{x_0, x_1} \; \mathrm{E}_{(y_1 \mid x_0, y_0)} \Big[U(x_0, y_1) + \mathrm{E}_{(y_2 \mid x_1, y_1)} [U(x_1, y_2)] \Big] \tag{13}$$

where $x_0^*, x_1^* \mid y_0$ denotes the optimal sequence of actions. Note that (13) specifies an open-loop control strategy that conditions only on the initial state y_0. However, adherence to a Bayesian approach demands that this strategy be updated as new information is acquired over time, and Bayes Theorem provides the formal mechanism for updating. For example, continuing with the problem of $T = 2$ decisions, suppose management action $x_0 = x_0^*$ is selected at $t = 0$ and the next system state y_1 is observed. Then the new observations (x_0, y_1) are used to update both our belief about the model parameters (α_1, α_2, and β) as summarized in their posterior distribution *and* the predictive distribution of future states $(y_2 \mid x_1, y_1, x_0, y_0)$, which integrates over posterior uncertainty in the model parameters. These updated distributions are then used to compute the optimal state-dependent action $x_1^* \mid y_1, x_0, y_0$ at time $t = 1$ as follows:

$$\overline{U}(x_1^* \mid y_1, x_0, y_0) = \max_{x_1} \; \mathrm{E}_{(y_2 \mid x_1, y_1, x_0, y_0)} \left[U(x_1, y_2) \right] \tag{14}$$

Therefore, at each decision time, maximization of the expected utility of a sequence of proposed management actions yields an open-loop control strategy; however, by updating the model parameters and its predictions at each time step and by conditioning on the observed state at each time step, we effectively generate a closed-loop control strategy that adapts to the observed consequences (changes in state) and costs of the management actions that are actually selected.

Strictly speaking, the optimal solutions provided by SDP and Bayesian updating are different. However, when the optimal Bayesian decisions are updated adaptively (as described in the previous 2 paragraphs), there are few practical differences between the two solutions. In addition, the similarity in these solutions is partially attributed to a strong similarity in their objective functions. For example, substituting right-hand-side of (3) into the recursive formula (5) used in backward induction yields

$$\overline{U}(x_{t-1}^* \mid y_{t-1}) = \max_{x_{t-1}} \; \mathrm{E}_{(y_t \mid x_{t-1}, y_{t-1})} \left[U(x_{t-1}, y_t) + \max_{x_t} \mathrm{E}_{(y_{t+1} \mid x_t, y_t)} \left[U(x_t, y_{t+1}) \right] \right] \tag{15}$$

which is equivalent to

$$\overline{U}(x_0^* \mid y_0) = \max_{x_0} \; \mathrm{E}_{(y_1 \mid x_0, y_0)} \left[U(x_0, y_1) + \max_{x_1} \mathrm{E}_{(y_2 \mid x_1, y_1)} \left[U(x_1, y_2) \right] \right] \tag{16}$$

in a problem of $T = 2$ decisions. Note the similarity between (16) and the optimal sequence of actions computed using Bayesian updating (13). Both objective functions include a time-ordered nesting of conditional expectations. The primary difference is that in (13) only one maximization is computed over the entire sequence of management actions, whereas in (16) there is a time-ordered nesting of maximizations that automatically yields a closed-loop control strategy.

4 Numerical Examples

In this section numerical examples are used to illustrate the theoretical results developed in Section 3.

4.1 Example 1

Suppose an optimal policy for a sequence of $T = 5$ decisions is required and the parameters in the model of state dynamics (Section 2.1) are known: $\alpha_1 = 0$, $\alpha_2 = 1$, and $\beta = 0.5$. For these parameter values a decision to implement management action $x = 1$ is expected to increase the odds that $y = 1$ to a greater extent than a decision to implement management action $x = 0$, and the effects of serial dependence on changes in system state are expected to be intermediate in magnitude relative to the effects of the management actions. We also assume in this example that the optimal policy maximizes the sum of state- and action-dependent utilities (described in Section 2.2) that are expected to result from the sequence of actions. Table 1 contains a sequence of optimal state-dependent management actions computed using SDP. Notice that the optimal policy appears relatively insensitive to system state, recommending action $x = 1$ at almost every decision time. The reason, of course, is because higher utilities are realized if $y_t = 1$ and the transitional probabilities of achieving this state are higher for action $x_{t-1} = 1$ (0.62 and 0.82) than for action $x_{t-1} = 0$ (0.38 and 0.62), regardless of whether the current state y_{t-1} is 0 or 1.

To compute an optimal sequence of management actions using Bayesian updating, the prior distribution of model parameters is assumed to have a point mass at the values $\alpha_1 = 0$, $\alpha_2 = 1$, and $\beta = 0.5$, which were specified earlier. Therefore, predictions of state dynamics can ignore any uncertainty in these values (at least initially at time $t = 0$). The optimal sequence of management actions computed using Bayesian updating is insensitive to the initial system state y_0 and recommends management action $x = 1$ at each of the 5 decision times (Table 2). This policy is obviously consistent with that computed using SDP. In fact, the expected utilities summed over the sequence of optimal management actions are nearly identical when we compare SDP *vs* Bayesian updating (2.6831 *vs* 2.6793 for $y_0 = 0$, 2.9255 *vs* 2.9217 for $y_0 = 1$).

For any sequence of $T > 1$ decisions, the backward-induction algorithm used in SDP ensures that the expected sum of utilities associated with its optimal policy will equal or exceed the expected sum of utilities associated with a policy of actions that does not rely on the determination of future system states. This is the essential distinction between a closed-loop control strategy, which requires a sequential evaluation of system state, and an open-

Table 1: Optimal management actions and expected utilities computed using SDP. Parameters in the model of state dynamics are $\alpha_1 = 0$, $\alpha_2 = 1$, and $\beta = 0.5$.

		$x_t^* \mid y_t$					$\overline{U}(x_t^* \mid y_t)$				
y_t	$t =$	0	1	2	3	4	0	1	2	3	4
0		1	1	1	1	1	0.4225	0.9694	1.5376	2.1100	2.6831
1		1	1	1	1	0	0.6225	1.2035	1.7784	2.3521	2.9255

Table 2: Optimal management actions and expected utility computed using Bayesian updating. Parameters in the model of state dynamics are $\alpha_1 = 0$, $\alpha_2 = 1$, and $\beta = 0.5$.

y_0	$x_0^*, x_1^*, \ldots, x_4^* \mid y_0$					$\overline{U}(x_0^*, x_1^*, \ldots, x_4^* \mid y_0)$
0	1	1	1	1	1	2.6793
1	1	1	1	1	1	2.9217

loop control strategy, which only requires knowledge of the initial system state (Intriligator, 1971). The optimal policy computed using Bayesian updating qualifies (at least initially) as an open-loop control strategy because the optimal sequence of actions is based on a *probable*, as opposed to an *observed*, sequence of future system states. Therefore, it is not surprising that the expected sum of utilities associated with the Bayesian optimal policy (Table 2) is lower than that computed using SDP (Table 1). Generally speaking, the magnitude of the difference may increase with the number of decision times. Likewise, there are no differences in the optimal policies of SDP and Bayesian updating when only $T = 1$ decision is considered because both approaches condition solely on the initial (observed) state y_0.

4.2 Example 2

Consider a Markov decision problem identical to that described in Section 4.1 except that the parameters in the model of state dynamics (Section 2.1) are $\alpha_1 = 0$, $\alpha_2 = 1$, and $\beta = 3$. Therefore, the only difference is that the effects of serial dependence on changes in system state are expected to be high in magnitude relative to the effects of the management actions.

Table 3 contains a sequence of optimal state-dependent management actions computed using SDP. In contrast to the previous example, the optimal policy recommends action $x = 0$ (i.e., "do nothing") at most of the decision times, with the exception that action $x = 1$ has some value early in the decision sequence if $y_t = 0$. Although higher utilities are realized if $y_t = 1$, the transitional probabilities of achieving this state when $y_{t-1} = 0$ are very low (0.05 and 0.12), regardless of whether management action 0 or 1 is selected at $t - 1$. The reason

Table 3: Optimal management actions and expected utilities computed using SDP. Parameters in the model of state dynamics are $\alpha_1 = 0$, $\alpha_2 = 1$, and $\beta = 3$.

		$x_t^* \mid y_t$					$\overline{U}(x_t^* \mid y_t)$				
y_t	$t =$	0	1	2	3	4	0	1	2	3	4
0		1	1	0	0	0	0.0474	0.1378	0.2670	0.4802	0.7675
1		0	0	0	0	0	0.9526	1.8622	2.7330	3.5686	4.3747

Table 4: Optimal management actions and expected utility computed using Bayesian updating. Parameters in the model of state dynamics are $\alpha_1 = 0$, $\alpha_2 = 1$, and $\beta = 3$.

y_0	$x_0^*, x_1^*, \ldots, x_4^* \mid y_0$					$\overline{U}(x_0^*, x_1^*, \ldots, x_4^* \mid y_0)$
0	1	1	0	0	0	0.7558
1	0	0	0	0	0	4.3724

is that the high value of β relative to α_1 and α_2 greatly reduces the odds of a change in state. Of course, this also implies that if $y_{t-1} = 1$ the probabilities of remaining in this state at time t are very high (0.95 and 0.98). In this situation the optimal management action is $x_{t-1} = 0$ owing to the higher expected cost of implementing action $x_{t-1} = 1$.

The optimal sequence of management actions computed using Bayesian updating (Table 4) is identical to the optimal policy computing using SDP provided no changes in state occur from the initial (observed) state y_0. Again, this agreement is not surprising given the relatively high level of serial dependence in states. As in the previous example, the expected utilities summed over the sequence of optimal management actions are nearly identical when we compare SDP *vs* Bayesian updating (0.7675 *vs* 0.7558 for $y_0 = 0$, 4.3747 *vs* 4.3724 for $y_0 = 1$).

5 Discussion

In this paper we show that the optimal solutions to Markov decision problems provided by Bayesian updating without backward induction are open-loop control strategies, which differ from the optimal time- and state-dependent solutions provided by SDP. However, we believe that the differences in these solutions may be safely ignored for 2 reasons. First, a strong similarity in the objective functions of these 2 approaches is induced by combining Markovian state dynamics with a utility function that equals a sum of time-dependent utilities (e.g., compare (16) and (13)). Thus, an optimal solution computed by Bayesian updating at the beginning of a sequence of decisions is unlikely to differ dramatically from an optimal policy computed using SDP. Second, and even more important, is that strict adherence to a Bayesian philosophy effectively yields a closed-loop control strategy because the new information obtained after making each decision is used to update both the model parameters and the predicted consequences of current and future actions, given the observed (current) state of the system. An advantage of Bayesian updating is that Bayes Theorem provides a coherent procedure for updating the model parameters and its predictions as decisions are made and as new system states are observed.

Thus far we have only described the use of Bayesian updating in Markov decision problems with discrete state spaces. However, there are other classes of decision problems of great practical importance that may benefit by applying Bayesian approaches in their solution. For example, consider problems where landscapes are manipulated for the purpose of habitat management. These problems include spatial management units, and the collection of observable states relevant to decision making may include both continuous and discrete measurements of habitat. Furthermore, observations in different management units (though made simultaneously) are almost certain to be more correlated as the distance between units decreases owing to similarities in vegetation and other characteristics of the habitat. Therefore, the models of state dynamics needed in habitat management are likely to be considerably complex. Fortunately, there are virtually no limits to the complexity of models that can be entertained using Bayesian updating. Technological advancements in Bayesian computation and modern computers currently permit sophisticated, hierarchical models of spatial and temporal dependence to be fitted with relative ease (Wikle, Berliner, and Cressie, 1998; Datta, Ghosh, and Waller, 2000).

Bayesian updating is also likely to be useful in sequential decision problems where the relative effects of different management actions are poorly understood. In these problems managers initially may place greater value on learning about the magnitude of these effects than on achieving a particular management objective. These competing objectives must be specified in the utility function, which will include both model parameters (to quantify learning) and model predictions of observable system states (to quantify specific management objectives). A Bayesian treatment of this problem is reasonably straightforward. The benefits of learning can be formulated in the utility function as a discrepancy between the posterior distribution of model parameters and updates of the posterior computed from the distribution of predicted outcomes associated with a proposed set of management actions.

References

Bonney, G. E. (1987), "Logistic regression for dependent binary observations," *Biometrics*, 43, 951–973.

Carlin, B. P., Kadane, J. B., and Gelfand, A. E. (1998), "Approaches for optimal sequential decision analysis in clinical trials," *Biometrics*, 54, 964–975.

Datta, G., Ghosh, M., and Waller, L. A. (2000), "Hierarchical and empirical Bayes methods for environmental risk assessment," in *Handbook of Environmental Statistics*, eds. Rao, C. R. and Sen, P. K., North Holland, pp. 223–245.

Dorazio, R. M. and Johnson, F. A. (2003), "Bayesian inference and decision theory: a framework for decision making in natural resource management," *Ecological Applications*, 13, 556–563.

Intriligator, M. D. (1971), *Mathematical optimization and economic theory*, Englewood Cliffs, New Jersey: Prentice Hall.

Müller, P. (1999), "Simulation based optimal design," in *Bayesian Statistics 6*, eds. Berger, J. O., Bernardo, J. M., Dawid, A. P., and M., S. A. F., Oxford University Press, pp. 459–474.

Müller, P., Berry, D., Grieve, A., Smith, M., and Krams, M. (2000), "Simulation-based sequential Bayesian design," Tech. rep.

Puterman, M. L. (1994), *Markov decision processes: discrete stochastic dynamic programming*, New York: John Wiley and Sons.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998), "Hierarchical Bayesian space-time models," *Environmental and Ecological Statistics*, 5, 117–154.