

REVIEW ARTICLE

Partial observability and management of ecological systems

Byron K. Williams  | Eleanor D. Brown 

U.S. Geological Survey, Science and Decisions Center, Reston, Virginia, USA

Correspondence

Eleanor D. Brown, U.S. Geological Survey, Science and Decisions Center, 12201 Sunrise Valley Drive, Reston, VA 20192, USA.

Email: ebrown@usgs.gov

Funding information

U.S. Geological Survey, Grant/Award Number: G19AC00146; U.S. Geological Survey, Grant/Award Number: G19AC00146

Abstract

The actual state of ecological systems is rarely known with certainty, but management actions must often be taken regardless of imperfect measurement (partial observability). Because of the difficulties in accounting for partial observability, it is usually treated in an ad hoc fashion, or simply ignored altogether. Yet incorporating partial observability into decision processes lends a realism that has the potential to improve ecological outcomes significantly. We review frameworks for dealing with partial observability, focusing specifically on dynamic ecological systems with Markovian transitions, i.e., transitions among system states that are influenced by the current system state and management action over time. Fully observable states are represented in an observable Markov decision process (MDP), whereas obscure or hidden states are represented in a partially observable process (POMDP). POMDPs can be seen as a natural extension of observable MDPs. Management under partial observability generalizes the situation for complete observability, by recognizing uncertainty about the system's state and incorporating sequential observations associated with, but not the same as, the states themselves. Decisions that otherwise would depend on the actual state must be based instead on state probability distributions ("belief states"). Partial observability requires adaptation of the entire decision process, including the use of belief states and Bayesian updates, valuation that includes expectations over observations, and optimal strategy that identifies actions for belief states over a continuous belief space. We compare MDPs and POMDPs and highlight POMDP applications to some common ecological problems. We clarify the structure and operations, approaches for finding solutions, and analytic challenges of POMDPs for practicing ecologists. Both observable and partially observable MDPs can use an inductive approach to identify optimal strategies and values, with a considerable increase in mathematical complexity with POMDPs. Better understanding of POMDPs can help decision makers manage imperfectly measured ecological systems more effectively.

KEYWORDS

decision strategy, Markov decision process, partial observability, system dynamics, uncertainty

TAXONOMY CLASSIFICATION

Applied ecology; Conservation ecology; Ecosystem ecology; Population ecology

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Published 2022. This article is a U.S. Government work and is in the public domain in the USA. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Many important issues in ecology and management of ecological systems concern the behavior of dynamic systems in the presence of uncertainty. But changing ecological status and associated uncertainties can present real challenges to effective management (Nicol et al., 2015; Williams et al., 2002). For example, with dynamic systems smart decision making over an extended time must account for the potential effects of both current and future actions. A large body of theory and methodology has been developed over many decades to assess the optimal control of dynamic systems, although the size and complexity of problems to which it can be applied remain limited (Bertsekas, 2017).

Uncertainty about the actual state of an ecological system and its processes presents additional challenges. In ecology, a major source of uncertainty is partial observation (or imperfect measurement) of status over time. System dynamics are almost always tracked with sampling that leaves much of the system unobserved and subject to estimator imprecision (Williams & Brown, 2019). This is the case even with the most carefully designed and intensive sampling effort.

The challenges presented by imperfect observability are clearly seen in animal ecology and conservation. For example, the inadequacy of treating counts of animals as if they are known abundances has become widely recognized. That counts reflect the degree of detection during sampling is by now universally accepted, and much of the methods literature in recent years has dealt with sampling processes that account for partial detectability. In contrast, imperfect observability has been integrated into ecological management decision methods only to a very limited extent, largely because of the complexity of decision processes that incorporate estimated (imperfectly known) state and other variables, and the computational difficulties of implementing associated methods even with relatively small problems. There is a clear need to go beyond treating partial observability in sampling and estimation, by expanding its integration further into actual decision making.

We focus on ecological systems that are managed and tracked over time, and specifically on Markov decision processes, i.e., processes for which the probability of transition between successive states at any point depends only on the state and action taken at that time (Puterman, 1994). We use a standard objective for decision making of maximizing the accumulation of discounted returns over time.

The observability of the actual state of an ecological system when decisions are being made determines the type of Markov process. Markovian transitions among observable states are represented in an observable Markov decision process (MDP), whereas transitions among partially observable states are represented in a partially observable process (POMDP). Most aspects of a Markovian control problem must be adapted to partial observability, including transitions among states, valuation, and status tracking.

Many ecological problems lend themselves to a treatment with POMDPs. A common situation involves a partially observable resource that is subject to sequential decision making and

monitoring over an extended time. To date, POMDPs have been applied to a limited number of ecological management and conservation problems for which accurate measurement is difficult or expensive. Among the most common of these are management of cryptic endangered species (Chadès et al., 2008; McDonald-Madden et al., 2011; Tomberlin, 2010a); control of invasive plant species (Regan et al., 2011) and animal species (Kling et al., 2017; Peron et al., 2017; Rout et al., 2014), especially invasive forest pests (Fackler & Haight, 2014; Fackler & Pacifici, 2014; Haight & Polasky, 2010); and commercial fisheries (Lane, 1989; Memarzadeh et al., 2019; Memarzadeh & Boettiger, 2018). These and other examples are highlighted in Table 1.

Importantly, incorporating partial observability into decision processes lends a realism that has the potential to improve ecological outcomes. For example, McDonald-Madden et al. (2011) showed that accounting for partial observability led to better strategic outcomes in conservation planning to save the last remaining wild Sumatran tigers (*Panthera tigris sumatrae*). Realism can be especially important in a regulatory context such as commercial fisheries, where standard models that assume perfect measurements of a stock can lead to harvest decision rules that cause fishery collapse, as in the case of the Argentine hake *Merluccius hubbsi* (Memarzadeh & Boettiger, 2018). In contrast, Memarzadeh et al. (2019) demonstrated that POMDP-based decision methods could avoid unintentional extinctions, and lead to consistently higher rates of recovery of depleted fish stocks.

In this paper, we compare completely and partially observed Markov decision processes for dynamic ecological systems that are managed and tracked over time. A comparison of MDPs and POMDPs highlights analytic and operational similarities between these two situations and clarifies the increased complexity one confronts when realistically accounting for limited observability. We build on recent ecological literature (e.g., Chadès et al., 2021; Williams, 2009, 2011) and provide additional detail for ecologists who wish to understand the mechanics of POMDPs. We describe specifications, policies, valuations, and solution approaches for observable and partially observable MDPs. In addition, we discuss model extensions, infinite versus finite time horizons, mixed observability processes, adaptive management with POMDPs, nonstationary models, and continuous states in considerable detail.

In the following sections, we illustrate the concepts of POMDPs with examples from long-term sport hunting of waterfowl in North America. Waterfowl hunting has been regulated for over a century by U.S. federal law and international agreement, and managed since 1995 through the annual setting of hunting regulations under the rubric of “adaptive harvest management” (Johnson et al., 2015; Williams & Johnson, 1995). Harvest management relies on simple models of waterfowl population dynamics that are based on hypotheses about the impact of harvest on annual survivorship and the importance of density dependence in recruitment (Figure 1). Models incorporating different hypotheses produce different population trajectories, and model effectiveness can be evaluated by comparing these trajectories against

TABLE 1 Applications of partially observable Markov decision processes in ecology

Author(s)	Resource Context	Species	Actions	Features	Solution approach
Chadès et al. (2008)	Endangered species	Sumatran tiger <i>Panthera tigris sumatrae</i>	Manage habitat, survey, do nothing	2 states (extant/extinct), 3 actions, 3 time steps	Analytical solution, incremental pruning
Fackler and Pacifici (2014)	Pest infestation	General pest species	Unspecified	2 hidden models, environment signals used for updating model uncertainty	State space discretization, stochastic dynamic programming
Haight and Polasky (2010)	Invasive species	Forest pest	Monitoring only, treatment only, both, or neither	Discretized approximation of belief space for 3 states	Customized FORTRAN program for 4 actions over 20 time periods
Lane (1989)	Commercial fishing	Salmon (<i>Oncorhynchus</i> spp.)	Intra-seasonal fishing within geographic fishing zones	Finite states for each zone; actions are fishing within a zone, or not	Sondik one-pass algorithm
McDonald-Madden et al. (2011)	Endangered species	Sumatran tiger	Survey, management, do nothing	Considers 2 populations, compares results with perfect observability	Incremental pruning
Memarzadeh and Boettiger (2018)	Marine fisheries	Argentine hake <i>Merluccius hubbsi</i>	Fish harvest	Handles either or both structural uncertainty and partial observability	Point-based value iteration with SARSOP
Regan et al. (2011)	Invasive species	Branched broomrape <i>Orobancha ramosa</i>	Low cost and inefficient action, high cost but efficient action, do nothing	3 states, 3 actions, 2 observations	Stochastic dynamic programming
Rout et al. (2014)	Invasive species	Black rat <i>Rattus rattus</i>	Quarantine, control, mixture	3 states, 3 actions	Incremental pruning
Tomberlin (2010b)	Endangered seabird habitat	Marbled murrelet <i>Brachyramphus marmoratus</i>	Monitoring, habitat management	2 occupancy states, 2 actions, 5 time steps	Eagle/Monahan algorithms
Pascal et al. (2020)	Threatened and endangered species	Sumatran tiger	Manage, monitor	Framed in terms of the termination of management or surveying, 2 observations	Point-based value iteration with SARSOP
Tomberlin (2010a)	Erosion control on roads in redwood forests	Forestry	High or low monitoring effort, erosion treatment, no action	High- and low-level erosion condition	Backward recursion with α - vectors over finite time
Nicol et al. (2015)	Management of shorebird habitats under changing sea-level conditions	10 species of migratory shorebirds	Protection of non-breeding sites against sea-level change	Factored MOMDP with observed states, uncertain and non-stationary transition structure	Point based value iteration with Symbolic Perseus
Nicol and Chadès (2012)	Marine oil spill risks	Sea otter <i>Enhydra lutris</i>	Reduction or cleanup of oil spills, reintroduction, monitoring	Discretization of a continuous state with CU-tree ^a	Point-based value iteration with Perseus
Kling et al. (2017)	Marine species in a preserve	Lionfish <i>Pterois</i> spp.	Lionfish removal, monitoring	Observations based on removal effort and presence/absence of monitoring	Projected belief MDP
Memarzadeh et al. (2019)	Commercial fisheries	Multiple fish species	Fishing catch quotas	Compares MSY, MDP, and POMDP solutions	Point-based value iteration with SARSOP

(Continues)

TABLE 1 (Continued)

Author(s)	Resource Context	Species	Actions	Features	Solution approach
Sloggy et al. (2020)	Commercial forestry	Loblolly pine <i>Pinus taeda</i>	Monitoring, harvest, regeneration, delay	Valuation based on stochastic price and harvest amounts	Projected belief MDP
Fackler and Haight (2014)	Invasive species	Forest pest	Treatment only, monitoring, both monitoring and treatment, no action	Discretization of a continuous state	Monahan's exact method, variations on Lovejoy's discretization method
Fackler et al. (2014)	Recreation management	Golden eagle <i>Aquila chrysaetos</i>	Management of recreation near nesting sites	Observable occupancy, uncertain disturbance effect	Unspecified
MacLachlan et al. (2017)	Bovine tuberculosis in cattle herds	Domestic cattle <i>Bos taurus</i>	Herd testing for infection, followed by isolation if found	3 states for a herd, testing and control of multiple herds	Projected belief MDP
Sethi et al. (2005)	Commercial fisheries	Commercial fish species	Fish harvest quotas	Discrete observations, continuous states and actions; spline interpolations	Value function iteration for infinite time horizon
Chadès et al. (2021)	Australian birds threatened by habitat degradation and predation	Gouldian finch <i>Erythrura gouldiae</i>	Fire and grazing management, feral cat control, provide nesting boxes, do nothing	Observable states, stationary but unknown state transitions	MOMDP ("hidden model MDP") with discrete states and models, several solution algorithms

^aSee Uthé and Veloso (1998).

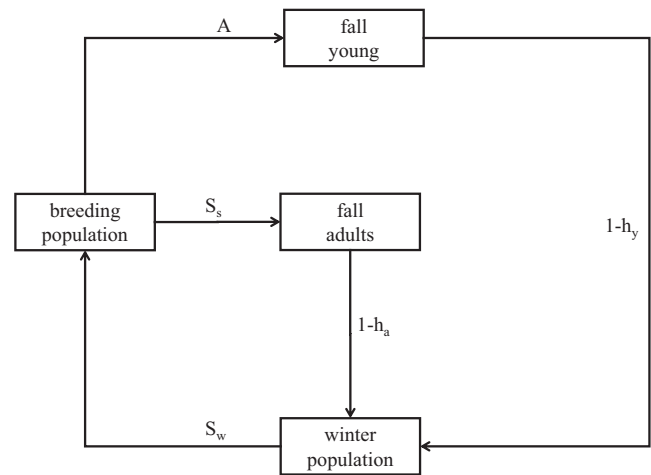


FIGURE 1 Model of waterfowl population dynamics. Includes survival rates for spring–summer (S_s) and fall–winter (S_w), along with harvest rates for young (a_y) and adults (a_a) and age ratio (A) for reproduction/recruitment.

observations from annual population monitoring. Such a framework can be used to investigate optimal harvest strategies in the presence of partial observability, as well as imperfect understanding of population dynamics (Williams, 2011).

2 | PROCESS SPECIFICATION

In this section, we introduce the general elements of Markov decision processes, including system states, transitions among states, observations, management actions, returns (or rewards), discount factors, and time horizons. These elements provide a foundation for describing dynamic ecological systems that are managed over time.

In an ecological context, decision making over time builds on transitions among states, as influenced by management actions in concert with ecological processes such as mortality, reproduction, and movement. Stochastic variation in the transitions can be described with transition probabilities in a stochastic process, or in the case of Markovian transitions, a Markov decision process. In our example of waterfowl harvests, the change in population size from 1 year to the next is held to be influenced by the current population size, environmental conditions and the amount of harvest in the fall. Stochasticity in population size the ensuing year is a result of environmental fluctuations, randomness in the influence of hunting regulations, and stochastic biological processes that produce change.

A formal specification of a Markov decision process, whether partially or completely observable, must account for system dynamics and management returns over some time horizon. More specifically, it includes the duration of the process, a characterization of system state, probabilities of transition among states at each time step, and a value function that aggregates returns to management over time. Ecological status is assumed to be tracked as decisions are made at discrete times. We assume initially that there are

finitely many possible states and actions at each point over a finite time horizon, and later consider continuous-state and infinite-time POMDPs.

2.1 | Specification of observable MDPs

A controlled process with Markovian transitions among observable states is characterized as follows.

Notation:

x , state of an ecological system, which for an MDP is observable.

a , action that influences returns and transitions among states ("state transitions")

State transitions:

$$x' = F(x, a, z)$$

with random environmental conditions z , from which are produced probabilities of transition

$$P(x' | x, a)$$

from state x to state x' , given that action a is taken.

Returns:

Immediate returns $R(a|x)$ are assumed to depend on the system's state and the action taken in that state. If returns are based on transitions, then $R(a|x) = \sum_{x'} P(x' | x, a) R(a | x, x')$.

MDP specification:

An observable MDP is specified by the tuple $\{X, A, P, R, T, \lambda\}$, where

- X is the set of system states x . Examples could include population size or density, population vital rate, spatial distribution, biodiversity, and habitat features.

- A is the set of actions a that are available to a manager, potentially including monitoring as well as conservation actions. Examples could include selection of hunting limits, introduction or removal of species, habitat manipulation, contaminant clean-up, adaptations to climate change, regulatory actions, and field sampling designs.
- P is a transition probability function specifying probabilities $P(x' | x, a)$ of transition from state x to state x' , given that action a is taken. The conditional probability $P(x | x, a)$ corresponds to no change, and $\sum_{x'} P(x' | x, a) = 1$.
- R is a return or reward function, with $R(a|x)$ the immediate return when action a is taken and the system is in state x . For example, returns could be measured in terms of population survival rate, number of animals, increase in biodiversity, risk abatement, economic profit, and opportunity cost.
- T is the terminal time of a time horizon consisting of equal time steps between an initial time and T , which could be infinite.
- λ is a discount factor between 0 and 1 that relates future returns to present value. As λ declines from unity, future returns become less important relative to immediate returns.

In an observable MDP, observations coincide with actual states. At any time, the state affects the selection of an action and influences returns and transitions to subsequent states (Figure 2). Actions in turn influence state transitions and returns.

The observable MDP framework can be applied to our example of the sport harvest of waterfowl. Thus, state x represents population size at a given point in time, x' is the population size at the next time, a is the harvest rate targeted by current regulations, z represents environmental conditions (e.g., spring precipitation), and $R(a|x)$ is the amount of harvest for harvest rate a , given the population size x . The state transition function $x' = F(x, a, z)$ describing population change from one time to the next is held to be understood and well specified,

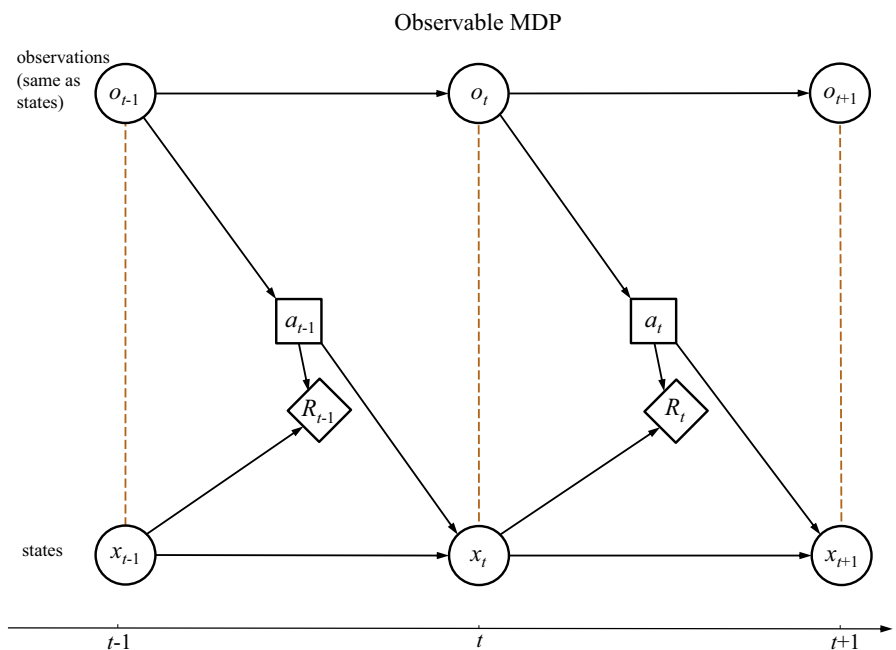


FIGURE 2 Influence diagram for an observable Markov decision process

and the population size is assumed to be known with certainty (or can effectively be treated as such) at each point in time.

2.2 | Specification of partially observable MDPs

POMDPs extend the framework of observable MDPs by including observations that differ from, but are related to, the unobservable system states. A common situation is for the observations to produce estimates of the system state (Nichols & Williams, 2006), although in general the only requirement is a statistical association between observations and the process state. Like the observable states in an MDP, observations in a POMDP are used to track changes in status over time. A Markov decision process with transitions among unobservable states is characterized by the following additional features and adaptations.

Notation:

x , system state, which is unobservable

a , action that influences returns, state transitions, and (possibly) observations

o , observation (usually assumed to be discrete) that is associated with, but not the same as, system state

b , belief state, with $b(x)$ the probability associated with state x

Observation function:

$$o' = G(x', a, \epsilon)$$

producing random observations o' , with probabilities

$$f(o'|x', a).$$

Actions may or may not influence observations; if not, the observation probabilities reduce to $f(o'|x')$. Initially, we assume observation o' is tied to the posterior system state x' after implementation of prior action a . Later, we consider a different order for observations and state updates. In some but not all cases, observations can be expressed as data-based estimators.

Returns:

Immediate returns are averaged over belief state b :

$$R(a|b) = \sum_x b(x)R(a|x).$$

POMDP specification:

A POMDP generalizes $\{X, A, P, R, T, \lambda\}$ for observable MDPs, by allowing states to be only partially observable and appending a probability distribution for observations in an observation space O . Thus, a POMDP is specified by the tuple $\{X, A, O, P, f, R, T, \lambda\}$, where

- O is a set of potential observations o , obtainable through activities such as field sampling, modeling, or laboratory assessments.
- f is an observation function, with $f(o'|x', a)$ the probability that o' is observed, given state x' and action a .

Because the states are themselves unobservable, ecological status must be tracked with belief states. At any time the actual state of the system influences immediate returns, transitions to subsequent states, and observations, but not actions (Figure 3). Observations are used to update belief states, which in turn inform the selection of actions. Finally, actions control transitions, returns, and (possibly) observations. A comparison of Figures 2 and 3 makes it clear that the framework for POMDPs extends that of an observable MDP, by incorporating observations that differ from the actual system states and introducing belief states to track the system's status over time.

In our waterfowl example, the only difference in the frameworks for partial and complete observability concerns the observability of population size x . For the POMDP framework, x cannot be observed directly and must be tracked with data o that are obtained through field sampling. The data are combined into an estimator of population size that is associated with the actual population size, with randomness inherited from sampling and estimation protocols. For this situation, the estimator distribution serves as the population belief state.

The use of belief states to track the status of the system is a critical feature distinguishing POMDPs from observable MDPs. The states in an observable MDP typically are discrete and countable, and define a finite state space. Given finitely many actions, it is theoretically possible to list all state/action combinations and compare them in evaluating MDP policies. For a POMDP with finitely many actions and observations, it also is possible to identify all action/observation combinations for a particular belief state. However, any effort to do so over all action/belief state combinations is defeated by the continuous nature of a belief space comprising infinitely many belief states. As discussed later, a different approach from that for MDPs must be taken to evaluate a POMDP, i.e., one that explicitly accounts for a continuous belief space.

3 | PROCESS POLICY

In this section we describe policies for a Markov decision process in terms of time-specific states, observations, and actions, and characterize policies for both observable and partially observable MDPs in terms of *policy trees*. The notation for policy trees highlights the linkages between observable MDPs and POMDPs.

The trajectory of a Markov decision process over its time horizon is controlled by the temporal sequence of decisions imposed on the process, i.e., the process policy. A policy extends the notion of a time-specific action influencing system transitions, to include actions and transitions over the duration of the process. Thus, it identifies actions that are tied to the status of the system at every point in the time horizon. The sequence of state-based decisions for a Markov process is a defining part of the process, in that state trajectories, values, patterns of actions, and recurrences among states are all influenced by the process policy.

FIGURE 3 Influence diagram for a partially observable Markov decision process (after Chadès et al., 2021).

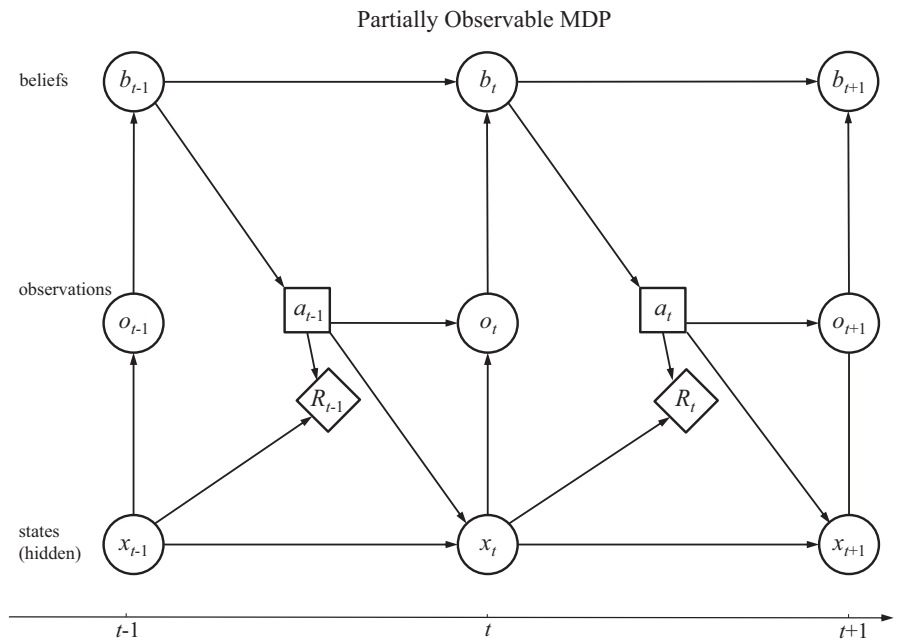
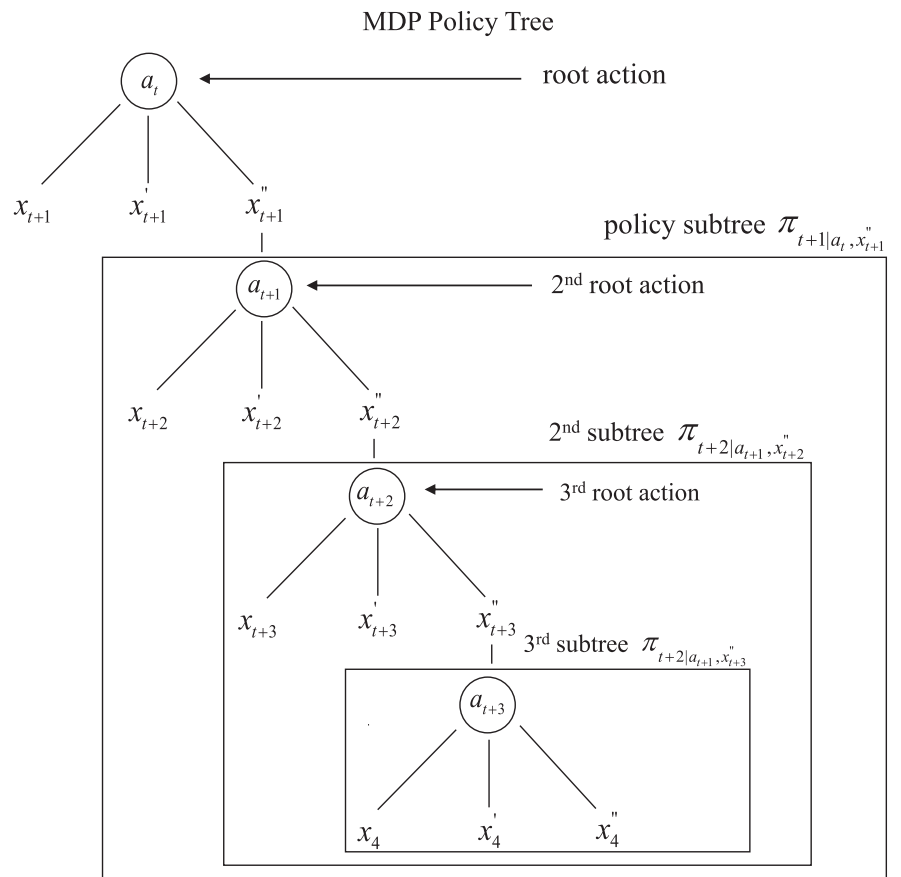


FIGURE 4 Policy tree for an observable Markov decision process



For observable MDPs, a policy essentially assigns an action a for every system state $x \in X$ at every time over the duration of the process. On the other hand, a policy for partially observable MDPs assigns an action a for every belief state b at every point in time. Policies for both MDPs and POMDPs can be described with actions that are hierarchically organized in policy trees (Kaelbling et al., 1998).

3.1 | Policy for observable MDPs

A policy tree for an observable Markov decision process displays actions and (observable) states over the course of the process time horizon $\{t, \dots, T\}$. A tree is arranged temporally, with a root action followed in sequence by states and actions at later times (Figure 4).

If action a_t is taken at t , the sub-tree $\pi_{t+1|a_t, x'}$ consists of actions for states over the remainder of the time horizon. By construction, policy tree π_t is simply a root action a_t and sub-trees for all subsequent states x' , that is,

$$\pi_t = \{a_t, \pi_{t+1|a_t, x'} | x' \in X\}.$$

Because of the hierarchical nature of a policy tree, any state at any time could be thought of as a starting point, with the action for that state considered to be the root action of a policy tree. This allows one to essentially “decompose” a policy into a temporal hierarchy, in which the decision-making framework at a given time subsumes all decisions for later times, and is itself subsumed in the decision-making frameworks for earlier times. As discussed in the next section, this hierarchical clustering allows a concise representation of iterative valuation and policy determination.

In our waterfowl example, a policy tree under full observability simply consists of hunting regulations each year for each population size. A particular trajectory of population sizes over time will have an associated sequence of hunting regulations, which fluctuate over time as the population does. And at any particular time, the range of regulations for a policy will be tied to the possible population sizes at that time. Regulatory variation across sizes and times is expressed in the notation $\pi_t = \{a_t, \pi_{t+1|a_t, x'} | x' \in X\}$.

3.2 | Policy for partially observable MDPs

Because system states are not observed under partial observability, policy trees for a POMDP must be based on observations rather than the (unobservable) states themselves. A POMDP policy tree has a root action followed in sequence by observations and actions at later times (Figure 5). If action a_t is taken at t , the sub-trees $\pi_{t+1|a_t, o'}$ consist of actions for later observations over the duration of the process. By construction, policy tree π_t is simply the combination of a root action a_t and sub-trees for all possible observations o' , that is,

$$\pi_t = \{a_t, \pi_{t+1|a_t, o'} | o' \in O\}.$$

As with observable MDPs, the clustering of policy trees for POMDPs allows iterative valuation and policy determination to be concisely represented.

In our waterfowl hunting example, a policy tree under partial observability consists of hunting regulations each year for each estimate of population size based on the field data. A particular trajectory of data-based estimates over time will have an associated sequence of hunting regulations. And at any particular time, the range of regulations will be tied to the possible population estimates at that time. Regulatory variation across data and times is expressed by the notation $\pi_t = \{a_t, \pi_{t+1|a_t, o'} | o' \in O\}$.

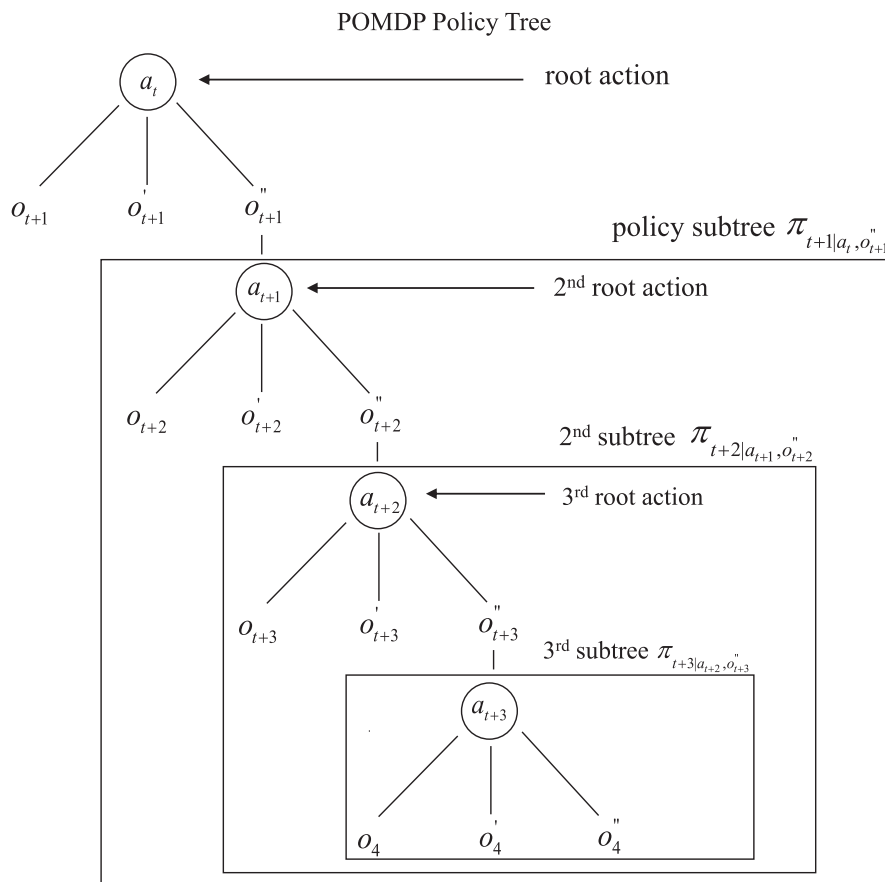


FIGURE 5 Policy tree for a partially observable Markov decision process

4 | PROCESS VALUATION

In this section, we discuss valuation for observable MDPs and POMDPs, including optimal valuation. We clarify how valuation is actually determined with step-by-step procedures for finding policy-based values, and we describe some procedural alternatives found in the literature for optimal policy and valuation.

The value function serves as a metric for comparing as well as measuring performance of policies for a decision process. For observable MDPs, it aggregates returns for an MDP policy tree, starting in state x at time t . For partially observable MDPs, it aggregates returns for a POMDP policy tree starting in belief state b at time t . In both cases, the value function can be used to compare policies and identify an optimal policy.

4.1 | Valuation with observable MDPs

Valuation for completely observable Markov decision processes can be described in terms of policy trees $\pi_t = \{a_t | \pi_{t+1}|a_t, x', x' \in X\}$, each tree having an associated vector of state-specific components

$$V_{\pi_t}(x) = R(a_t|x) + \lambda \sum_{x'} P(x'|x, a_t) V_{\pi_{t+1}|a_t, x'}(x') \quad (1)$$

(see Appendix S1). The value function in Equation (1) includes an immediate return $R(a_t|x)$ along with future values $V_{\pi_{t+1}|a_t, x'}(x')$ that are averaged over the system states x' . Calculation of $V_{\pi_t}(x)$ thus involves two steps:

- averaging the posterior values $V_{\pi_{t+1}|a_t, x'}(x')$ with transition probabilities $P(x'|x, a_t)$; and
- discounting the average posterior value with λ and adding the immediate return $R(a_t|x)$ to get $V_{\pi_t}(x)$.

A more concise expression for the value function is

$$V_{\pi_t}(x) = R(a_t|x) + \lambda V'_{\pi_{t+1}}(x|a_t), \quad (2)$$

where $V'_{\pi_{t+1}}(x|a_t)$ represents a transformation of future values in Equation (1) by the transition probabilities, i.e.,

$$V'_{\pi_{t+1}}(x|a_t) = \sum_{x'} P(x'|x, a_t) V_{\pi_{t+1}|a_t, x'}(x').$$

The assessment of a decision process typically involves a search for policies that can produce the highest value. To obtain optimal valuation with observable MDPs, the values and policies in Equation (1) can be optimized at each time with the Bellman equation (Bellman, 1957), by means of backward recursion (Bertsekas, 2012). From Equation (1), optimal valuation can be expressed as

$$V_t[x] = \max_a \left\{ R(a|x) + \lambda \sum_{x'} P(x'|x, a) \max_{\pi_{t+1}} V_{\pi_{t+1}}(x') \right\}$$

(see Appendix S1). Thus, the optimal value for a state x is produced in a two-step procedure:

- optimize future returns $V_{\pi_{t+1}}(x')$ over the possible trees at $t+1$; and
- optimize the sum $R(a|x) + \lambda \sum_{x'} P(x'|x, a) V_{\pi_{t+1}}(x')$ over a

(see Williams et al., 2002; Marescot et al., 2013 for details). Optimal valuation can also be expressed in terms of Equation (2) by

$$V[x] = \max_a \left\{ R(a_t|x) + \lambda \max_{\pi_{t+1}} V'_{\pi_{t+1}}(x|a_t) \right\}. \quad (3)$$

In our waterfowl example, with observable population status, the value function for a population of size x_0 starting at time $t = 0$ can be represented simply as the expected sum of current and future harvest amounts over the problem time horizon, $V(x_0) = E \left[R(a_0|x_0) + \sum_{t=1}^T \lambda^t R(a_t|x_t) \right]$ where future population states are described in terms of Markov transitions as above. We note that such a value function is intrinsically conservation oriented, in that current harvest, by influencing the status of future populations, must account for future harvest yields.

4.2 | Valuation with partially observable MDPs

Valuation for partially observable Markov decision processes is based on policy trees $\pi_t = \{a_t, \pi_{t+1}|a_t, o' | o' \in O\}$. Every tree π_t has associated with it a vector of state-specific values

$$V_{\pi_t}(x) = R(a_t|x) + \lambda \sum_{x'} P(x'|x, a_t) \sum_{o'} f(o'|x', a_t) V_{\pi_{t+1}|a_t, o'}(x') \quad (4)$$

(see Appendix S1). The value function in Equation (4) includes an immediate return $R(a_t|x)$ for a prior state x , along with future values $V_{\pi_{t+1}|a_t, o'}(x')$ averaged over observations o' as well as posterior states x' . A comparison of Equations (1) and (4) shows that valuation of a POMDP has the same general form as that of an MDP, except $V_{\pi_{t+1}|a_t, x'}(x')$ in Equation (1) is replaced by the average value

$$\sum_{o'} f(o'|x', a) V_{\pi_{t+1}|a_t, o'}(x')$$

in Equation (4).

Because the state x of a partially observable process is not known, actual valuation must be based on a belief state b , with $V_{\pi_t}(x)$ averaged over b :

$$V_{\pi_t}(b) = \sum_x b(x) V_{\pi_t}(x).$$

In the Appendix S1, we describe two useful forms for computing $V_{\pi_t}(b)$. One uses a transformation of future values with the transition probabilities

$$V'_{\pi_{t+1}}(x|a_t, o') = \sum_{x'} P(x', o' | x, a_t) V_{\pi_{t+1}|a_t, o'}(x')$$

to express valuation as

$$\begin{aligned} V_{\pi_t}(b) &= R(a_t | b) + \lambda \sum_{o'} \sum_x b(x) V'_{\pi_{t+1}}(xa_t | o'), \\ &= R(a_t | b) + \lambda \sum_{o'} V'_{\pi_{t+1}}(b | a_t, o'). \end{aligned} \quad (5)$$

Note that Equation (5) has the same general form as Equation (2) for observable MDPs, except $V'_{\pi_{t+1}}(x|a)$ in Equation (2) is replaced by the aggregated value

$$\sum_{o'} V'_{\pi_{t+1}}(b|a_t, o')$$

in Equation (5). The effect of partial observability is thus to require an aggregation of values over the observations.

An alternative but equivalent form for $V_{\pi_t}(b)$ uses Bayesian updating of beliefs,

$$\begin{aligned} b_{a, o'}(x') &= \frac{P(o', x' | b, a)}{\frac{P(o' | b, a)}{\sum_x P(o', x' | x, a) b(x)}} \\ &= \frac{P(o', x' | b, a)}{P(o' | b, a)}, \end{aligned}$$

to get

$$V_{\pi_t}(b) = R(a|b) + \lambda \sum_{o'} P(o' | b, a) V_{\pi_{t+1}|a, o'}(b_{a, o'}). \quad (6)$$

The forms in Equations (5) and (6) produce the same values for all belief states in the belief space.

Value expressions (5) and (6) both can be used to compute optimal values for a POMDP. Optimal values based on Equation (5) are given by

$$V_t[b] = \max_{a_t} \left\{ R(a_t | b) + \lambda \sum_{o'} \max_{\pi_{t+1}} V'_{\pi_{t+1}}(b | a_t, o') \right\}, \quad (7)$$

and optimal values based on Equation (6) are given by

$$\begin{aligned} V_t[b] &= \max_{\pi_t} \sum_x b(x) V_{\pi_t}(x) \\ &= \max_{a_t} \left\{ R(a_t | b) + \lambda \sum_{o'} P(o' | b, a_t) \max_{\pi_{t+1}} V_{\pi_{t+1}}(b_{a_t, o'}) \right\} \end{aligned} \quad (8)$$

(see Appendix S1).

As with observable MDPs, expressions (7) and (8) both involve two optimizations, one over trees for time $t+1$ and one over actions at time t . A comparison of Equations (3) and (7) shows that MDPs and POMDPs have analogous formats for optimization, except the latter equation includes an aggregation of optimal future values across observations.

In our waterfowl example, with harvest regulations based on partially observable populations, the value function for a population with belief state b_0 starting at time $t=0$ can be represented simply as the expected sum of current and future harvest amounts over

the problem time horizon, $V(b_0) = E \left[R(a_0 | b_0) + \sum_{t=1}^T \lambda^t R(a_t | b_t) \right]$. In this case, future belief states are tied to observations through Bayes' theorem, as above. As with complete observability, accounting for future harvests means that the current harvest, by influencing future population status, must account for future harvest yields.

4.3 | Standard versus extended models

In the standard POMDP model for state transitions, observations are held to occur after state transitions, without directly affecting the state transition probabilities. An alternative model allows observations to occur before state transitions. By incorporating a different sequencing of observations and state transitions, an alternate or extended model allows one to consider many problems not easily accommodated by the standard model, namely those in which observations can influence the transition probabilities. In our waterfowl hunting example, observations of waterfowl harvest in the fall can produce updated beliefs before winter mortality and spring reproduction affect next year's population state, and thus can influence the transitions used in the valuation of harvest strategies.

The operational difference between the standard and extended models is seen by a comparison of belief-updating and the respective value functions. With the standard model, observations occur after the state transitions,

$$x, a \rightarrow x' \rightarrow o',$$

so that observations o' do not influence the transition probabilities $P(x' | x, a)$. Belief states are updated by

$$\begin{aligned} b_{a, o'}(x') &= \frac{P(x', o' | b, a)}{\frac{P(o' | b, a)}{f(o' | x', a) P(x' | b, a)}} \\ &= \frac{P(x', o' | b, a)}{P(o' | b, a)}, \end{aligned}$$

and the process value function averages immediate and future value over observations o' :

$$V_{\pi_t}(b) = R(a|b) + \lambda \sum_{o'} P(o' | b, a) V_{\pi_{t+1}|a, o'}(b_{a, o'})$$

(see Appendix S1).

On the other hand, with the extended model the observations occur before the state transitions,

$$x, a \rightarrow o \rightarrow x',$$

so that observations o can influence the transition probabilities $P(x' | x, a, o)$. Belief states are updated by

$$\begin{aligned} b_{a, o}(x') &= \frac{P(x', o | b, a)}{\frac{f(o | b, a)}{f(o | x, a) P(x' | b, a, o)}} \\ &= \frac{P(x', o | b, a)}{f(o | b, a)}, \end{aligned}$$

and the process value function averages immediate and future value over observations o :

$$V_{\pi_t}(b) = R(a|x) + \lambda \sum_o f(o | b, a) V_{\pi_{t+1|a,o}}(b_{a,o}) \quad (9)$$

(see Appendix S1). The value function shown in Equation (9) for the extended model differs from that for the standard model only in the use of prior and posterior observations in the updating of beliefs and weighting of future values.

The extended model allows for assessment of many ecological problems that otherwise would be difficult or impossible to assess with the standard model. Fackler and Pacifici (2014) describe three examples representing different levels of dependence between observations and future states. One involves the observed harvest of an unobserved population, where the future population state is directly influenced by the observation of harvest in the prior year. Another example involves a treatment to reduce an unobserved pest infestation, where observed environmental conditions in the previous year influence future infestation. A third involves the control of avian nest predation, where observed predator numbers in the previous year influence predation and thus the future status of an avian population. Assessment in these and other cases is facilitated by the extended model, in which observations informing and possibly influencing management actions that affect future ecological conditions occur before the ecological transitions themselves.

5 | SOLUTION APPROACHES

In this section, we consider the mechanics of different approaches to finding policies with optimal value. We discuss valuation by means of value iteration for both observable and partially observable MDPs. We describe the construct of α -vectors for POMDPs, and outline iterative approaches to optimal policy and valuation that use α vectors.

A key challenge in managing dynamic systems involves the number of decisions that can potentially be made over time. The number of possible policy trees for an observable MDP increases exponentially with an increasing number of states, actions, and length of the time horizon. Even more troubling for POMDPs is that a listing and evaluation of trees is not possible because of the continuous belief space. In fact, finite-horizon POMDPs are PSPACE-complete (Papadimitriou & Tsitsiklis, 1987), and infinite-horizon POMDPs are undecidable (Madani et al., 2003). Thus, approximations of optimal solutions must be used for most problems.

5.1 | Solution approaches with observable MDPs

The solution of an observable MDP yields optimal values $V_t[x]$ across a discrete state space at each time t . With finitely many states and

actions, values for every policy tree could at least conceivably be listed for all states at each time, and optimal actions and values could be identified. However, such an exhaustive enumeration is prohibitively costly in terms of computing resources for all but small problems.

Finding optimal values and policies is greatly facilitated by value iteration, in which optimal valuation begins at the terminal time and proceeds backward to find optimal values that build on those previously identified (Marescot et al., 2013). Value iteration involves the following steps:

- determine the optimal value $V_T[x] = \max_a R(a|x)$ and optimal action $a_T^* = \arg \max_a R(a|x)$ for each state x at time T ;
- determine optimal values $V_{T-1}[x] = \max_a \{R(a|x) + \lambda \sum_{x'} P(x'|x, a) V_T[x']\}$ and optimal actions $a_{T-1}^*(x) = \arg \max_a \left\{ R(a|x) + \lambda \sum_{x'} P(x'|x, a) V_T[x'] \right\}$ for each state at time $T-1$; and
- determine $V_t[x] = \max_a \{R(a|x) + \lambda \sum_{x'} P(x'|x, a) V_{t+1}[x']\}$ in reverse sequence for each time $t = 0, 1, \dots, T-2$.

The final result is a policy that identifies optimal actions and values for all states over the time horizon. This approach, known as value iteration or dynamic programming, helps to alleviate the “curse of dimensionality” that otherwise can defeat attempts to find a solution (Bellman, 1957).

Dynamic programming has been used for a wide range of ecological problems (see, e.g., Marescot et al., 2013; Williams et al., 2002). In most cases, an ecological system is described in terms of Markovian transitions among finitely many observable states, and management actions that influence the transitions over an extended, often indefinite, time horizon. Objectives often optimize combinations of ecological production costs, management costs, and metrics of system status.

5.2 | Solution approaches with partially observable MDPs

The solution of a POMDP consists of the optimal values $V_t[b]$ across a continuous belief space at each time t . With finitely many system states, actions, and observations, all combinations of these factors could be listed for any belief state. However, it is not possible to do so for all the infinitely many belief states in the continuous belief space of a POMDP, and thus not possible to enumerate values over the continuous space. This contrasts with the situation for observable MDPs over a space of finitely many states and requires a substantially different method.

A standard approach with POMDPs takes advantage of the fact that only finitely many policy trees are needed at any given time to define an optimal policy across the belief space (Smallwood & Sondik, 1973). Each tree defines a linear function, and optimization over the linear functions partitions the belief space into a finite number of segments such that optimal values are produced with the

same linear function for all belief states in a given segment. One consequence is an optimal value function that is piecewise linear over the belief space (see, e.g., Figure 6). The vectors defining the piecewise linearity are called α -vectors, and those for a particular time t are denoted in aggregate by L_t .

By working inductively from the terminal time, it is possible to derive α -vectors (and their partitioning of belief space) at each time t , on the basis of previously identified α -vectors. The procedure for doing so begins at the terminal time T , where the optimal terminal value for belief state b is the maximum of $\sum_x b(x)R(a|x)$ for the possible actions a . The α -vectors for terminal time consist of the return vectors $R(a)$ with components $R(a|x)$, $x \in X$ that produce a maximum average return for at least one belief state:

$$\begin{aligned} V_T[b] &= \sum_x b(x)\alpha(x) \\ &= \max_{a \in A} \sum_x b(x)R(a|x). \end{aligned}$$

Maximization leads to a partition of the belief space into segments, such that the same action (and vector of returns) is optimal for all belief states in a segment (Figure 6). The set of all α -vectors for terminal time T is denoted by L_T .

Building on L_T , an inductive argument for time $t \leq T - 1$ utilizes previously identified α -vectors for stage $t + 1$ to construct the α -vectors for t . With the form in Equation (8), $V_t[b]$ can be written as

$$V_t[b] = \sum_x b(x) \left\{ R(a^*|x) + \lambda \sum_{o'} P(o' | x, a^*) \max_{\alpha \in L_{t+1}} \sum_{x'} b_{\alpha, o'}(x') \alpha(x') \right\},$$

which allows one to identify for a belief state b the α -vector at time t with components

$$R(a^*|x) + \lambda \sum_{o'} P(o' | x, a^*) \max_{\alpha \in L_{t+1}} \sum_{x'} b_{\alpha, o'}(x') \alpha(x').$$

Operationally, the inductive task is to find α -vectors in L_T at the terminal time T as described above, then use L_T to find the α -vectors in L_{T-1} for time $T - 1$, then use L_{T-1} to find the α -vectors in L_{T-2} for time $T - 2$, and so on to the beginning of the timeframe.

Because an α -vector can be constructed as above for any belief state, the challenge at each time becomes one of selecting a limited number of belief states that will produce all the α -vectors needed to define $V_t[b]$ over the whole belief space. Most approaches to exact solutions for POMDPs are distinguished by the method of finding a set of belief states that will produce all the α -vectors. Two general approaches (Cassandra, 1994) are:

- at each time generate a superset L_t^+ of vectors that includes the set L_t of α -vectors, then reduce L_t^+ to L_t (e.g., Cassandra et al., 1997; Monahan, 1982; Zhang & Liu, 1997); and
- at each time create subsets L_t^- of vectors that approximate the optimal value function, then grow the sets while eliminating dominated vectors to get L_t (e.g., Cheng, 1988; Kaelbling et al., 1998).

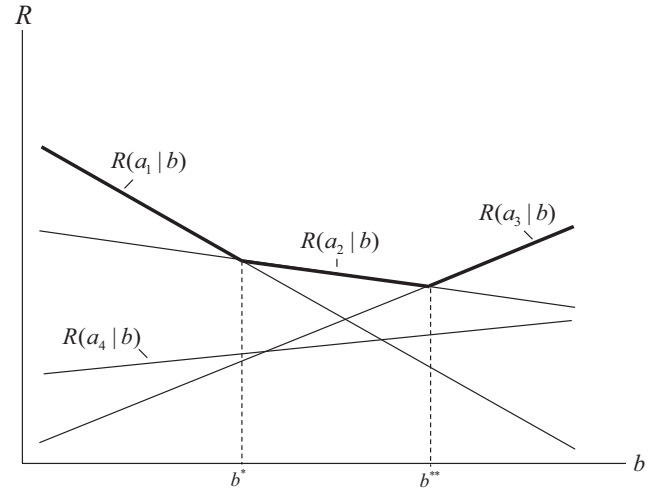
Optimal Policy for Time T 

FIGURE 6 Value functions for terminal time T , with 2 states, 4 actions, and belief state $(b, 1 - b)$. Each action generates a different return function $R(a|b) = bR_1(a) + (1 - b)R_2(a)$. Partitioning of belief space into 3 segments and the optimal actions for each are determined by which return function produces the largest value at each belief state. Optimal value function is indicated by darkened line segments.

In large part, methods for finding exact POMDP solutions do not scale well, and are tractable only for fairly small problems over a limited time (Littman, 2009). Fortunately, some ecological problems can be framed in ways that make them amenable to exact solutions. For larger problems, approximation methods that limit the search for optimal valuation are required (see Discussion).

6 | INFINITE TIME HORIZONS

In this section, we extend the time horizon to allow for decision making over an unlimited amount of time. This is an important consideration because many problems are framed in terms of decision making that can sustain ecological systems indefinitely. Here we describe policy valuation that at any given time is based on expected values that accumulate over infinitely many future time steps. We show how policy and value differ between observable and partially observable MDPs with infinite time horizons.

The development thus far has been based on a time horizon $\{0, 1, \dots, T\}$ with a known and finite terminal time T . Because conservation is so often framed in terms of sustaining ecological systems into the indefinite future, it is useful to consider management that continues over infinitely many decision periods, and identify steady-state management policies that sustain resources indefinitely. In our waterfowl harvest example, we may wish to consider harvest strategies over an indefinitely long time horizon. With full observability and time discounting, the value function $V(x_0) = E[R(a_0|x_0) + \sum_{t=1}^{\infty} \lambda^t R(a_t|x_t)]$ has finite values, so optimal policies and values can be identified. Under partial observability optimal valuation can be approximated, and possibly

determined exactly, depending on the structure of the harvest problem.

6.1 | Infinite time horizon for observable MDPs

Optimal valuation for an observable process with infinitely many time steps can be obtained with a stationary policy consisting of state-specific actions that are invariant to the time at which they are taken (Howard, 1960; Puterman, 1994). Let $\pi = [a(x)]$ represent such a policy, where the same action $a(x)$ is taken for state x irrespective of the time of its occurrence.

A process with stationary policy π can be represented in matrix form by a return vector $R_\pi = [r(a(x))]$ and a matrix $P_\pi = [P(x'|x, a(x))]$ of action-specific transition probabilities. Optimal valuation is given in matrix form by

$$V_{\pi^*} = \max_{\pi} (I - \lambda P_\pi)^{-1} R_\pi, \quad (10)$$

with a corresponding optimal policy

$$\pi^* = \arg \max_{\pi} (I - \lambda P_\pi)^{-1} R_\pi$$

(see Appendix S1). A straightforward procedure for identifying optimal values and policies starts with the selection of an arbitrary policy π to approximate π^* , followed by the determination of values $V_\pi(x)$ by

$$V_\pi = (I - \lambda P_\pi)^{-1} R_\pi.$$

The values V_π then are used to identify a new policy π' by

$$\pi' = \arg \max_{\pi} \{R_\pi + \lambda P_\pi V_\pi\},$$

and the new policy is used in turn to determine new values

$$V_{\pi'} = (I - \lambda P_{\pi'})^{-1} R_{\pi'}.$$

Under mild conditions, recursive policy approximation and value determination can be shown to converge to π^* and V_{π^*} , irrespective of the initial policy choice (Howard, 1960; Ross, 1970).

6.2 | Infinite time horizon for partially observable MDPs

Value iteration for POMDPs, in which the α - vectors for one time are used to find α - vectors for the immediately preceding time, can be used to approximate, and sometimes identify, optimal policies and value functions for infinite time horizons (Poupart, 2005). Repeated value iteration produces values (and policies) that begin to converge, as increasingly discounted values for later rewards add less and less to the accumulated value. That is, the longer the duration of the

system process, the smaller the difference between successive valuations, and the closer the value function gets to a stationary value function and policy (Cassandra, 1994).

In some but not all cases, the optimal value function for infinitely many time steps can be determined exactly in a limited number of steps, and described as a piecewise convex function with a limited set of α - vectors (Hansen, 1998; Sondik, 1978). In other cases, value iteration converges to the infinite horizon optimal value function only in the limit as the number of time steps increases without bound. For this situation the optimal value function will be convex in b , but not necessarily piecewise linear (Kaelbling et al., 1998; White & Harrington, 1980). In the latter case, repeated value iteration provides an approximation of the optimal infinite horizon value function, but the approximation can be arbitrarily close with enough iterations (Sawaki & Ichiwaka, 1978; Sondik, 1978).

7 | MIXED OBSERVABILITY

In this section, we describe mixed observability models for situations in which only some state variables are observable. This is especially important in ecology because ecological systems often include both observable and unobservable attributes, and both can be important in ecological assessment and management. Here we develop adaptive management in the context of mixed observability, and further extend adaptive decision making to include nonstationarity over time.

It may be that some state variables in a system are observable and some are not. For example, the management of a nature preserve might involve conserving a threatened species that is not observable, and managing its wetland habitats that are. It is useful to account for such a mixture of observability conditions in designing management strategies.

Thus, consider a framework for a POMDP in which the system is characterized by two states (x, y) with process transition probabilities $P(x', y' | x, y, a)$ and observations $o = (o_x, o_y)$ with observation probabilities $f(o'_x, o'_y | x', y', a)$. Assuming x and y are discrete with dimensions n_1 and n_2 , one can treat this problem as a classical POMDP of dimension $n = n_1 \times n_2$. The process probabilities

$$P(x', y', o'_x, o'_y | x, y, a) = P(x', y' | x, y, a) f(o'_x, o'_y | x', y', a)$$

can be used for valuation as described above.

This framework can be used to define a mixed observability MDP or MOMDP (Araya-Lopez et al., 2010; Ong et al., 2010), in which the system state is separated into observable states x and unobservable states y . The observation probabilities for known states are given by

$$f(o'_x | x', y', a, o'_y) = \begin{cases} 1 & \text{if } o'_x = x' \\ 0 & \text{otherwise,} \end{cases}$$

that is, observation o'_x coincides with the state x' . On the other hand, observation o'_y is stochastically related to the unobservable state y' by

$$f(o'_y | x', y', a, o'_x) = f(o'_y | x', y', a, x).$$

Assuming observations for y are not influenced by x , the transition probabilities are

$$P(x', y', o'_y | x, y, a) = P(x', y' | x, y, a) f(o'_y | y', a)$$

with belief updates

$$b(y') = \frac{P(y' | x, x', o'_y, b_y, a) P(x', y', o'_y | x, b_y, a)}{\sum_{y'} P(x', y', o'_y | x, b_y, a)}.$$

In the absence of an unobservable state y , a MOMDP problem is seen to reduce to an observable MDP, for which the system state x is observed (Figure 2). Alternatively, in the absence of an observable state x the problem reduces to a POMDP in state y , with an observation function $f(o'_y | y', a)$ (Figure 3). An important effect of factorization into observable and unobservable components is to reduce the dimensionality of the belief state space, which in turn reduces the computation time for finding solutions with POMDP solvers (Nicol et al., 2015).

7.1 | MOMDPs and adaptive management

The MOMDP framework can be applied to adaptive management problems, which involve structurally uncertain systems and the reduction of structural uncertainty about system processes through management actions. Adaptive management is commonly described in terms of observable MDPs for which there is uncertainty about the transition structure or its parameters (Walters, 1986; Williams, 2009). For example, system dynamics may be characterized by one of several models, with uncertainty as to which is the most appropriate. Alternatively, there may be an accepted model but uncertainty about one of more model parameters, such as a population model with uncertain survival or reproduction rates. In either case, state transitions can be characterized with transition probabilities $P(x' | x, a, y)$, where y denotes a particular model (or parameter value) and process uncertainty is expressed in terms a belief state b_y over a discrete space of models or parameters (Williams, 2011).

This situation can be treated as a special case of a MOMDP, in which x represents the observable system state and y represents the unknown model or parameter value. When the process model is only partially observable and the system state is known, the decision process is sometimes called a hidden model MDP or hmMDP (Chadès et al., 2014; Pozzi et al., 2017).

In many adaptive management applications, the true process is held to be stationary over time and included in the model or parameter set. Monitoring of system status over time is assumed to

reveal the actual state x at each monitoring event, with no other observations to inform b_y besides the sequential monitoring of system status. In this situation, valuation becomes

$$V_{\pi_t}(x, b_y) = R(a | x, b_y) + \sum_{x'} P(x' | x, a, b_y) V_{\pi_{t+1}}(x', b'_y) \quad (11)$$

with optimal valuation

$$V_t[x, b_y] = \max_a \left\{ R(a | x, b_y) + \sum_{x'} P(x' | x, a, b_y) V_{t+1}[x', b'_y] \right\},$$

where

$$R(a | x, b_y) = \sum_y b(y) R(a | x, y),$$

$$P(x' | x, a, b_y) = \sum_y b(y) P(x' | x, a, y)$$

and

$$b'_y \propto \sum_y P(x' | x, a, y) b(y)$$

(Williams, 2011). Like POMDPs in general, this problem is PSPACE-complete over finite horizons (Chadès et al., 2014), and thus is difficult to solve for any but small problems.

7.2 | Nonstationary models

A useful generalization of hidden model MDPs allows for nonstationarity in the model structure, such that the true model (or parameter) is itself subject to change through time. For example, climate change can produce such nonstationarity, as climate trends alter system dynamics over time. Pollution, habitat fragmentation, disturbances and other factors can similarly affect ecological processes and lead to nonstationary dynamics.

Nonstationarity can be incorporated by allowing for the model structure to change through time as environmental and other factors change. One approach is to model the structural change (Nicol et al., 2015), by characterizing a change from a model (or parameter) y to y' by transition probabilities $P(y' | y)$ and including the probabilities as an added source of change along with the state dynamics. An intuitive expression that includes both sources of change consists of the probabilities

$$P(x', y' | x, a, y) = P(y' | y) P_y(x' | x, a),$$

where state transitions from x to x' are based on model y' once a model change occurs with probability $P(y' | y)$. Because there are two sources of structural uncertainty in this expression, namely model uncertainty for the prior and posterior models, it is necessary to account for both in valuation:

$$V_{\pi_t}(x | y, y') = R_y(a | x) + \sum_{x'} P_{y'}(x' | x, a) V_{\pi_{t+1}}(x' | y').$$

TABLE 2 Immediate return for conservation action a given state x

		Action		
		a_1 (preserve)	a_2 (alter habitat)	a_3 (trap and remove)
State	x_1 (low)	$R(a_1 x_1) = 14.5$	$R(a_2 x_1) = 12.0$	$R(a_3 x_1) = 10.0$
	x_2 (high)	$R(a_1 x_2) = 5.0$	$R(a_2 x_2) = 7.5$	$R(a_3 x_2) = 5.5$

Letting $P(y'|y) = b_y(y')$ and $b_y = [b(y'|y)]$, the average value over the models y' is

$$V_{\pi_t}(x|y, b_y) = R_y(a|x) + \sum_{x'} P(x'|x, a, b_y) V_{\pi_{t+1}}(x'|y, b'_y)$$

where

$$b'_y(y') = \frac{b_y(y')P_{y'}(x'|x, a)}{P(x'|x, a, b_y)}$$

with $P(x'|x, a, b_y) = \sum_{y'} b_y(y')P_{y'}(x'|x, a)$ and $b'_y = [b'_y(y')]$ (see Appendix S1). A second averaging over the models y produces

$$V_{\pi_t}(x|b, b_y) = R(a|b) + \sum_{x'} P(x'|x, a, b_y, b) V_{\pi_{t+1}}(x'|b', b'_y) \quad (12)$$

where

$$b'(y) = \frac{b(y)P(x'|x, a, b_y)}{P(x'|x, a, b_y, b)}$$

with $P(x'|x, a, b_y, b) = \sum_y b(y)P(x'|x, a, b_y)$ and $b' = [b'(y)]$ (see Appendix S1).

Equation (12) can be seen as a generalization of Equation (11) for valuation under stationarity; if $P(y'|y) = b_y(y')$ is eliminated, Equation (12) reduces to valuation under stationarity as in Equation (11).

Mixed observability models offer opportunities to account for multiple uncertainty factors in ecological assessment and management, especially under current conditions of rapid environmental change due to climate change and other factors. In particular, there is real potential for advances in learning-based adaptive management under nonstationary conditions. Additional features for consideration include the incorporation of partially observable states as well as system models (Fackler & Pacifici, 2014), and autocorrelations in trajectories of model structure over time (Memarzadeh et al., 2019).

8 | CONTINUOUS STATES

In this section, we address the complexity added in POMDPs with a continuous state space. Although much of the modeling and analysis of POMDPs is based on an assumption that state variables range over discrete values, many ecological problems focus on states such as density rate and size, which can vary over a continuous range of

values. Such a situation presents serious difficulties in formulating and evaluating policies under partial observability. We describe approaches for policy valuation under these conditions.

The restriction to discrete and finite states and observations clearly limits the range of ecological applications for POMDPs, since many ecological problems involve continuous state variables for which the solution methods for discrete decision processes are not applicable (Zhou et al., 2010). For example, our waterfowl harvest problem may be described in terms of continuous rather than discrete population status, where the population is modeled as a continuous Markov process with transitions from states over a continuous range to other states in that range. A different approach must be used to assess such a problem.

A key issue in the propagation and updating of a continuous belief state is that posterior belief states typically do not have the same functional form as the prior belief states. A possible solution is to approximate a continuous-state POMDP with one over a discretized state space, and use the optimal policy for the resulting discrete-state POMDP as a proxy for the continuous process (Hauskrecht, 2000; Zhou & Hansen, 2001). Other approaches involve gradient ascent (Meuleau et al., 1999; Ng & Jordan, 2000), neural networks (Bertsekas & Tsitsiklis, 1996; Sallans, 2000), and Monte Carlo simulation (Brooks & Williams, 2010; Thrun, 1999).

A promising new approach for handling continuous-state POMDPs is "density projection," so named because it involves the projection of belief states onto a set of parametrically defined probability distributions. With density projection, the belief states share a common functional form, and thus can be characterized by their parameters rather than by the probability masses for individual system states. Though Bayesian updating produces a posterior belief state that differs in form from its prior, the posterior is approximated with a proxy that is close to it and in the same family as the prior belief state.

The practical challenge of finding the best approximation for a posterior belief is achieved in density projection by identifying distribution parameters of the proxy that minimize the Kullback–Leibler divergence between the true and proxy distributions (Zhou et al., 2010). Zhou et al. (2010) show that for distributions in the exponential family, minimization of Kullback–Leibler divergence is obtained by matching the sufficient statistics of the true and approximate distributions. With the additional step of discretizing the parameter space and using a nearest-neighbor approach to represent transitions between discrete parameter values, one can use solution approaches for discrete-state POMDPs to find approximate solutions to the continuous-time MDP (see Appendix S1).

By allowing continuous belief states to be characterized by probability density function parameters taking only a limited number of values, density projection goes a long way toward addressing the curse of dimensionality and expands dramatically the range of POMDP applications. The approach has been used to address structural uncertainty (Springborn & Sanchirico, 2013) as well as partial observability, where it was first applied informally to wildlife management by Moore (2008). Since then, there have been a number of biological examples (see Table 1 for examples).

9 | EXAMPLES

In this section, we use simple examples involving control of a nuisance species to show how POMDPs build upon the framework and calculations for observable MDPs and produce piece-wise linear optimal valuations.

9.1 | Observable MDP example

To illustrate assessment of an observable MDP, consider a simple problem of controlling the abundance of a nuisance animal species, involving two states (x_1 for low abundance, x_2 for high abundance); three potential actions (no investment in conservation (a_1), temporary habitat alteration (a_2), and trapping and removal of animals (a_3)); and a model describing the consequences of these actions on the population status. The transition probabilities for each action are

$$P(a_1) = [P(x' | x, a_1)] = \begin{bmatrix} .3 & .7 \\ 0 & 1 \end{bmatrix} \begin{array}{cc} \text{low} \Rightarrow \text{low} & \text{low} \Rightarrow \text{high} \\ \text{high} \Rightarrow \text{low} & \text{high} \Rightarrow \text{high} \end{array}$$

$$P(a_2) = [P(x' | x, a_2)] = \begin{bmatrix} .8 & .2 \\ .3 & .7 \end{bmatrix}$$

$$P(a_3) = [P(x' | x, a_3)] = \begin{bmatrix} .6 & .4 \\ .8 & .2 \end{bmatrix}$$

Some patterns are noteworthy. In the absence of any conservation action, there is a high probability of transition from low to high abundance, but no chance of transition from high to low abundance. Habitat alteration produces smaller probabilities of transition from high to low abundance than trapping. And there are substantial

TABLE 3 Optimal time-specific values and conservation actions for state x

	Time		
	T-2	T-1	T
State x_1 (low)	$V_t[x_1] = 32.2$; $a^* = a_2$	$V_t[x_1] = 23.8$; $a^* = a_2$	$V_t[x_1] = 14.5$; $a^* = a_1$
State x_2 (high)	$V_t[x_2] = 25.8$; $a^* = a_3$	$V_t[x_2] = 17.3$; $a^* = a_3$	$V_t[x_2] = 7.5$; $a^* = a_2$

probabilities that high abundance will remain unchanged even when a conservation action is undertaken.

Returns for this problem include immediate costs and benefits of conservation actions, as well as social perceptions about the appropriateness of an action. It is assumed that the cost of trapping is greater than that of temporary habitat alteration, that positive values accrue to both the reduction of abundance and the retention of low abundance, and that social perceptions and values vary with costs, success, and the type of action taken. The average return when action a is taken in state x is shown in Table 2.

It is easy to see that at terminal time T the optimal value for a low population is $V_T[x_1] = \max_a R(a|x_1) = 14.5$ with optimal action $a^* = a_1$. For a large population the optimal value is $V_T[x_2] = \max_a R(a|x_2) = 7.5$ with optimal action $a^* = a_2$.

At time $T-1$ optimal valuation with discount factor $\lambda = 0.9$ is given by

$$V_{T-1}[x] = \max_a \left\{ R(a|x) + 0.9 \sum_{x'} P(x' | x, a) V_T[x'] \right\},$$

with optimal value.

$$V_{T-1}[x_1] = \max\{23.1, 23.8, 22.5\} = 23.8 \text{ for } a^* = a_2$$

for state x_1 and

$$V_{T-1}[x_2] = \max\{11.8, 16.1, 17.3\} = 17.3 \text{ for } a^* = a_3$$

for state x_2 . At time $T-2$ optimal valuation is given by

$$V_{T-2}[x] = \max_a \left\{ R(a|x) + 0.9 \sum_{x'} P(x' | x, a) V_{T-1}[x'] \right\},$$

with

$$V_{T-2}[x_1] = \max\{31.8, 32.2, 29.1\} = 32.2 \text{ for } a^* = a_2$$

and

$$V_{T-2}[x_2] = \max\{18.3, 24.8, 25.8\} = 25.8 \text{ for } a^* = a_3$$

A summary of the optimal strategy and valuation for three time steps is shown in Table 3.

Backward recursion beyond $T-2$ generates a stationary policy $\pi = [a_2, a_3]$ with habitat conservation (a_2) for a small population and removal (a_3) for a large population. These actions attempt to maintain the size of a small population and reduce the size of a large population over indefinitely many time steps. From Equation (10), the state-specific optimal values for an infinite time horizon are $V[x_1] = 126.5$ and $V[x_2] = 130.1$.

9.2 | Partially observable MDP example

An observable MDP can be extended to create a POMDP by allowing for partial observability with an observation function. For example,

three possible observations, $o_1, o_2,$ and o_3 (for, e.g., observed population counts that are low, medium, or high) might be associated with state-specific probabilities (Table 4):

The observation probabilities combine with Markov transitions to define the POMDP transitions $P(o', x' | x, a) = f(o' | x')P(x' | x, a)$. With only two states, the belief state at any time can be described by a vector with a scalar value b for state x_1 and $(1-b)$ for state x_2 .

To illustrate optimal decision making with a POMDP, we again consider two states but allow a fourth action, for example, a combination of habitat alteration and removal. At terminal time T , there

are no future values to consider, so the optimal value function for a given belief state is the maximum of the linear functions

$$V_T(a|b) = bR(a|x_1) + (1 - b)R(a|x_2)$$

where action a can be a_1, a_2, a_3 or a_4 . Figure 6 displays four lines corresponding to value functions for the actions over the belief space $[0,1]$. Optimization over the actions partitions the belief space $[0,1]$ into three segments that are defined by the intersections of three of the four lines (the function $V_T(a_4|b)$ is dominated over $[0,1]$, and thus is not needed to describe the optimal value function). The figure makes clear that optimization produces a convex optimal value function $V_T[b]$ that is piecewise linear in b . Thus, $V_T[b]$ is given by $V_T(a_1|b)$ for belief states less than b^* ; by $V_T(a_3|b)$ for belief states greater than b^{**} ; and by $V_T(a_2|b)$ for belief states between b^* and b^{**} .

The return vectors for the three value functions defining the optimal value function constitute the α -vectors for time T , with an α -vector corresponding to each of the three partition segments. With more actions the number of intersections tends to increase, so the number of segments in the partition of $[0,1]$ and the number of α -vectors does as well. Countering this tendency is the fact that more dominated lines typically occur, which tends to reduce the count of α -vectors.

At time $T-1$, the optimal value $V_{T-1}[b]$ is produced with the algorithm for Equation (7) in the following steps:

1. for each action a_{T-1} and combination (o', a_T) , transform the return vector with components $R(a_T|x)$ into a vector with component $V'(a_T | x, a_{T-1}, o') = \sum_{x'} P(o', x' | x, a_{T-1}) R(a_T | x')$;
2. maximize $V'(a_T | b, a_{T-1}, o') = \sum_x b(x) V'(a_T | x, a_{T-1}, o')$ over the actions a_T ;

TABLE 4 Probabilities corresponding to observation a for a given state x

	Observation		
	o_1 (low)	o_2 (medium)	o_3 (high)
State x_1	$f(o_1 x_1) = 0.1$	$f(o_2 x_1) = 0.6$	$f(o_3 x_1) = 0.3$
State x_2	$f(o_1 x_2) = 0.5$	$f(o_2 x_2) = 0.4$	$f(o_3 x_2) = 0.1$

TABLE 5 Immediate returns for two actions, given two states. $R(a|b)$ corresponds to returns averaged over belief state b

	State		$R(a b)$
	$x = x_1$	$x = x_2$	
Action a_1	$R(a_1 x_1) = 2.3$	$R(a_1 x_2) = 7.9$	$R(a_1 b) = 7.9 - 5.6b$
Action a_2	$R(a_2 x_1) = 8.1$	$R(a_2 x_2) = 2.5$	$R(a_2 b) = 2.5 + 5.6b$

Optimal Policy for $a_{T-1} = a_1$

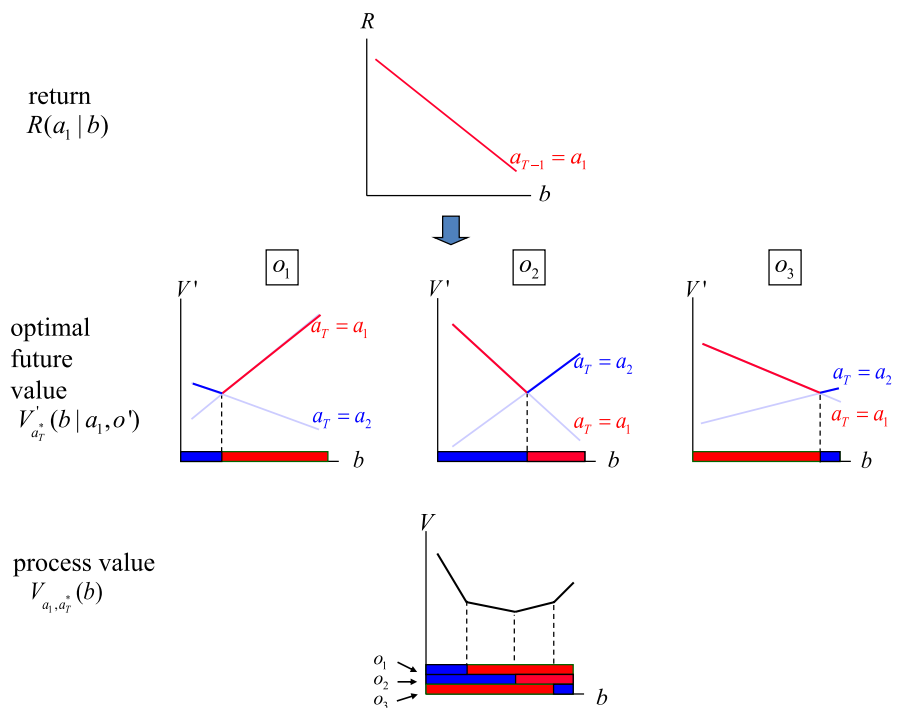


FIGURE 7 Valuation at time $T-1$ for a policy tree with root action a_1 and optimal sub-policies thereafter. Graphs display (i) immediate returns $R(a_1|b)$; (ii) backcast values $V'(a_T|b, a_1, o')$ for each observation, along with partition segment cutpoints; and (iii) the accumulation of immediate returns and optimal backcast values over observations to get $V_{a_1, a_T^*}(b)$.

Optimal Policy for $a_{T-1} = a_2$

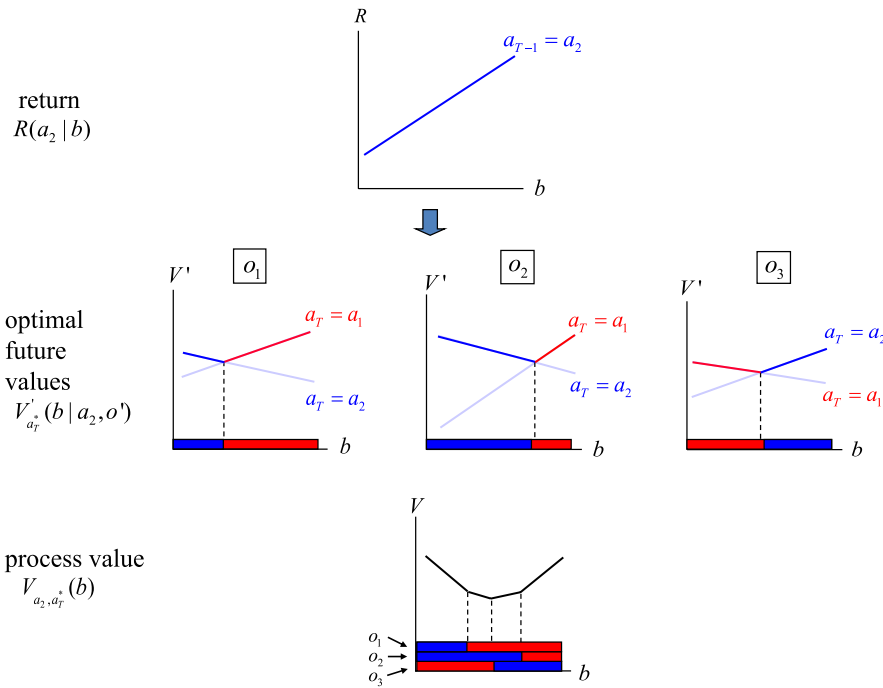


FIGURE 8 Valuation at time $T-1$ for a policy tree with root action a_2 and optimal sub-policies thereafter. Graphs display (i) immediate returns $R(a_2|b)$; (ii) backcast values $V^*(a_T|b, a_2, o')$ for each observation, along with partition segment cutpoints; and (iii) the accumulation of immediate returns and optimal backcast values over observations to get $V_{a_2, \pi_T^*}(b)$.

TABLE 6 Values $V_{a_T}^*(b | a_{T-1}, o')$ at time $T-1$ for actions (a_{T-1}, a_T) and observation o' following a_{T-1}

		Observation		
		$o' = o_1$	$o' = o_1$	$o' = o_1$
Action $a_{T-1} = a_1$	$a_T = a_1$	$2.8 + 5.2b$	$7.9 - 6.8b$	$7.5 - 3.4b$
	$a_T = a_2$	$4.8 - 3.1b$	$0.2 + 6.4b$	$2.5 + 2.1b$
Action $a_{T-1} = a_2$	$a_T = a_1$	$4.3 + 3.8b$	$0.3 + 7.7b$	$6.2 - 2.4b$
	$a_T = a_2$	$6.2 - 2.3b$	$7.9 - 2.4b$	$3.9 + 2.1b$

- accumulate the results of step 2 over all observations o' and add the immediate return $R(a_{T-1}|b)$; and
- maximize the result of step 3 over the actions a_{T-1} to get

$$V_{T-1}[b] = \max_{a_{T-1}} \left\{ R(a_{T-1}|b) + \lambda \sum_{o'} \max_{a_T} V^*(a_T | b, a_{T-1}, o') \right\}.$$

Though the arithmetic in these steps can be tedious, the computations are actually simple. Because the functions $V^*(a_T | b, a_{T-1}, o')$ are simply lines in two dimensions, the solution of the optimization simplifies to a piecewise linear value function in two dimensions.

For illustrative purposes consider only two actions a_1 and a_2 , with immediate and average returns shown in Table 5 (also see Figures 7 and 8).

For each action a_{T-1} and observation o' , the returns can be transformed with the probabilities $P(x', o' | x, a_{T-1})$ as indicated in the Appendix S1, to produce linear functions $V_{a_T}^*(b | a_{T-1}, o')$ shown in Table 6.

Conditional on action a_{T-1} and each observation o' , optimal values for time T are then obtained by optimizing $V_{a_T}^*(b | a_{T-1}, o')$ over a_T (Figures 7 and 8), and a subsequent optimization over the actions at

$T-1$ identifies the optimal value function and final partition of belief space (Figure 9).

The optimal partition of belief space $[0,1]$ shown in Figure 9 includes several segments, with the same optimal policy for all belief states in a segment. The number of segments defined by the optimizations can be expected to increase with the number of potential actions.

For time t prior to $T-1$, the optimal value function for a general time identifies the maximum accumulated returns over the remaining time horizon for each belief state b starting at time t . Thus, the value function

$$V_t[b] = \max_{a_t} \left\{ R(a_t | b) + \lambda \sum_{o'} P(o' | b, a_t) V_{\pi_{t+1|a_t, o'}}^*(b_{a_t, o'}) \right\} \\ = \max_{a_t} \left\{ R(a_t | b) + \lambda \sum_{o'} P(o' | b, a_t) V[b_{a_t, o'}] \right\}$$

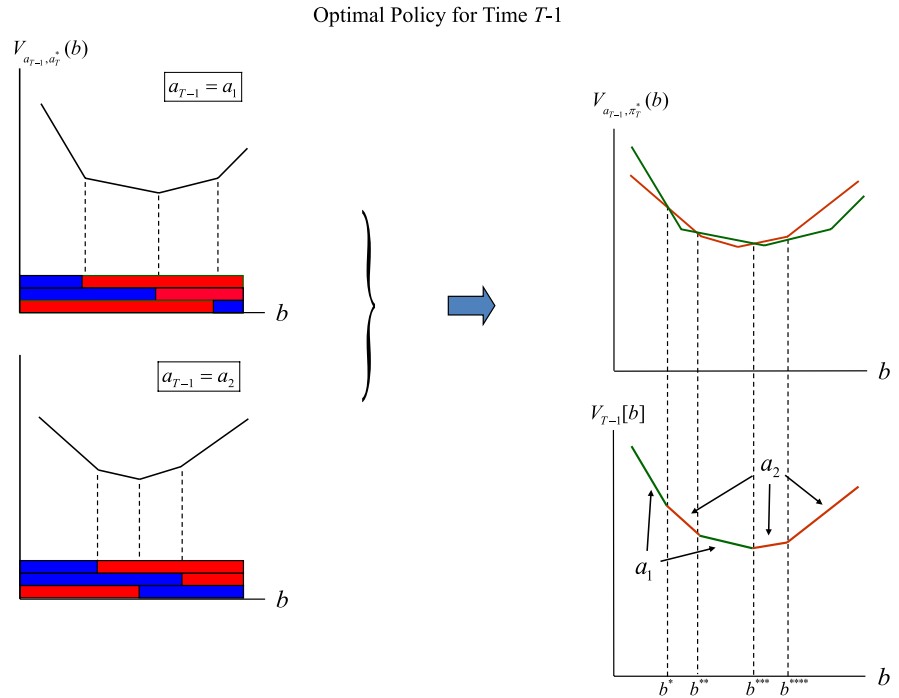
is optimized by a two-step procedure to get $V_{t+1}[b] = \max_{\pi_{t+1}} V_{\pi_{t+1}}(b)$ at time $t+1$, followed by a second optimization over the actions a_t . The solution gives an optimal action a_t^* and associated optimal value $V_t[b]$ for each belief state b for each time.

The identification of optimal values and policies in the foregoing invasive species problem is greatly simplified by the small number of population sizes, actions and observations. However, even with this simplification the number of segments defined by the optimizations can become exponentially large as the duration of the process is extended.

10 | DISCUSSION

We have focused on partially observable Markov decision processes in the context of managing and monitoring ecological systems, when

FIGURE 9 Combining the value functions $V_{a_1, \pi_T^*}(b)$ and $V_{a_2, \pi_T^*}(b)$ to produce optimal valuation $V_{T-1}[b]$ for time $T-1$. Partitioning of belief space is determined by the time T partitions for $V_{a_1, \pi_T^*}(b)$ and $V_{a_2, \pi_T^*}(b)$, and the intersection points of the 2 functions. The optimal action for belief states in each partition segment is determined by which of the 2 value functions produces the larger value.



there is only limited understanding of ecological status. Markov transitions are usually assumed to occur in a discrete state space, with controls that influence both rewards and transitions among states. All aspects of the control problem must be adapted to partial observability, including state transitions, valuation, and the tracking of system status by means of belief states. These features add considerable complexity, in large part because of the expansion of a discrete state space under complete observability into a continuous state belief space under partial observability.

A technical treatment of partial observability with POMDPs is rarely undertaken in ecology and ecological assessments, despite the almost universal presence of uncertainty about a system's status. In fact, a POMDP framework is applicable across a broad spectrum of ecological problems involving populations, communities, ecosystems, and habitats. It also can be applied naturally to decision making about monitoring protocols and programs, by including actions in the observation function $f(o'|x',a)$ that allow a manager to address whether, when and how to conduct monitoring so as to maximize conservation value.

Several factors contribute to the limited use of POMDPs in ecology and ecological management. Challenges include the complexity of the POMDP framework and the notation needed to characterize it; difficulties in interpreting solutions for all but very simple problems; the inability to scale up exact methods to problems with large numbers of states and lengthy time horizons; and importantly, the lack of explanatory documentation and examples that can help potential users (Chadès et al., 2021).

All combinations of finitely many states, actions, and observations can be listed for any belief state in a POMDP. However, it is not possible to do so for all the infinitely many belief states in the continuous belief space of a POMDP, and thus it is not possible to enumerate values over the continuous space. Most approaches for solving POMDPs

utilize the piecewise linear structure of the optimal value function, which allows the partitioning of belief space into segments and the use of a single linear function to produce optimal values for all belief states in a given segment. The challenge is then to identify the partition segments and associated linear functions for each time step.

Numerous solution methods have been formulated for POMDPs, each with its own advantages and limitations. Several approaches, such as the witness algorithm (Kaelbling et al., 1998; Littman, 1996) and incremental pruning (Cassandra et al., 1997; Zhang & Liu, 1997), produce exact solutions, but scale poorly and generally can be used for only a limited class of small problems. Ad hoc procedures (e.g., use of observation moments as if they are actual system states, gridding of belief space and valuation at grid points to approximate $V[b]$) are relatively straightforward, but may perform poorly even for small problems (Cassandra, 1994). Point-based value iteration (Pineau et al., 2006; Spaan & Vlassis, 2005), a popular approach that approximates the value function with a limited number of systematically identified belief states, has become increasingly available via recent web applications (Pascal et al., 2020). Outstanding issues are the range and density of the belief states that are included, and convergence rates and costs of the approach with increasing scale.

There are some key assumptions underlying POMDPs that limit their use. One is that transitions among states are Markovian, which restricts the usefulness of POMDPs to ecological systems not exhibiting hysteresis and other lags in resource processes and valuations. Another is that the sets X, A , and O of process states, actions, and observations are assumed to be finite. One approach for problems with continuous actions and observations is to discretize their range of values (Nicol & Chadès, 2012), but the solutions produced may be sensitive to the discretization rules. Another uses density projection to approximate solutions, as described earlier.

Additional assumptions are that the structure of the ecological system is fixed and fully known. Structural uncertainty can be accommodated in a POMDP framework as discussed in Section 7.1 (Memarzadeh & Boettiger, 2018; Williams, 2009, 2011), which allows for adaptive learning as management is pursued (Fackler et al., 2014; Peron et al., 2017). Structural nonstationarity can also be modeled in terms of mixed observability, as suggested in Section 7.2. Artificial intelligence shows promise for nonstationary decision processes (Nicol et al., 2015).

For problems that meet the basic assumptions, POMDPs add realism in framing the management of ecological systems, by recognizing that they are almost never observed in their entirety and that sampling produces only stochastic estimators of ecological status (Williams & Brown, 2019). Though relatively few in number, applications of POMDPs in ecology have grown in recent years, as resource analysts and managers increasingly seek to account for uncertainty. Applications are aided by ongoing developments in theory, solution techniques, and computing capacity (e.g., Dujardin et al., 2017), as well as improvements in the display of policy graphs (Ferrer-Mestres et al., 2020, 2021). In particular, finding efficient approaches to approximate optimal solutions for large problems is a rapidly growing area of research. Coupled with advances in the fast-evolving field of ecological sampling and estimation, POMDPs hold considerable promise for more effective ecological management.

AUTHOR CONTRIBUTIONS

Byron Williams: Conceptualization (lead); methodology (lead); writing – original draft (lead); writing – review and editing (lead). **Eleanor Brown:** Conceptualization (supporting); funding acquisition (lead); writing – review and editing (supporting).

ACKNOWLEDGMENTS

We thank the USGS Science and Decisions Center for support for B.K.W. during preparation of this paper. We thank reviewers Iadine Chadès and James Nichols for helpful comments. This material is based on work supported by the U.S. Geological Survey under Cooperative Agreement No. G19AC00146.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

No data were used in the preparation of this manuscript. Therefore, no data are available for access.

ORCID

Byron K. Williams  <https://orcid.org/0000-0001-7644-1396>

Eleanor D. Brown  <https://orcid.org/0000-0001-7798-830X>

REFERENCES

- Araya-López, M., Thomas, V., Buffet, O., & Charpillat, F. (2010). A closer look at MOMDPs. 22nd International Conference on Tools with Artificial Intelligence - ICTAI 2010, 197–204.
- Bellman, R. (1957). *Dynamic programming*. Princeton University Press.
- Bertsekas, D. P. (2012). *Dynamic programming and optimal control* (Vol. II, 4th ed.). Athena Scientific.
- Bertsekas, D. P. (2017). *Dynamic programming and optimal control* (Vol. I, 4th ed.). Athena Scientific.
- Bertsekas, D. P., & Tsitsiklis, J. N. (1996). *Neuro-dynamic programming* (1st ed.). Optimization and Neural Computation Series, Athena Scientific.
- Brooks, A., & Williams, S. (2010). A Monte Carlo update for parametric POMDPs. In *Robotics Research*. Springer. https://doi.org/10.1007/978-3-642-14743-2_19
- Cassandra, A. R. (1994). *Optimal policies for partially observable Markov decision processes*. Technical report CS-94-14. Brown University.
- Cassandra, A. R., Littman, M. L., & Zhang, N. L. (1997). Incremental pruning: A simple fast, exact method for partially observable Markov decision processes. *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*. arXiv preprint arXiv:1302.1525.
- Chadès, I., Carwardine, J., Martin, T., Nicol, S., Sabbadin, R., & Buffet, O. (2014). MOMDPs: A solution for modelling adaptive management problems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1). <https://doi.org/10.1609/aaai.v26i1.8171>
- Chadès, I., McDonald-Madden, E., McCarthy, M. A., Wintle, B., Linkie, M., & Possingham, H. P. (2008). When to stop managing or surveying cryptic threatened species. *Proceedings of the National Academy of Sciences*, 105, 13936–13940.
- Chadès, I., Pascal, L. V., Nichol, S., Fletcher, C. S., & Ferrer-Mestres, J. (2021). A primer on partially observable Markov decision processes (POMDPs). *Methods in Ecology and Evolution*, 12, 2058–2072.
- Cheng, H.-T. (1988). *Algorithms for partially observable Markov decision processes*. PhD Dissertation. University of British Columbia.
- Dujardin, Y., Dietterich, T., & Chadès, I. (2017). Three new algorithms to solve N- POMDPs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), 4495–4501.
- Fackler, P., & Pacifici, K. (2014). Addressing structural and observational uncertainty in resource management. *Journal of Environmental Management*, 133, 27–36.
- Fackler, P. L., & Haight, R. G. (2014). Monitoring as a partially observable decision problem. *Resource and Energy Economics*, 37, 226–241.
- Fackler, P. L., Pacifici, K., Martin, J., & McIntyre, C. (2014). Efficient use of information in adaptive management with an application to managing recreation near golden eagle nesting sites. *PLoS One*, 9, e102434. <https://doi.org/10.1371/journal.pone.0102434>
- Ferrer-Mestres, J., G. Dietterich, T. Buffet, O., and Chadès, I. 2020. Solving K-MDPs. *Proceedings of the International Conference on Automated Planning and Scheduling*30: 110–118.
- Ferrer-Mestres, J., G. Dietterich, T., Buffet, O., and Chadès, I. 2021. K-N-MOMDPs: Towards interpretable solutions for adaptive management. *Proceedings of the AAAI Conference on Artificial Intelligence*35: 14775–14784.
- Haight, R. G., & Polasky, S. (2010). Optimal control of an invasive species with imperfect information about the level of infestation. *Resource and Energy Economics*, 32, 519–533.
- Hansen, E. A. (1998). An improved policy iteration algorithm for partially observable MDPs. *Advances in Neural Information Processing Systems*, 10, 1015–1021.
- Hauskrecht, M. (2000). Value-function approximations for partially observable Markov decision processes. *Journal of Artificial Intelligence Research*, 13, 33–95.
- Howard, R. A. (1960). *Dynamic programming and Markov processes*. Wiley.
- Johnson, F. A., Boomer, G. S., Williams, B. K., Nichols, J. D., & Case, D. J. (2015). Multi-level learning in the adaptive management of waterfowl harvests: 20 years and counting. *Wildlife Society Bulletin*, 39, 9–19.
- Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101, 99–134.

- Kling, D. M., Sanchirico, J. N., & Fackler, P. L. (2017). Optimal monitoring and control under state uncertainty: Application to lionfish management. *Journal of Environmental Economics and Management*, 84, 223–245.
- Lane, D. E. (1989). A partially observable model of decision making by fishermen. *Operations Research*, 37, 240–254.
- Littman, M. L. (1996). *Algorithms for sequential decision making*. PhD Dissertation. Brown University.
- Littman, M. L. (2009). A tutorial on partially observable Markov decision processes. *Journal of Mathematical Psychology*, 53, 119–125.
- MacLachlan, M. J., Springborn, M. R., & Fackler, P. L. (2017). Learning about a moving target in resource management: Optimal Bayesian disease control. *American Journal of Agricultural Economics*, 99, 140–162.
- Madani, O., Hanks, S., & Condon, A. (2003). On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence*, 147, 5–34.
- Marescot, L., Chapron, G., Chadès, I., Fackler, P. L., Duchamp, C., Marboutin, E., & Gimenez, O. (2013). Complex decisions made simple: A primer on stochastic dynamic programming. *Methods in Ecology and Evolution*, 4, 872–884.
- McDonald-Madden, E., Chadès, I., McCarthy, M. A., Linkie, M., & Possingham, H. P. (2011). Allocating conservation resources between areas where persistence of a species is uncertain. *Ecological Applications*, 21, 844–858.
- Memarzadeh, M., & Boettiger, C. (2018). Adaptive management of ecological systems under partial observability. *Biological Conservation*, 224, 9–15.
- Memarzadeh, M., Britten, G. L., Worm, B., & Boettiger, C. (2019). Rebuilding global fisheries under uncertainty. *Proceedings of the National Academy of Sciences*, 116, 15985–15990.
- Meuleau, N., Peshkin, L., Kim, K.-E., & Kaelbling, L. P. (1999). Learning finite-state controllers for partially observable environments. *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*. arXiv preprint arXiv:1301.6721.
- Monahan, G. E. (1982). State of the art - a survey of partially observable Markov decision processes: Theory, models and algorithms. *Management Science*, 28, 1–16.
- Moore, A. L. (2008). *Managing populations in the face of uncertainty: Adaptive management, partial observability and the dynamic value of information*. PhD Dissertation. University of Melbourne.
- Ng, A. Y., & Jordan, M. (2000). PEGASUS: A policy search method for large MDPs and POMDPs. *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*. arXiv preprint arXiv:1301.3878.
- Nichols, J. D., & Williams, B. K. (2006). Monitoring for conservation. *Trends in Ecology and Evolution*, 21, 668–673.
- Nicol, S., & Chadès, I. (2012). Which states matter? An application of an intelligent discretization method to solve a continuous POMDP in conservation biology. *PLoS One*, 7, e28993. <https://doi.org/10.1371/journal.pone.0028993>
- Nicol, S., Fuller, R. A., Iwamura, T., & Chadès, I. (2015). Adapting environmental management to uncertain but inevitable change. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20142984. <https://doi.org/10.1098/rspb.2014.2984>
- Ong, S. C. W., Png, S. W., Hsu, D., & Lee, W. S. (2010). Planning under uncertainty for robotic tasks with mixed observability. *International Journal of Robotics Research*, 29, 1053–1068.
- Papadimitriou, C. H., & Tsitsiklis, J. N. (1987). The complexity of Markov decision processes. *Mathematics of Operations Research*, 12, 441–450.
- Pascal, L., Memarzadeh, M., Boettiger, C., Lloyd, H., & Chadès, I. (2020). A shiny R app to solve the problem of when to stop managing or surveying species under imperfect detection. *Methods in Ecology and Evolution*, 11, 1707–1715. <https://doi.org/10.1111/2041-210X.13501>
- Peron, M., Becker, K. H., Bartlett, P., & Chadès, I. (2017). Fast-tracking stationary MOMDPs for adaptive management problems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.11173>
- Pineau, J., Gordon, G., & Thrun, S. (2006). Anytime point-based approximations for large POMDPs. *Journal of Artificial Intelligence Research*, 27, 335–380.
- Poupart, P. (2005). *Exploiting structure to efficiently solve large scale partially observable Markov decision processes*. PhD Dissertation. University of Toronto.
- Pozzi, M., Memarzadeh, M., & Klima, K. (2017). Hidden-model processes for adaptive management under uncertain climate change. *Journal of Infrastructure Systems*, 23, 4017022.
- Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. Wiley.
- Regan, T. J., Chadès, I., & Possingham, H. (2011). Optimally managing under imperfect detection: A method for plant invasions. *Journal of Applied Ecology*, 48, 76–85.
- Ross, S. M. (1970). *Applied probability models with optimization applications*. Holden-Day.
- Rout, T. M., Moore, J. L., & McCarthy, M. A. (2014). Prevent, search or destroy? A partially observable model for invasive species management. *Journal of Applied Ecology*, 51, 804–813.
- Sallans, B. (2000). Learning factored representations for partially observable Markov decision processes. *Advances in Neural Information Processing Systems*, 12, 1050–1056.
- Sawaki, K., & Ichiwaka, A. (1978). Optimal control of partially observable Markov decision processes over an infinite horizon. *Journal of the Operations Research Society of Japan*, 21, 1–14.
- Sethi, G., Costello, C., Fisher, A., Hanemann, M., & Karp, L. (2005). Fishery management under multiple uncertainty. *Journal of Environmental Economic and Management*, 50, 300–318.
- Sloggy, M. R., Kling, D. M., & Plantinga, A. J. (2020). Measure twice, cut once: Optimal inventory and harvest under volume uncertainty and stochastic price dynamics. *Journal of Environmental Economic and Management*, 103, 102357.
- Smallwood, R. D., & Sondik, E. J. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21, 1071–1088.
- Sondik, E. J. (1978). The optimal control of partially observable Markov decision processes over the infinite horizon: Discounted costs. *Operations Research*, 26, 282–304.
- Spaan, M. T. J., & Vlassis, N. (2005). Perseus: Randomized point-based value iteration for POMDPs. *Journal of Artificial Intelligence Research*, 24, 195–220.
- Springborn, M., & Sanchirico, J. N. (2013). A density projection approach for non-trivial information dynamics: Adaptive management of stochastic natural resources. *Journal of Environmental Economics and Management*, 66, 609–624.
- Thrun, S. (1999). Monte Carlo POMDPs. *Advances in Neural Information Processing Systems*, 12, 1064–1070.
- Tomberlin, D. (2010a). Decision support for environmental monitoring and restoration: Application of the partially observable Markov decision process. *International congress on Environmental Modelling and Software*, 286. <https://scholarsarchive.byu.edu/iemssconference/2010/all/286>
- Tomberlin, D. (2010b). Endangered seabird habitat management as a partially observable Markov decision process. *Marine Resource Economics*, 25, 93–104.
- Uthé, W. T., & Veloso, M. M. (1998). Tree based discretization for continuous state space reinforcement learning. *AAAI/IAAI*, 98, 769–774.
- Walters, C. J. (1986). *Adaptive management of renewable resources*. Blackburn Press.
- White, C. C., & Harrington, D. (1980). Application of Jensen's inequality for adaptive suboptimal design. *Journal of Optimization Theory and Applications*, 32, 89–99.

- Williams, B. K. (2009). Markov decision processes in natural resources management: Observability and uncertainty. *Ecological Modelling*, 220, 830–840.
- Williams, B. K. (2011). Resolving structural uncertainty in natural resources management using POMDP approaches. *Ecological Modelling*, 222, 1092–1102.
- Williams, B. K., & Brown, E. D. (2019). Sampling and analysis frameworks for inference in ecology. *Methods in Ecology and Evolution*, 10, 1832–1842.
- Williams, B. K., & Johnson, F. A. (1995). Adaptive management and the regulation of waterfowl harvests. *Wildlife Society Bulletin*, 23, 430–436.
- Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). *Analysis and management of animal populations*. Academic Press.
- Zhang, N. L., & Liu, W. (1997). A model approximation scheme for planning in partially observable stochastic domains. *Journal of Artificial Intelligence Research*, 7, 199–230.
- Zhou, E., Fu, M. C., & Marcus, S. I. (2010). Solving continuous-state POMDPs via density projection. *IEEE Transactions on Automatic Control*, 55, 1101–1116.
- Zhou, R., & Hansen, E. A. (2001). An improved grid-based approximation algorithm for POMDPs. *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, 1, 707–714.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Williams, B. K., & Brown, E. D. (2022). Partial observability and management of ecological systems. *Ecology and Evolution*, 12, e9197. <https://doi.org/10.1002/ece3.9197>