# MOMDPs: a Solution for Modelling Adaptive Management Problems

Iadine Chadès, Josie Carwardine, Tara Martin, Samuel Nicol, Régis Sabbadin, Olivier Buffet

# MOMDPs: a Solution for Modelling Adaptive Management Problems

**Iadine Chadès** and **Josie Carwardine** and **Tara G. Martin**
CSIRO Ecosystem Sciences
{iadine.chades, josie.carwardine, tara.martin}@csiro.au

**Samuel Nicol**
University of Alaska Fairbanks
scnicol@alaska.edu

**Régis Sabbadin**
INRA
sabbadin@toulouse.inra.fr

**Olivier Buffet**
INRIA / Université de Lorraine
olivier.buffet@loria.fr

## Abstract

In conservation biology and natural resource management, adaptive management is an iterative process of improving management by reducing uncertainty via monitoring. Adaptive management is the principal tool for conserving endangered species under global change, yet adaptive management problems suffer from a poor suite of solution methods. The common approach used to solve an adaptive management problem is to assume the system state is known and the system dynamics can be one of a set of pre-defined models. The solution method used is unsatisfactory, employing value iteration on a discretized belief MDP which restricts the study to very small problems. We show how to overcome this limitation by modelling an adaptive management problem as a restricted Mixed Observability MDP called hidden model MDP (hmMDP). We demonstrate how to simplify the value function, the backup operator and the belief update computation. We show that, although a simplified case of POMDPs, hmMDPs are PSPACE-complete in the finite-horizon case. We illustrate the use of this model to manage a population of the threatened Gouldian finch, a bird species endemic to Northern Australia. Our simple modelling approach is an important step towards efficient algorithms for solving adaptive management problems.

## Introduction

In conservation biology and natural resource management, adaptive management or 'learning by doing' is an iterative process of improving management by reducing uncertainty via monitoring. Coined by Walters and Hilborn (1978), adaptive management has gained notoriety as an approach to manage ecosystems to conserve biodiversity. The key elements of adaptive management include clear definition of objectives, specification of alternative models to achieve these objectives, implementation of two or more models in a comparative experimental framework, monitoring to assess the relative merits and limitations of alternative models, and iterative modification of management to determine the true model, if any (Keith et al. 2011). Despite its virtues, there are few examples of the application of adaptive management in practice. Several factors have been proposed to explain the widespread implementation difficulties in adaptive management programs. Amongst them is the inefficiency of methods used to solve adaptive management problems and the inability to experimentally manipulate endangered populations.



To date, adaptive management problems have been tackled using discretized belief MDP methods (Williams 2009). The belief space is discretized using $p$ subintervals for each belief variable. The updating rule does not guarantee that the updated belief falls on one of these grid points and therefore an interpolation rule must be used to define the transition probabilities for the belief states. This technique has also been studied to sidestep the intractability of exact POMDP value iteration, using either a fixed grid (Lovejoy 1991; Bonet 2002) or a variable grid (Brafman 1997; Zhou and Hansen 2001). The grid based methods differ mainly in how the grid points are selected and what shape the interpolation function takes. In general, regular grids do not scale well in problems with high dimensionality and non-regular grids suffer from expensive interpolation routines.

In this paper, we propose a transparent and formal way of representing an adaptive management problem as a special case of POMDP. We demonstrate for the first time how to model and solve an adaptive management problem as a restricted Mixed Observability MDP (MOMDP) called hidden model MDP (hmMDP). Our approach benefits from recent developments in the field of robotics and decision-making under uncertainty (Ong et al. 2010; Araya-López et al. 2010). Our framework is particularly relevant in the case of endangered species where it may not be possible to undertake a replicated experiment to learn the true model.

### Case Study

We illustrate our method on managing a population of a threatened bird species, the Gouldian finch. The most pervasive threats to wild populations of Gouldian finches are habitat loss and degradation caused by inappropriate fire and grazing regimes and introduced predators such as feral cats. The response of the population to different management actions is uncertain. Each of four experts provided a possible model, which are probability distributions describing how the population might respond to four alternative threat management actions. Our objective is to implement the management action that is most likely to lead to a high persistence probability for the Gouldian finch population. An optimal

adaptive strategy will provide the best decision by determining which model (1 to 4) is most likely the real model, and hence which action is optimal, over time.

The next section presents background on POMDPs and their solution. Then we present the hmMDP model and study its complexity. The following section shows how the specificities of hmMDPs allow for simplified computations. Our simple case study is finally described and used in experiments to show the benefit of advanced solution techniques.

## POMDPs

Partially Observable Markov Decision Processes (POMDPs) are a convenient model for solving sequential decision-making optimization problems when the decision-maker does not have complete information about the current state of the system. Formally, a discrete POMDP is specified as a tuple $\langle S, A, O, T, Z, r, H, b_0, \gamma \rangle$, where:

- $S$ is the set of states $s$ that might be partially observed or imperfectly detected by a manager;
- $A$ is the set of actions (or decisions) $a$ the manager needs to choose from at every time step;
- $O$ is the set of observations $o$ the manager perceives;
- $T$ is a probabilistic transition function describing the stochastic dynamics of the system; an element $T(s, a, s')$ represents the probability of being in state $s'$ at time $t+1$ given $(s, a)$ at time $t$, $T(s, a, s') = p(s_{t+1} = s' | s_t = s, a_t = a)$;
- $Z$ is the observation function, with $Z(a, s', o') = p(o_{t+1} = o' | a_t = a, s_{t+1} = s')$ representing the conditional probability of a manager observing $o'$ given that action $a$ led to state $s'$;
- $r : S \times A \to \Re$ is the reward function identifying the benefits and costs of being in a particular state and performing an action;
- $H$ is the —finite or infinite— horizon (this section focuses on the infinite case);
- $b_0$ is an initial belief, a probability distribution over states;
- $\gamma \in [0, 1]$ is a discount factor (may be 1 if $H$ is finite).

The optimal decision at time $t$ may depend on the complete history of past actions and observations. Because it is neither practical nor tractable to use the history of the action-observation trajectory to compute or represent an optimal solution, belief states, i.e., probability distributions over states, are used to summarize and overcome the difficulties of imperfect detection. A POMDP can be cast into a fully observable Markov decision process defined over the (continuous) belief state space. In our case, solving a POMDP means finding a strategy (or policy) $\pi : B \to A$ mapping the current belief state ($b \in B$) to an allocation of resources. An optimal strategy maximizes the expected sum of discounted rewards over an infinite time horizon $E[\sum_t \gamma^t R(b_t, a_t)]$ where $b_t$ and $a_t$ denote the belief state and action at time $t$, $R(b, a) = \sum_s b(s) r(s, a)$. For a given belief state $b$ and a given policy $\pi$ this expected sum is also referred to as the value function $V_\pi(b)$. A value function allows us to rank strategies by assigning a real value to each belief $b$. An optimal strategy $\pi^*$ is a strategy such that, $\forall\ b \in B$, $\forall\ \pi$, $V_{\pi^*}(b) \geq V_\pi(b)$. Several strategies can be optimal and share a same optimal value function $V^*$ which can be computed using the dynamic programming operator for a POMDP represented as a belief MDP (Bellman 1957), i.e., $\forall b \in B$:

$$V^*(b) = \max_{a \in A} \left[ \sum_{s \in S} r(s, a) b(s) + \gamma \sum_{o'} p(o' | b, a) V^*(b^{ao'}) \right],$$

where $b^{ao'}$ is the updated belief given that action $a$ was performed and $o'$ is observed. This function can be computed recursively using Bellman's principle of optimality (Bellman 1957):

$$V_{n+1}(b) = \max_{a \in A} \left[ \sum_{s \in S} r(s, a) b(s) + \gamma \sum_{o'} p(o' | b, a) V_n(b^{ao'}) \right], \tag{1}$$

where $b^{ao'}$ is updated using Bayes' rule:

$$b^{ao'}(s') = \frac{p(o' | a, s')}{p(o' | b, a)} \sum_{s \in S} p(s' | s, a) b(s). \tag{2}$$

Sondik has shown that the finite time horizon value function is piecewise linear and convex (PWLC) and that the infinite time horizon value function can be approximated arbitrary closely by a PWLC function (Sondik 1971). An alternative way of representing $V$ is to use vectors:

$$V(b) = \max_{\alpha \in \Gamma} \alpha \cdot b, \tag{3}$$

where $\Gamma$ is a finite set of vectors called $\alpha$-vectors, $b$ is the belief represented as a finite vector, and $\alpha \cdot b$ denotes the inner product of an $\alpha$-vector and $b$. The gradient of the value function at $b$ is given by the vector $\alpha_b = \arg\max_{\alpha \in \Gamma} b \cdot \alpha$. The policy can be executed by evaluating (3) at $b$ to find the best $\alpha$-vector: $\pi(b) = a(\alpha_b)$. Exact methods like Incremental Pruning (IP) rely on regularly pruning dominated hyperplanes to reduce their number (Cassandra 1998).

While various algorithms from the operations research and artificial intelligence literatures have been developed over the past years, exact resolution of POMDPs is intractable: finite-horizon POMDPs are PSPACE-complete (Papadimitriou and Tsitsiklis 1987) and infinite-horizon POMDPs are undecidable (Madani, Hanks, and Condon 2003).

In recent years, approximate methods have been developed successfully to solve large POMDPs. Amongst them, point based approaches approximate the value function by updating it only for some selected belief states (Pineau, Gordon, and Thrun 2003; Spaan and Vlassis 2005). Typical point-based methods sample belief states by simulating interactions with the environment and then update the value function and its gradient over a selection of those sampled belief states.

## From MOMDPs to hmMDPs

In this section we demonstrate how to model adaptive management problems using a specific factored POMDP model (Boutilier and Poole 1996). A factored POMDP is a POMDP that explicitly represents the independence relationships between variables of the system.

## MOMDPs

A Mixed Observability MDP (MOMDP) (Ong et al. 2010) is specified as a tuple $\langle X, Y, A, O, T_x, T_y, Z, r, H, b_0, \gamma \rangle$ where:

- $S = X \times Y$ is the factored set of states of the system with $X$ a random variable representing the completely observable components and $Y$ a random variable representing the partially observable components. A pair $(x, y)$ specifies the complete system state;
- $A$ is the finite set of actions;
- $O = O_x \times O_y$ is set of observations with $O_x = X$ the completely observable component, and $O_y$ the set of observations of the hidden variables;
- $T_x(x, y, a, x', y') = p(x'|x, y, a)$ gives the probability that the fully observable state variable takes the value $x'$ at time $t + 1$ if action $a$ is performed in state $(x, y)$ at time $t$ and has already led to $y'$; $T_y(x, y, a, x', y') = p(y'|x, y, a, x')$ gives the probability that the value of the partially observable state variable changes from $y$ to $y'$ given that action $a$ is performed in state $(x, y)$ and the fully observable state variable has value $x'$;
- $Z$ is the observation function with $Z(a, x', y', o'_x, o'_y) = p(o'_x, o'_y|a, x', y')$ the probability of observing $o'_x, o'_y$ given action $a$ was performed, leading to system state $(x', y')$. In a MOMDP we assume the variable $X'$ is perfectly observable so we have $p(o'_x|a, x', y', o'_y) = 1$ if $o'_x = x'$, and 0 otherwise;
- $r$, $H$, $b_0$ and $\gamma$ are defined as for POMDPs.

The belief space $B$ now represents our beliefs on $y$ only since the state variable $X$ is fully observable. Any belief $b$ in $B$ on the complete system state $s = (x, y)$ is represented as $(x, b_y)$. In algorithms based on PWLC approximations, accounting for visible state variables allows for substantial speed-ups by reasoning on multiple low-dimensional belief spaces instead of the original high-dimensional one, as demonstrated with MO-SARSOP (Ong et al. 2010)—based on SARSOP, a state of the art point-based solver—and MO-IP (Araya-López et al. 2010)—based on IP.

## Hidden Model MDPs

Our adaptive management problem assumes managers can perfectly observe the state of the studied system but are uncertain about its dynamics. In ecology, this problem has traditionally been solved by assuming that the real but unknown MDP model is one of a finite set of known models. This can be treated as a *hidden model MDP* (hmMDP), i.e., a MOMDP where the partially observable state variable corresponds to the hidden model. In this setting, the following assumptions can be incorporated to simplify the solution:

1. The real model of the dynamics of the system $y_r$ is an element of a finite set $Y$ of predefined models;
2. The finite set of actions $A$ effects on the completely observable variable $X$ and has no effect on $Y$; $Y$ is also independent from $X$, i.e., $p(y'|x, y, a, x') = p(y'|y)$;

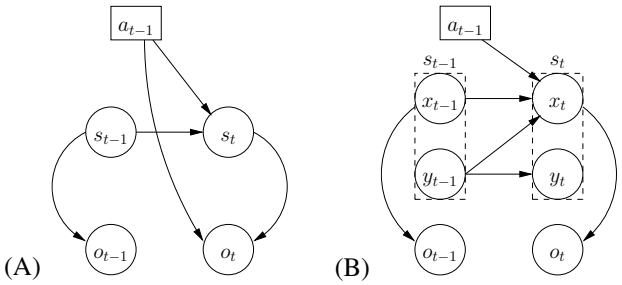


Figure 1: Standard POMDP (A) vs. hidden model MDP (B)

3. The real model $y_r$—although unknown—will not change over time and therefore $T_y$ is the identity matrix, i.e., $p(y'|y) = 1$ if $y = y'$ and 0 otherwise;
4. The hidden variable $Y$ cannot be observed and the observation function $Z$ is only defined on the completely observable variable $X$, i.e., $O = O_x$ and $p(o'_x|a, x') = 1$ if $o'_x = x'$ and 0 otherwise.

Figure 1 illustrates the comparison of the standard POMDP and hmMDP models.

## Definition of an hmMDP

An hmMDP problem is thus a simplified version of MOMDP. In fact, it can be seen as a finite set of (completely observed) MDP models, differing only in their transition and reward functions. Namely, an hmMDP is a tuple $\langle X, Y, A, \{p_y\}_{y \in Y}, \{r_y\}_{y \in Y}, H, \gamma, b_0 \rangle$, where $Y$ is the finite set of potential models, $p_y(x'|x, a) = p(x'|x, a, y)$ and $r_y(x, a) = r(x, a, y)$. Here, $b_0$ is our initial belief on which model is the true one.

When solving an hmMDP problem, we are looking for a policy $\pi(x, b) \to a$ maximizing the usual value function:

$$V^\pi(x, b) = E\left[\sum_{t=0}^{H} \gamma^t r(X^t, \pi(X^t, B^t), B^t)|b_0, \pi\right],$$

$$= \sum_{x^1 \ldots x^H, y} b_0(y) P_y(x^1 \ldots x^H|x^0 = x) \sum_{t=0}^{H} \gamma^t r_y(x^t, a^t),$$

where the $X^t$ are random variables over $X$ and the $B^t$ are random vectors over possible beliefs.

## Complexity of hmMDPs

Although hmMDPs are simplified MOMDPs, they are still computationally complex. We show that solving an hmMDP in the finite-horizon case is a PSPACE-complete problem, thus as hard to solve as a classical finite-horizon POMDP.

**Proposition 1.** *Deciding whether a finite-horizon hmMDP problem admits a solution policy of value greater than a predefined threshold is a* PSPACE*-complete problem.*

Finite-horizon hmMDP clearly belongs to PSPACE, as a particular case of POMDP. The hardness proof results from a reduction of the PSPACE-hard *State Disambiguation* problem (SD) (Conitzer and Sandholm 2003) to the hmMDP problem.

**Definition 1** (STATE-DISAMBIGUATION)**.** *We are given:*

- *A set* $\Theta = \{\theta_1, \ldots, \theta_n\}$ *of possible states of the world and a uniform probability distribution* $p$ *over* $\Theta$*.*
- *A utility function* $u : \Theta \to [0; +\infty[$*.* $u(\theta_i)$ *is the utility of knowing* for sure *that the state of the world is* $\theta_i$*.*
- *A set* $\mathcal{Q} = \{q_1, \ldots, q_m\}$ *of queries.* $q_j = \{q_{j_1}, \ldots, q_{jp_j}\}$ *is a set of subsets of* $\Theta$*, such that* $\bigcup_{1 \leq k \leq p_j} q_{jk} = \Theta$*. If the true state of the world is* $\theta_i$ *and* $q_j$ *is asked, an answer is chosen (uniformly) randomly among the answers* $q_{jk}$ *containing* $\theta_i$*.*
- *A maximum number* $N$ *of queries that can be asked and a target real value* $G > 0$*.*

*The* STATE DISAMBIGUATION *problem consists in deciding whether there exists a policy asking at most* $N$ *queries that gives expected utility at least* $G$*. If* $\pi_\delta(\theta_i)$ *denotes the probability of identifying* $\theta_i$ *by using policy* $\delta$*, the SD problem amounts to deciding whether there exists* $\delta$ *such that* $v(\delta) = \sum_{1 \leq i \leq n} p(\theta_i)\pi_\delta(\theta_i)u(\theta_i) \geq G$*, where p is the uniform distribution over* $\Theta$*.*

*Proof.* The following transformation is a reduction from SD to hmMDP:

- $X = \{q_{jk_j}\}_{j=1\ldots m, k_j=1\ldots p_j}$, $Y = \Theta$.
- $A = \mathcal{Q} \cup \{\{\theta_1\}, \ldots, \{\theta_n\}\}$.
- If $a \in \mathcal{Q}$, $p(x'|x, a, y) = p(q_{rk'}|q_{jk}, q_r, \theta_i) = p(q_{rk'}|q_r, \theta_i) = \frac{1}{|q_r|}$ if $\theta_i \in q_{rk'}$, and 0 else.
  If $a = \{\theta_i\}$, $p(x|x, \{\theta_i\}, \theta_i) = 1, \forall x, \theta_i$.
- $\forall t < N, r^t(x, y, a) = r^t(q_{jk}, \theta_i, q_l) = 0, \forall q_{jk}, \theta_i, q_l$ and $r^t(x, y, a) = r^t(q_{jk}, \theta_i, \{\theta_j\}) < 0$ (this, to prevent choosing $a \in \Theta$ for $t < N$).
- $r^N(x, y, a) = r^N(q_{jk}, \theta_i, \{\theta_i\}) = u(\theta_i), \forall q_{jk}, \theta_i$ and $r^N(q_{jk}, \theta_i, a) = -\max_{\theta_j} u(\theta_j), \forall a \neq \{\theta_i\}$ (this ensures that, if a policy does not disambiguate $\theta_i$, it has expected value less than or equal to 0).
- $H = N, b^0$ is a uniform distribution over $Y$.
- value threshold $G$ (identical to SD threshold).

One can easily check that the policy spaces of the SD and hmMDP problems are isomorphic and there exists a policy $\delta$ such that $v(\delta) \geq G$ in the SD problem if and only if there is a policy $\pi$ which has value at least $G$ in the corresponding hmMDP. □

## Solution Methods

We have provided a new framework to model adaptive management problems in ecology. This framework is simpler than the MOMDP framework and sufficient, as we will show in the application section. However, we have also shown that the computational complexity of this new problem is the same as general POMDPs. In this section we demonstrate that solution algorithms, such as the MO-SARSOP algorithm, can benefit from the model simplification to gain computational efficiency, even though exact resolution remains time-exponential.

Building on Ong et al. (2010) and Araya-López et al. (2010), we show how to simplify the calculation of the value function (1) and belief update (2) for existing POMDP algorithms. First, Ong et al. (2010) have shown that $V(b) = \bar{V}(x, b_y)$ for any MOMDP, and Eq. (1) can be rewritten:

$$V_{n+1}(x, b_y) = \max_{a \in A}[\sum_{y \in Y} r_y(x, a)b_y(y) + \gamma \sum_{y,x',y',o'} b_y(y) \; p(x'|x, y, a, y')p(y'|y, a)p(o'|a, x', y')V_n(x', b_y^{ao'})]. \tag{4}$$

Accounting for the stationarity assumptions (2-3) and $p(o'|x', y') = 1$, $o' = x'$, (assumption 4) leads to:

$$V_{n+1}(x, b_y) = \max_{a \in A}[\sum_{y \in Y} r_y(x, a)b_y(y) + \gamma \sum_{y,x'} b_y(y)p(x'|x, y, a)V_n(x', b_y^{a,o'=x'})],$$

which can be rewritten substituting $V_n(x', b_y^{a,o'=x'}) = \max_{\alpha \in \Gamma_n} b_y^{a,o'=x'} \cdot \alpha$:

$$V_{n+1}(x, b_y) = \max_{a \in A}[\sum_{y \in Y} r_y(x, a)b_y(y) + \gamma \sum_{y,x'} b_y(y)p(x'|x, y, a) \max_{\alpha \in \Gamma_n} b_y^{a,o'=x'} \cdot \alpha] \tag{5}$$

with $b_y^{a,o'=x'}$ the updated belief given $a, o'$ and $b_y$ simplified as follows:

$$b_y^{a,o'=x'}(y) = \frac{p(x'|x, y, a)b_y(y)}{\sum_{y''} b_y(y'')p(x'|x, y'', a)}. \tag{6}$$

Substituting (6) in $\max_{\alpha \in \Gamma_n} b_y^{a,o'=x'} \cdot \alpha$ we can rearrange (5):

$$\begin{aligned} V_{n+1}(x, b_y) &= \max_{a \in A}[\sum_{y \in Y} r_y(x, a)b_y(y) + \gamma \sum_{y,x'} b_y(y) \times \\ &\quad p(x'|x, y, a)\frac{\max_{\alpha \in \Gamma_n} \sum_{y'} p(x'|x, y', a)b_y(y')\alpha(y')}{\sum_{y''} b_y(y'')p(x'|x, y'', a)}] \\ &= \max_a[\sum_{y \in Y} r_y(x, a)b_y(y) + \\ &\quad \gamma \sum_{x'} \max_{\alpha \in \Gamma_n} \sum_{y'} p(x'|x, y', a)b_y(y')\alpha(y')] \\ &= \max_a[b_y \cdot \alpha_0^a + \gamma \sum_{x'} \max_{\alpha \in \Gamma_n} b_y \cdot G_\alpha^{ax'}(x)], \end{aligned} \tag{7}$$

where $G_\alpha^{ax'}(x) = \sum_{y'} p(x'|x, y', a)\alpha(y')$ and the reward function $r(s, a)$ is represented as a set of $|A|$ vectors $\alpha_0^a = (\alpha_0^a(1), ..., \alpha_0^a(|S|))$, one for each action $a$, $\alpha_0^a(s) = r(s, a)$.

Using the identity $\max_{y_j} x \cdot y_j = x \cdot \arg\max_{y_j} x \cdot y_j$ twice we get:

$$V_{n+1}(x, b_y) = b_y \cdot \arg\max_a b_y \cdot [\alpha_0^a + \gamma \sum_{x'} \arg\max_{\alpha \in \Gamma_n} b_y \cdot G_\alpha^{ax'}(x)].$$

**Algorithm 1**: MO-SARSOP $\alpha$-vector backup at a node $(x, b_y)$ of $T_R$

1 BACKUP($T_R, \Gamma, (x, b_y)$);
2 **forall** $a \in A, x' \in X, o' \in O$ **do**
3 $\quad \alpha_{a,x',o'} \leftarrow \arg\max_{\alpha \in \Gamma_y(x')}(\alpha \cdot \tau(x, b_y, a, x', o'))$;
4 **forall** $a \in A, y \in Y$, **do**
5 $\quad \alpha_a(y) \leftarrow r(x, y, a) + \gamma \sum_{x',o',y'} T_x(x, y, a, x') T_y(x, y, a, x', y') \times Z(x', y', a, o') \alpha_{a,x',o'}(y')$;
6 $\alpha' \leftarrow \arg\max_{a \in A}(\alpha_a \cdot b_y)$;
7 Insert $\alpha'$ into $\Gamma_y(x)$;

**Algorithm 2**: hm-SARSOP $\alpha$-vector backup at a node $(x, b_y)$ of $T_R$

1 BACKUP($T_R, \Gamma, (x, b_y)$);
2 **forall** $a \in A, x' \in X$ **do**
3 $\quad \alpha_{a,x'} \leftarrow \arg\max_{\alpha \in \Gamma_y(x')}(\alpha \cdot \tau(x, b_y, a, x'))$;
4 **forall** $a \in A, y \in Y$, **do**
5 $\quad \alpha_a(y) \leftarrow r_y(x, a) + \gamma \sum_{x'} T_x(x, y, a, x') \alpha_{a,x'}(y)$;
6 $\alpha' \leftarrow \arg\max_{a \in A}(\alpha_a \cdot b_y)$;
7 Insert $\alpha'$ into $\Gamma_y(x)$;

Finally we can define the vector backup $\beta(b_y)$ as the vector whose inner product with $b_y$ yields $V_{n+1}(b_y)$:

$$V_{n+1}(x, b_y) = b_y \cdot \beta(b_y), \text{ where} \tag{8}$$
$$\beta(b_y) = \arg\max_a b_y \cdot [\alpha_0^a + \gamma \sum_{x'} \arg\max_{\alpha \in \Gamma_n} b_y \cdot G_\alpha^{ax'}(x)].$$

By simplifying the backup operator calculation (8) and the belief update operation (6) we have shown how existing POMDP solving algorithms based on a PWLC approximation can be adapted when looking at problems with hidden models. These two procedures are common to most exact algorithms such as Witness or Incremental Pruning (Cassandra 1998), and most point-based POMDP algorithms, such as PBVI (Pineau, Gordon, and Thrun 2003), Perseus (Spaan and Vlassis 2005), symbolic Perseus (Poupart 2005) and SARSOP (Ong et al. 2010).

Algorithm 1 presents the backup procedure used at each belief point by MO-SARSOP, and Algorithm 2 shows how—under the adaptive management assumptions—it can be simplified at two levels: i) we do not need to consider the set of observations $O$, and the belief update calculation $\tau$ is simplified (line 2); ii) the observation function $Z$ and the model dynamics $T_y$ are not required (line 5). The sampling procedure also benefits from the simplified belief update (6).

## Management of a Threatened Bird

We apply our adaptive management model to a threatened Gouldian finch population in the Kimberley, Australia. Our management objective is to maximize the likelihood of a high persistence probability of this population over the medium term, in an area where we can measure the population state but are uncertain about the state response to the management actions. We asked four experts to assess the likelihood of a high (and conversely low) probability of persistence under four plausible management actions. Following our hidden model MDP model, we define the set of states $S = X \times Y$ where $X$ represents the local probability of persistence $X = \{\text{Low, High}\}$. We pose $Y = \{\text{Expert 1, Expert 2, Expert 3, Expert 4}\}$ the set of possible true models that predict the state dynamics of our studied species. The set of actions $A = \{\text{DN,FG,C,N}\}$ represents the management actions to choose from at every time step: do nothing (DN), improve fire and grazing management (FG), control feral cats (C) and provide nesting boxes (N). We assume that the same amount of funds are spent on each action. We elicited from the experts the transition probabilities $T_x$ for each model and state transition. Finally we define the reward function so that $r(\text{Low,DN}) = 0$, $r(\text{High,DN}) = 20$, $r(\text{Low,}\{\text{FG,C,N}\}) = -5$ and $r(\text{High,}\{\text{FG,C,N}\}) = 15$. We assume that each expert is as likely to be correct at the beginning of our adaptive management program.

| | Problem 1: $\lvert S\rvert = 8$ $\lvert X\rvert = 2, \lvert Y\rvert = 4$ | Problem 2: $\lvert S\rvert = 162$ $\lvert X\rvert = 81, \lvert Y\rvert = 2$ |
|---|---|---|
| IP | h= 5 $\lvert\alpha\rvert = 3753$ t=349.2s | h=4 $\lvert\alpha\rvert = 1181$ t=703.7s |
| MO-IP | h=5 $\lvert\alpha\rvert = 2052$ t=106.8s | h=4 $\lvert\alpha\rvert = 218$ t=0.59s |
| Grid bMDP$^+$ | h=26 $\lvert\alpha\rvert = 4402$ t=1831s err=1.28 | h=23 $\lvert\alpha\rvert = 426$ t=1849s err= 3.84 |
| SARSOP$^+$ | $\lvert\alpha\rvert = 25153$ t=1831s err=0.066 | $\lvert\alpha\rvert = 6995$ t=422.85s err<0.001 |
| MO-SARSOP$^+$ | $\lvert\alpha\rvert = 38137$ t=1831s err=0.055 | $\lvert\alpha\rvert = 3861$ t=19.36s err<0.001 |

Table 1: Performance for the Gouldian finch problems. Experiments conducted on a 2.40 GHz Core 2 computer with 3.45 GB of memory. $(^+)$ err represents the Bellman residual error between 2 successive value functions (precision).

We first attempted to solve our adaptive management problem using IP[1]. Due to a lack of memory space we were not able to provide an exact solution for a time horizon greater than 5 time steps with 3753 $\alpha$-vectors computed (Problem 1, Table 1). We solved the same problem using MO-IP (Araya-López et al. 2010) and reached the 5th horizon 3 times faster than IP. We then computed approximate solutions using the algorithms Grid-bMDP[1], SARSOP and MO-SARSOP (Ong et al. 2010). We interrupted Grid-bMDP after $\sim$30 mins (horizon 26, 4402 $\alpha$-vectors and an estimated error of 1.28). Using the same computational time, SARSOP and MO-SARSOP provided the best performances by far with a higher precision for MO-SARSOP. Experiments conducted with a version of MO-IP implementing the modifications specific to adaptive management showed no speed up at horizon 5 (106.69s vs. 106.80s for the original MO-IP). We also ran these algorithms on a version of our Gouldian finch problem with a larger set of states ($|X| = 81$)

[1]Using Cassandra's pomdp-solver toolbox.

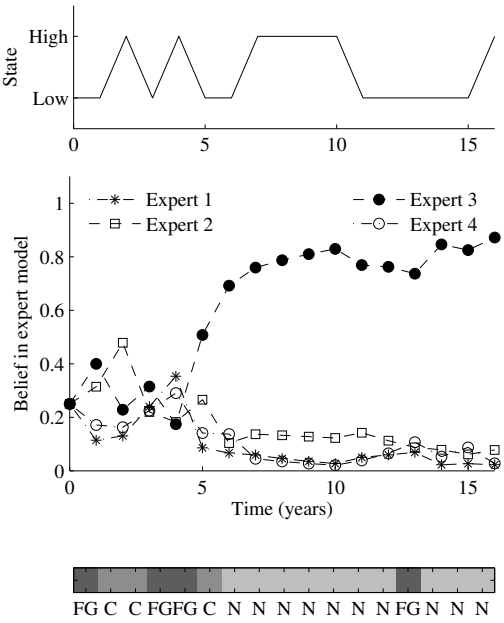


Figure 2: Simulation of our adaptive management strategy when Expert 3 holds the real model. The graphs represent the state dynamics (top), the belief in each expert's predictions over time (middle), and the action performed (bottom).

and a smaller set of experts (Experts 1 and 2). In this case we explicitly distinguish how the interactions between 4 species (dingo, cats, long-tailed finch and Gouldian finch) could drive the management success (Problem 2, Table 1).

The solutions of Problem 1 provide guidance on which management action should be performed in absence of precise knowledge of what the real model is. Expert 1 and 4 provided similar transition probabilities; both believe that the appropriate management of fire and grazing (FG) will generate a higher probability of the population reaching and remaining in the 'High' persistence state. Consequently, in absence of prior information on what the real model is, the first action to perform is FG. Expert 2 believes the management of feral cats (C) and the provision of nesting boxes (N) have more benefits than fire and grazing management. Expert 3 believes that habitat loss and intra-species competition is the limiting factor of persistence and therefore attributes the highest probabilities of persistence when action 'provide nesting boxes' (N) is implemented. Note that Expert 3 values the management of cats and fire and grazing as 'good' actions. We assessed the quality of our adaptive strategy by simulations. When we assume that cats are the main threat and Expert 2 holds the real model, our adaptive strategy quickly finds the best action to perform. However when we assume that fire and grazing are the main threats, our adaptive strategy provides the best action to perform but has trouble identifying the real model, because Experts 1 and 4 are similarly supportive of the combined FG action. When we assume Expert Model 3 holds the real model (see Figure 2), the simulations first lead to believing that the management of cats is most beneficial and Expert 2 holds the real model. Again this ambiguity is due to the high success rate given by Expert 3 when feral cats are managed. After several time steps, our adaptive strategy favors Expert 3.

## Discussion

Adaptive management problems as currently solved in the ecology literature suffer from a lack of efficient solution methods. We demonstrate and explain how to model a classic adaptive management problem as a POMDP with mixed observability focusing on the special case of hidden model MDPs (hmMDPs). We show that hmMDPs are PSPACE-complete in the finite-horizon case. However, if data were available, MO-SARSOP could approximately solve AM problems with up to $\sim$250,000 states in 20–30 minutes (Ong et al. 2010). The assumption that the real model is contained within the model set is simplistic (Assumption 1), however it is currently the way AM is solved. Point based methods help us account for a large set of models so that we do not risk being too far from the real model, but, when managing threatened species, having many models makes it difficult to be confident on the real model as observations are few. In this situation, we would recommend incorporating a model for each of the management options so that the best action can be chosen in the face of uncertainty.

While our examples may seem simplistic in AI they are complex and reasonably realistic in ecology (Chadès et al. 2008). Practitioners are reluctant to adopt new methods. Our aim is to bridge the gap between AI and ecological management by solving real, tractable problems at a scale defined by practitioners. Conservation practitioners need to understand the model and the management rules need to be simple enough to be implemented (Chadès et al. 2011). Moreover, we need to be able to populate these models with data, which in ecology is often lacking. Both examples are based on a real application with real experts, with only 4 management actions currently possible for the Gouldian finch. Example 2 is complex due to interacting species and requires eliciting more information to populate the corresponding DBN.

POMDPs are a powerful model that does not require the restrictive assumptions of the traditional adaptive management literature and offers new and exciting adaptive management applications. First, there is no need to assume a static model (Assumption 3). If we allow the real model to change over time we can tackle systems that are influenced by climate change or seasonal fluctuations. Second, there is no need to assume perfect detection of the state ($X$) (Assumption 4) and detection probabilities can be included. Not assuming that the real model is one of a predefined set of models (Assumption 1) remains an unsolved challenge. To our knowledge this cannot be tackled with current POMDP solutions but model-based Bayesian Reinforcement Learning (Poupart et al. 2006) could be a relevant alternative.

## Acknowledgements

This research was funded through an Australian Government's NERP, an Australian Research Council's Centre of Excellence (I.C., J.C., T.G.M.) and a CSIRO Julius Career Award (T.G.M.).

## References


Araya-López, M.; Thomas, V.; Buffet, O.; and Charpillet, F. 2010. A closer look at MOMDPs. In *Proc. of the 22nd Int. Conf. on Tools with Artificial Intelligence*.

Bellman, R. E. 1957. *Dynamic Programming*. Princeton, N.J.: Princeton University Press.

Bonet, B. 2002. An e-optimal grid-based algorithm for partially observable Markov decision processes. In *Proc. of the 19th Int. Conf. on Machine Learning (ICML-02)*.

Boutilier, C., and Poole, D. 1996. Computing optimal policies for partially observable decision processes using compact representations. In *Proc. of the Nat. Conf. on Artificial Intelligence*.

Brafman, R. 1997. A heuristic variable grid solution method for POMDPs. In *Proc. of the Nat. Conf. on Artificial Intelligence*.

Cassandra, A. R. 1998. *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes*. Ph.D. Dissertation, Brown University, Dept. of Computer Science.

Chadès, I.; McDonald-Madden, E.; McCarthy, M. A.; Wintle, B.; Linkie, M.; and Possingham, H. P. 2008. When to stop managing or surveying cryptic threatened species. *PNAS* 105:13936–13940.

Chadès, I.; Martin, T. G.; Nicol, S.; Burgman, M. A.; Possingham, H. P.; and Buckley, Y. M. 2011. General rules for managing and surveying networks of pests, diseases, and endangered species. *PNAS* 108(20):8323–8328.

Conitzer, V., and Sandholm, T. 2003. Definition and complexity of some basic metareasoning problems. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03)*, 1099–1106.

Keith, D. A.; Martin, T. G.; McDonald-Madden, E.; and Walters, C. 2011. Uncertainty and adaptive management for biodiversity conservation. *Biological Conservation* 144(4):1175–1178.

Lovejoy, W. 1991. Computationally feasible bounds for partially observed Markov decision processes. *Operations research* 39(1):162–175.

Madani, O.; Hanks, S.; and Condon, A. 2003. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence* 147(1-2):5–34.

Ong, S. C. W.; Png, S. W.; Hsu, D.; and Lee, W. S. 2010. Planning under uncertainty for robotic tasks with mixed observability. *The International Journal of Robotics Research* 29(8):1053–1068.

Papadimitriou, C. H., and Tsitsiklis, J. N. 1987. The complexity of Markov decision processes. *Journal of Mathematics of Operations Research* 12(3):441–450.

Pineau, J.; Gordon, G.; and Thrun, S. 2003. Point-based value iteration: An anytime algorithm for POMDPs. In *Proc. of the Int. Joint Conf. on Artificial Intelligence*.

Poupart, P.; Vlassis, N.; Hoey, J.; and Regan, K. 2006. An analytic solution to discrete Bayesian reinforcement learning. In *Proc. of the 23rd Int. Conf. on Machine Learning*.

Poupart, P. 2005. *Exploiting Structure to Efficiently Solve Large Scale Partially Observable Markov Decision Processes*. Ph.D. Dissertation, University of Toronto.

Sondik, E. 1971. *The Optimal Control of Partially Observable Markov Decision Processes*. Ph.D. Dissertation, Stanford University, California.

Spaan, M., and Vlassis, N. 2005. Perseus: Randomized point-based value iteration for POMDPs. *Journal of Artificial Intelligence Research* 24:195–220.

Walters, C. J., and Hilborn, R. 1978. Ecological optimization and adaptive management. *Annual Review of Ecology and Systematics* 9:pp. 157–188.

Williams, B. 2009. Markov decision processes in natural resources management: Observability and uncertainty. *Ecological Modelling* 220(6):830–840.

Zhou, R., and Hansen, E. 2001. An improved grid-based approximation algorithm for POMDPs. In *Proc. of the 17th Int. Joint Conf. on Artificial Intelligence*.