Received: 15 September 2020 | Accepted: 15 July 2021

DOI: 10.1111/2041-210X.13692

REVIEW ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

A primer on partially observable Markov decision processes (POMDPs)

Iadine Chadès[1] | Luz V. Pascal[2] | Sam Nicol[1] | Cameron S. Fletcher[3] | Jonathan Ferrer-Mestres[1]

[1]CSIRO, Dutton Park, Queensland, Australia
[2]ENSTA, Palaiseau, France
[3]CSIRO, Atherton, Queensland, Australia

Correspondence
Iadine Chadès
Email: iadine.chades@csiro.au

Funding information
CSIRO Research Office Postdoctoral Fellowship; CSIRO MLAI Future Science Platform; CSIRO Julius Career Award

Handling Editor: Satu Ramula

Abstract

1. Partially observable Markov decision processes (POMDPs) are a convenient mathematical model to solve sequential decision-making problems under imperfect observations. Most notably for ecologists, POMDPs have helped solve the trade-offs between investing in management or surveillance and, more recently, to optimise adaptive management problems.
2. Despite an increasing number of applications in ecology and natural resources, POMDPs are still poorly understood. The complexity of the mathematics, the inaccessibility of POMDP solvers developed by the Artificial Intelligence (AI) community, and the lack of introductory material are likely reasons for this.
3. We propose to bridge this gap by providing a primer on POMDPs, a typology of case studies drawn from the literature, and a repository of POMDP problems.
4. We explain the steps required to define a POMDP when the state of the system is imperfectly detected (state uncertainty) and when the dynamics of the system are unknown (model uncertainty). We provide input files and solutions to a selected number of problems, reflect on lessons learned applying these models over the last 10 years and discuss future research required on interpretable AI.
5. Partially observable Markov decision processes are powerful decision models that allow users to make decisions under imperfect observations over time. This primer will provide a much-needed entry point to ecologists.

KEYWORDS

AI, decisions, partially observable Markov decision processes, stochastic dynamic programming, uncertainty

1 | INTRODUCTION

Changes in species population size, habitat quality, presence or absence of threats, and environment and climate are attributes that make ecological systems dynamic (Brown et al., 2001). In conservation and applied ecology, making an informed decision to manage a dynamic system would ideally be based on perfect knowledge of the state of the system, as one management action rarely fits all situations. State-transition models provide a way of representing these dynamics as discrete set of states and transitions (Bestelmeyer et al., 2003, 2017). When looking for the optimal sequence of decisions to achieve an objective, stochastic dynamic approaches and their mathematical implementation, Markov decision processes (MDPs), are the go-to model for applied ecologists (Bellman, 1957; Marescot et al., 2013; Sigaud & Buffet, 2013). However, the application of these approaches implicitly or explicitly assumes that the state of a system is or can be accurately identified. Unfortunately, ecological systems are difficult to monitor and our ability to identify the state of the systems varies widely (Nichols & Williams, 2006; Norouzzadeh et al., 2018). While multiple monitoring approaches

are available to help identify the state of an ecological system, in practice many situations require managers to make decisions in the absence of complete information on the study system (Field et al., 2005). Such a situation may occur when there is no time to collect more data (urgent decision-making), or when collecting additional data is costly and time-consuming, or simply impossible in practice (Chadès & Nicol, 2016a,b). When the state of the system is unknown or partially unknown, partially observable Markov decision processes (POMDPs) are the mathematical model that will guide the decision-making process.

Introduced by Åström (1965), POMDPs generalise MDPs (Bellman, 1957) by incorporating the idea that decision-makers might not be able to perfectly monitor the world state. Because POMDPs are common to Artificial Intelligence (AI) and Machine Learning (ML) problems, for example, to design smart autonomous robots and cars (Cassandra, 1998), the AI and ML scientific community has developed many exact, approximate (with performance guarantee) and heuristic (without performance guarantee) algorithms to tackle the formidable computational problem POMDPs presents (Kaelbling et al., 1998; Pineau et al., 2003). However, while AI and ML research focuses on optimally solving large problems efficiently for autonomous systems (Russell & Norvig, 2002), applied ecology is a human-operated system for which interpretation and explanation of models and results are perhaps more important than overall performance. Here, drawing on our experience designing POMDP algorithms and solving decision problems, we present a primer on POMDPs for ecologists.

Partially observable Markov decision processes have been applied to a range of ecological applications. In conservation, POMDPs have been used to explore the dilemma between investing resources in either on-ground management or surveillance of cryptic-threatened species (Chadès et al., 2008; Dujardin et al., 2015; McDonald-Madden et al., 2011), to manage threats and reintroductions of a listed species (Nicol & Chadès, 2012), and to decide to survey nesting sites or allow human use of endangered seabird habitat in coastal forest (Tomberlin, 2010). In invasive species management, POMDPs have informed how long to manage for invasive weeds with difficult to detect microscopic seeds (Regan et al., 2011), or to decide between preventing, searching for and destroying infestations when the severity of infestation is only uncertainly known (Rout et al., 2014). In natural resource management and economics, POMDPs have also helped determine the influence of monitoring costs for reaching a target natural vegetation community (White, 2005), or the cost of management to further guide invasive species management (Fackler & Haight, 2014; Haight & Polasky, 2010). Recently, Memarzadeh et al. (2019) showed that POMDP-based management can avoid over-exploitation of fisheries while also generating increased economic value. Accounting for space, POMDPs have been used to derive general management and monitoring priorities for small networks representing meta-populations of threatened or invasive species or diseases assuming Susceptible-Infected-Susceptible (SIS) dynamics (Chadès et al., 2011). More recently, it was shown that POMDPs were also useful for decision-making to solve adaptive management problems (Walters, 1986) when the dynamics of the system are unknown (Williams, 2011). Indeed, POMDPs can also model the uncertainty about the system dynamics as state uncertainty and help optimise adaptive management programmes (Chadès et al., 2012; Williams, 2011). Applications include protecting habitat for migratory shorebirds under uncertain consequences of sea-level rise assuming non-stationary dynamics (Nicol et al., 2013, 2015), and with state uncertainty to assess over-exploitation of fisheries (Memarzadeh & Boettiger, 2018) or inform recreation management to simultaneously preserve an abundant eagle population in Denali National Park and hiker access (Fackler et al., 2014).

Despite a growing literature on the topic, guidance on how to define and solve POMDP problems remains scarce—but see attempts in operations research (Lovejoy, 1991b; White, 1991), management sciences (Monahan, 1982), AI (Kaelbling et al., 1998) and mathematical psychology (Littman, 2009). Here, we fill this gap for ecological readers. We first introduce MDPs and outline when POMDPs are useful. We then formally define POMDPs and provide a typology of problems that POMDP can help solve. We explain some of the underlying theory that makes solving POMDPs a challenging problem and present ways of solving POMDPs using a selected number of POMDP toolboxes. We introduce the github repository https://github.com/conservation-decisions/POMDPproblems that provide a one-stop shop for examples of POMDP problems. We discuss the need to understand POMDP solutions that will lead to further uptake of POMDPs in ecology. Finally, we reflect on more than 10 years of research applying POMDPs and discuss future research directions. The Supporting Information provides a much-needed outline of the steps involved in installing and running POMDP solvers (in R, C/C++ and MATLAB).

2 | MARKOV DECISION PROCESSES

Markov decision processes (Bellman, 1957; Puterman, 2014) are a convenient model for solving sequential decision-making optimisation problems when the decision-maker has complete information about the current state of the system. Formally, an MDP is specified as a tuple $\langle S, A, T, r, H, \gamma \rangle$, where:

- S is the finite set of states that describe the system.
- A is the finite set of actions (or decisions) the manager needs to choose from at every time step.
- 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, $T(s, a, s') = p(s_{t+1} = s' \mid s_t = s, a_t = a)$ (we adopt the AI notations where $'$ represents the time step $t + 1$). The probability distribution over the next state s' defined by p follows the Markov property which gives its name to Markov decision processes. The probability of reaching state s' given action a is implemented only depends on action a and the previous state of the system s.
- $r: S \times A \to \mathfrak{R}$ is the reward function identifying the benefits and costs of being in a particular state and performing a particular action.

- H is the time horizon (potentially infinite) over which actions are implemented.
- $\gamma \in [0, 1]$ is a discount factor relating future rewards and costs to their net present value. A discount factor lower than one expresses that immediate rewards are more valuable than later ones (Koopmans, 1960). In practice, a discount factor lower than one ensures the convergence of the optimisation criterion (e.g. the expected sum of discounted rewards over an infinite time horizon) and helps solvers converge more rapidly towards a solution (Kaelbling et al., 1998).

To help understand MDP and POMDP models, we use the *Sumatran tiger* problem as an illustrative example (Chadès et al., 2008; Pascal et al., 2020). At each time step, managers have to allocate their limited resources to keep a population of threatened species extant. The set of states is defined as locally extinct or extant ($S = \{extinct, extant\}$) and represents the status of a local population at a given time step. In the completely observable case, we assume that managers perfectly detect the state of a local population. At each time step, managers can choose between two actions *do_nothing* or invest in *management* of threats ($A = \{do\_nothing, manage\}$). The transition probabilities between states given an action represent the dynamics of the system. The probability of a population becoming locally extinct at a given time step following the implementation of the action *do_nothing* is $p(s_{t+1} = extinct \mid a_t = do\_nothing, s_t = extant) = 0.1$. Under management, this probability drops to $p(s_{t+1} = extinct \mid a_t = manage, s_t = extant) = 0.05816$. We assume that when the local population is extinct there is no possible recovery, and thus the population remains extinct $p(extinct \mid extinct, .) = 1$. The transition matrices for both actions are (data from Chadès et al., 2008):

do_nothing	*extinct*	*extant*
extinct	1	0
extant	0.1	0.9

manage	*extinct*	*extant*
extinct	1	0
extant	0.05816	0.94184

In the case of the *Sumatran tiger*, it is assumed that the species attracts funding when the species is in state *extant*. The action *manage* comes at a cost. Implementing an action while the population is extant provides a value corresponding to the difference between the economic value of the species and the cost of implementing the action. The reward matrix $r(s, a)$ is defined as:

	extinct	*extant*
do_nothing	0.0	175.133
Manage	−18.784	156.349

Figure 1a provides a compact graphical and mathematical representation of an MDP as an influence diagram. An influence diagram is a graphical structure for modelling uncertain variables (e.g. s_t) and decisions (a_t) and explicitly revealing probabilistic dependence (T) and the flow of information (arrows, Shachter, 1986).

Solving an MDP problem means finding the best decision to implement in a given state at each time step. These decision rules are called an MDP strategy or policy. A policy is a function which associates a deterministic action to a state configuration of the system and can be seen as a set of rules a decision-maker would follow to choose the action to perform in each state. In most cases, managers are interested in implementing policies that will provide the highest value over time. To determine which policy has the highest value, an optimisation criterion (or objective) must be defined. Here, we consider two criteria:

- The expected sum of discounted rewards over a finite horizon $\mathbf{E}\left(\sum_{t=0}^{H} \gamma^t r_t\right)$. From an ecological perspective, this criterion assumes that decisions need to be made within a finite time frame H. Such a situation can occur when a programme has a fixed amount of time to achieve a specific outcome. In this case, an optimal policy is time-dependent (non-stationary, $\pi_t: S \rightarrow A$). Algorithms such as backwards induction can help finding an optimal solution (Puterman, 2014);
- The expected sum of discounted rewards over an infinite time horizon $\mathbf{E}\left(\sum_{t=0}^{\infty} \gamma^t r_t\right)$. From an ecological perspective, we assume

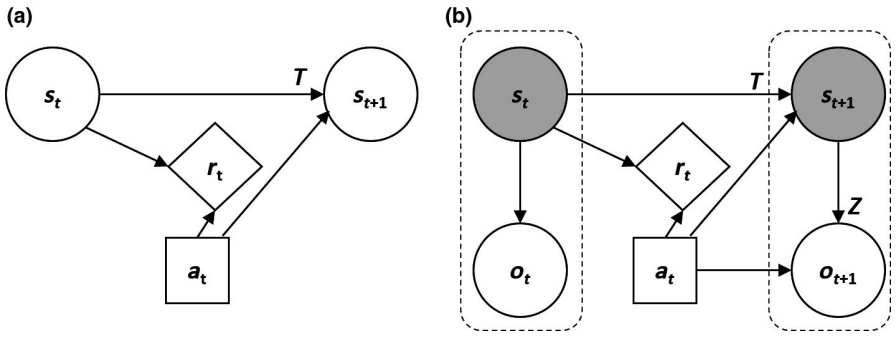

FIGURE 1 (a) MDPs and (b) POMDPs influence diagrams representing the decision situation graphically over one time step (t to $t + 1$). (a) In the MDP case (Section 2), the state s_t is fully observable to managers that must act (a_t) to achieve a desired objective. Actions may come at a cost and states may generate a reward (r_t). Dynamics between states are stochastic and assumed Markovian (T). (b) In the POMDP case (Section 3), the state (s_t, shaded) is partially observable, that is, managers only get an incomplete observation of the state (o_t). The observation function (Z) provides the probability of getting an observation (o_{t+1}) given a state configuration (s_{t+1}) and an action (a_t)

that the time horizon of a programme should not influence the decision-making process and only the state of the system does. In this case, an optimal policy is independent of time (stationary, $\pi: S \rightarrow A$). Algorithms such as value iteration or policy iteration are among the most used approaches to find a solution (Puterman, 2014).

While the time horizon of the criterion can matter, computational challenges also influence the choice of criteria. Finite time horizon problems require accounting for a time dimension and are more complex to solve. In addition, more efficient algorithms have been designed for infinite horizon problems (Kaelbling et al., 1998). For the sake of simplicity, we will assume an expected sum of discounted rewards over an infinite time horizon. Marescot et al. (2013) provide step-by-step MDP examples for the finite time horizon case.

Finding an optimal policy, that is, a policy that has the highest value, requires solving Bellman's stochastic dynamic programming equations (Bellman, 1957). This can be done by directly calculating an optimal policy π^*, for all s:

$$\pi^*(s) = \arg\max_{a \in A} r(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V_{\pi^*}(s').$$

with V_{π^*} the optimal value function, for all s:

$$V_{\pi^*}(s) = \max_{a \in A} r(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V_{\pi^*}(s').$$

In essence, the Bellman optimality equations state that an optimal policy can be calculated by finding, for each state, an action a that maximises the immediate expected reward, given an optimal sequence of actions will be implemented in future states (s'). Several exact and approximate algorithms have been implemented to solve the Bellman optimality equations for MDPs (Chadès et al., 2014; Fackler, 2011; Marescot et al., 2013). For example, the *value iteration* algorithm repeatedly applies the Bellman update over value functions V_t, starting at any initial value function (V_0):

$$V_{t+1}(s) = \max_{a \in A} r(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V_{\pi^*}(s').$$

The algorithm converges towards an optimal value function when the difference between two successive value functions $|V_{t+1}(s) - V_t(s)|$ is less than an ε for all states s.

We solved the MDP for the *Sumatran tiger* example (using a discount factor of 0.95). In that case, the optimal value function was $V_{\pi^*}(extant) = 1485.469$ and $V_{\pi^*}(extinct) = 0$. The optimal policy was $\pi^*(extant) = manage$ and $\pi^*(extinct) = do\_nothing$. This means that while the species is known to be extant locally, the optimal decision is to keep managing. However, when the species is known to be locally extinct, the best decision is to stop managing and do nothing.

3 | PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES

3.1 | When are POMDPs useful?

Partially observable Markov decision processes augment MDPs by accounting for state uncertainty (Åström, 1965). POMDPs are a convenient model for solving sequential decision-making optimisation problems when the decision-maker does not have complete information about the current state of the system.

A challenging first step is to identify when POMDPs are needed. We have identified three key elements, all of which must be relevant to make a problem suitable to be modelled as a POMDP:

1. States of the system can be defined and not one decision (action) fits all states. For example, for threatened species, likely upper and lower population abundance estimates under different environmental conditions and threats may provide a natural definition of system states. In this system, we would expect the best action to be state-dependent, for example, near-extinction population sizes to require reintroduction actions, low–moderate population sizes to require threat management actions and close to carrying capacity populations to require no action.
2. Natural variability and/or control uncertainty make the system dynamics stochastic. These dynamics are assumed to be Markovian, that is, the probability of transitioning to a state at time $t + 1$ only depends on the state of the system and action implemented at time t. For example, we expect that population dynamics are stochastic processes and that the success of management actions such as reintroduction is likely to have uncertain outcomes. Given an appropriate time step, and by capturing relevant information into the definition of state, the state transition dynamics follow the Markov property.
3. Managers are unable to perfectly observe the state of the system (state uncertainty). Contrary to the perfect observable case, the best decision to implement given an incomplete observation of the system state is not straightforward and requires additional consideration such as the history of past observations and actions. State uncertainty changes the problem from a MDP into a more challenging POMDP.

Together, steps (a) and (b) define an MDP, while step (c) makes the problem partially observable. MDPs are both conceptually and computationally easier to solve than POMDPs and have a much longer history in ecology than POMDPs (Clark et al., 2000; Marescot et al., 2013).

In the language of statistics, POMDPs are models that describe optimal control of hidden Markov models (HMMs). In this framing, the partially observed states of the model are latent variables that are inferred from the (fully observable) observation states. Table 1 identifies the different Markov models by differentiating whether

TABLE 1 A classification of Markov models based on complete observability of the states and control over the state transitions

Markov Models		Do we have control over state transitions?	
		No	Yes
Are the states completely observable?	Yes	Markov chain	Markov decision process (MDP)
	No	Hidden Markov model (HMM)	Partially observable Markov decision process (POMDP)

a manager has control over the state transitions of the systems and whether the states of the system are completely observable.

3.2 | Definition

Formally, a finite POMDP is specified as a tuple $\langle S, A, O, T, Z, r, H, b_0, \gamma \rangle$, where:

- S, A, T, r, H and γ are defined as in the MDP case; However, the set of actions (or decisions) may include actions such as monitoring which increases the ability of managers to detect the state of the system more accurately or reduce state uncertainty.
- O is the finite set of observations o the manager perceives;
- Z is the observation function, with $Z(a, s', o') = p(o_{t+1} = o' \mid a_t = a, s_{t+1} = s')$ representing the conditional probability of a manager observing o' given that action a led to state s'. In problems where state uncertainty cannot be reduced (e.g. increasing surveillance effort is impossible), the observation function will be independent from actions, that is, $Z(a, s', o') = p(o_{t+1} = o' \mid s_{t+1} = s')$.
- b_0 is an initial belief, a probability distribution over states. Intuitively, a belief state represents where we think we are at a given time step. In some cases, b_0 can be omitted.

In the case of the *Sumatran tiger* example, the problem can be modelled as a POMDP by accounting for the imperfect detection of the status of the species. Due to their low numbers and cryptic nature, evaluating the status of a threatened species is challenging. In the absence of detection, a population could be locally extinct or extant. In the absence of sighting, managers need to decide when to stop managing and switch their limited resources to surveillance, and ultimately when to surrender (Chadès et al., 2008; Pascal et al., 2020). As in the MDP setting, a *Sumatran tiger* population can be locally extinct or extant ($S = \{$*extinct*, *extant*$\}$). The set of possible observations is defined as absent or present ($O = \{$*absent*, *present*$\}$) and denote if tigers were observed at a given time step. The set of actions that managers can decide to implement now include an opportunity to invest in camera traps to increase detection of tigers (*survey*) ($A = \{$*do_nothing*, *manage*, *survey*$\}$)—for sake of simplicity, the action 'manage and survey' is not included in this example but see McDonald-Madden et al. (2011). The system dynamics is stochastic and the probability transition functions $T(s, a, s')$ for *manage* and *do_nothing* are defined similarly to the MDP case. It is assumed that *survey* has the same management effectiveness as *do_nothing*, that is, $T(s, do\_nothing, s') = T(s, survey, s')$. The reward function $r(s, a)$ also has to include the cost of *survey* in a given state s:

		extinct	*extant*
$r(s, a) =$	*do_nothing*	0.0	175.133
	manage	−18.784	156.349
	survey	−10.840	164.293

The observation function $Z(a, s', o')$ represents the probability of detecting the *Sumatran tiger* given that action a was applied and led to state s'. For example, the probability of observing that the tiger is *present* given that the population is *extant* and action *do_nothing* or *manage* was applied is 0.01. If action *survey* is implemented, the probability of detecting an extant population of *Sumatran tiger* is 0.78. It is assumed that no false positive can occur, that is, $P(present \mid extinct, .) = 0$:

do_nothing||*manage*	*absent*	*present*
extinct	1	0
extant	0.99	0.01

survey	*absent*	*present*
extinct	1	0
extant	0.21807	0.78193

When modelling a POMDP problem, we recommend representing the problem as an influence diagram to clarify the dependence between variables (Figure 1b; Shachter, 1986). Defining each element of a POMDP can be challenging. Because POMDPs augment MDPs, we recommend first defining the problem as an MDP before adding the state uncertainty components that define a POMDP problem. This step will help better understanding the mechanistic insights that state uncertainty adds to the decision problem (e.g. (Chadès et al., 2011)).

4 | TYPOLOGY OF POMDP PROBLEMS

We have identified three types of problems that POMDPs can solve. The first type of problem is the classic dilemma where managers have to choose between investing in reducing state uncertainty, usually through monitoring, and exploitation of current knowledge (Type 1). The second type of problem assumes managers have no means of reducing uncertainty, that is, monitoring is not an option, but managers can invest in management

actions for which effectiveness vary across the state space (Type 2). The third type of problem is the family of adaptive management problems, where the dynamics of the system are not observable (Type 3).

4.1 | Type 1: Exploiting current knowledge or reducing uncertainty

The *Sumatran tiger* example illustrates the exploration (reduce state uncertainty) and exploitation (manage under current uncertainty) dilemma that most POMDP applications address. As previously discussed, every year managers must decide whether to invest their limited resources in: on-ground management activities to abate threats (manage); or monitor the status of a cryptic species' local population (survey); alternatively, they might choose to surrender (do nothing). Because the site has a local population of a relatively charismatic threatened species (e.g. *Sumatran tiger*), visitors of the area pay a small fee to visit the park, which generates a source of income every year. Finding the optimal decisions that maximise the revenue of the park over the long term (and therefore the persistence of the local population) is a POMDP problem. This problem was explored on one population (Chadès et al., 2008; see `SumatranTiger.pomdp` in our repository), two populations (McDonald-Madden et al., 2011, `Tiger2pop.pomdp`) and generalised to meta-populations of threatened and invasive species, as well as diseases, structured as networks ((Chadès et al., 2011; Dujardin et al., 2015), `tiger-metapopT10.pomdp`).

The key element of this type of problem is the ability to reduce uncertainty through at least one action that increases the ability to detect the state of the system. However, increasing detection usually comes at an additional cost which has to be balanced with resources that could target abatement of threats. This problem was further developed as a webapp (https://conservation-decisions.shinyapps.io/smsPOMDP/) to help managers of the Saving our Species program, in New South Wales (Australia) decide when to stop managing or surveying threatened species (Pascal et al., 2020).

4.2 | Type 2: Managing under imperfect detection

The second type of problem illustrates the dilemma that arises when state uncertainty cannot be resolved by implementing a specific monitoring action, rather it is usually assumed that some level of monitoring always occurs and provides a background detection rate (Memarzadeh & Boettiger, 2019; Regan et al., 2011). For example, in Regan et al. (2011), the presence of an invasive plant is uncertain. POMDP is used to inform when to (a) use a more efficient and more costly action (fumigation); (b) when to use a less efficient and less costly action (host denial crop) and finally (c) when to do nothing (pasture crop). The authors found that the optimal strategy depended on the ability to detect the invasive species and the risk of outside colonisation. This problem is defined as `weeds.pomdp` in our repository.

Because type 2 problems do not include actions that resolve uncertainty, it is particularly important to ensure that a POMDP formulation is worthwhile. This can be done by first solving the MDP formulation of this type of problem. If the MDP solution identifies several actions as optimal for the states that are uncertain (partially observable) and the difference in performance or costs between these actions matters significantly then POMDPs are likely to be a useful approach. If these conditions are not met, we do not recommend modelling the problem as a POMDP. This is because POMDPs can be significantly more complex to solve and interpret than MDPs.

4.3 | Type 3: Adaptive management problems

The third type of problems are those related to adaptive management (Chadès et al., 2012). Adaptive management problems can be modelled using POMDPs by augmenting the state space with a state variable that represents the unknown parameter (parameter uncertainty) or unknown system dynamics (model uncertainty). In most adaptive management problems, it is often assumed that a state variable is completely observable (Chadès et al., 2017). This type of POMDP is called Mixed Observability MDP (MOMDP; Ong et al., 2010). The state space now includes the unknown parameter or model as state uncertainty (see Supporting Information for modelling details).

For example, in Chadès et al. (2012), the authors illustrate how POMDP models can help adaptively manage a population of a threatened bird species, the *Gouldian finch*. The most pervasive threats to wild populations of *Gouldian finch* 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 were probability distributions describing how the population might respond to four alternative threat management actions. The objective was to implement the management action that was most likely to lead to a high persistence probability for the *Gouldian finch* population. In an adaptive management context, an optimal adaptive strategy will provide the best decision at each point in time by determining which action is optimal given updated beliefs on which expert's model is most likely the 'true' model, as observations are made over time (see Figure 2 for a simulation example). Solving this problem was possible by representing the state space with two state variables $S = X \times Y$, with $X = \{Low, High\}$ a completely observable state variable representing the local persistence of the *Gouldian finch* population, and $Y = \{Expert1, \ldots, Expert4\}$ a hidden state variable representing the state dynamics provided by experts (Chadès et al., 2012, see `gouldian4exp.txt`).

Other examples of POMDP applications to adaptive management problems include the protection of habitat for migratory shorebirds under uncertain consequences of sea-level rise (Nicol et al., 2015) and assessment of over-exploitation of fisheries (Memarzadeh

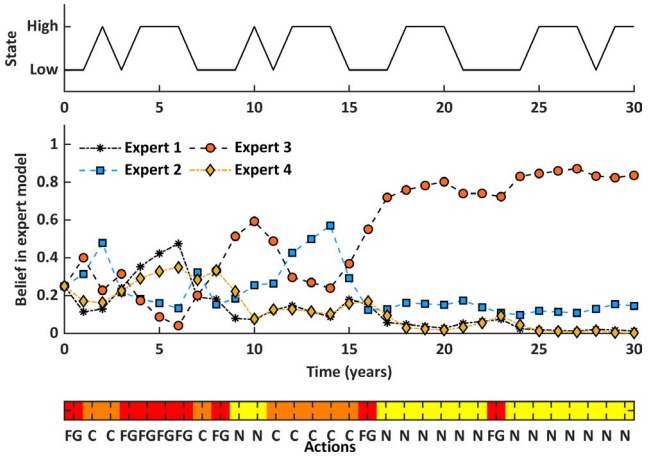

FIGURE 2 Simulation of an optimal policy for the *Gouldian finch* adaptive management problem solved using Symbolic Perseus (`gouldian4exp.txt`). Here, the simulation assumed a scenario where expert 3 was the true model. The top panel represents the observable state ('Low', 'High'). The middle panel represents the belief in experts' models updated over time. The last panel represents the optimal actions recommended and implemented over time (FG: Fire and grazing; C: Feral cats; N: Nesting box)

et al., 2019) or recreation management (Fackler et al., 2014). Readers interested in studies that include state uncertainty into adaptive management problems can refer to Fackler and Pacifici (2014) and Memarzadeh and Boettiger (2018).

5 | HOW TO SOLVE A POMDP PROBLEM?

5.1 | Casting a POMDP as a belief MDP

A consequence of including state uncertainty components into the problem definition is that making decisions based solely on a single observation leads to poor decision-making. Rather, the optimal decision at time t now depends on the complete history of past actions and observations. Optimisation requires us to search over all possible state-action histories to select the best policy; however, the space of possible histories grows exponentially with time, making it impossible to store explicitly. Because it is neither practical nor tractable to use the history of the action-observation trajectory to compute or represent an optimal solution, we keep track of belief states to summarise and overcome the difficulties of imperfect detection. A belief state b is a probability distribution over states. Intuitively, a belief state represents where we think we are at a given time. Given a belief state, $b(s)$ is the assigned probability of being in state s. B represents the set of belief states b. For example, in the *Sumatran tiger* example, a belief $b = \{0.5, 0.5\}$ means that there are equal probabilities for a local population to be extinct or extant. For the *Gouldian finch* problem (Figure 2), the initial belief is weighted at 0.25 across the models provided by the four experts.

Åström (1965) has shown that belief states are sufficient statistics to summarise all the observable history of a POMDP without loss of optimality. A POMDP can be cast into a fully observable MDP defined over the (continuous) belief state space. Intuitively, an optimal decision will depend on the probability distribution over the states of the system. Given an observation at time $t+1$ and an action implemented at time t, a belief state will be updated accordingly at each time step by applying Bayes' rule. Formally, for all $s', b^{ao'}(s')$ is the updated probability of being in state s' given that action a was performed and o' is observed:

$$b^{ao'}(s') = \frac{p(o' \mid a, s')}{p(o' \mid b, a)} \sum_{s \in S} p(s' \mid s, a) b(s). \quad (1)$$

with $p(o' \mid a, s')$ the POMDP observation function and $p(o' \mid b, a)$ the probability of observing o' given b and a calculated as $\sum_{s,s'} b(s)\, p(s' \mid s, a) p(o' \mid a, s')$, also called the normalising factor. The process of updating the belief state is conveniently Markovian as the next belief only depends on the current belief state, action and observation. Going back to the *Sumatran tiger* example, if we assume $b(\textit{extant}) = 1$ and action *manage* is implemented, two observations are possible. In the case of observation *present* then $b'(\textit{extant}) = 1$, that is, we are certain that the population is extant. However, if observation *absent* is received, the new belief state is now $b'(\textit{extant}) = 0.942$ and $b'(\textit{extinct}) = 1 - 0942 = 0.058$. In other words, in the absence of sighting, there is a small probability that the population might be extinct.

Solving a POMDP means finding a function $\pi: B \rightarrow A$ mapping a belief state ($b \in B$) to an action (e.g. an allocation of resources). Similar to the MDP case, this function is called policy in AI or strategy in applied ecology. We will assume that the optimisation criterion is to maximise the expected sum of discounted rewards over an infinite time horizon, that is, $\mathbf{E}\left[\sum_t^{\infty} \gamma^t R(b_t, a_t)\right]$ where b_t and a_t denote the belief state and action at time t, and $R(b_t, a_t) = \sum_s b(s)\, r(s, a)$.

For a given belief state b and a given policy π, the expected sum of discounted rewards is also referred to as the value function $V_\pi(b)$. A value function allows us to rank policies by assigning a value to each belief state b. A policy is optimal (denoted π^*), if its value function (V_{π^*}) is greater than or equal to the value of any other policy for any belief state. The value function can be computed using the dynamic programming equations for a POMDP represented as a belief MDP, that is, $\forall\, b \in B$:

$$V_{\pi^*}(b) = \max_{a \in A} \left[\sum_{s \in S} r(s, a)\, b(s) + \gamma \sum_{o' \in O} p(o' \mid b, a) V_{\pi^*}\left(b^{ao'}\right) \right], \quad (2)$$

where $b^{ao'}$ is the updated belief (Equation 1), $p(o' \mid b, a)$ the probability of observing o' given b and a calculated as $\sum_{s,s'} b(s)\, p(s' \mid s, a) p(o' \mid a, s')$.

An equivalent formulation uses the belief transition function τ, that is, $\tau(b, a, b')$ is the probability of transitioning to belief state b', given action a was implemented in belief state b:

$$V_{\pi^*}(b) = \max_{a \in A} \left[\sum_{s \in S} r(s, a)\, b(s) + \gamma \sum_{b' \in B} \tau(b, a, b')\, V_{\pi^*}(b') \right], \quad (3)$$

where the belief transition function can be calculated as $\tau(b, a, b') = \sum_{o' \in O} p(o' \mid b, a) 1(b' = b^{a,o'})$, that is, the sum of probabilities of all observations that lead to b'.

In both formulations, the optimisation problem consists of finding an optimal action for all values of b that maximises the sum of the immediate rewards (first term of the equation) and the expected future rewards (second term of the equation). While the first term of the equation is easy to calculate, the second term requires knowing the optimal values for all future belief states of all current belief states. Interested readers can refer to Williams (2011) for a step-by-step example of how to apply the stochastic dynamic programming algorithm *value iteration* to POMDPs. The main challenge that prevents efficient application of classic dynamic programming algorithms for MDPs is the continuous (belief) state space. At each iteration, we need to enumerate all the possible next belief states $b^{ao'}$ from a given belief state b. While we know that for each b there are a maximum of $|A|\,|O|$ next belief states, there is an infinite amount of belief states b because B is continuous. For this reason, algorithms have focused on proposing approximate or heuristic approaches to solve POMDPs rather than exact solutions (see Section 6).

5.2 | Solution representation

Most POMDP solvers provide two ways of representing a POMDP solution: through a set of α-vectors that represent the value function or through a graph that represents the policy directly. To understand the representation of these solutions, we first need to understand the concept of α-vectors and why they are critical to represent both value functions and the policy.

The value function for a POMDP is piecewise linear convex. This value function can be modelled arbitrarily closely as the upper envelope of a finite number of linear functions, known as α-vectors (Smallwood & Sondik, 1973; Sondik, 1978).

Figure 3 represents this concept with three α-vectors and the upper envelope is represented with a grey colour. Because of this, we can simply write $V_\pi = \{\alpha_1, \ldots, \alpha_n\}$, the value function defined over the full belief state space. Using this representation, we can also compute the value at a given belief b:

$$V_\pi(b) = \max_{\alpha \in \{\alpha_1, \ldots, \alpha_n\}} \alpha \cdot b, \tag{4}$$

where $\alpha \cdot b = \sum_{s \in S} b(s)\alpha(s)$ is the standard inner product operation in vector space. Given a set of α-vectors $\Gamma = \{\alpha_1, \ldots, \alpha_n\}$, the α-vector that maximises the value of belief state b is given by $\alpha_b = \arg\max_{\alpha \in \Gamma} b \cdot \alpha$. To each α-vector is associated an action, $a(\alpha)$.

Once a POMDP solution is calculated, solvers provide the set of α-vectors which can be used to plot the value function directly or queried for a given belief. In the case of the *Sumatran tiger*, we obtained a policy with 13 α-vectors (see Supporting Information APPL/SARSOP POMDP solver). Given the set of 13 α-vectors representing a POMDP solution ($\Gamma = \{\alpha_1, \ldots, \alpha_{13}\}$) and the current belief state b_t, we can derive the optimal action a^* to implement. If

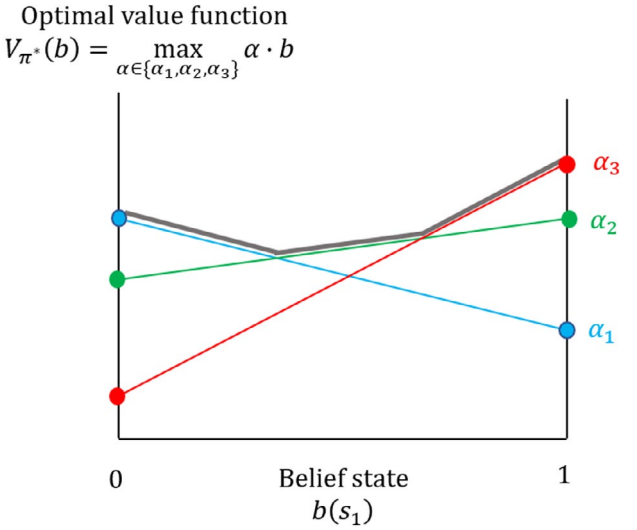

FIGURE 3 A two-state POMDP solution can be represented with the optimal value function (grey line). The x-axis represents the probability of being in state $s_1(b(s_1))$ from which we can easily derive the probability of being in state $s_2(b(s_2) = 1 - b(s_1))$. The y-axis represents the expected value of being in belief state b and implementing an optimal action $(V_{\pi^*}(b))$. Because the optimal value function is Piecewise Linear Convex (PWLC), it can be represented by a set of alpha vectors $(\{\alpha_1, \alpha_2, \alpha_3\})$. To each of these α vectors is associated an action $a(\alpha)$

the current belief state is $b_t = (0.8, 0.2)$, this means that we believe that with probability 0.8 the local tiger population is extant and with probability 0.2 it is extinct. Given Equation 4, we can derive the α-vector that maximises the value of belief state b_t. In this case, the α-vector that maximises the value of belief state b_t is α-vector 9 ($\alpha = (1400.18, -115.06)$) and $V_{\pi^*}(b_t) = 1097.132$. The action associated with this α-vector is *manage*.

A second way to represent a POMDP solution is a policy graph. Policy graphs are derived from the set of α-vectors and are usually provided by POMDP solvers as a text file coding a graph (e.g. Cassandra, 2015). Policy graphs are directed graphs with $|\Gamma|$ nodes (one node per α-vector) and at most $|\Gamma| \times |O|$ edges (one edge per observation per node, Figure 4).

Implementing an optimal policy now requires finding the starting node and following the policy graph:

1. Given an initial belief state b_0, identify the starting node α_{b_0} (Equation 4).
2. Implement the corresponding optimal action by evaluating $a(\alpha_{b_0})$.
3. At the next time step, get observation o';
4. Go to the node that corresponds to the observation and so on.

In the case of the *Sumatran tiger*, the policy graph has a conveniently simple form (Figure 5a) which can be interpreted by humans and further simplified (Figure 5b).

In summary, α-vectors are an important concept because many algorithms provide the set of α-vectors as solution. In addition, the total number of α-vectors $|\Gamma|$ is a good indicator of the difficulty of interpreting a POMDP solution using a policy graph (Figures 4 and 5).

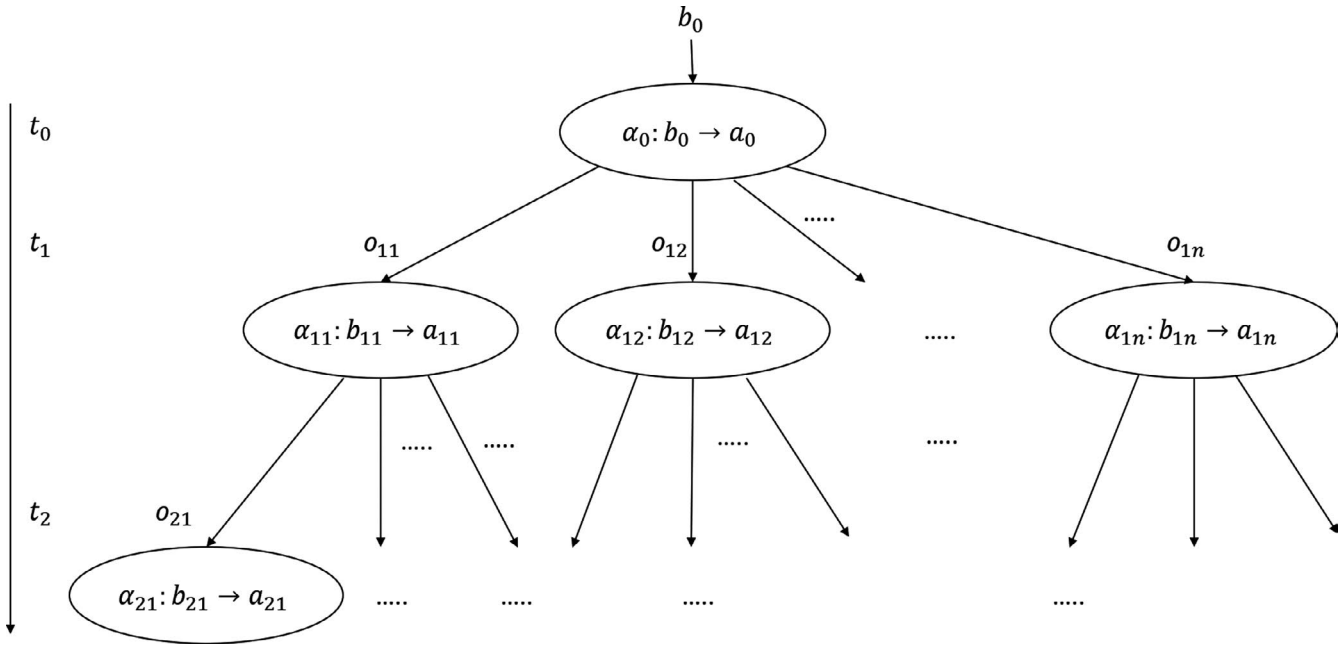

FIGURE 4 General representation of a POMDP policy graph. A policy graph is a directed graph where nodes identify the action to perform (also corresponding to an α-vector, Figure 3) and edges represent observations. In its general form, assuming an initial belief b_0 at t_0, the first action to implement is given by α_{b_0}. Once action a_0 is implemented, n observations are possible at time t_1. To each observation corresponds a new belief state from which an optimal action can be derived. While this figure represents the general form of a policy graph, some applications will exhibit policy graphs of small dimensions. For example, the *Sumatran tiger* policy graph has 13 nodes (α-vectors, see Figure 5). Unfortunately, policy graphs can easily reach several thousand nodes which makes interpretation of POMDP solutions challenging (e.g. 18,815 for the problem *Gouldian finch*, Figure 2; Table 3)

5.3 | Interpretation and visualisation

Although critical for uptake, visualising and interpreting POMDP solutions is a difficult endeavour. In low-dimensional problems (2 or 3 states), the value function can be represented graphically (see Figure 3). For problems with more than three states, where graphically representing the value function becomes impossible, a policy graph can provide an alternative (Figure 4). Unfortunately, policy graphs with large amounts of α-vectors rapidly become too dense to visualise and interpret. Of interest to readers are the published attempts at interpreting and representing optimal policies through study of the optimal solution. In Chadès et al. (2008), authors identified that the optimal policy had a structure that could be summarised in three action nodes (manage, survey and surrender) and links corresponding to thresholds related to 'not seen for *x* years' (Figure 5). In Regan et al. (2011), the authors also exploited the structure of the optimal policy using 'time since detection' as the main explanatory variable of the optimal solution. In Nicol and Chadès (2012), the authors simplified the optimal policy graph to 25 nodes by not accounting for low probability occurrences. In most cases, studying the sequence of optimal actions for a set of plausible scenarios (scenario analysis) was key to deriving a compact representation of the optimal policy. In some cases, modelling problems in a factored representation can help with interpreting models and solutions (Chadès et al., 2011; Hoey et al., 1999). Essentially, a factored representation identifies and takes advantage of conditional independence between variables akin to Bayesian network representation. Factored formulations result in (a) compact representation of the optimisation problems using trees and dynamic Bayesian networks, (b) implementation of efficient computation and (c) easier interpretation of solutions with decision trees used to represent optimal solutions in a structured way (see Methods section of Chadès et al., 2011). However, when the number of state variables becomes too large, factored policies also become too complex for humans to interpret.

6 | POMDP SOLVERS

6.1 | Family of algorithms

Perhaps the simplest approach to solve POMDPs consists in using discretised belief MDP methods. The belief space is discretised 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. Once discretised, the grid belief MDP can be solved using stochastic dynamic programming approaches (e.g. applying *value iteration* to solve Equation 3, Fackler, 2011). This technique has been studied to sidestep the intractability of exact POMDP *value iteration*, using either a fixed grid (Fackler, 2011; Lovejoy, 1991a) or a variable grid (Zhou

& 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.

One inefficiency of the grid approaches comes from the optimisation of the value function for a set of beliefs that will never be visited during policy implementation. That is, in some cases, there is no possible sequence of actions and observations that leads from an initial belief b_0 to a regular grid point. Hence, approaches such as point-based approaches optimise the value function for a set of reachable belief points (Kurniawati et al., 2008; Pineau et al., 2003; Shani et al., 2013).

6.2 | POMDP solvers

Here, we introduce a selected number of POMDP solvers. Explaining the algorithms implemented in these solvers is beyond the scope of this manuscript, rather we invite interested readers to refer to the excellent technical reviews already published (Kaelbling et al., 1998; Shani et al., 2013). Our Supporting Information explains how to install and use the solvers we selected (Supporting Information; Table 2).

Historically, one of the first toolboxes available, Cassandra's pomdp-solve, proposes five exact algorithms and an approximate grid-based algorithm. Of interest to users, pomdp-solve solves finite horizon problems with or without discounting. It provides several stopping criteria options. Most notably, Cassandra's toolbox provides the first definition of the POMDP input file '.pomdp' format called 'Tony's POMDP file format' (Cassandra, 2015). Most POMDP toolboxes have a parser that reads Tony's POMDP file format. Output files are of two types, value functions represented as a set of alpha vectors '.alpha' and a policy graph represented as a graph extension '.pg'. Both files can be opened in text editors. Cassandra's toolbox is implemented in C and benefits from a recent R wrapper (Table 2). pomdp-solve suits small size problems.

MDPSolve (Fackler, 2011) is a MATLAB toolbox for solving MDPs and POMDPs using dynamic programming. The toolbox was created to be used for a variety of management problems (Fackler & Haight, 2014; Fackler & Pacifici, 2014) and it applies the *value iteration* and *policy iteration* algorithms. POMDPs are solved by discretising the belief state space and interpolating over rectangular and simplex grids (Lovejoy, 1991a; Zhou & Hansen, 2001). Then, the discretised belief state problem is solved as an MDP. MDPSolve provides an extended documentation and user's guide. MDPSolve also suits small size problems.

Point-based approaches approximate the value function by updating it only for some selected belief states (Pineau et al., 2003; Shani et al., 2013; Spaan & Vlassis, 2005). Typical point-based methods sample belief states by simulating interactions with the environment and then updating the value function and its α-vectors over a selection of those sampled belief states. Point-based approaches are designed to tackle relatively large size problems.

Perseus is a point-based approach that performs approximate iterative improvement of the value function, ensuring that in each updating stage the value of each point in the belief set is improved. The key observation is that a single updating stage may improve the

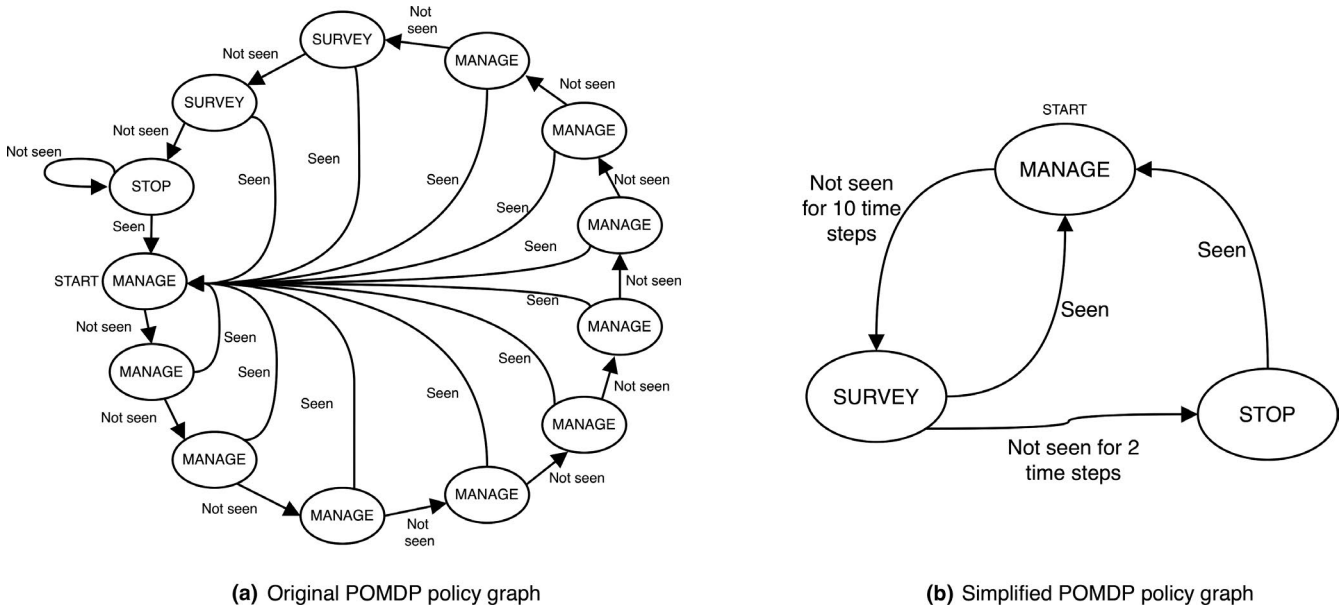

FIGURE 5 Two representations of the *Sumatran tiger* solution (Chadès et al., 2008; Pascal et al., 2020). (a) Policy graph generated by POMDP solvers with each vertex representing an α-vector and identifying an action. Edges correspond to observations. (b) The *Sumatran tiger* policy graph can be simplified because it exhibits an easy to understand solution, which can be expressed as 'how long should I manage or survey for in absence of sighting?'

TABLE 2 List of POMDP toolbox and brief description. We provide more information on how to use these toolboxes in supplementary information

Name	Description	Input format	Output
pomdp-solve	Cassandra's toolbox. Exact and approximate value iteration algorithms to solve POMDPs; http://www.pomdp.org/code/ (C) https://github.com/farzad/pomdp (R)	.pomdp	Text files .alpha .pg
MDPSolve	Discretise the belief state space and interpolate over rectangular and simplex grids Fackler (2011); https://github.com/PaulFackler/MDPSolve	.m files to specify models	MATLAB structure variables
Perseus	Perseus randomised point-based approximate value iteration algorithm to solve POMDPs Spaan and Vlassis (2005); http://www.st.ewi.tudelft.nl/mtjspaan/software/ (MATLAB)	Pomdp	global variables and .mat file
Symbolic Perseus	Perseus randomised point-based approximate value iteration algorithm to solve Factored POMDPs Poupart (2005); https://cs.uwaterloo.ca/ppoupart/software.html	.txt	.mat file
APPL/SARSOP	Efficient Point-Based POMDP Planning by Approximating (Successive Aproximations of the Reachable Space under Optimal Policies) Kurniawati et al. (2008); https://github.com/AdaCompNUS/sarsop (C++) https://github.com/boettiger-lab/sarsop/ (R)	.pomdpx (XML)	XML file

TABLE 3 Selected list of ecological POMDP problems and their characteristics (type of POMDP problems, the number of states (S), observations (O) and actions (A)). Problems were solved and evaluated using the APPL toolbox (SARSOP algorithm) over an infinite time horizon. $V_{\pi*}(b_0)$ represents the expected discounted sum of rewards calculated through simulations and $|\Gamma|$ represents the number of α-vectors that contribute to the optimal value function and policy graph. Input files defining problems are available at https://github.com/conservation-decisions/POMDPproblems

Problem	Type	$\lvert S\rvert, \lvert O\rvert, \lvert A\rvert$	$V_{\pi^*}(b_0)$	$\lvert\Gamma\rvert$	Comments
Tiger (Chadès et al., 2008)	1	2,2,3	1,405	13	Analytic approximation and webapp Pascal et al. (2020)
Tiger2pop (McDonald-Madden et al., 2011)	1	4,4,4	6.61	6	Comparison with MDP solutions
Weeds (Regan et al., 2011)	2	3,2,3	20.0	2	The finite horizon problem has 3 α-vectors
Gouldian4Exp (Chadès et al., 2012)	3	8,2,4	95.97	18,815	The *Gouldian finch* problem is the first application of POMDPs to adaptive management problems
Gouldian2Exp (Chadès et al., 2012)	3	162,81,4	75.32	1,782	Version of the *Gouldian finch* problem with larger completely observable state space and smaller number of hidden states

value of many belief points. In contrast to other point-based methods, Perseus updates only a (randomly selected) subset of points in the belief set, sufficient for improving the value of each belief point in the set (Spaan & Vlassis, 2005). Perseus is written in MATLAB, and has a MATLAB parser for Tony's POMDP file format. A closely related tool is Symbolic Perseus (Poupart, 2005), which uses a similar point-based algorithm to Perseus, but has more efficient mechanisms for factored POMDPs which are useful to model complex problems with several state variables (Williams et al., 2005).

Finally, APPL (Approximate POMDP Planning Toolkit) implements the SARSOP algorithm (Successive Approximations of the Reachable Space Under Optimal Policies). This solver is a point-based POMDP solver that samples the optimal reachable belief space (Kurniawati et al., 2008). APPL is coded in C++ and is recognised for its efficiency (SARSOP won the ICAPS 2011 planning competition, a major algorithm challenge for AI researchers). An R wrapper is available (Boettiger et al., 2020), making SARSOP an attractive option for many ecologists already familiar with the R environment. Of interest to readers is MO-SARSOP, the version of SARSOP that allows solution of Mixed Observability MDP by factoring state variables into completely observable state variables and partially observable variables (see Supporting Information). MOMDPs have been shown to be particularly useful in solving adaptive management problems (Chadès et al., 2012; Nicol et al., 2015; Péron et al., 2017). Table 3 shows the experimental results derived using SARSOP on a selected set of ecological problems.

7 | DISCUSSION

In this manuscript, we have identified three types of ecological problems that can be solved using POMDPs: (a) when to invest in reducing uncertainty? (b) what action is best given current uncertainty? and (c) how to optimise an adaptive management problem? Solving a POMDP is time-consuming and creates formidable challenges in developing both more efficient algorithms and more interpretable solutions. To increase the chance of success, the following steps summarise some of the lessons we have learned over the years.

Solving the completely observable version (MDP) of a POMDP problem is a very valuable step, especially for complex problems (Chadès et al., 2011). We found that there is little to no value in developing a partially observable version without a clear understanding of the simpler MDP version. For example, some important insights that can be gained from the MDP include: whether a single action dominates the optimal MDP policy (in which case there is no need for either MDP or POMDP); whether the reward function is appropriate or whether the states are appropriately discretised.

Reducing the state variables and observation variables to the smallest ensemble possible is critical to finding solutions that can be manageably represented, interpreted and explained. Gauging how far to reduce the ensemble is difficult because modelling is an art that improves with experience. The question about how best to develop a minimum yet accurate model has inspired some of our work. Our first attempt asked 'which states matter?' (Nicol & Chadès, 2012) where we used an automated online algorithm devised by Uther and Veloso (1998) to automatically select a tractable set of discrete states from a continuous space. The algorithm only split states where the additional state would significantly increase the quality of the optimal MDP policy, resulting in a compact discrete state space that was tractable to solve using POMDP approaches. More recently, Ferrer-Mestres et al. (2020) took the opposite approach of starting with all states of an MDP and aggregating states to a smaller number K while minimising the loss of performance.

Assessing performance through simulation of scenarios remains an essential step to demonstrate the value of POMDPs. Good practice includes comparing POMDP solutions with rules of thumb (Chadès et al., 2011), fixed decisions (Nicol et al., 2015) and completely observable conditions (McDonald-Madden et al., 2011; Memarzadeh et al., 2019). It is also worth discussing obvious links between the recommendations that POMDPs and value of information analysis (VoI) provide. VoI evaluates the expected gain from reducing uncertainty through some form of data collection exercise. As such, it is a tool which can be used to assess the cost-effectiveness of alternative research projects (Pratt et al., 1995; Wilson, 2015). Recently, applications of VoI in ecology have increased substantially (Canessa et al., 2015; Nicol et al., 2019; Runge et al., 2011; Xiao et al., 2019). VoI evaluates the expected gains assuming a two-stage process: first reduce uncertainty, then manage. The time and cost spent reducing uncertainty are often unaccounted for. In comparison, POMDPs assume reducing uncertainty could be an optimal action at any step of the decision process, that is, POMDPs solve a dynamic value of information problem. In this sense, POMDPs are more flexible and powerful than expected value of perfect or sampled information (EVPI/EVSI). We found that VoI calculations are more suited to provide fast prototyping recommendations (Nicol et al., 2018).

Several extensions of POMDPs have been proposed. Of interest to advanced readers, ρ-POMDPs define problems where the objective is also to maximise the gain of information (Araya et al., 2010). In that case, ρ-POMDPs differ from POMDPs in their reward function $\rho(b, a)$—rather than $r(s, a)$—that allows defining control and information-oriented criteria (see Fehr et al., 2018 for proposed algorithms). Another extension of POMDPs was proposed by Fackler and Pacifici (2014) to include observation causality (xPOMDP). This extension was motivated by problems that have either structural or observational uncertainty or both (Baggio & Fackler, 2016). Fackler and Haight (2014) have also investigated how the timing and informativeness of monitoring influence optimal strategies which we have not discussed in this manuscript.

We have limited this primer to finite and discrete state and action space POMDPs and we have not covered the fast increasing literature on solving continuous state, action or observation space POMDPs. These continuous POMDP approaches include density projection (Kling et al., 2017; Zhou et al., 2010), reinforcement learning and discretisation (Brechtel et al., 2013; Nicol & Chadès, 2012), point-based approaches (Porta et al., 2006), value approximation (Sunberg & Kochenderfer, 2018) or deep learning approaches (Igl et al., 2018; Karkus et al., 2017).

We hope that our primer provides a valuable entry point into understanding and solving more complex POMDPs. Automatically building interpretable solutions is an emerging area of research in AI also called explainable artificial intelligence (XAI). Traditionally, algorithms to solve POMDPs have focused on solving large state or action spaces. However, POMDP solutions with thousands of states are, in practice, difficult to represent and too complex for human decision-makers to understand. With the increasing demand on AI and ML algorithms to provide interpretable solutions (Rudin, 2019; Rudin & Radin, 2019), we believe the time has come to provide benchmarks of problems that can help assess the quality of XAI approaches for the ecology literature. Our recent efforts published in AI have focused on proposing algorithms to simplify POMDP models and solutions by reducing the size of the policy graph by pruning less important α-vectors (Dujardin et al., 2015, 2017; Ferrer-Mestres et al., 2021). But more needs to be done. To facilitate future work in this domain, we have created a github repository https://github.com/conservation-decisions/POMDPproblems to collate POMDP problems. We invite readers to contribute problems and ideas that will influence XAI research efforts and help design algorithms suited for the field of ecology.

ACKNOWLEDGEMENTS

A first unsuccessful attempt was made to write this primer a long time ago, we thank TJ. Regan, C. Hauser, YM. Buckley and TG. Martin for their support. We thank Gwen Iacona for providing feedback on an earlier version of this manuscript. We thank Paul Fackler and

anonymous reviewer 2 for their thoughtful comments. J.F.-M. was supported by a CSIRO Research Office Postdoctoral Fellowship to I.C. I.C. was supported by the CSIRO MLAI Future Science Platform (Activity Decisions). S.N. was supported by a CSIRO Julius Career Award.

CONFLICTS OF INTEREST

The authors have no conflict of interest to declare.

AUTHORS' CONTRIBUTIONS

I.C. designed the research; I.C. led writing of original manuscript with support from all authors; L.V.P., J.F.-M. and I.C. ran and performed the simulations. All authors edited and improved the manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13692.

DATA AVAILABILITY STATEMENT

We have created a github repository https://github.com/conservation-decisions/POMDPproblems to collate POMDP problems and Zenodo https://doi.org/10.5281/zenodo.5234598 (Chades et al., 2021).

ORCID

Iadine Chadès https://orcid.org/0000-0002-7442-2850
Luz V. Pascal https://orcid.org/0000-0002-5389-4461
Sam Nicol https://orcid.org/0000-0002-1160-7444
Cameron S. Fletcher https://orcid.org/0000-0001-5543-4330
Jonathan Ferrer-Mestres https://orcid.org/0000-0002-6647-9937

REFERENCES

Araya, M., Buffet, O., Thomas, V., & Charpillet, F. (2010). A POMDP extension with belief-dependent rewards. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems* (pp. 64–72). Curran Associates, Inc.

Åström, K. J. (1965). Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, *10*, 174–205. https://doi.org/10.1016/0022-247X(65)90154-X

Baggio, M., & Fackler, P. L. (2016). Optimal management with reversible regime shifts. *Journal of Economic Behavior & Organization*, *132*, 124–136. https://doi.org/10.1016/j.jebo.2016.04.016

Bellman, R. (1957). A Markovian decision process. *Journal of Mathematics and Mechanics*, *6*(4), 679–684. https://doi.org/10.1512/iumj.1957.6.56038

Bestelmeyer, B. T., Ash, A., Brown, J. R., Densambuu, B., Fernández-Giménez, M., Johanson, J., Levi, M., Lopez, D., Peinetti, R., Rumpff, L., & Shaver, P. (2017). *State and transition models: Theory, applications, and challenges* (pp. 303–345). Springer International Publishing.

Bestelmeyer, B. T., Brown, J. R., Havstad, K. M., Alexander, R., Chavez, G., & Herrick, J. E. (2003). Development and use of state-and-transition models for rangelands. *Rangeland Ecology & Management/Journal of Range Management Archives*, *56*, 114–126. https://doi.org/10.2307/4003894

Boettiger, C., Ooms, J., & Memarzadeh, M. (2020). *sarsop: Approximate POMDP Planning Software*. R package version 0.6.1. Retrieved from https://github.com/boettiger-lab/sarsop

Brechtel, S., Gindele, T., & Dillmann, R. (2013). Solving continuous POMDPs: Value iteration with incremental learning of an efficient space representation. In D. Sanjoy & M. David (Eds.), *International Conference on Machine Learning* (pp. 370–378). JMLR.org.

Brown, J. H., Whitham, T. G., Ernest, S. M., & Gehring, C. A. (2001). Complex species interactions and the dynamics of ecological systems: Long-term experiments. *Science*, *293*, 643–650. https://doi.org/10.1126/science.293.5530.643

Canessa, S., Guillera-Arroita, G., Lahoz-Monfort, J. J., Southwell, D. M., Armstrong, D. P., Chadès, I., Lacy, R. C., & Converse, S. J. (2015). When do we need more data? A primer on calculating the value of information for applied ecologists. *Methods in Ecology and Evolution*, *6*, 1219–1228. https://doi.org/10.1111/2041-210X.12423

Cassandra, A. R. (1998). A survey of POMDP applications. In *Working notes of AAAI 1998 fall symposium on planning with partially observable Markov decision processes* (Vol. 1724). Retrieved from http://www.cassandra.org/arc/papers/applications.pdf

Cassandra, A. R. (2015). The POMDP page. Retrieved from https://www.pomdp.org/

Chadès, I., Carwardine, J., Martin, T. G., Nicol, S., Sabbadin, R., & Buffet, O. (2012). MOMDPs: A solution for modelling adaptive management problems. In *AAAI Conference on Artificial Intelligence*. AAAI'12. AAAI press. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/4990

Chadès, I., Chapron, G., Cros, M.-J., Garcia, F., & Sabbadin, R. (2014). MDPtoolbox: A multi-platform toolbox to solve stochastic dynamic programming problems. *Ecography*, *37*, 916–920. https://doi.org/10.1111/ecog.00888

Chadès, I., Martin, T. G., Nicol, S., Burgman, M. A., Possingham, H. P., & Buckley, Y. M. (2011). General rules for managing and surveying networks of pests, diseases, and endangered species. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 8323–8328. https://doi.org/10.1073/pnas.1016846108

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. https://doi.org/10.1073/pnas.0805265105

Chadès, I., & Nicol, S. (2016a). Small data, big ideas. https://doi.org/10.5281/zenodo.164443

Chadès, I., & Nicol, S. (2016b). Small data call for big ideas. *Nature*, *539*, 31. https://doi.org/10.1038/539031e

Chadès, I., Nicol, S., Rout, T. M., Peron, M., Dujardin, Y., Pichancourt, J. B., Hastings, A., & Hauser, C. E. (2017). Optimization methods to solve adaptive management problems. *Theoretical Ecology*, *10*, 1–20. https://doi.org/10.1007/s12080-016-0313-0

Chades, I., Pascal, L., & Ferrer-Mestres, J. (2021). Conservation-decisions/POMDPproblems: Initial (v1.0). *Zenodo*, https://doi.org/10.5281/zenodo.5234598

Clark, C. W., Mangel, M. et al (2000). *Dynamic state variable models in ecology: Methods and applications*. Oxford University Press.

Dujardin, Y., Dietterich, T., & Chades, I. (2015). α-min: A compact approximate solver for finite-horizon POMDPs. In *International Joint Conference on Artificial Intelligence*. IJCAI'15 (pp. 2582–2588). AAAI Press.

Dujardin, Y., Dietterich, T., & Chadès, I. (2017). Three new algorithms to solve N-POMDPs. In *AAAI Conference on Artificial Intelligence*. AAAI'17 (pp. 4495–4501). AAAI Press.

Fackler, P. (2011). MDPSolve. https://github.com/PaulFackler/MDPSolve

Fackler, P. L., & Haight, R. G. (2014). Monitoring as a partially observable decision problem. *Resource and Energy Economics*, *37*, 226–241. https://doi.org/10.1016/j.reseneeco.2013.12.005

Fackler, P., & Pacifici, K. (2014). Addressing structural and observational uncertainty in resource management. *Journal of Environmental Management*, *133*, 27–36. https://doi.org/10.1016/j.jenvman.2013.11.004

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*, 1–14. https://doi.org/10.1371/journal.pone.0102434

Fehr, M., Buffet, O., Thomas, V., & Dibangoye, J. (2018). rho-POMDPs have lipschitz-continuous epsilon-optimal value functions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (pp. 6933–6943). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2018/file/de7f47e09c8e05e6021ababdf6bc58e7-Paper.pdf

Ferrer-Mestres, J., Dietterich, T. G., Buffet, O., & Chades, I. (2020). Solving K-MDPs. In *Proceedings of the International Conference on Automated Planning and Scheduling* (Vol. 30, pp. 110–118). AAAI press. Retrieved from https://ojs.aaai.org/index.php/ICAPS/article/view/6651

Ferrer-Mestres, J., Dietterich, T. G., Buffet, O., & Chades, I. (2021). *K-N*-MOMDPs: Towards interpretable solutions for adaptive management. In *Proceedings of the Association for the Advancement of Artificial Intelligence, Virtual conference* (17th edn, Vol. 35, pp. 14775–14784). AAAI Press.

Field, S. A., Tyre, A. J., & Possingham, H. P. (2005). Optimizing allocation of monitoring effort under economic and observational constraints. *The Journal of Wildlife Management*, *69*, 473–482.

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. https://doi.org/10.1016/j.reseneeco.2010.04.005

Hoey, J., St-Aubin, R., Hu, A., & Boutilier, C. (1999). SPUDD: Stochastic planning using decision diagrams. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. UAI'99 (pp. 279–288). Morgan Kaufmann Publishers.

Igl, M., Zintgraf, L., Le, T. A., Wood, F., & Whiteson, S. (2018). Deep variational reinforcement learning for POMDPs. In J. Dy & K. Andreas (Eds.), *Proceedings of the 35th International Conference on Machine Learning*. Proceedings of Machine Learning Research (Vol. 80, pp. 2117–2126). PMLR. Retrieved from http://proceedings.mlr.press/v80/igl18a/igl18a.pdf

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*, 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X

Karkus, P., Hsu, D., & Lee, W. S. (2017). QMDP-Net: Deep learning for planning under partial observability. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4694–4704). Curran Associates Inc.

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. https://doi.org/10.1016/j.jeem.2017.01.001

Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society*, *28*(2), 287–309. https://doi.org/10.2307/1907722

Kurniawati, H., Hsu, D., & Lee, W. S. (2008). Sarsop: Efficient point-based POMDP planning by approximating optimally reachable belief spaces.

Littman, M. L. (2009). A tutorial on partially observable Markov decision processes. *Journal of Mathematical Psychology*, *53*, 119–125. https://doi.org/10.1016/j.jmp.2009.01.005

Lovejoy, W. S. (1991a). Computationally feasible bounds for partially observed Markov decision processes. *Operations Research*, *39*, 162–175. https://doi.org/10.1287/opre.39.1.162

Lovejoy, W. S. (1991b). A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, *28*, 47–65. https://doi.org/10.1007/BF02055574

Marescot, L., Chapron, G., Chades, 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. https://doi.org/10.1111/2041-210X.12082

McDonald-Madden, E., Chades, 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. https://doi.org/10.1890/09-2075.1

Memarzadeh, M., & Boettiger, C. (2018). Adaptive management of ecological systems under partial observability. *Biological Conservation*, *224*, 9–15. https://doi.org/10.1016/j.biocon.2018.05.009

Memarzadeh, M., & Boettiger, C. (2019). Resolving the measurement uncertainty paradox in ecological management. *The American Naturalist*, *193*, 645–660. https://doi.org/10.1086/702704

Memarzadeh, M., Britten, G. L., Worm, B., & Boettiger, C. (2019). Rebuilding global fisheries under uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 15985–15990. https://doi.org/10.1073/pnas.1902657116

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. https://doi.org/10.1287/mnsc.28.1.1

Nichols, J. D., & Williams, B. K. (2006). Monitoring for conservation. *Trends in Ecology & Evolution*, *21*, 668–673. https://doi.org/10.1016/j.tree.2006.08.007

Nicol, S., Brazill-Boast, J., Gorrod, E., McSorley, A., Peyrard, N., & Chadès, I. (2019). Quantifying the impact of uncertainty on threat management for biodiversity. *Nature Communications*, *10*, 1–14. https://doi.org/10.1038/s41467-019-11404-5

Nicol, S., Buffet, O., Iwamura, T., & Chadès, I. (2013). Adaptive management of migratory birds under sea level rise. In *Twenty-Third International Joint Conference on Artificial Intelligence*. IJCAI'13 (pp. 2955–2957). AAAI Press. Retrieved from http://ijcai.org/Proceedings/13/Papers/434.pdf

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*. 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

Nicol, S., Ward, K., Stratford, D., Joehnk, K. D., & Chadès, I. (2018). Making the best use of experts' estimates to prioritise monitoring and management actions: A freshwater case study. *Journal of Environmental Management*, *215*, 294–304. https://doi.org/10.1016/j.jenvman.2018.03.068

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Ong, S. C., Png, S. W., Hsu, D., & Lee, W. S. (2010). Planning under uncertainty for robotic tasks with mixed observability. *The International Journal of Robotics Research*, *29*, 1053–1068. https://doi.org/10.1177/0278364910369861

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.

Péron, M., Becker, K. H., Bartlett, P., & Chadès, I. (2017). Fast-tracking stationary MOMDPs for adaptive management problems. In *Thirty-First AAAI Conference on Artificial Intelligence*. AAAI'17 (pp. 4531–4537). AAAI press.

Pineau, J., Gordon, G., Thrun, S. et al (2003). Point-based value iteration: An anytime algorithm for POMDPs. In *IJCAI* (Vol. 3, pp. 1025–1032).

Porta, J. M., Vlassis, N., Spaan, M. T., & Poupart, P. (2006). Point-based value iteration for continuous POMDPs. *Journal of Machine Learning Research*, *7*, 2329–2367.

Poupart, P. (2005). *Exploiting structure to efficiently solve large scale partially observable Markov decision processes* (Ph.D. thesis). University of Toronto.

Pratt, J. W., Raiffa, H., Schlaifer, R. O. (1995). *Introduction to statistical decision theory*. MIT Press.

Puterman, M. L. (2014). *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons.

Regan, T. J., Chades, I., & Possingham, H. P. (2011). Optimally managing under imperfect detection: A method for plant invasions. *Journal of Applied Ecology*, *48*, 76–85. https://doi.org/10.1111/j.1365-2664.2010.01915.x

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. https://doi.org/10.1111/1365-2664.12234

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x

Rudin, C., & Radin, J. (2019). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*, *1*. https://doi.org/10.1162/99608f92.5a8a3a3d

Runge, M. C., Converse, S. J., & Lyons, J. E. (2011). Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. *Biological Conservation*, *144*, 1214–1223. https://doi.org/10.1016/j.biocon.2010.12.020

Russell, S., & Norvig, P. (2002). Artificial intelligence: A modern approach.

Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research*, *34*, 871–882. https://doi.org/10.1287/opre.34.6.871

Shani, G., Pineau, J., & Kaplow, R. (2013). A survey of point-based POMDP solvers. *Autonomous Agents and Multi-Agent Systems*, *27*, 1–51. https://doi.org/10.1007/s10458-012-9200-2

Sigaud, O., & Buffet, O. (2013). *Markov decision processes in artificial intelligence*. John Wiley & Sons.

Smallwood, R. D., & Sondik, E. J. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, *21*, 1071–1088. https://doi.org/10.1287/opre.21.5.1071

Sondik, E. J. (1978). The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs. *Operations Research*, *26*, 282–304. https://doi.org/10.1287/opre.26.2.282

Spaan, M. T., & Vlassis, N. (2005). Perseus: Randomized point-based value iteration for POMDPs. *Journal of Artificial Intelligence Research*, *24*, 195–220. https://doi.org/10.1613/jair.1659

Sunberg, Z., & Kochenderfer, M. (2018). Online algorithms for POMDPs with continuous state, action, and observation spaces. In *Proceedings of the International Conference on Automated Planning and Scheduling* (Vol. 28).

Tomberlin, D. (2010). Endangered seabird habitat management as a Partially Observable Markov Decision Process. *Marine Resource Economics*, *25*, 93–104. https://doi.org/10.5950/0738-1360-25.1.93

Uther, W., & Veloso, M. (1998). Tree based discretization for continuous state space reinforcement learning. In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence*. AAAI '98/IAAI '98 (pp. 769–774). American Association for Artificial Intelligence. Retrieved from https://www.aaai.org/Papers/AAAI/1998/AAAI98-109.pdf

Walters, C. J. (1986). *Adaptive management of renewable resources*. Macmillan Publishers Ltd.

White, B. (2005). An economic analysis of ecological monitoring. *Ecological Modelling*, *189*, 241–250. https://doi.org/10.1016/j.ecolmodel.2005.03.010

White, C. C. (1991). A survey of solution techniques for the partially observed Markov decision process. *Annals of Operations Research*, *32*, 215–230. https://doi.org/10.1007/BF02204836

Williams, B. K. (2011). Resolving structural uncertainty in natural resources management using POMDP approaches. *Ecological Modelling*, *222*, 1092–1102. https://doi.org/10.1016/j.ecolmodel.2010.12.015

Williams, J. D., Poupart, P., & Young, S. (2005). Factored partially observable Markov decision processes for dialogue management. In *Proc. IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems* (pp. 76–82). AAAI press.

Wilson, E. C. (2015). A practical guide to value of information analysis. *Pharmacoeconomics*, *33*, 105–121. https://doi.org/10.1007/s40273-014-0219-x

Xiao, H., McDonald-Madden, E., Sabbadin, R., Peyrard, N., Dee, L. E., & Chadès, I. (2019). The value of understanding feedbacks from ecosystem functions to species for managing ecosystems. *Nature Communications*, *10*, 1–10. https://doi.org/10.1038/s41467-019-11890-7

Zhou, E., Fu, M. C., & Marcus, S. I. (2010). Solving continuous-state POMDPs via density projection. *IEEE Transactions on Automatic Control*, *55*, 1101–1116. https://doi.org/10.1109/TAC.2010.2042005

Zhou, R., & Hansen, E. A. (2001). An improved grid-based approximation algorithm for POMDPs. In *IJCAI* (pp. 707–716). Citeseer.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Chadès, I., Pascal, L. V., Nicol, S., Fletcher, C. S., & Ferrer-Mestres, J. (2021). A primer on partially observable Markov decision processes (POMDPs). *Methods in Ecology and Evolution*, 00, 1–15. https://doi.org/10.1111/2041-210X.13692