23rd International Congress on Modelling and Simulation, Canberra, ACT, Australia, 1 to 6 December 2019
mssanz.org.au/modsim2019

POMDPs for Sustainable Fishery Management

Jerzy A. Filar [a], **Zhihao Qiao** [b] **and Nan Ye** [c]

[a] *Centre for Applications in Natural Resource Mathematics, School of Mathematics and Physics, The University of Queensland*

[b] *Centre for Applications in Natural Resource Mathematics, School of Mathematics and Physics, The University of Queensland*

[c] *School of Mathematics and Physics, The University of Queensland*
Email: zhihao.qiao@uq.edu.au

Abstract: The challenge of sustainable fishery management is to design harvest policies that attain the dual objectives of: (a) protecting the species from over fishing, and (b) ensuring adequate economic return to fishers. It is clear that a suitable compromise between these two, conflicting, objectives must be achieved. However, a major difficulty stems from the need to deal with various sources of uncertainty associated with the fluctuations of the population, such as sea-surface temperature, pollution, or levels of nutrients. This is further complicated by the uncertainties associated with the effects of the management decisions and fishing pressure.

Partially Observable Markov Decision Processes (POMDPs) provide a natural mathematical framework for incorporating these uncertainties in the decision making process. This was already recognised by several authors. However, the promise of POMDPs has not yet been realised because they are provably computationally hard to solve in general, and for many years were considered to be solvable only for toy problems. In addition, the underlying dynamics of fish populations are normally described by deterministic difference or differential equations and it is not entirely clear how these should be incorporated into the stochastic dynamics of POMDPs.

This paper summarizes a, still preliminary, study that tackles both of the above problems. In particular, the computational complexity problem is tackled with the help of suitable discretization of state and action spaces and DESPOT; a state-of-the-art POMDP solver. In addition, the deterministic dynamics of the widely used Beverton-Holt model are modified to incorporate stochasticity in both the proliferation rate and in the observations based on catch and the outputs of the latter model.

The resulting POMDP formulation takes into account some of the uncertainties in managing fisheries, and shows that an adaptive management policy can be more advantageous than a simple fixed action policy. We also report on experiments with various modelling choices and their effects on the resulting policy.

Finally, recognising that POMDP policies are sometimes hard to interpret, we demonstrate that our adaptive management policy possesses an attractive feedback (or closed-loop) structure. Namely, the actions selected by that policy depend on the current expected biomass of the harvested species. Effectively, the policy maps the current expected biomass to a decision to use certain harvest levels in prescribed proportions. Naturally, when the expected biomass is low the more conservative (i.e., lower) harvest actions are preferred. On the other hand, when the expected biomass is high, actions corresponding to higher harvest levels are selected. Nonetheless, the most intensive (i.e., greedy) harvest levels are never selected because of the sustainability concerns.

Keywords: *POMDPs, sustainability, fishery management*

1 INTRODUCTION

The fishing industry plays an important socioeconomic role in Australia. However, developing predictable sustainable management policies is challenging and the decision-making process is slow. Thus there is a need to develop efficient decision-making techniques for sustainable fishery management (Queensland Department of Agriculture and Fisheries, 2017).

The challenge of sustainable fishery management is to design harvest policies that attain the dual objectives of: (a) protecting the species from over fishing, and (b) ensuring adequate economic return to fishers. It is clear that a suitable compromise between these two, conflicting, objectives must be achieved.

A fundamental difficulty lies in quantifying the uncertain impact of harvest policies on the population of the harvested species. This is compounded by other sources of uncertainty associated with the growth of the population and abiotic environmental factors such as seas surface temperature, pollution or turbidity. Partially Observable Markov Decision Processes (POMDPs) provide a natural and rigorous mathematical framework for incorporating such uncertainties in decision making. This was already recognised in (Lane, 1989). However, that promise has not yet been realised because POMDPs are provably computationally hard to solve in general, and for many years were considered to be solvable only for toy problems.

Fortunately, many advanced algorithms have been developed in the past two decades (Kurniawati et al., 2008; Silver and Veness, 2010; Ye et al., 2017). We believe that these advances have made POMDPs ripe for addressing various application challenges in general (Péron et al., 2017), and the problem of sustainable fishery management in particular.

In this paper, we investigate the application POMDPs for developing sustainable fishery management techniques. We review POMDP basics in Section 2 before introducing our POMDP models for sustainable fishery management in Section 3. Importantly, these incorporate the biological dynamics of a classical population model (Beverton and Holt, 1957). We present and discuss results of the simulation study in Section 4.

2 POMDPs

POMDPs provide a rigorous mathematical framework for modelling how an agent can make decisions in an uncertain environment. Fig. 1 illustrates interactions between the agent and the environment in a POMDP. At

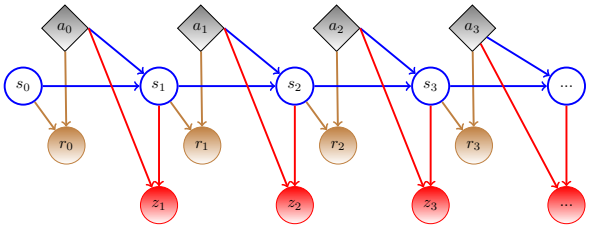


Figure 1. A POMDP models how an agent interacts with an uncertain environment.

each time step t, the environment is assumed to be in a state s_t belonging to a set of all feasible states S. The state is not fully observable, and thus at time t, the agent maintains a belief b_t on the state, which is a probability distribution on the feasible states. At time step t, the agent takes an action a_t from a set of feasible actions A, then the environment transits from current state s_t to a new state s_{t+1} with the probability $T(s_{t+1} \mid s_t, a_t)$. The agent receives a reward $r_t = R(s_t, a_t)$, and an observation o_t from a set of feasible observations O for the new state s_{t+1} with probability $Z(o_t \mid s_{t+1}, a_t)$. When the decision horizon is infinite, the agent's objective is to maximize the expected total discounted reward

$$V(b_0) = \mathbb{E}\left(\sum_{t=0}^{\infty} \gamma^t r_t \mid b_0\right), \tag{1}$$

where $\gamma \in (0, 1)$ is a discount factor, and the expectation is taken with respect to the uncertainties in the initial belief b_0, the state transition dynamics T, and the observation dynamics Z.

Given a current belief b, after the agent executes an action a and receives an observation o, the agent can compute the updated belief b' using

$$b'(s') = Pr(s' \mid a, o, b) = \frac{Z(o \mid s', a)Pr(s' \mid a, b)}{Pr(o \mid a, b)} \tag{2}$$

where $Pr(s' \mid a, b) = \sum_{s \in S} T(s' \mid s, a)b(s)$, and $Pr(o \mid a, b) = \sum_{s' \in S} Z(o \mid s', a)Pr(s' \mid a, b)$.

In this paper, we shall use the DESPOT algorithm (Ye et al., 2017), a current state-of-the-art solver for discrete-action discrete-observation POMDPs, to solve our proposed models.

3 POMDPs For Sustainable Fishery Management

In this section, we describe our fishery management POMDP model. We use both the biomass m and the proliferation rate ρ of the species as our state variable $s = (m, \rho)$, and the catch as the observation variable o. Both the biomass and the proliferation rate are often unknown or uncertain in practice. While fishery management authorities may implement various control measures, such as catch limits and limits to fishing vessel size, we focus on controlling the harvest rate as a preliminary study, that is, the target harvest rate is our action variable a. Below, we provide details on how we choose S, O, A, T, Z and R. We first describe the more natural setting with continuous states, actions and observations, and then describe how we can convert this to the case of discrete states, actions and observations, so that we can leverage current efficient POMDP solvers.

3.1 Continuous State-Action, Discrete Time Models

We describe the transition dynamics, the observation model, and the reward function below.

Transition dynamics. A key building block of our methodology is the Beverton-Holt model (BHM), a classic model used in modelling population growth of bio-species when there is no human intervention (Beverton and Holt, 1957). In the case of fisheries, BHM dictates that the biomass m_{t+1} at time $t+1$ is related to the biomass m_t at time t by

$$m_{t+1} = \frac{K\rho m_t}{(\rho - 1)m_t + K}, \quad t = 0, 1, 2, \ldots, \tag{3}$$

where $K \in \mathbb{R}^+$ is known as the carrying capacity (i.e. maximum population that can be supported by the ecosystem), and $\rho \in \mathbb{R}^+$ is the proliferation rate per generation. Often it is assumed that K and ρ are either known or can be reasonably accurately estimated. In this paper, we only consider the case $\rho \geq 1$. For the case $\rho < 1$, the population decreases even without any harvest, and to keep the population sustainable, we would actually need to consider "population boost" actions instead.

For POMDP modelling, we need to extend the above deterministic model to capture several sources of uncertainties. Both the biomass m_t and the proliferation rate ρ_t are usually not perfectly known for the state $s_t = (m_t, \rho_t)$. The proliferation rate of a species is generally assumed to be a constant, and thus we initially assume that ρ_0 is first chosen from some initial distribution, but remains unchanged afterwards, that is, $\rho_t = \rho$ for some constant ρ for all $t \geq 0$. For the population growth model, we extend the deterministic BHM to take into account actions and the fact that the population growth may deviate slightly from that prescribed by the BHM due to stochastic environmental factors. We thus extend the BHM to incorporate the multiplicative harvest rate $a_t \in [0, 1]$ and introduce the stochasticity to the following modified transition dynamics

$$m_{t+1} = \frac{K\tilde{\rho}(1 - a_t)m_t}{K + (\tilde{\rho} - 1)m_t}, \quad t = 0, 1, 2, \ldots, \tag{4}$$

where we now assume that $\tilde{\rho} = \rho e^{\xi}$, with ρ being the default constant proliferation rate, and $\xi \sim N(0, \sigma_d^2)$ being used to capture the uncertainty in population growth. Thus $\tilde{\rho}$ is now a random variable. This is derived under the simplifying assumption that we allow the population to first grow under the BHM, and then harvest a portion of a_t at the end of the time step t.

Observation model. We assume that the observation (i.e. the catch) $C_{t+1} = o_t$ is related to the state $s_{t+1} = (m_{t+1}, \rho_{t+1})$ and action a_t by

$$C_{t+1} = o_t(s_{t+1}, a_t) := \min\left(\frac{a_t m_{t+1}}{1 - a_t} e^{\psi_t}, \frac{m_{t+1}}{1 - a_t}\right), \tag{5}$$

where $\frac{a_t}{1-a_t} m_{t+1}$ is the amount of fish harvested assuming that we can achieve a harvest rate of exactly a_t, and $\psi_t \sim N(0, \sigma_o^2)$. The latter implies that the catch is assumed to be lognormally distributed, as is often done in fishery science, (e.g., see (Linton and Bence, 2011) or Haddon (2011) p. 88).

Reward function. For the reward function, in the first instance, we use a reward function that is simply proportional to the catch, that is,

$$R(s_t, a_t, s_{t+1}, o_t) = co_t, \tag{6}$$

where c measures the net earning per unit catch, and is equal to the price per unit catch minus the cost per unit catch. We can also use more sophisticated reward functions that explicitly take sustainability objectives into account, by incorporating a penalty term when the biomass falls below an undesirable threshold and a cost for switching actions. The resulting reward function has the form

$$R(s_t, a_t, s_{t+1}, o_t) = co_t + \text{pen}(m_{t+1}, \delta_p) - c_s \times |a_t - a_{t+1}|. \tag{7}$$

Here $\text{pen}(m_{t+1}, \delta_p)$ denotes a penalty incurred whenever the biomass m_{t+1} is below the target threshold δ_p. In this paper, we use the following penalty:

$$\text{pen}(m_{t+1}, \delta_p) = -c_p \sigma(m_{t+1} - \delta_p)\, \mathbb{I}(m_{t+1} < \delta_p). \tag{8}$$

where $c_p > 0$ is a constant and $\sigma(x) = 1/(1 + e^{-0.01x})$ is the sigmoid function. Note that a reward function of the form $R(s_t, a_t, a_{t+1}, s_{t+1}, o_t)$ is computationally expensive to deal with, and also differs from the simple reward function $R(s_t, a_t)$ introduced in Section 2. However, it can be easily reduced to the same functional form by taking expectations, as follows

$$R(s_t, a_t) = \mathbb{E}_{s_{t+1}, o_t | s_t, a_t} \left[R(s_t, a_t, s_{t+1}, o_t) \right]. \tag{9}$$

These two reward functions appear to be very different, but any policy has the same expected total discounted reward for both of them.

While Eq. (6) can take fuel cost and human labor cost into account, there are also costs associated with changing from one harvest rate to another, which is not proportional to the size of catches. For example, to enforce a change of a targeted harvest rate, the regulator may need to restrict the length of the fishing season or the number of permits, and this may require subsidies. We can also easily incorporate such action-switching cost into the reward function. We tried an additional cost term of the form $-c_s|a_{t+1} - a_t|$ where $c_s > 0$ is a constant. Naturally, this requires us to augment the state with a component storing the previous action, thereby increasing computational burden even further.

3.2 Discretized Models

Next, we describe how we discretize the continuous state-action POMDP. We use notations with hats to indicate that they are the discrete counterparts of the continuous POMDP. We first specify the set of discretized states, actions and observations.

Under BHM, the biomass will never exceed the carrying capacity K. Thus we divide $(0, K]$ into n_m equal intervals of length K/n_m, and use the discretized state $\hat{m}_i$ to denote the interval $\left(\frac{(i-1)K}{n_m}, \frac{iK}{n_m} \right]$ for $1 \le i \le n_m$. For the default proliferation rate ρ, we choose n_ρ proliferation rates defining the fecundity set $\mathcal{F} = \{ \hat{\rho}_1, \ldots, \hat{\rho}_{n_\rho} \}$. For the actions, we choose n_a harvest rates $\hat{a}_1, \ldots, \hat{a}_{n_a}$, where $\hat{a}_i = \frac{2i-1}{2n_a}$. For the observation, we follow a similar discretization as that for the state space. Specifically, we have n_o discretized observations $\hat{o}_1, \ldots, \hat{o}_{n_o}$, where $\hat{o}_i := \left(\frac{(i-1)K}{n_o}, \frac{iK}{n_o} \right]$, for $1 \le i \le n_o$.

Now we describe how we discretize the initial belief, the transition dynamics, the observation probabilities, and the reward function. Our discretization method can be applied to general POMDPs with continuous states, actions, and observations, and does not rely on the specifics of our continuous POMDP model for fishery management.

For the initial belief, the probability of the discrete state $\hat{s} = (\hat{m}, \hat{\rho})$ is the probability that the continuous state $s = (m, \rho)$ is consistent with the $\hat{s}$, or formally

$$\hat{b}_0(\hat{s}) = \mathbb{E}_{s \sim b_0} \big[\mathbb{I}(s \in \hat{s}) \big], \tag{10}$$

where $m \sim b_0$ indicates that s follows the distribution b_0 for the continuous biomass m. Here we use the notation $s \in \hat{s}$ to denote that s belongs to the discretized state $\hat{s}$, or specifically, $m \in \hat{m}$ and $\hat{\rho} \in \mathcal{F}$.

For the transition dynamics, the probability that $\hat{s} = (\hat{m}, \hat{\rho})$ transitions to $\hat{s}' = (\hat{m}', \hat{\rho}')$ is defined as the probability that a continuous state $s = (m, \rho)$ uniformly sampled from $\hat{s}$ transitions to a state $s = (m', \rho')$ belonging to $\hat{s}$, upon executing action $\hat{a}$. That is,

$$\hat{T}(\hat{s}, \hat{a}, \hat{s}') = \mathbb{E}_{s \sim U(\hat{s}), s' \sim p(\cdot|s,\hat{a})} \left[\mathbb{I}(s' \in \hat{s}') \right], \tag{11}$$

where $U(\hat{s})$ is the uniform distribution on the interval represented by $\hat{s}$, and $p(s' \mid s, \hat{a})$ is transition model for the continuous case, and is determined by Eq. (4) under the assumption that ρ in $\tilde{\rho} = \rho e^{\xi}$ remains constant.

For the observation model, we define

$$\hat{Z}(\hat{o} \mid \hat{s}', \hat{a}) = \frac{1}{D} \sum_{\hat{s}} \mathbb{E}_{s \sim U(\hat{s}), (s', o) \sim p(\cdot, \cdot|s, \hat{a})} \left[\mathbb{I}(s' \in \hat{s}', o \in \hat{o}) \right], \tag{12}$$

with the normalization constant $D = \sum_{\hat{s}} \mathbb{E}_{s \sim U(\hat{s}), (s', o) \sim p(\cdot, \cdot|s, \hat{a})} \left[\mathbb{I}(s' \in \hat{s}') \right]$, and $p(s', o \mid s, a)$ is the probability of transitioning to s' and observing o after executing action a in state s in the continuous model.

The reward function $\hat{R}(\hat{s}, \hat{a})$ is defined as the expected reward obtained when action $\hat{a}$ is executed on a continuous state s uniformly sampled from the discretized state $\hat{s}$, that is,

$$\hat{R}(\hat{s}, \hat{a}) = \mathbb{E}_{s \sim U(\hat{s}), (s', o) \sim p(\cdot, \cdot|s, \hat{a})} \left[R(s, \hat{a}, s', o) \right]. \tag{13}$$

There are no closed formulae for the expectations in general. We use simple Monte Carlo simulations to estimate the probability values, but we make a few improvements to efficiently compute the discretization. First, it is inefficient to estimate each value in $\hat{T}(\hat{s}, \hat{a}, \cdot)$ separately. Instead, we estimate all values together by drawing a large number of s and s', and counting the frequencies that s' falls into each discretized state, then we normalize the frequencies to obtain the discretized transition probabilities for $\hat{s}$ and $\hat{a}$. We can use a similar computation procedure to efficiently discretize the observation model as well. Second, while we presented our discretization in a very generic way, some of the computation can be simplified in our case. In particular, we can write down the discretized initial belief $\hat{b}_0$ in a closed form when we use a uniform distribution on the biomass as the initial belief in the continuous model. In addition, for the observation model, if o depends on s' only in the continuous model, then we can simplify the discretization procedure for the observation model so that we only need to sample s' from $\hat{s}'$, instead of sampling all the random variables s, s' and o.

4 SIMULATION STUDY AND DISCUSSION

In the simulation study we experimented with a wide range of modelling choices concerning rewards, beliefs and parameter values. We carefully chose the model parameters such that they are consistent with the values that are determined from real data reported in the literature. Due to space constraints, we describe the outputs of only a small sample of four experiments.

We first consider a basic model using the following initial belief and reward functions.

- B1: This scenario considers the case that we start with little knowledge about the species. The biomass m is uniformly drawn from $[0, K]$, and the proliferation rate ρ is equally likely to be one of the n_ρ values.
- R1: This is a simple profit model in which we simply take profit as proportional to the catch. The reward function is defined by Eq. (6) with $c = 1$.

We also consider three variants obtained by changing either the initial belief of the reward functions.

- B2: This scenario considers the case where there is already a very good estimate of the biomass obtained from extensive stock assessment studies. Initial biomass is assumed to be exactly equal to one half of the carrying capacity.
- R2: This scenario considers the case when action switching costs are taken into account. The reward function is given by Eq. (7) with $c = 1, c_s = 1500, c_p = 0$.
- R3: This scenario considers the case when incentives are given out to keep the biomass mass above a threshold. The reward function is defined by Eq. (7) with $c = 1, c_s = 0, c_p = 3000, \delta_p = 6000$.

Model	(R1, B1)	(R2, B1)	(R3, B1)	(R1, B2)
Total discounted reward	35532.3 (118.7)	35078.1 (134.7)	21113.7 (166.6)	35794.6 (101.4)

Table 1. Average total discounted rewards obtained by DESPOT over 500 simulations for the four different models. The number in the bracket is the standard error.

For the discrete models, we use $K = 10,000$, $n_m = 10$, $n_\rho = 3$, $n_o = 10$, and $n_a = 10$. We chose the 3 default proliferation rates as 5.8, 6.8 and 7.8, which are similar to the values used by Quinn II (2012). For the actions, a1 corresponds to 5% harvest rate, a2 to 15%, and so on.

To evaluate the total discounted value of a policy, we run each policy 500 times. Each run is over 10 time steps, and uses the discount factor of 0.95. For each time step DESPOT is given 5 seconds to search for an action.

Table 1 shows the total discounted rewards obtained by DESPOT on the above four models. Unsurprisingly, adding switching action costs and penalty costs decreases the discounted reward (see results for (R1, B1), (R2, B1) and (R3, B1)). The results also show if we start with more certainty in the initial belief, we can also achieve higher reward with more certainty, as seen by the slightly larger total discounted reward and slightly smaller standard error for (R1, B2) as compared to (R1, B1). Note that, the initial half carrying capacity assumption about the stock corresponds to maximum sustainable stock level in the classical logistic model. Since Eq. (3) is a form of discretization of the logistic model (e.g., see (Bohner and Warth, 2007)) this could, perhaps, be a reason why B2 leads to high performance.

4.1 Comparison of Adaptive Policy and Fixed Action Policies

Next, we compare the performance of the best adaptive policy produced by DESPOT under the base model (R1, B1) to fixed action policies which always harvest a fixed proportion of the biomass.

As expected, the optimal adaptive policy (blue curve) outperforms all the fixed action policies and that gap is increasing with the time horizon. Note that a greedy myopic policy, like harvesting at a constant rate of 75% (purple curve) performs well initially, but is not sustainable and eventually performs poorly when the time horizon is large. Similarly, a conservative policy, like harvesting at a constant rate of 45% (orange curve) is not able to fully exploit the resources available, and does not perform well either. Interestingly, there is a moderate fixed action policy of harvesting at 65% (red curve) that performs quite well as compared to the optimal adaptive policy, but is still sub-optimal. We expect that a fixed action policy will perform less favourably when there is more uncertainty over factors such as the proliferation rate of the species, and the environmental factors such as temperature and water acidity.

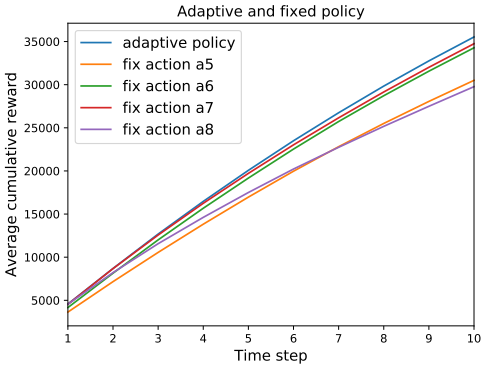


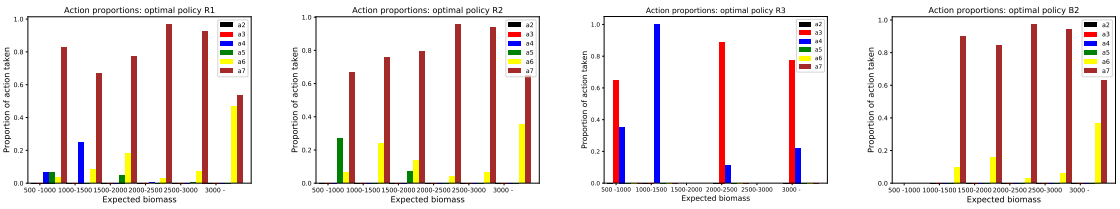
Figure 2. Comparison between the adaptive and fixed action policies for the base model (R1, B1).

4.2 Simplified Decision-Making Based on Expected Biomass

In general, it is difficult to interpret an optimal POMDP policy. We conjectured that for our application, there is a simple interpretation that the optimal policy is mainly making a decision based on the expected biomass at each time step.

To investigate this hypothesis, we visualize a policy by plotting the proportions of different actions taken by DESPOT for different expected biomass levels. These plots are shown in Fig. 3. The horizontal axis in the plots is labelled with the level of the expected biomass, and we only show the proportions of actions which are executed by DESPOT.

Fig. 3 (a) shows the plot for our basic model. We can see that when the expected biomass is small, some low harvest rate actions are executed. As the expected biomass increases, the proportion of low harvest rate actions gradually decreases, and the proportion of higher harvest rate actions increase. Eventually, the policy only uses



(a) (R1,B1) (b) (R2,B1) (c) (R3,B1) (d) (R1,B2)

Figure 3. Action proportions with different expected biomass

moderate harvest actions a6 and a7. In fact, the policy tries to maintain an equilibrium by executing actions a6 and a7. This is also consistent with our finding that fixed action policies of a6 or a7 are actually very close to the optimal policy.

Fig. 3 (b) and Fig. 3 (c) show the plots for when the reward function includes action switching cost and penalty for low biomass, respectively. We observe similar trend as for the base model. In addition, we can see that including an action switching cost reduces the number of different actions selected, and the inclusion of low biomass penalty eliminated the more greedy harvest actions a6 and a7. Finally, Fig. 3 (d) shows that the fixed initial biomass belief at half the carrying capacity level resulted in a relatively more stable use of actions a6 and a7. Surprisingly, perhaps, the relative frequencies of the use of a6 versus a7 fluctuate with the increases in the expected biomass. We suspect that this is because in this case, a6 and a7 are generally very similar, or possibly both are optimal actions for certain beliefs.

ACKNOWLEDGEMENT

We are indebted to Drs Sabrina Streipert, Yoni Nazarathy, Thomas Taimre and Marijn Jansen for several helpful discussions. This work was also partially supported by the ARC Discovery grant DP180101602.

REFERENCES

Beverton, R. J. H. and S. J. Holt (1957). *On the dynamics of exploited fish populations*, Volume 19 of *Fishery investigations (Great Britain, Ministry of Agriculture, Fisheries, and Food)*. London: H. M. Stationery Off.

Bohner, M. and H. Warth (2007). The beverton-holt dynaminc equation. *Applicable Analysis 86*, 1007–1015.

Haddon, M. (2011). *Modelling and quantitative methods in fisheries*, Volume 2nd e d. Boca Raton: CRC Press.

Kurniawati, H., D. Hsu, and W. S. Lee (2008). SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In *Proc. Robotics: Science and Systems*, Volume 62.

Lane, D. E. (1989). A partially observable model of decision making by fishermen. *Operations Research 37*(2), 240–254.

Linton, B. C. and J. R. Bence (2011). Catch at age assessment in the face of time-varying selectivity. *ICES Journal of Marine Science 68*, 611–618.

Péron, M., K. H. Becker, P. Bartlett, and I. Chadès (2017). Fast-tracking stationary momdps for adaptive management problems. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Queensland Department of Agriculture and Fisheries (2017). Sustainable fisheries strategy 2017–2027.

Quinn II, T. J. (2012). Population dynamics. In A.-H. El-Shaarawi and W. Piegorsch (Eds.), *Encyclopedia of Environmentrics* (2nd ed.). John Wiley & Sons Ltd.

Silver, D. and J. Veness (2010). Monte-Carlo planning in large POMDPs. *Advances in Neural Information Processing Systems 23*, 2164–2172.

Ye, N., A. Somani, D. Hsu, and W. S. Lee (2017). Despot: Online pomdp planning with regularization. *Journal of Artificial Intelligence Research 58*, 231–266.