Managing Interacting Species: A Reinforcement Learning Decision Theoretic Approach

Chadès[1], I., T.G. Martin[2], J.M.R. Curtis[3], and C. Barreto[1]

[1]INRA, UR875, Unité de Biométrie et Intelligence Artificielle, Toulouse France
[2]Centre for Applied Conservation Research, Forest Sciences, University of British Columbia, Canada
[3]Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada
Email: chades@toulouse.inra.fr

Keywords: *Decision Theory, Predator-prey, Reinforcement Learning*

EXTENDED ABSTRACT

Persistence of threatened species relies heavily on the effectiveness of conservation decisions. Yet, conservation strategies may generate positive and/or negative impacts on non-target species through direct (e.g. competition, predation), or indirect (e.g. habitat use) species interactions. Accounting for such interactions rarely occurs in conservation planning due to high biological uncertainty as well as the computational challenge of solving problems of this magnitude. Consequently, the simultaneous implementation of single-species management strategies for species that interact may jeopardize the recovery of one or more of the threatened species. Here we address these obstacles using a simulator and reinforcement learning approach. Reinforcement learning simplifies the representation of complex stochastic processes, and provides an intelligent way of exploring the solution search space. We apply this approach to two threatened species and compare optimal management strategies for ensuring species recovery and coexistence through their ranges.

INTRODUCTION

Recovery of threatened species relies in part on the effectiveness of management decisions. However, determining the most effective strategy remains a key challenge. Many recovery plans under relevant legislation focus on single species (Clark and Harvey 2002). In cases where threatened species interact directly through competition or predation and/or indirectly through sharing habitat, the recovery goals for one species may counteract those of another. A case in point is that of the sea otter *Enhydra lutris* and northern abalone *Haliotis kamtschatkana*, both listed under the Species at Risk Act (SARA).

Abalones are the preferred prey of sea otters and conflicts between recovery goals of sea otters and abalone are recognized (Gardner et al. 2000, Watson 2000, Gerber 2004). Yet to date, this conflict has not been examined quantitatively in an optimal management framework (e.g., Martin et al 2007). Here we employ a novel application of Reinforcement Learning to determine optimal management strategies for two threatened species which interact as predator and prey.

1. MARKOV DECISION PROCESSES AND REINFORCEMENT LEARNING

In the Artificial Intelligence community, Markov Decision Processes (MDP) and Reinforcement Learning (RL) are used to solve sequential decision making problems under uncertainty. In this paper we consider the case of non stationary Markov decision problems in a finite horizon.

Given a state-space X and an action-space A, the dynamic of a Markov decision process is characterized as follows: as a result of choosing action $a_t \in A$ in state $x_t \in X$ at decision epoch $t \in N$, the decision maker receives a reward $r_t(x_t,a_t)$ and the system state at the next decision epoch $x_{t+1} \in X$ is determined by the probability $p_t(x_{t+1}|x_t,a_t)$. A Markov decision problem is defined by adding to that process a performance criterion to maximize over a set of decisional policies. This criterion is a measure of the expected sum of the rewards along a trajectory, and policies are functions that indicate the action a_t to execute given information about the past trajectory at time t. For stationary infinite-horizon Markov decision problems, most performance criteria lead to the existence of stationary optimal policies, i.e. functions π that map states in X to actions in A (Puterman 1994). In finite-horizon problems such as ours, trajectories are sequences of exactly N transitions. The performance criterion considered in this case is the finite total expected reward criterion $V_\pi^N(x) \equiv E\left(\sum_{t=1}^{N-1} r_t(x_t,a_t) + r_N(x_N)\middle| x_1 = x\right)$ where E is the expected value given the policy π and starting state x.

When dealing with finite-horizon MDPs, the stationary assumption cannot be considered anymore; optimal policies are functions of time and state into action $\pi : N \times X \to A$. For this particular kind of MDP a policy π can be decomposed into a set $\{\pi_1,\pi_2,\ldots,\pi_N\}$ of policies $\pi_t : X_t \to A_t$. For each decision step t, a value function associated to π is defined as

$$V_\pi^i(x) = E\left(\sum_{t=i}^{t=N-1} r_t(x_t,\pi_t(x_t)) + r_N(x_N)\middle| x_i = x\right)$$

A policy π is optimal if it maximizes the value function V^{π}_1 on X_1. For this criterion, the classical Bellman optimality equations that characterize optimal policies are:

$$V_i^*(x) = \max_a \left\{ r_i(x,a) + \sum_{y \in X_{i+1}} p_i(y|x,a) V_{i+1}^*(y) \right\} \text{ for all}$$

$x \in X_i$, and $V^*_{N+1}=0$. This optimality equation has a single solution $V^*=\{V^*_1,\ldots,V^*_N\}$ that can be computed by a dynamic programming algorithm in $O(T.|A||S|^2)$ complexity when transition probabilities and the reward function are known (Puterman, 1994). However, the complexity of these approaches precludes its use when state or action spaces are large. Moreover, it cannot be applied when the transition probabilities or rewards are unknown. Here we deal with a complex ecological system with a non stationary stochastic process where estimating the transition probabilities by simulation is time consuming if not computationally impossible. RL algorithms are designed to overcome these difficulties by applying two principles: unknown quantities are estimated by means of simulation; large state or action spaces are handled through function approximation. The learning process estimates the optimal value functions V_t^* and the corresponding policies π_t^* from observed transitions and rewards.

We use the Q-learning (Watkins 1989) and R-learning (Schwartz 1993) algorithms adapted to non-stationary and finite horizon problems (Garcia et al, 1998). Both algorithms have similar general structure and can make use of function approximation. They iteratively compute the optimal value function: for each state x at each instant t, the optimal value $Q_t^*(x, a)$ of each decision a is estimated on the basis of simulated transitions. When all these values have been correctly estimated, the optimal policy can be derived through $\pi_t^*(x) = \arg\max Q_t^*(x,a)$. Q-learning is based on the Bellman optimality equation, replacing the $V_t(x)$ value function of a policy by a new function

$$Q_i^{\pi}(x,a) = r_i(x,a) + \sum_{y \in X_{i+1}} p_i(y|x,a) V_{i+1}^{\pi}(y)$$

and the optimality equation becomes

$$Q_i^{*}(x,a) = r_i(x,a) + \sum_{y \in X_{i+1}} p_i(y|x,a) \max_b Q_{i+1}^{*}(y,b)$$

To estimate these state/action values, the algorithm performs Bellman's updates on the basis of a sample of simulated transitions instead of the actual probabilities and rewards (Algorithm 1). If every action in each state at each instant is assessed an infinite number of times (*Kmax*) and if the learning rate (α) decreases, then $Q \rightarrow Q^*$ with probability 1. The SelectAction function which handles the classic exploration-exploitation trade-off (Sutton & Barto 1998) provides search efficiency by focusing on the most relevant state/action pairs.

R-learning can be seen as a parallel version of Q-learning. The updating rules take into account an estimate of the average expected reward per time step ρ defined as

$$\rho_{\pi}(x) = \frac{1}{N} E\left(\sum_{t=1}^{t=N-1} r_t(x_t, \pi_t(x_t)) + r_N(x_N) \middle| x_1 = x \right)$$

Algorithm 2 presents the corresponding updating rules. Although both algorithms converge to optimal policies, R-learning learns faster than Q-learning (Garcia et al 1998).

Algorithm 1 Finite horizon Q-Learning

for k=0 to *Kmax* **do**
  $x \leftarrow$ StartingState
  for t=1 to N **do**
    $a \leftarrow$ SelectAction()
    $(y, r) \leftarrow$ SimulateTransition(x,a)
    $Q_t(x,a) \leftarrow Q_t(x,a) + \alpha_t(x,a)\,(r + \max_b Q_{t+1}(y, b) - Q_t(x,a))$
    $x \leftarrow y$

Algorithm 2 Finite horizon R-Learning updating rules

$R_t(x,a) \leftarrow R_t(x,a) + \alpha_t(x,a)\,[r - \rho + max_b\, R_{t+1}(y, b) - R_t(x,a)]$
If $R_t(x,a) = max_a\, R_t(x, a)$ **then**
  $\rho \leftarrow \rho + \beta[r - \rho + max_b\, R_{t+1}(y, b) - max_a\, R_t(x,a)]$

2. ABALONE AND SEA OTTER MODELS

In order to learn the best multi-species conservation strategy for abalones and sea otters we built a simulator that incorporates the population dynamics of each species, their interaction, and how management decisions influence their threat level.

2.1. Abalone population model

We use a general size-structured matrix model for abalone published by Bardos et al (2006). Population structure at time-step n is represented as a vector $x(n)$ and is related to population structure at time n+1 using the stage-structured projection matrix:

$$\vec{x}(n+1) = \begin{pmatrix} g_{1,1}\,s_1 & 0 & 0 & 0 & f_5\,s_5 & f_6\,s_6 & f_7\,s_7 \\ g_{2,1}\,s_1 & g_{2,2}\,s_2 & 0 & 0 & 0 & 0 & 0 \\ g_{3,1}\,s_1 & g_{3,2}\,s_2 & g_{3,3}\,s_3 & 0 & 0 & 0 & 0 \\ g_{4,1}\,s_1 & g_{4,2}\,s_2 & g_{4,3}\,s_3 & g_{4,4}\,s_4 & 0 & 0 & 0 \\ g_{5,1}\,s_1 & g_{5,2}\,s_2 & g_{5,3}\,s_3 & g_{5,4}\,s_4 & g_{5,5}\,s_5 & 0 & 0 \\ g_{6,1}\,s_1 & g_{6,2}\,s_2 & g_{6,3}\,s_3 & g_{6,4}\,s_4 & g_{6,5}\,s_5 & g_{6,6}\,s_6 & 0 \\ g_{7,1}\,s_1 & g_{7,2}\,s_2 & g_{7,3}\,s_3 & g_{7,4}\,s_4 & g_{7,5}\,s_5 & g_{7,6}\,s_6 & s_7 \end{pmatrix} \begin{pmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \\ x_4(n) \\ x_5(n) \\ x_6(n) \\ x_7(n \end{pmatrix}$$

This model reflects the fecundity, mortality and growth of seven abalone size classes, where the population of stage i at time n is represented by $x_i(n)$, the growth matrix g whose elements $g_{i,j}$ are transition probabilities from class j to class i and s_j is the survival probability for an individual spending one time-step in class j. The f_i represent the fecundity of class i multiplied by the probabilities of fertilization and larval survival. The f_i are then multiplied by adult survival probabilities s_i, reflecting an assumption that spawning occurs at the end of each time-step. The resulting transition matrix is then modified to take into account density-dependent fecundity processes at various stages. We use Bardos et al (2006) fecundity values and the survival vector s=*(0.207,0.689,0.771,0.804,0.824,0.838,0.848)* to represent abalone dynamics in the absence of predation. For simplicity we express the abalone x_i in the projection matrix as densities.

Poaching of abalone is a key threat to population recovery (Gardner et al 2000, Jubinville 2000). In our model, we assume the pressure of poaching to be similar to that of commercial fishing, targeting only the largest size class (Bardos et al 2006). A stochastic process governs the probability of success of poaching removing 90% of class 7 with probability 0.75 and 70% with probability 0.25 every year. Simulation with poaching threat leads to 0.31 abalone per m² in the absence of sea otter predation.

2.2. Sea otter population model

Sea otter population dynamics were modelled using a Beverton-Holt model described by Gerber et al (2004). This model includes an asymptotic relationship between density and recruitment: $N_{t+1} = \frac{e^r K N_t}{e^r N_t - N_t + K}$ where K is the carrying capacity, N_t current population size, and r, intrinsic rate of increase. The parameters are based on a population of sea otters occurring in Washington State with K=612 and r=0.26 (Gerber et al 2004). While predators such as orcas, sharks and bald eagles do impact sea otters (e.g. Estes et al 1998), oil spills pose the single largest threat to sea otter populations

Table 1. Threat levels and three hypothetical functional responses defined for different densities of abalone and sea otter abundance. Functional response of abalones to sea otter predation is defined by the decrease in abalone survival rate (s) of stages 3-7, as L (Low) 5%, M (Medium) 15%, and H (High) 25%.

Abalone status (abalone m^{-2})	Endangered (<0.1)			Threatened (≥0.1-<0.3)			Special Concern (≥0.3-<0.5)			Not at Risk (≥0.5)		
Functional Response	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
Sea otter status (% of K)												
Not at Risk (>60)	L	H	L	L	H	M	M	H	H	H	H	H
Special concern (≤60->40)	L	H	L	L	H	L	M	H	M	H	H	H
Threatened (≤40->30)	L	M	L	L	M	L	M	M	M	H	M	H
Endangered (≤30)	L	L	L	L	L	L	M	L	L	H	L	M

(Ralls & Siniff 1990). In our sea otter model we implement a stochastic process by which oil spills may occur every 10 years on average with intensity varying from 0.4 to 0.2 reducing N by 20-40% (Gerber et al 2004).

2.3. States definition

We define X the set of states of our problem $X=\{X_a,X_s\}$ where X_a (*resp.* X_s) represents the set of states of our abalone population (*resp.* sea otter).

It is not feasible to consider continuous state MDP because of the computational complexity required for this problem. We therefore model the adult abalone population using a finite set of 20 states (X_a). Each state represents a range of 0.05 density which corresponds to one of four hypothetical categories of threat (Table 1).

We define X_s the set of states representing 10 levels of the sea otter population. Each state represents a 10% increment of the population's carrying capacity e.g. when K = 612, sea otter population is in state zero when its population abundance is in between 0 and 61 individuals. Additionally, sea otters are assumed to be in one of 4 threat classes (Table 1).

2.4. Decisions

Five management actions are considered and define the set A of actions: do nothing, reintroduce sea otters, enforce anti-poaching, control sea otters and the combined action of sea otter control and anti-poaching.

Because sea otters have yet to colonize much of their former range our conservation action for their recovery is the re-introduction of 93 sea otters (state 1 status *endangered*) into a hypothetical ecosystem. Subsequent to reintroduction, the population grows as described by our population model. This action can only occur once.

Anti-poaching enforcement is one of our conservation actions for abalone recovery. We model the effects of anti-poaching enforcement by stochastically reducing the survival of size class 7 by either 10% (with probability 0.75) or 30% (with probability 0.25, rather than the 90 - 70% reduction in survival assumed when poaching occurs). Thus, even when anti-poaching measures are implemented, poaching still occurs but at a reduced intensity.

We considered a third and more controversial conservation action; the direct removal of sea otters (Gardner et al. 2000), for example through the reinstatement of First Nation subsistence hunting. This action involves reducing the sea otter population by 3% of the carrying capacity each year, only when sea otters were in a *not at risk* state.

2.5. Interaction between sea otters and abalone

In the absence of a mathematical model describing the interaction between sea otters and abalone, we derived three hypothetical functional responses based on the literature. The first response assumes the threat from sea otters on abalone increases as abalone density increases. This response is assumed to be independent of sea otter density because at low sea otter densities, sea otters tend to prey heavily on large, high energy prey such as abalone (Table 1, F1). As sea otter density increases, individual otters tend to specialise on one of several prey species, however the impact on abalone remains significant (Estes et al 2003). The second functional response assumes sea otter quantities not abalone density drives the response. Here the level of predation imposed on abalone irrespective of abalone density will increase with increasing sea otter abundance (Table 1, F2). The third function is influenced by the abundance of both sea otter and abalone. It resembles a sigmoid response where the predation rate accelerates at first as prey density increases and then decelerates towards satiation at high prey densities (Table 1, F3). Sigmoid functional responses are typical of generalist predators, like sea otters, which readily switch from one prey species to another and/or which concentrate

Table 2. Two reward structures (R1, R2) for different combinations of species status.

Abalone status	Endangered		Threatened		Special Concern		Not at Risk	
Reward	R1	R2	R1	R2	R1	R2	R1	R2
Sea otter status								
Not at Risk	0	10	0	10	17	17	20	20
Special concern	0	7	0	7	0	14	17	17
Threatened	0	0	0	0	0	7	0	10
Endangered	0	0	0	0	0	7	0	10
Extirpated	0	0	0	0	0	7	0	10

their feeding in areas where certain resources are most abundant.

2.6. Rewards

To determine an optimal management strategy we describe a set of rewards that are achieved when certain states (i.e. threat levels) are realised (Table 2). Two types of arbitrary reward structures are used. The first provides a reward only when both sea otter and abalone populations are in *not at risk* or *special concern* states, with highest rewards when both are *not at risk*. Conversely the second reward structure provides rewards when at least one population is in a *not at risk* or *special concern* state and reflects a trade-off that may ensue if both populations cannot be maintained at levels outlined in the respective recovery strategies.

3. RESULTS

3.1. Management conditions under which both species can co-exist at various status levels

We used Q-learning and R-learning algorithms to learn optimal conservation strategies over a 50-year time horizon. Both reinforcement learning algorithms were run 500,000 times with 10 time-steps each to learn near optimal policies. Decisions were taken every 5 years. For all the following results we set the starting state by simulating abalone population dynamics in the presence of poaching but absence of sea otters.

Management scenario 1: Sea otter reintroduction and anti-poaching enforcement

We first examine two management actions currently being utilized in the northeast Pacific Ocean: reintroduction of sea otters and anti-poaching enforcement.

Under functional response 1, reward 1 and 2, we find it is optimal to introduce sea otters at the first time step (R). Anti-poaching enforcement (A) is then the optimal decision for the remaining time-horizon. For R1, only two valuable states are reached: when both species are at *special concern* (+14) or when abalones are at *special concern* and sea otters are at *not at risk* (+17). For R2, rewards are dominated by sea otter status. The abalone population oscillates between *threatened* and *special concern* due to changes in predation impact from low to medium at these threat levels (Table 1). Oil spills have no impact on the status of abalone.

Under functional response 2, reward 1 and 2, the optimal strategy is to first increase the level of abalone to *not at risk*. This level can be reached after 10 or 15 years of anti-poaching enforcement. Sea otters are then reintroduced. For reward 1 only, +14 and +17 are obtained and occur when both species are at *special concern* or when abalones are at *special concern* and sea otters are at *not at risk*. Here, the abalone population fluctuates with changes in sea otter population density as a result of oil spills because under functional response 2 predation pressure is related to sea otter density.

Under functional response 3, reward 1 and 2, the optimal strategy is always to introduce sea otters at the first time-step. Anti-poaching enforcement is then the optimal decision for the remaining time-horizon. Only two valuable states are reached: when both species are at *special concern* (+14) or when abalones are at *special concern* and sea otters are at *not at risk* (+17). Abalone density decreases with increases in sea otter population until the former reaches *threatened* or *endangered* status, when it becomes less impacted by sea otters. As defined by functional response 3, abalone density fluctuates in response to changes in sea otter population abundance which are driven by oil spills.

Table 3. Comparative performance for three management scenarios of Q-Learning (QL) and R-Learning (RL) algorithms.

	F. response 1		F. response 2		F. response 3	
	R1	R2	R1	R2	R1	R2
QL1	66.06	**111.85**	19.90	98.96	53.63	108.28
RL1	66.11	111.44	20.50	98.55	53.94	108.33
QL2	**66.29**	111.36	19.47	99.11	53.07	108.72
RL2	66.10	111.78	19.48	99.12	54.92	108.65
QL3	66.17	111.56	**25.39**	**100.06**	**84.67**	**119.95**
RL3	66.05	111.44	24.73	99.90	84.53	119.64

Management scenario 2: Including actions of sea otter reintroduction and control

Implementing sea otter reintroduction and removal does not improve the overall performance of optimal strategies (Table 3 QL2, RL2).

Management scenario 3: Including action of sea otter reintroduction and control, and anti-poaching simultaneously

The combined actions of sea otter reintroduction, control and anti-poaching outperform all other management scenarios. After sea otters are reintroduced during the first time step, population control is implemented when sea otters are *not at risk* with functional response 2 and 3 (Table 3, QL3, RL3; Figure 1). This scenario does not change the performance with functional response 1 as the threat to abalone is independent of sea otter density. This combined strategy allows abalone to increase in density until density-dependant effects come into play as defined under functional response 3.

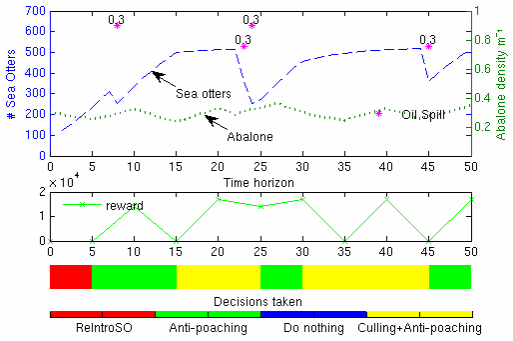

Figure 1. Simulation of optimal management under Functional response 3, reward 1.

Expected cumulated rewards for three management scenarios

The comparative performance of these strategies illustrates the differences between the different functional responses (Table 3). If our objective is to maintain both populations at the same time at *special concern* or better (reward 1) then strategies perform better with functional response 1. Optimal strategies under functional response 2 performed badly: it is unlikely that both species could be conserved at viable levels simultaneously. If our objective is to conserve at least one species and preferably both species (reward 2) best results are obtained with functional response 1 & 3.

4. DISCUSSION

We fail to find a management strategy which allows both species to co-exist at *not at risk* levels. There are several possible reasons for this. Firstly our models and assumptions may be unrealistic. In the absence of published information on abalone and sea otter functional responses we have assumed three contrasting responses based on the literature.

A second possibility for not achieving co-existence at *not at risk* levels is that the densities specified for recovery of abalone may be unrealistic in the presence of sea otters. Interestingly, Watson (2000) argues that sea otter recovery and abalone fisheries are mutually exclusive corroborating what we find here. We do find however, with anti-poaching enforcement and sea otter removal co-existence at *not at risk* (sea otter) and *special concern* (abalone) can be achieved and outperforms all other scenarios assessed here.

To our knowledge this is the first application of reinforcement learning for optimal management of interacting species at risk. There is a need to develop these methods further in order to increase our capacity to solve the complex management problems arising from species interactions such as those described here.

5. ACKNOWLEDGEMENTS

We are grateful to D. Bardos, L. Convey, L. Gerber, L. Nichol, T. Tinker, and J. Watson for their insights into sea otter and abalone populations, and N. Peyrard, F. Garcia, R. Sabbadin and O. Buffet for methodological discussions. This work has been supported by NSERC and Parks Canada (TM, JC), INRA (IC, CB).

6. REFERENCES

Bardos, D.C., Day, W.D., Lawson, N.T. & Linacre, N.A. (2006) Dynamical response to fishing varies with compensatory mechanism: An abalone population model. *Ecol. Modelling*, 192, 523-42.

Clark, J. and Harvey, E. (2002) Assessing Multi-Species Recovery Plans under the Endangered Species Act *Ecol. Applications*, Vol. 12(3),,655-662

CITES (2007). Appendices I, II, and III. CITES, Geneva, Switzerland. Available at:

http://www.cites.org/eng/app/appendices.pdf (Accessed 6 August 2007).

COSEWIC. (2000). COSEWIC assessment and update status report on the sea otter Enhydra lutris in Canada. In, p v + 17pp. Committee on the Status of Endangered Wildlife in Canada, Ottawa.

Estes, J.A., Tinker, M.T., Williams, T.M. & Doak, D.F. (1998) Killer whale predation on sea otters liking oceanic and nearshore ecosytems. *Science*, 282(16), 473-76.

Estes, J.A., Riedman, M.L., Staedler, M.M., Tinker, M.T. and Lyon, B.E. (2003). Individual variation in prey selection by sea otters: patterns, causes, and implications. *J An. Ecol.* 72,144-155.

Fanshawe, S., Vanblaricom, G.R. & Shelly, A.A. (2003) Restored Top Carnivores as Detriments to the Performance of Marine Protected Areas Intended for Fishery Sustainability: a Case Study with Red Abalones and Sea Otters. *Cons. Bio.* 17(1), 273-83.

Fisheries and Oceans Canada (2007). Recovery strategy for the Northern Abalone (*Haliotis kamtschatkana*) in Canada [Proposed]. Species at Risk Act Recovery Series. In (ed Fisheries and Oceans. Can), p vi + 31pp.

Garcia, F., Ndiaye, S.M. (1998) A learning rate analysis of reinforcement learning algorithms in finite-horizon. In *Proc. of ICML'98*.

Gardner, J., Griggs, J. & Campbell, A. (2000). Summary of a strategy for rebuilding abalone stocks in British Columbia. In Workshop on rebuilding abalone stock in British Columbia (ed A. Campbell), Vol. 130, pp. 151-55. *Can Spec. Publ. Fish. Aquat. Sci.*

Gerber, L.R., Buenau, K.E. & VanBlaricom, G.R. (2004) Density dependence and risk of extinction in a small population of sea otters. *Biod. and Cons.*, 13, 2741-57.

Jubinville, B. (2000) Enforcing the fishery closure for northern (pinto) abalone (*Haliotis kamtschatkana*) in British Columbia. *In* Workshop on Rebuilding Abalone Stocks in British Columbia. *Can Spec. Publ. Fish. Aquat. Sci.* 130. pp. 52.

Laidre, K.L., Jameson, R.J., Jeffries, S.J., Hobbs, R.C. & Bowlby, C.E. (2002) Estimates of carrying capacity for sea otters in Washington state. *Wild. Soc. Bull.*, 4, 1172-81.

Martin, T.G., Chadès, I., Arcese, P., Possingham, H.P., Marra, P. & Norris, D.R. (2007) Optimal conservation of migratory species. *PLoS ONE*.

Puterman, L.M. (1994). *Markov decision processes.* John Wiley and Sons, New York.

Ralls, K. & Siniff, D.B. (1990). Sea otters and oil: Ecological perspective. In *Sea mammals and oil: confronting the risk*, pp. 199-209. Academic Press Inc., San Diego.

Schwartz, A. (1993). A Reinforcement Learning Method for Maximizing Undiscounted rewards. In *ICML*, vol. 10.

Sutton, R.S., Barto, A.G. (1998) *Reinforcement Learning: An introduction*. MIT Press, Cambridge.

Taylor, B. L., and DeMaster, D. P. (1993) Implications of non- linear density dependence. *Mar. Mamm. Sci.* 9(4):360- 371.

Watkins, C. (1989) *Learning from delayed rewards*. PhD Thesis, Cambridge University, Cambridge, England.

Watson, J. C. (2000). The effects of sea otters (*Enhydra lutris*) on abalone (*Haliotis* spp.) populations. *In* Workshop on rebuilding abalone stocks in British Columbia. *Edited by* A. Campbell. *Can. Spec. Publ. Fish. Aquat. Sci. 130*. pp. 123-132.