Optimizing Sea Lice Management in Norwegian Salmon Aquaculture Using POMDPs

Olivia Beyer Bruvik
Stanford Univeristy
oliviabb@stanford.edu

***Abstract*—Sea lice (*Lepeophtheirus salmonis*) infestations are a persistent threat to salmon aquaculture, contributing to fish mortality, economic loss, and environmental harm. While treatments such as mechanical removal and chemical baths can reduce parasite load, they are costly, stressful to fish, and risk fostering resistance. Selecting optimal treatment strategies remains challenging due to uncertain infestation dynamics and noisy monitoring data.**

In this work, we develop a model-based reinforcement learning framework using Partially Observable Markov Decision Processes (POMDPs) to optimize sea lice treatment policies at the pen level. We evaluate three dynamic programming solvers—Value Iteration, QMDP, and SARSOP—against heuristic and random baselines. Simulations across a range of cost-weight trade-offs demonstrate that POMDP-based policies significantly reduce lice levels while maintaining economic efficiency. This study highlights the potential of data-driven decision support for precision aquaculture, offering scalable strategies tailored to site-specific dynamics using real-world data.

I. Introduction

A. Background on Sea Lice

Sea lice, parasitic copepods such as *Lepeophtheirus salmonis*, are a significant challenge in aquaculture, particularly salmon farming. These ectoparasites are persistent ectoparasite copepods that attach to Atlantic salmon worldwide and cause lesions, leading to reduced fish welfare and harvest biomass. The life cycle of sea lice includes planktonic and parasitic stages, with population growth often exhibiting exponential dynamics. Effective management of sea lice is critical to maintaining fish health and ensuring the sustainability of salmon farming.

The impacts of sea lice on salmon aquaculture have been extensively studied due to their significant economic and environmental implications [2]. A study of sea lice infestations in all Norwegian salmon farms over 84 months suggested that an average infestation over a typical central region spring-release cycle generates damages of USD 0.46 per kg of harvested biomass, equivalent to 9% of farm revenues [1]. These losses are driven mainly by treatment costs, reduced fish growth, and reduced food conversion efficacy [1]. Beyond economics, sea lice pose ecological risks by transferring from farmed to wild fish populations, threatening biodiversity. These impacts demonstrate the need for robust, cost-effective management strategies that balance fish welfare with operational constraints.

B. Problem Significance

Managing sea lice is inherently complex due to uncertainty in monitoring and the high costs of interventions. Lice counts are typically obtained by computer vision methods or manual sampling, which introduces observation noise, making it difficult to accurately assess infestation levels. Traditional monitoring methods rely on manual inspection of small samples, often yielding inaccurate estimates.

Farmers operate under constraints set by the Norwegian government that regulate the max average adult female sea lice count per fish allowed without mandated treatment regiments. Mechanical delousing offers a key treatment method, but is costly, has variable efficacy, and harms fish welfare. The need to balance the benefits of keeping sea lice levels below with the economic and ecological impacts of treatment methods underscore the need for optimizing treatment timing and frequency.

The emergence of precision aquaculture has led to increased availability of monitoring data [15], [21]. However, translating these data into effective treatment decisions remains a challenge. Consequently, despite advances in AI and computer vision for lice monitoring, no analytical decision-making frameworks have been developed for treatment optimization. As a result, Norwegian salmon farmers continue to rely heavily on intuition and observational insights, often relying on heuristic approaches such as threshold-based treatments. Although simple, these methods fail to account for observation uncertainty or the stochastic nature of lice population dynamics, leading to suboptimal decision that either overtreat (increasing costs) or undertreat (allowing infestations to escalate). Lice level management is, consequently, a complex problem that necessitates dynamic decision-making in response to the current lice levels and lice growth rates.

C. Research Gap

Partially Observable Markov Decision Processes (POMDPs) provide a powerful mathematical framework for sequential decision making under uncertainty, combining stochastic transitions, noisy observations, and a reward structure to optimize actions. While POMDPs have been applied to invasive species management [22], their potential to optimize sea lice treatment strategies remains unexplored. Nonetheless, the nature of lice prevention and treatment strategy as a series of decisions for an environment with high levels of uncertainty renders it a suitable application for POMDPs. [3]

Despite their potential, POMDPs have not been applied to sea lice management. Existing studies often rely on mathematical models and Markov Decision Processes (MDPs), which assume perfect state knowledge despite the observational noise inherent to sea lice monitoring. Furthermore, while exponential growth is a hallmark of sea lice populations, few models explore log-normal state spaces to capture this dynamic accurately. This research addresses these gaps by developing a POMDP-based approach that incorporates observation uncertainty, exponential growth dynamics, and data-driven parameterization.

D. Research objectives

The primary objective is to develop a POMDP framework for sea lice management, modeling sea lice levels as states, treatment decisions as actions, and noisy counts as observations. We implement discrete POMDPs for policy computations and continuous POMDPs for simulations.

A secondary objective is to determine the optimal treatment policies using POMDP solvers and compare their performance through simulations. We explore three dynamic programming algorithms to determine the optimal treatment strategy for the given POMDP, including Value Iteration with a fully observable state space, Successive Approximations of the Reachable Space under Optimal Policies (SARSOP), and Q-value Markov Decision Processes (QMDP). We compare these policies against a heuristic baseline, assessing tradeoffs between treatment costs and lice levels across a range of economic priorities.

II. Literature Review

A. Prior Modeling Efforts

Recent years have seen increasing application of mathematical optimization techniques to aquaculture management. Studies have employed various methods including dynamic programming [16], mixed-integer programming [23], additive models [20], and stochastic optimization [24] to address challenges in fish farming. However, these approaches often rely on simplified models of complex biological systems that may not capture the full dynamics of sea lice populations.

Recent work has begun to explore machine learning approaches for predicting sea lice abundance [13] [7], state-space modeling for intersite spread of sea lice [8], [6], and simple models to inform area management of sea lice parasitism [14]. However, few studies have addressed optimization of treatment strategies based on these predictions.

We can look to adjacent fields like agricultural systems to appreciate the utility of reinforcement learning to inform control methods in farming. Several studies have demonstrated the potential of reinforcement learning in livestock management [19], crop disease control [9], and invasive species management [22], suggesting its applicability to parasite management in aquaculture.

B. Overview of Decision Making Under Uncertainty

An MDP provides a mathematical framework for sequential decision making under uncertainty, where all the uncertainty arises from outcomes that are partially random and partially under control of a decision maker. A POMDP is a more general decision making problem in which the agent is not sure what state they are in. In POMDPs, exact observations of the state are replaced with a probabilistic relationship with the state [10]. POMDPs have important applications across domains, including aircraft collision avoidance, automated driving, breast cancer screening, financial consumption and portfolio allocation, and distributed wildfire surveillance [10]. The success of POMDPs lies in their ability to handle partial observability and optimize long-term rewards, which enable them to model complex systems with uncertain state and transition dynamics more realistically.

POMDPs have been applied to invasive species management [22], where the population size is rarely precisely known. By incorporating uncertainty in the population estimate, POMDPs led to more realistic and effective management policies. Despite a known high noise in sea lice population measurements, POMDP models for sea lice management have not been widely explored.

C. Policy Optimization Methods

Solving POMDPs remains a computations challenge, especially in continuous and high-dimensional belief spaces. Several algorithmic approaches have been developed to generate optimal policies, each with distinct assumptions and tradeoffs.

Traditional value iteration for POMDPs operates on discretized representations of belief states, computing the exact value function through dynamic programming updates. When applied to the underlying MDP, it yields exact policies for fully observable problems. While value iteration guarantees convergence to the optimal value function through repeated updates, it does not handle partial observability directly [10]. Furthermore, value iteration does not scale well to larger or continuous state spaces.

In 1995, Littman et al. addressed the scaling issues associated with traditional value iteration by introducing a novel hybrid approach that uses the underlying MDP optimal value function as an aid in POMDP solution, but still appreciates partial observability at the decision step while assuming full observability after that step [12]. While Littman et al. were able to obtain high quality policies for a class of POMDP's with nearly 100 states, the authors note that other techniques will be needed to handle the thousands of states needed to address realistic problems [12].

SARSOP seeks to overcome the curse of dimensionality by exploiting the notion of *optimally reachable belief spaces* to improve computation efficiency. It handles partial observability at every step but offers efficient point-based POMDP planning by approximating optimally reachable belief spaces [11]. While SARSOP produces near-optimal policies with significant computational savings over full belief space enumeration, it applies only to discrete state, action, and observation spaces, necessitiating discretization of continuous spaces.

III. Problem Formulation

We model the process as a POMDP, enabling the modeling of uncertainty in both system dynamics and observations. The POMDP model consists of a one-dimensional state space that represents the level of sea lice in a log space. We utilized log space to linearize the transition dynamics given the exponential nature of the growth of sea lice populations.

We discretize the state space to accommodate the discrete nature of the SARSOP and QMDP solvers; however, we utilize a continuous state space during simulations. We will refer to these two POMDP models as our "Discrete POMDP" and "Continuous POMDP" going forward.

We define the POMDP as a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{O}, \gamma)$, where

- $\mathcal{S}$: The state space, representing the level of the sea lice in log space.
- $\mathcal{A}$: The action space, representing treatment decisions.
- $T(s'|s, a)$: The transition model, giving the probability of transitioning to state s' after taking action a in state s.
- $R(s, a)$: The reward function, representing the expected reward received when executing action a from state s.
- $O(o|a, s')$: The observation function, representing the probability of observing o, given that we took action a and transitioned to state s'.
- $\gamma \in [0, 1)$: The discount factor.

A. State Space Definition

Let state s_t at time t be defined as the current sea lice level in log space at time t: $s_t = \log(N_t)$, where N_t is the absolute sea lice level. We constrain the sea lice level range from $1e - 3$ to 10, which reflects the most common range of sea lice values observed in historical data from 2014 to 2025 under the assumption that the complete absence of sea lice is impossible. Since log is a monotonically increasing function, we have a state space ranging from $\log(1 - 3)$ to $\log(10)$.

Due to the discrete nature of the solvers, we discretize the state space of sea lice levels when solving policies. We do so by rounding each log sea lice level to two decimal places, effectively creating discrete buckets with a resolution of 0.01 log sea lice per fish. We conversely utilize a continuous state space during simulations.

B. Action Space Definition

Let action $a_t \in \{0, 1\}$ at time t be defined as a binary element that represents whether mechanical treatment was applied ($a_t = 1$) or not ($a_t = 0$) in week t.

C. Transition Function

We assume a growth function

$$\ell_{t+1} = \begin{cases} e^r * \ell_t, & \text{without treatment} \\ (1 - \rho) * e^r * \ell_t, & \text{with treatment} \end{cases} \tag{1}$$

where

- ℓ_t is the absolute sea lice level at time t
- ρ is the treatment efficacy of mechanical treatment, representing the percentage of average adult female lice per fish removed during mechanical treatment,
- r is the growth rate of lice in a week, and
- σ^2 is the variance capturing noise.

We can linearize the growth function by applying the log transformation on both sides

$$\log(\ell_{t+1}) = \begin{cases} r + \log(\ell_t), & \text{without treatment} \\ \log(1 - \rho) + r + \log(\ell_t), & \text{with treatment} \end{cases} \tag{2}$$

Since we model sea lice level in log space and $s_t = \log(\ell_t)$, we can use this growth function to determine the expected next level of sea lice in log space $\mu_{s_{t+1}} = \mu(s_{t+1} \mid s_t, a_t)$.

$$\mu_{s_{t+1}} = \begin{cases} r + s_t, & \text{without treatment}(a_t = 0) \\ \log(1 - \rho) + r + s_t, & \text{with treatment}(a_t = 1) \end{cases} \tag{3}$$

1) Discrete POMDP: When solving the problem in discretized state space, we utilize the sparse categorical **SparseCat** distribution from the POMDPs.jl package that takes a list of outcomes and a list of associated probabilities as arguments [17]. We then draw the next state from the following distribution:

$$T(s_{t+1} \mid s_t, a_t) = SparseCat(s_{t+1,i}, P(s_{t+1,i}|s_t, a_t)), \tag{4}$$

where $s_{t+1,i}$ are discretized values derived from a Normal distribution $\mathcal{N}(\mu_{s_{t+1}}, \sigma^2)$, with variance σ^2. We sample five points around the mean $\mu(s_{t+1} \mid s_t, a_t)$, so the transition function becomes a categorical distribution with five values in the following states $[\mu_{s_{t+1}} - 2\sigma, \mu_{s_{t+1}} - \sigma, \mu_{s_{t+1}}, \mu_{s_{t+1}} + \sigma, \mu_{s_{t+1}} + 2\sigma]$. Each observation $o_{t,i}$ is interpolated to the closest state space with one decimal place precision, and clamped to be within the range $[\log(1e-3), \log(10)]$ to ensure numerical stability.

2) Continuous POMDP: When we simulate the problem in continuous state space, we sample randomly from the normal distribution $\mathcal{N}(s_{t+1}, \sigma^2)$, with variance σ^2, using the **ImplicitDistribution** function in POMDPs.jl [17].

D. Reward Function

Our goal is to optimize both the cost of the treatment and the level of sea lice, so we consider a convex combination of these two variables in the reward function. The reward function $R(s_t, a_t)$ is defined as:

$$R(s_t, a_t) = \begin{cases} -\lambda \ell_t, & \text{without treatment}(a_t = 0) \\ -\lambda \ell_t - (1 - \lambda) * c, & \text{with treatment}(a_t = 1) \end{cases} \tag{5}$$

where:

1) ℓ_t is the raw sea lice level at time step t in episode i with $\ell_t = e^{s_t}$

2) c: The treatment cost of mechanical lice removal
3) λ: Weight parameter indicating the relative importance of optimizing for sea lice level and treatment cost.

E. Observation Function

We define the observation function as a Gaussian distribution centered at the true sea lice level.

1) Discretized POMDP: When solving the problem in discretized state space, the observation $o \in O$ is drawn from the following distribution:

$$O(o_t \mid s_t, a_t) = SparseCat(o_{t,i}, P(o_{t,i}|s_t)), \quad (6)$$

where $o_{t,i}$ are discretized values derived from a normal distribution $N(s_t, \sigma^2)$, with variance σ^2. We sample five points $[s_t - 2\sigma, s_t - \sigma, s_t, s_t + \sigma, s_t + 2\sigma]$. Like in the transition function, each observation $o_{t,i}$ is interpolated to the closest state space with one decimal place precision, and clamped to be within the range $[\log(1e-3), \log(10)]$ to ensure numerical stability.

2) Continuous POMDP: When simulating the problem in continuous state space, we sample randomly from the normal distribution $N(s_t, \sigma^2)$ using the **ImplicitDistribution** in POMDPs.jl [17].

F. Discount factor

We utilize a discount factor of $\gamma = 0.95$ for all experiments.

IV. Methods and Algorithms

To explore trade-offs between computation costs and performance, we explored three different policy solvers: value iteration with full observability, QMDP, and SARSOP. We compared these algorithms against a heuristic policy with a threshold-based stochastic treatment rule.

A. Heuristic Policy

We implemented a heuristic policy that takes in two parameters: a sea lice threshold l_{thres} and a belief threshold P_{thres}. The agent applies treatment if the cumulative probability of the belief states where the sea lice level is above the sea lice threshold is above the belief threshold. Otherwise, the policy arbitrarily picks any action $a \in A$. The heuristic model utilizes a discrete updater.

B. Value Iteration

We next implemented value iteration with full observability. Value iteration is an exact solution method that updates the value function directly. It begins with any bounded value function $\mathcal{U}$, which is improved by applying the Bellman update:

$$\mathcal{U}_{k+1}(s_t) = max_{a_t}(R(s_t, a_t) + \gamma \sum_{s_{t+1}} T(s_{t+1}|s_t, a_t)\mathcal{U}_k(s_{t+1})), \quad (7)$$

where

- $\mathcal{U}_k(s_t)$ is the value of state s_t at iteration k
- $R(s_t, a_t)$ is the expected reward received when executing action a_t from state s_t.

In our implementation, we transformed our POMDP into an MDP using the **UnderlyingMDP** function in the **POMDP-Tools.jl** package [18]. We solved the MDP using a vanilla **ValueIterationSolver** in the **DiscreteValueIteration.jl** package with a max iterations set to 30 and a default Bellman Residual of $1e - 3$ [5].

C. QMDP

The QMDP algorithm scales better by using the optimal Q-values of the underlying MDP to create the Q_{MDP} value function for a POMDP:

$$Q_{MDP}(b) = max_a Q(s, a)b(s) \quad (8)$$

where $Q(s, a)$ is the action value function representing the expected return when starting in state *s*, taking action *a*, and then continuing with the greedy policy with respect to Q [10]:

$$Q(s, a) = R(s, a) + \gamma \sum_{s'} T(s'|s, a)U(s') \quad (9)$$

We solved the POMDP using the **QMDPSolver** function in **POMDPs.jl** using the default max iterations parameter of 30 [17].

D. SARSOP

We next implemented SARSOP, which is a point-based POMDP algorithm that exploits the notion of *optimally reachable belief spaces* to improve high computational complexity often associated with POMDPs [11].

The algorithm considers $R(b_0)$, the subset of belief points reachable from a given initial point $b_0 \in B$, where B is the full belief space, under optimal sequences of actions. The optimal sequences of actions consitute the POMDP solution itself, so SARSOP compute successive approximations of $R^*(b_0)$ through heuristic exploration to sample $R(b_0)$ [11].

We implemented SARSOP using the the **SARSOPSolver** function in **POMDPs.jl** with the default max time parameter of 10 seconds [17].

E. Belief Updates

There are various algorithms for updating our belief based on the observation and action taken by the agent. When solving the policies, the state space is discrete, allowing us to perform exact belief updates with the following equation [10].

$$b'(s_{t+1}) = O(o|a_t, s_{t+1}) \sum_{s_t} T(s_{t+1}|s_t, a_t)b(s_{t+1}) \quad (10)$$

To handle belief updates in continuous state spaces during simulations, we utilize the *Extended Kalman Filter* (EKF) and *Unscented Kalman Filter* (UKF), which extend the traditional *Kalman filter*, which provides an exact update under the assumption that T and O are linear The belief state in our POMDP is represented as a Gaussian distribution over sea lice levels, parametrized by a mean μ_{s_t} and variance $\sigma^2_{s_t}$. The EKF approximates non-linear state transitions and observation functions using first-order Taylor expansions. The

EKF extends the standard Kalman filter to problems whose dynamics are nonlinear with Gaussian noise with the following equations [10]:

$$T(\boldsymbol{s_{t+1}}|\boldsymbol{s_t}, \boldsymbol{a_t}) = N(\boldsymbol{s_{t+1}}|\boldsymbol{f}_T(\boldsymbol{s_t}, \boldsymbol{a_t}), \boldsymbol{\Sigma_{s_t}}) \tag{11}$$

$$O(\boldsymbol{o_t}|\boldsymbol{s_{t+1}}, \boldsymbol{a_t}) = N(\boldsymbol{o_t}|\boldsymbol{f}_O(\boldsymbol{s_{t+1}}), \boldsymbol{\Sigma_{o_t}}), \tag{12}$$

where $\boldsymbol{f}_T(\boldsymbol{s_t}, \boldsymbol{a_t})$ and $\boldsymbol{f}_O(\boldsymbol{s_{t+1}})$ are differentiable functions [10].

Conversely, UKF employs a deterministic sampling approach to campute the non-linear effects more accurately. The UKF is preferred for log-normal-space models where exponential growth introduces non-linearities.

The choice of EKF or UKF depends on the trade-off between computational cost and accuracy, with UKF offering improved performance for log-normal dynamics at a higher computation expense.

We implemented both the EKF and UKF utilizing the **runKalmanFilter** function in the **GaussianFilters.jl** package with parametric process and observation noise [4].

F. Simulation Setup

For all POMDP models, we utilized the following set of parameters:

Parameter	**Value**
Growth rate (r)	1.26
Treatment effectiveness (ρ)	0.7
Treatment cost ($C_{\text{treatment}}$)	10
Discount factor (γ)	0.95
Transition noise (σ^2)	0.04
Observation noise (σ^2)	0.04
Log sea lice range	$[\log(1e-3), \log(10)]$
Initial log sea lice range	$[\log(1e-3), \log(1)]$
Heuristic belief threshold (P_{thres})	0.5
Heuristic sea lice threshold (l_{thres})	$\log(5)$
Reward function weight (λ)	$[0, 1]$ in steps of 0.05

TABLE I: Model and simulation parameters used in all experiments.

We ran simulations of the generated policies using the **RolloutSimulator** function in the **POMDPs.jl** package with 1000 episodes and 52 steps per episode [17].

G. Metrics

To evaluate the performance of different policy optimization strategies, we key performance metrics: average reward, average treatment cost, and average sea lice level per week.

Given a configuration with E episodes and T time steps per episode, we caculate the following metrics over the total simulation horizon of $E \times T$ steps:

We calculate the average reward per week as follows:

$$\bar{R} = \frac{1}{E \times T} \sum_{i=1}^{E} \sum_{t=1}^{T} r_{i,t} \tag{13}$$

$$\bar{C} = \frac{1}{E \times T} \sum_{i=1}^{E} \sum_{t=1}^{T} a_{i,t} \times C_{\text{treatment}} \tag{14}$$

$$\bar{L} = \frac{1}{E \times T} \sum_{i=1}^{E} \sum_{t=1}^{T} \ell_{i,t}, \tag{15}$$

where

- $r_{i,t}$ is the reward received at time step t in episode i
- $a_{i,t}$ is the action taken at time step t in episode i, equal to 1 if treatment is applied, 0 otherwise
- $C_{\text{treatment}}$ is the fixed cost of applying treatment
- $\ell_{i,t}$ is the raw sea lice level at time step t in episode i with $\ell_{i,t} = e^{s_{i,t}}$

V. Results and Analysis

This study evaluated the performance of different policies for managing sea lice populations in aquaculture. We looked at three algorithms: VI, SARSOP, and QMDP, and compared their generated policies against random and heuristic policies. We conducted simulations to assess policy effectiveness, and visualized the trade-offs between treatment costs and sea lice levels. Below, we present the key findings from our simulations and sensitivity studies.

Sea lice management offers a multiobjective optimization problem as salmon farmers need to optimize for both population management and cost-effectiveness. To analyze the trade-off between average treatment cost (MNOK/year) and average sea lice levels, we simulated a range of lambda values from 0 to 10 at intervals of 0.05. The lambda value decides the relative weighting of sea lice levels and treatment costs in the reward function (Equation 5). Low lambda values prioritize keeping sea lice levels low whereas high lambda values prioritize cost-effective treatment strategies.

A. Decision-making patterns

To compare the policies generated by value iteration, QMDP, and SARSOP, we analyze the treatment decisions across the range of lambda values and sea lice levels. Figure 1 highlights the probability of treatment for the five policies. As expected, the random policy is uniformly random across all lambdas and sea lice levels, whereas the heuristic policy always treats if the sea lice level is above 5 average adult female lice per fish with stochasticity under that level. Policies generated by VI, QMDP, and SARSOP all reflect a similar pattern of treating at lower sea lice levels for higher lambda values.

Looking at Figure 1, we see that when lambda values are very high or very low, the treatment choice is uniform across the sea lice level for VI, QMDP, and SARSOP. At $\lambda = 0$, we never treat, and at $\lambda = 1$, we always treat no matter the sea lice level. The sea lice level influences treatment decisions with a lambda in the approximate range of $\lambda \in [0.6, 0.9]$. To consider both sea lice range and treatment costs, this range for the lambda value to seems to be optimal to explore for generating an optimal policy. This is to be expected as a lambda value in this range considers both objectives with a focus on keeping sea lice level low. Furthermore, we observe that the treatment decision heatmaps are very similar for policies found via VI, QMDP, and SARSOP, with the treatment decision being more sensitive to the lambda parameter than the solver algorithm itself.

B. Policy Performance: Cost, Sea Lice Levels, and Rewards

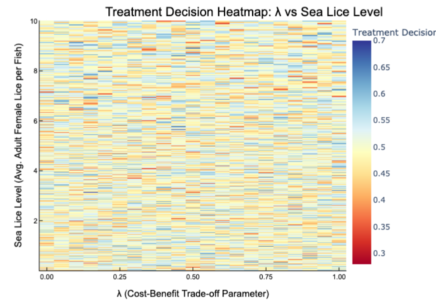
By looking at the decision heatmaps in Figure 1, we saw that determining an optimal lambda value is critical to establishing a balanced policy. In this section, we will analyze the policy performance on simulations across a range of lambda values.

Figure 2 shows the average rewards across λ values for five different policies. Higher rewards are desired, representing lower penalties for sea lice level and treatment costs. We can see that the policies generated using VI, QMDP, and SARSOP all outperform both the random and heuristic policies across all lambda values. We note that VI, QMDP, and SARSOP achieve the highest average rewards at the extremes of the range of lambda values; however, this is a result of uniform treatment decisions across sea lice levels. As such, as mentioned in the previous section, we are interested in the relative performance in the lambda range of $\lambda \in [0.6, 0.9]$. Within this range, we see that value iteration with full observability performs mildly better than QMDP and SARSOP at lambda ranges of $\lambda \in [0.6, 0.85]$ but worse at $\lambda \in [0.85, 1]$; however, these performance differences are minimal and within the confidence intervals, keeping us from making conclusive statements on any performance differences.

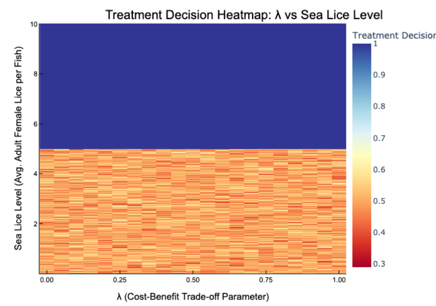
We can further break down the analysis of the differential rewards by looking at the subcomponents of sea lice levels and treatment costs. Figure 3 and Figure 4 show the average sea lice level and treatment cost over a range of lambda values for all policies. For the heuristic policy, average treatment costs ranged from 8 to 9 MNOK/year, with sea lice levels between 6.5 and 8 adult female lice per fish across lambda values. Conversely, the random policy resulted in lower treatment costs centered at around 5 MNOK / year at the cost of higher average sea lice levels between 8 and 9 average adult female lice per fish. In the lambda range of $\lambda \in [0.6, 0.9]$, the policies generated via VI, QMDP, and SARSOP led to lower average sea lice levels and treatment costs.

Figures 5a, 5b, and 5c showcase the Pareto frontier for the simulation performances of the policies generated by VI, QMDP, and SARSOP across lambda values. We want to minimize both average sea lice level and average treatment cost, rendering the optimal value at the lower left corner of the plot. Notably, all algorithms lead to similar Pareto curves. SARSOP has more spread in the average sea lice level at high lambda values, likely due to the observation uncertainty inherent to the SARSOP algorithm. Furthermore, the optimal values are achieved with a higher lambda value for VI, whereas points at the Pareto optimum for the QMDP and SARSOP plots are achieved at lower lambda values of around 0.7.

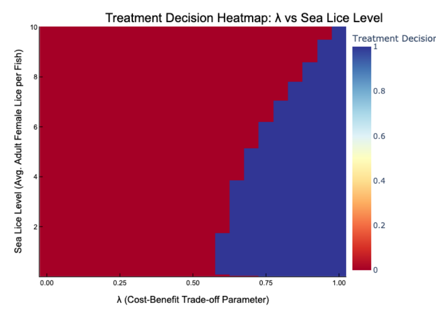
Figure 5 shows the Pareto frontier for all VI, QMDP, and SARSOP across lambda values. Here, we can see that there are marginal differences in the Pareto curves for the different algorithms.



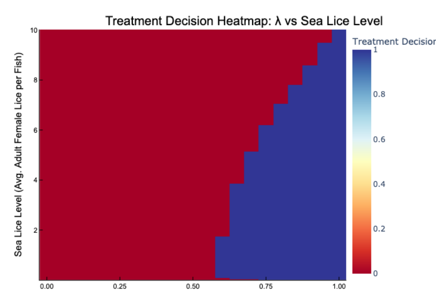
(a) Random Policy



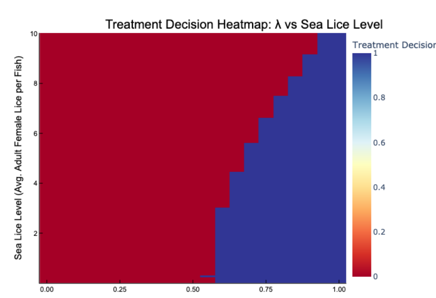
(b) Heuristic policy



(c) VI With Full Observability



(d) QMDP



(e) SARSOP

Fig. 1: Treatment decision heatmaps across λ and sea lice levels for three POMDP solvers. Each pixel indicates whether treatment is recommended (blue) or not (red).

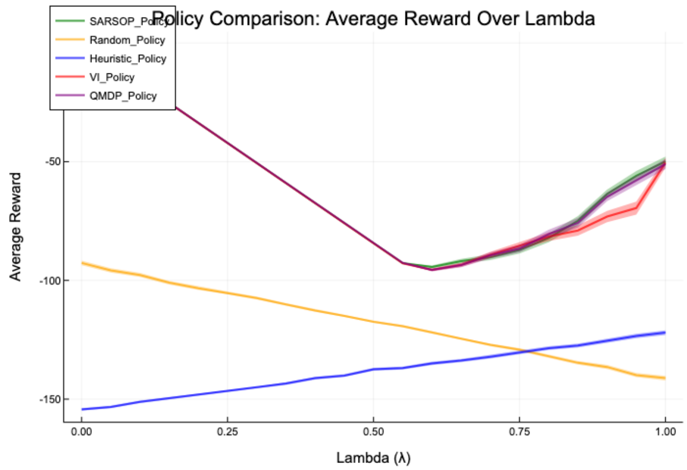


Fig. 2: Comparison of average rewards across λ values for five treatment policies. Shaded regions represent variability (e.g., 95% confidence intervals) across simulation episodes. Higher rewards are preferred, indicating better performance.

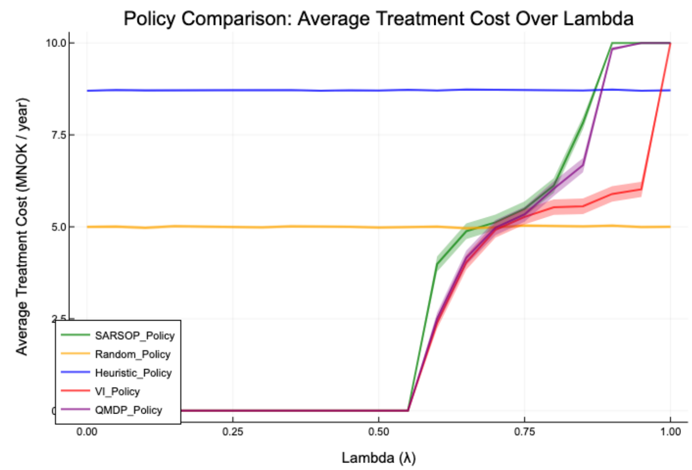


Fig. 4: Comparison of average treatment cost across λ values for five treatment policies. Shaded regions represent variability (e.g., 95% confidence intervals) across simulation episodes. Lower treatment costs are preferred, indicating better performance.

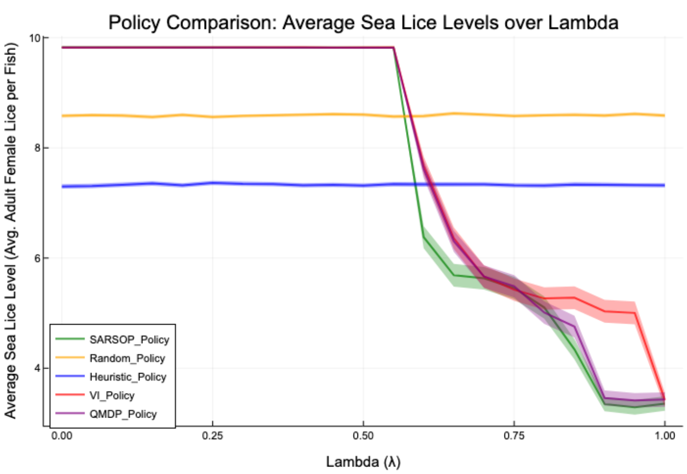


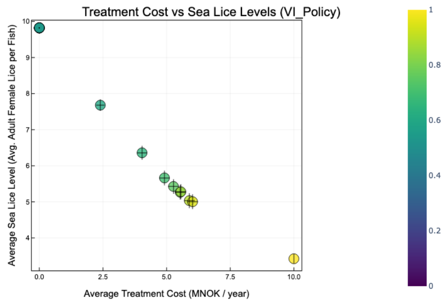
Fig. 3: Comparison of average sea lice levels across λ values for five treatment policies. Shaded regions represent variability (e.g., 95% confidence intervals) across simulation episodes. Lower lice levels are preferred, indicating better performance.

C. Optimized Policies

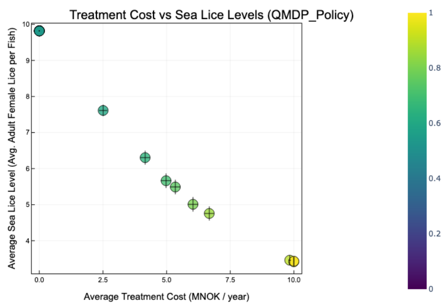
To compare the effectiveness of the policies, we compare the average levels of sea lice and the treatment costs achieved over the simulation period of 52 weeks for a chosen lambda of $\lambda = 0.6$. Figure 7 shows that the random and heuristic policies lead to the highest increase in sea lice levels in the first couple of weeks before leveling out at the 10-week mark. Comparatively, the policies generated via VI, QMDP, and SARSOP lead to a slower growth increase in sea lice levels. The average sea lice level continues to increase monotonically over the entire time period but slowing down as time passes. Notably, the SARSOP policy outperforms both the VI and QMDP policies in both growth rate and a lower plateau of an average of 8 adult female lice per fish. Conversely, VI and QMDP leads to very similar trajectory of sea lice levels over time, leveling out at around 9.5 average adult female lice per fish. The lower average levels of sea lice for SARSOP comes at the cost of higher treatment costs.

Figure 8 showcases the probability of treating per week a for the different policies. As expected, the random policy has a probability of treating centered around 50%. Conversely, the heuristic policy starts with a probability of treating of around 50%. Then, the probability increases linearly over time as the sea lice levels increase in the first couple of weeks as seen in figure 7, leveling out at a treatment-intensive level of around 90% likelihood of treating per week. Conversely, the VI, QMDP, and SARSOP policies have a high initial probability of treating, which then decreases sharply in the first ten weeks before continuing to decrease at a lower rate. The VI and QMDP policies follow a very similar trajectory with the lowest probability of treating in all weeks. SARSOP follows a similar trajectory but with a vertical shift: in every week, the SARSOP policy is around 20% more likely to treat than VI- and QMDP-generated policies.

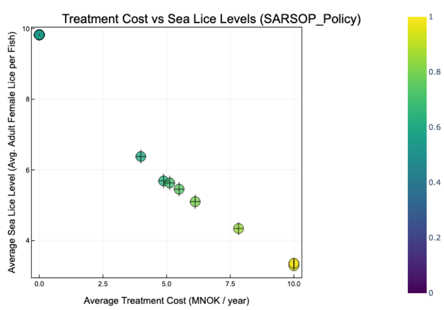
The performance difference of SARSOP in sea lice management is probably due to a lambda value of $\lambda = 0.6$ leading to a treatment boundary at a lower level of sea lice as observed in Figure 1. In other words, the SARSOP algorithm leads to policies that prioritize keeping sea lice levels low more than the VI and QMDP algorithms. One potential explanation for this involves the ability of SARSOP to account for partial observation, leading to better policies for controlling sea lice levels.



(a) Value Iteration With Full Observability



(b) QMDP



(c) SARSOP

Fig. 5: Policy performance in the treatment cost vs. sea lice trade-off space. Each point represents one value of λ; color indicates trade-off intensity.

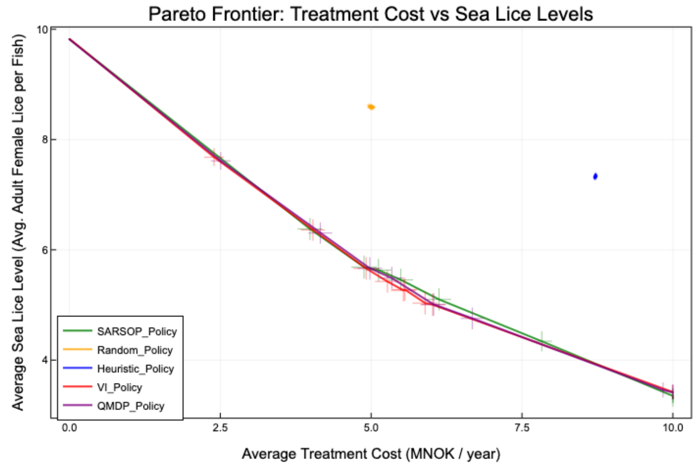


Fig. 6: Comparison of treatment cost vs sea lice levels across three POMDP-based policies (VI with full observability, QMDP, and SARSOP). Lower left represents the Pareto-efficient frontier.

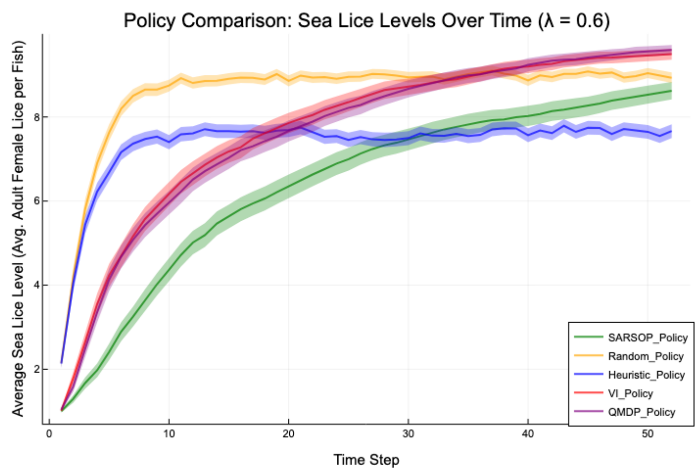


Fig. 7: Sea lice level trajectories over time for five policies at $\lambda = 0.6$. Shaded regions represent variability (e.g., 95% CI) across simulation episodes. Lower levels indicate more effective treatment strategies.

VI. Discussion

This study demonstrates the feasibility and potential of using POMDP-based algorithms to optimize sea lice treatment strategies in aquaculture. By incorporating observation noise, exponential population dynamics, and cost-sensitive reward structures, our framework provides a flexible and data-driven approach to decision-making under uncertainty.

A. Interpretation of Results

Our findings indicate that all three dynamic programming approaches—Value Iteration (VI), QMDP, and SARSOP—significantly outperform heuristic and random policies in reducing sea lice levels while maintaining reasonable treatment costs. Across all λ values, which determine the relative weighting of treatment costs and sea lice penalties, the learned policies consistently resulted in higher cumulative rewards. Notably, the differences between VI, QMDP, and SARSOP policies were minimal in the mid-range $\lambda \in [0.6, 0.9]$, suggesting that all three methods are viable for practical deployment depending on computational constraints and implementation requirements.

While VI and QMDP achieved slightly lower treatment frequencies, the SARSOP policy demonstrated improved lice suppression at the expense of higher treatment intensity. This outcome aligns with the SARSOP algorithm's ability to model partial observability throughout the belief space and suggests that SARSOP may be more suitable in scenarios where treatment effectiveness is prioritized over cost savings.

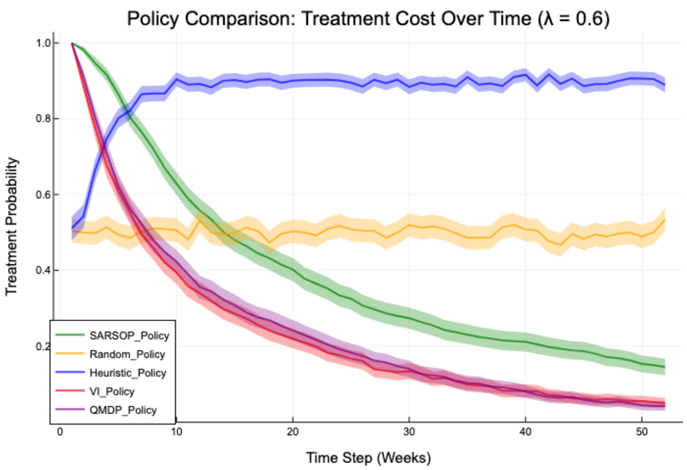


Fig. 8: Comparison of treatment probability over time for five policies at $\lambda = 0.6$. Shaded regions indicate variability across episodes. A lower or decaying treatment rate may signal a more confident or efficient policy.

B. Implications for Aquaculture Management

The practical significance of these findings lies in the ability to deploy adaptive treatment strategies based on observed sea lice levels and prior knowledge of infestation dynamics. Farmers traditionally rely on fixed thresholds or manual heuristics to initiate treatment, which do not account for observation noise or latent infestation trends. Our model suggests that near-optimal policies can be learned and adapted to individual site dynamics and risk preferences, potentially reducing both unnecessary treatments and the risk of large-scale infestations.

Further, this framework could support decision support tools integrated with modern sensor systems, including automated lice counting via computer vision or environmental DNA (eDNA) monitoring. The belief-based structure of POMDPs is well-suited to incorporate uncertainty from such sensor modalities, offering a principled alternative to reactive rule-based systems.

C. Limitations

This work is subject to several limitations. First, our simulations rely on a simplified 1D sea lice model with fixed parameters. While sufficient to demonstrate proof-of-concept, real-world sea lice dynamics are influenced by multiple interacting factors including water temperature, salinity, geographical location, and inter-farm interactions. Additionally, observation reliability is assumed constant; in practice, it may vary with lice monitoring methods, sea temperature, or human error.

Second, the transition function is based on a simple mathematical model of exponential growth of sea lice with a constant growth rate over time. An analysis of historical data from all Norwegian salmon farms from 2012-2025 using maximum likelihood estimation showed that the mean growth rate of sea lice varies from over time and location, with an average growth rate of 0.13 in 2015 versus 0.35 in 2024. The assumption of a constant growth rate limits the ability of the POMDP model to accurately reflect real-world dynamics, which can lead to suboptimal treatment policies.

Third, the reward function used in this study is a convex combination of treatment cost and sea lice penalty, with a scalar λ controlling the trade-off. While useful for algorithmic benchmarking, this formulation may not fully reflect stakeholder priorities (e.g., regulatory compliance, long-term resistance management). Future work should explore multi-objective optimization frameworks or include regulatory thresholds as constraints.

Fourth, the action space of the POMDP model is limited to the binary space of whether or not mechanical treatment was applied in a certain week. The efficacy of the treatment is assumed to be constant at $\rho = 0.7$, representing the percentage of sea lice removed in a single treatment. However, the true treatment efficacy varies with lice resistance, farm location, or treatment methods. The treatment efficacy assumption is enough for a proof-of-concept, but the treatment efficacy dynamics need to be validated empirically using historical data. Furthermore, there are several other treatment methods available such as water baths, feed treatment, lasers, etc., that the current model ignores. Adding these treatment methods, or a combination of them, to the action space will make the model more accurate at the cost of increased computational complexity.

D. Future Work

Several extensions would improve the ecological and operational realism of the model. First, incorporating environmental variables such as location and salinity could provide additional context for treatment decisions. Second, extending the model to account for spatial interactions between neighboring farms could improve regional-level management strategies. Third, investigating the impact of different discretization schemes for the state space could potentially improve the agent's ability to make fine-grained treatment decisions. Additionally, incorporating spatial and time information would allow us to utilize more granular empirical growth rates inferred from real data via maximum likelihood estimates by farm and year, enabling site-specific policy recommendations.

We also envision integrating high-resolution eDNA-based population monitoring data with real-time dynamic treatment planning. Recent breakthroughs in eDNA-based detection now offer precise, near real-time quantification of L. salmonis DNA, bridging the gap between observed and actual lice levels (Krolicka et al., 2022). Integrating these high-resolution population estimates into our dynamic programming framework will significantly enhance treatment accuracy and efficiency. Combining eDNA signals with a probabilistic inference framework like POMDPs would bridge the gap between true infestation dynamics and decision-making, offering a powerful tool for precision aquaculture.

VII. Conclusion

This study introduces a POMDP-based framework for optimizing sea lice treatment strategies in salmon aquaculture under uncertainty. By explicitly modeling observation noise, exponential parasite growth, and treatment costs, we enable principled decision-making that surpasses traditional heuristic and threshold-based methods. Through simulation studies, we show that dynamic programming algorithms—Value Iteration, QMDP, and SARSOP—consistently achieve higher rewards and more balanced trade-offs between treatment frequency and sea lice population suppression. SARSOP, in particular, yields improved sea lice control by better accounting for partial observability.

Our findings highlight the practical potential of reinforcement learning and sequential decision-making in aquaculture, especially when coupled with emerging sensor technologies such as eDNA-based monitoring. While the model simplifies real-world complexities, including variable environmental drivers and treatment efficacy, it lays a foundation for future work incorporating spatial dynamics, empirical parameterization, and multi-objective optimization. As aquaculture systems evolve toward increased automation and data availability, this framework offers a scalable and adaptive tool for sustainable sea lice management.

Acknowledgments

I thank Dr. Mykel Kochenderfer for his guidance and support throughout this research.

References

[1] Jay Abolofia, Frank Asche, and James E. Wilen. The cost of lice: Quantifying the impacts of parasitic sea lice on farmed salmon. *Marine Resource Economics*, 32(3):329–349, 2017.

[2] Annette S. Boerlage, Shailesh Shrestha, Ilkka Leinonen, Mona Dverdal Jansen, Crawford W. Revie, Aaron Reeves, and Luiza Toma. Sea lice management measures for farmed atlantic salmon (salmo salar) in scotland: Costs and effectiveness. *Aquaculture*, 580:740274, 2024.

[3] Annette S. Boerlage, Shailesh Shrestha, Ilkka Leinonen, Mona Dverdal Jansen, Crawford W. Revie, Aaron Reeves, and Luiza Toma. Sea lice management measures for farmed atlantic salmon (salmo salar) in scotland: Costs and effectiveness. *Aquaculture*, 580:740274, 2024.

[4] JuliaGaussianFilters contributors. Gaussianfilters.jl: Kalman filtering for nonlinear systems in julia. https://github.com/JuliaGaussianFilters/GaussianFilters.jl, 2022. Accessed: 2025-06-10.

[5] POMDPs.jl contributors. Discretevalueiteration.jl: A julia package for value iteration on discrete mdps. https://github.com/JuliaPOMDP/DiscreteValueIteration.jl, 2021. Accessed: 2025-06-10.

[6] P. De los Ríos-Escalante and E. and Ibáñez-Arancibia. Use of data science and machine learning techniques for study sea lice (caligus rogercresseyi) infestation on atlantic salmon (salmo salar). *International Journal of Veterinary Research*, 4(1), 2024.

[7] Adel Elghafghuf, Raphael Vanderstichel, Larry Hammell, and Henrik Stryhn. Estimating sea lice infestation pressure on salmon farms: Comparing different methods using multivariate state-space models. *Epidemics*, 31:100394, 2020.

[8] Adel Elghafghuf, Raphael Vanderstichel, Larry Hammell, and Henrik Stryhn. State-space modeling for inter-site spread of sea lice with short-term population predictions. *Ecological Modelling*, 452:109602, 2021.

[9] Romain Gautron, Odalric-Ambrym Maillard, Philippe Preux, Marc Corbeels, and Régis Sabbadin. Reinforcement learning for crop management support: Review, prospects and challenges. *Computers and Electronics in Agriculture*, 200:107182, 2022.

[10] Mykel J. Kochenderfer, Tim A. Wheeler, and Kyle H. Wray. *Algorithms for Decision Making*. MIT Press, 2022.

[11] Hanna Kurniawati, David Hsu, and Wee Sun Lee. *SARSOP: Efficient Point-Based POMDP Planning by Approximating Optimally Reachable Belief Spaces*, pages 65–72. 2009.

[12] Michael L. Littman, Anthony R. Cassandra, and Leslie Pack Kaelbling. Learning policies for partially observable environments: Scaling up. In Armand Prieditis and Stuart Russell, editors, *Machine Learning Proceedings 1995*, pages 362–370. Morgan Kaufmann, San Francisco (CA), 1995.

[13] De los Ríos-Escalante, E Ibáñez-Arancibia, et al. Use of data science and machine learning techniques for study sea lice (caligus rogercresseyi) infestation on atlantic salmon (salmo salar). *International Journal of Veterinary Research*, 4(1):55–59, 2024.

[14] Alexander G. Murray and Nabeil K.G. Salama. A simple model of the role of area management in the control of sea lice. *Ecological Modelling*, 337:39–47, 2016.

[15] Berit Rabe, Alexander G Murray, Stephen C Ives, Meadhbh Moriarty, and David J Morris. Searching for sea lice: surveillance to assess environmental infection pressures to model and inform sea lice infestation management. *Aquaculture Environment Interactions*, 16:241–262, 2024.

[16] Jagath W Rupasinghe and John OS Kennedy. Optimal batch lengths for barramundi farming under seasonal variations: A dynamic programming approach. 2006.

[17] Zachary N. Sunberg and collaborators. Pomdps.jl: A julia framework for partially observable markov decision processes. https://github.com/JuliaPOMDP/POMDPs.jl, 2022. Accessed: 2025-06-10.

[18] Zachary N. Sunberg and collaborators. Pomdptools.jl: Tools for solving and simulating pomdps in julia. https://github.com/JuliaPOMDP/POMDPTools.jl, 2022. Accessed: 2025-06-10.

[19] Xinyu Tian, Mahbuba Afrin, Sajib Mistry, Redowan Mahmud, Aneesh Krishna, and Yan Li. Mure: Multi-layer real-time livestock management architecture with unmanned aerial vehicles using deep reinforcement learning. *Future Generation Computer Systems*, 161:454–466, 2024.

[20] Ingunn Fride Tvete, Magne Aldrin, and Britt Bang Jensen. Towards better survival: Modeling drivers for daily mortality in norwegian atlantic salmon farming. *Preventive Veterinary Medicine*, 210:105798, 2023.

[21] Cooper Van Vranken et al. Towards a global fishing vessel ocean observing network (fvon): State of the art and future directions. *Frontiers in Marine Science*, June 20 2023.

[22] Thomas K. Waring, Vera L. J. Somers, Michael A. McCarthy, and Christopher M. Baker. When to monitor or control: Informed invasive species management using a partially observable markov decision process (pomdp) framework. *Methods in Ecology and Evolution*, 15(9):1667–1676, 2024.

[23] Yong-Suhk Wui and Carole R Engle. A mixed integer programming analysis of effluent treatment options proposed for pond production of hybrid striped bass, morone chrysops x m. saxatilis. *Journal of Applied Aquaculture*, 15(1-2):121–157, 2004.

[24] Hidekazu Yoshioka and Yuta Yaegashi. Optimization model to start harvesting in stochastic aquaculture system. *Applied Stochastic Models in Business and Industry*, 33(5):476–493, 2017.