Optimizing Sea Lice Management in Norwegian Salmon Aquaculture Using Q-Learning

Olivia Beyer Bruvik
Stanford Univeristy
oliviabb@stanford.edu

Ali Hindy
Stanford Univeristy
ahindy@stanford.edu

Will Denton
Stanford Univeristy
wdenton@stanford.edu

***Abstract*—Sea lice infestations represent one of the most significant challenges in salmon aquaculture, resulting in substantial economic losses and environmental concerns. While various treatment options exist, including chemical baths, in-feed treatments, and mechanical removal, determining optimal treatment policies remains challenging due to the complex dynamics of lice populations and treatment interactions. We present a model-free reinforcement learning approach that learns optimal treatment policies directly from historical farm data, circumventing the need to explicitly model complex biological processes. Our method considers eight distinct treatment combinations and demonstrates significantly better performance than random treatment policies, achieving stable rewards while maintaining lice levels below regulatory thresholds. This work represents a significant step toward data-driven decision support in aquaculture, offering a practical framework for optimizing sea lice management while balancing treatment costs and efficacy.**

I. Introduction

A. Problem statement

Currently, one-third of the global population depends on fish as a primary source of protein, with salmon being one of the most widely farmed species. [5] However, salmon farms worldwide face significant challenges from infestations of *Lepeophtheirus salmonis*, a parasitic lice species that causes lesions, rendering the fish unsuitable for consumption. Various treatment methods exist to combat these infestations, but their effectiveness is **uncertain**, influenced by environmental factors such as temperature, location, seasonal variations, and the treatment method itself.

B. Approach

Lice level management is a complex problem for salmon farmers that necessitates dynamic decision making in response to the current lice levels and lice growth rates. For example, a salmon farmer might experience an influx of lice into their pens due to a lice outbreak in a neighboring salmon farm and consequently decide to perform mechanic delousing to keep the lice levels below the official limit of average adult female lice count per fish. 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 problem for a sequential decision making algorithm. [1]

The ultimate goal of sea lice level management is to maximize the salmon revenue of the production cycle, which is dependent on the salmon biomass and the salmon quality at the time of harvest. The problem is challenging because sea lice levels evolve probabilistically, and we want to make sure that we address sea lice levels early enough to avoid a sea lice outbreak, but late enough so that we avoid unnecessarily stressing the salmon (which lowers salmon biomass and quality).

We present a model-free reinforcement approach to determine the optimal sequence of treatments that salmon farmers should apply on a weekly basis to minimize lice outbreaks given details about their specific farm. Our approach involves developing a robust policy derived from Q-learning optimization that proposes actionable solutions for salmon farmers, leveraging real-world data sourced from publicly available datasets. [13]

C. Appropriateness of Approach

The complex dynamics of sea lice populations in salmon farms present significant modeling challenges that make model-free approaches particularly attractive. While traditional approaches attempt to model population growth using partial differential equations, these models must account for numerous interacting factors including water temperature, salinity, seasonal fluctuations, and treatment effectiveness, along with spatial interactions between neighboring farms. [9] This modeling task becomes especially challenging when working with historical data where treatments were already applied, as it becomes difficult to disentangle the natural population dynamics from treatment effects. Moreover, the stochastic nature of treatment outcomes, environmental conditions, and inter-farm lice movements introduces multiple sources of uncertainty that are difficult to capture in explicit models. Model-free reinforcement learning circumvents these challenges by learning directly from observed state transitions, eliminating the need to explicitly model the underlying biological processes while still capturing their complex interactions through experienced outcomes.

II. Related Work and Literature Review

Sea lice management in salmon aquaculture has been extensively studied due to its significant economic and environmental implications. Furthermore, researchers have applied modern optimization techniques to different areas of aquaculture, such as fish farming and harvesting. While Norwegian government regulations have generated extensive historical data on sea lice in salmon farms, there remains a significant research gap in modeling sea lice populations. This gap is particularly pronounced when attempting to model population dynamics in the presence of treatments, as treatment effects in historical data confound the underlying biological processes. [7] We aim to combine existing techniques with this abundance of data to model the effects of treatment on fish populations.

A. Optimization in Aquaculture

Recent years have seen increasing application of mathematical optimization techniques to aquaculture management. Studies have employed various methods including dynamic programming [10], mixed-integer programming [14], and stochastic optimization [15] to address challenges in fish farming. However, these approaches often rely on simplified models of complex biological systems and may not capture the full dynamics of sea lice populations. We aim to model this problem as a Markov Decision Process (MDP), in order to accurately capture the complex effects of treatments on sea lice populations.

B. Reinforcement Learning in Agricultural Systems

While reinforcement learning has shown promise in agricultural applications [2], its use in aquaculture has been limited. Several studies have demonstrated the potential of RL in livestock management [11], crop disease control [4], and fish growth trajectory tracking [3], suggesting its applicability to parasite management in aquaculture. However, little research has been conducted focusing on optimizing treatment management for farmers, despite its real-world impact on the global economy. The success of model-free approaches in these domains is particularly relevant to our work, as they can learn optimal policies without requiring explicit models of complex biological systems.

C. Data-Driven Decision Making in Aquaculture

The emergence of precision aquaculture has led to increased availability of monitoring data [8], [13]. However, translating this data into effective treatment decisions remains challenging. Recent work has begun to explore machine learning approaches for predicting sea lice abundance [6], but few studies have addressed the optimization of treatment strategies based on these predictions.

III. Problem Formulation

We explicitly define our MDP as follows:

A. State Space Definition

Let state s_t at time t be defined as a vector: $s_t = [l_t, w_t]$ where:

1) $l_t \in \mathbb{R}^+$: Current sea lice level (continuous)
2) $w_t \in \mathbb{R}$: Water temperature (°C)

Therefore, $S = \mathbb{R}^+ \times \mathbb{R}$. We define the state space over one production cycle (one year), so $t \in [1, 52]$. Therefore, our MDP can be described with a finite time horizon. We discretize the continuous state space of sea lice levels by multiplying each lice count value by 100 and rounding to the nearest integer, effectively creating discrete buckets with a resolution of 0.01 lice per fish. For example, a lice level of 0.235 would be mapped to state index 23, while 0.239 would map to state index 24. The same approach was used for discretization of sea temperature in our Q-learning model without action value function approximation.

B. Action Space Definition

Let action a_t at time t be defined as a treatment combination vector: $a_t = [b_t, f_t, m_t]$ where:

- $b_t \in \{0, 1\}$: Bath treatment
- $f_t \in \{0, 1\}$: Feed treatment
- $m_t \in \{0, 1\}$: Mechanical removal

Each binary element represents whether that specific treatment is applied (1) or not (0). This yields a total of 7 possible treatment combinations plus the no-treatment option, giving us eight possible actions $a_t \in \{0, 1, \dots, 7\}$ where:

- $a_t = 0$: No treatment $[0, 0, 0]$
- $a_t = 1$: Bath only $[1, 0, 0]$
- $a_t = 2$: Feed only $[0, 1, 0]$
- $a_t = 3$: Mechanical only $[0, 0, 1]$
- $a_t = 4$: Bath + Feed $[1, 1, 0]$
- $a_t = 5$: Bath + Mechanical $[1, 0, 1]$
- $a_t = 6$: Feed + Mechanical $[0, 1, 1]$
- $a_t = 7$: All treatments $[1, 1, 1]$

In our binary action space experiment, we simplify this to $a_t \in \{0, 1\}$ where 0 represents no treatment and 1 represents mechanical removal only.

C. Transition Function

The transition probabilities $T(s_{t+1} \mid s_t, a_t)$ are empirically estimated from historical data rather than modeled theoretically. For each state-action pair (s_t, a_t), we maintain counts of observed transitions to next states s_{t+1} in our transition matrix. These counts are normalized to obtain probabilities:

$$T(s_{t+1} \mid s_t, a_t) = \frac{N(s_t, a_t, s_{t+1})}{\sum_{s'} N(s_t, a_t, s')}$$

where $N(s_t, a_t, s_{t+1})$ represents the number of times state s_{t+1} was observed following state s_t and action a_t in the historical data. If no transitions are observed for a state-action pair ($\sum_{s'} N(s_t, a_t, s') = 0$), the system remains in the current state ($s_{t+1} = s_t$).

D. Reward Function

The reward function $R(s_t, a_t)$ is defined as: $R(s_t, a_t) = -\beta_1 l_t - \beta_1 \max(0, l_t - 0.5) + \beta_1 \max(0, l_t - l_{t+1}) - \beta_2 c(a_t)$ where:

1) l_t: Current lice level at time t
2) l_{t+1}: Next observed lice level
3) $c(a_t)$: Cost of action a_t
4) β_1, β_2: Weight parameters for lice-related penalties and costs, respectively (experimentally determined)

The reward function consists of four terms that account for the following:

1) $\beta_1 l_t$: Base penalty proportional to current lice level
2) $\beta_1 \max(0, l_t - 0.5)$: Additional heavy penalty for exceeding threshold of 0.5 lice per fish
3) $\beta_1 \max(0, l_t - l_{t+1})$: Reward for reducing lice levels
4) $\beta_2 c(a_t)$: Treatment cost penalty

IV. Methods and Algorithms

We explored a variety of algorithms to compute an optimal policy, ultimately implementing Q-learning. While we explored model-based algorithms, we discovered that an exponential growth PDE transition function did not accurately represent the data set on which we trained our models. While the transition function assumes exponential growth of lice, the dataset demonstrated a more erratic pattern that was not captured by the proposed transition probabilities via a PDE $T(l_{t+1} \mid s_t, a_t) = (1-p_{a_t})e^{r_t} l_t$, where r_t is the growth rate of the population and p_{a_t} is the efficacy probability of treatment a_t. Model-based algorithms learn the transition and reward models through interaction with the environment. Therefore, as a result of the challenges of establishing a meaningful transition representation, we focused our efforts on model-free reinforcement and simulation methods.

A. Baseline

Our baseline model utilizes a random policy that arbitrarily picks any action $a \in A$ chosen independently from the current state space. The model serves as a baseline for comparison with our other methods.

B. Q-learning

We implemented model-free Q-learning, which incrementally estimates the action value function $Q(s, a)$ to evaluate an optimal policy without explicitly defining the transition and reward functions. Our algorithm uses the following incremental update rule to estimate the action value function:

$$Q(s, a) \leftarrow Q(s, a) + \alpha(r + \gamma max_{a'} Q(s', a') - Q(s, a))$$

We performed exploration with ϵ-greedy exploration with an epsilon value of 0.1. In a particular state s, we chose a random action with a probability of 10 percent and the action that optimized the Q-value for that state, $argmax_{a'} Q(s, a')$, 90 percent of the time.

V. Results and Analysis

Our empirical evaluation demonstrates the effectiveness of Q-learning for optimizing sea lice treatment strategies in salmon aquaculture. The agent learns directly from historical transition data, with state transitions sampled from observed outcomes rather than using a learned model of the environment dynamics. We elaborate on the difficulties of constructing a learned model in the discussion, and choose a model-free approach.

A. Action Space

In our first experiment, we consider eight possible treatment combinations: no treatment (control), bath treatment only, feed treatment only, mechanical removal only, bath and feed combined, bath and mechanical combined, feed and mechanical combined, and all three treatments combined. This represents the complete set of available interventions in salmon aquaculture. For our second experiment, we simplify to a binary choice between no treatment and mechanical removal, allowing us to analyze the effectiveness of one treatment (mechanical intervention) in isolation.

B. Performance Comparison

Figures 1 and 2 show that our Q-learning approach significantly outperforms a random treatment policy across both experiments. In the eight-action scenario (Figure 1), the Q-learning agent demonstrates remarkable stability, achieving a mean reward of -969.07 ± 940.66 with relatively small variance. In contrast, the random policy performs significantly worse with a mean reward of -52025.92 ± 3428.02 with high variance. Both experimental results are averaged over 3 simulations, for 10 epochs each.

When restricted to binary actions (Figure 2), we observe that while Q-learning still outperforms the random policy, the performance advantage is less pronounced. Q-learning achieves a mean reward of -11823.98 ± 5315.97, while the random policy achieves a mean reward of -32529.62 ± 9701.13. The binary action space leads to more variable performance in both policies, suggesting that the reduced flexibility in treatment options may limit the agent's ability to effectively manage sea lice levels.

C. Learning Stability and Convergence

The eight-action implementation shows superior stability and convergence characteristics compared to the binary version. With eight actions, the Q-learning agent quickly converges to a stable policy and maintains consistent performance across all epochs, as evidenced by the narrow confidence bands. This suggests that having access to a fuller range of treatment combinations allows the agent to learn more robust and reliable policies.

In contrast, the binary action space, while simpler, results in more volatile performance. This indicates that the restricted action space may be insufficient for handling the complex dynamics of sea lice populations, despite its apparent simplicity. The full treatment combination space not only offers better performance but also demonstrates more reliable learning characteristics.

VI. Discussion

As part of our exploration, we encountered several limitations to our dataset and approach that impacted our results. Notably, variations in the data, difficulties in defining a meaningful reward function, and challenges to exploring the entire state space with the given dataset proved challenges that we sought to overcome. In this section, we will discuss each

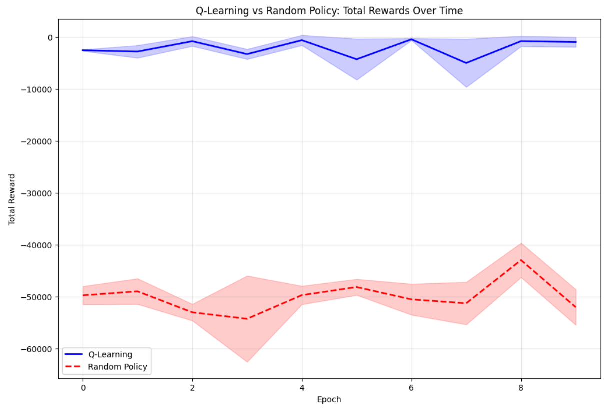

Fig. 1. Performance comparison between Q-learning and random policy with eight treatment combinations. The Q-learning agent demonstrates stable, superior performance with minimal variance.

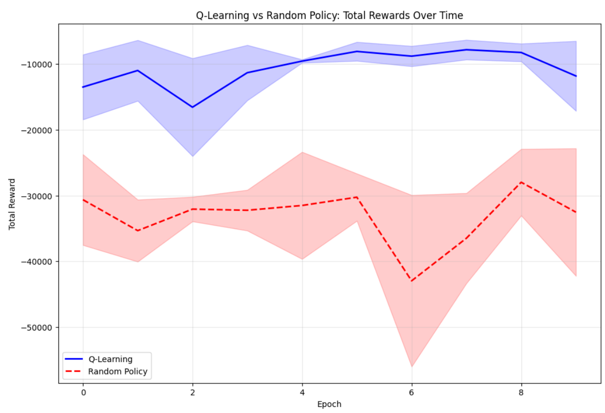

Fig. 2. Performance comparison between Q-learning and random policy with binary actions (no treatment vs. mechanical). While Q-learning outperforms random policy, both show increased variance compared to the eight-action scenario.

of these challenges in more detail, as well as provide some insights as to what a further exploration of the problem could look like.

A. Limitations in the data

For this project, we used public data provided by Lusedata to optimize sea lice management in Norwegian salmon farms. The dataset is extensive over all farms in Norway due to Norwegian regulations that mandate all salmon farmers to report sea lice levels per pen on a weekly basis. While this offered a valuable dataset to compute an optimal policy for sea lice management, the dataset offered some challenges as well.

Based on past literature as well as the nature of sea lice as an invasive species with exponential growth, we assumed that the sea lice levels would follow an exponential growth pattern; however, the dataset demonstrated a rather erratic pattern of sea lice levels across time, which was a challenge when simulating lice growth when computing an optimal policy with Q-learning. We have established several factors that might have contributed to this pattern.

First, sea lice levels in a particular farm are heavily affected by sea lice outbreaks in neighboring farms. Our state space does not account for the sea lice levels of neighboring farms, which is a limiting factor in our model. An extension to this project could involve mapping out the relative locations of different salmon farms and including the sea lice levels of neighboring salmon farms in the state space.

Second, sea lice monitoring methods vary drastically across farms. While some farms use manual counting procedures, other farms use more sophisticated equipment like lasers that use computer vision to detect sea lice on salmon and then apply artificial intelligence models to estimate sea lice levels in the pen. The variation in montioring methods and subsequent accuracy of reported sea lice levels will naturally affect the expected growth of sea lice as sea lice growth follows an exponential model.

Third, our model only considered a subset of all the available sea lice management methods due to data availability. Thus, our model did not take into account crucial factors such as the presence of electric fences and sea lice skirts in the pens, which will affect the influx of lice into pens as well as continued growth of the current sea lice population.

B. Limitations of the reward function

An ideal reward function would reflect the revenue of the salmon farm at the end of a production cycle. The revenue is dependent on the salmon biomass, salmon fish health, adherence to regulations on the maximum number of treatments. The types and number of treatments function as a proxy for these variables, as treatments significantly impact salmon biomass and fish health. Notably, total mortality could be reduced by 21 percent if avoiding all lice treatments, demonstrating the impact of lice treatments on salmon biomass in a production cycle [12]. Unfortunately, the dataset available for this project only provided data information on sea lice levels, treatments procedures, and treatment cost. Therefore, we had to make design choices for how we wanted to define a meaningful reward function to capture the expected revenue of salmon farmers as a function of treatments. Upon careful consideration, we define the reward function $R(s_t, a_t)$ as $R(s_t, a_t) = -\beta_1 l_t - \beta_1 \max(0, l_t - 0.5) + \beta_1 \max(0, l_t - l_{t+1}) - \beta_2 c(a_t)$ where $c(a_t)$ is the cost function for chosen treatments, $\max(0, l_t - l_{t+1})$ is the expected change in lice levels, and β_1, β_2 are weighting parameters. As expected, the choice of costs and weighting parameters greatly affected the resultant policy. In order for the policy to be meaningful, we would need additional data on the actual impact of treatment costs and lice levels on the expected revenue of a production cycle. As such, further exploration would likely necessitate partnering with specific farms to access the financials associated with sea lice management, as we estimated these costs ourselves.

C. Limitations in the state space exploration

Any problem in which a continues state space is discretized inherently implies limitations to the state representation. Although empirically this discretization helped performance, likely because similar states would imply similar preferred treatment options, in a further exploration of this problem working in the continuous state space may prove advantageous. Furthermore, additional factors that we did not have access to in the data, such as neighboring farm lice levels and more specific water conditions, may have proved useful in predicting lice levels and also treatment effectiveness. In practice, greater data collection may result in more successful policies.

VII. Conclusion

This paper demonstrates the effectiveness of model-free Q-learning for optimizing sea lice treatment strategies in salmon aquaculture. Our approach learns effective policies directly from historical treatment data without requiring explicit modeling of the complex biological dynamics governing sea lice populations. This is particularly advantageous given the challenges of modeling population growth in datasets where treatments have already been applied. Furthermore, the experimental results show that our Q-learning agent significantly outperforms random treatment policies across both eight-action and binary action spaces. Notably, the agent achieves more stable and superior performance when given access to the full range of treatment combinations, suggesting that flexibility in treatment options is valuable for effective sea lice management.

Several promising directions exist for future work. 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. Finally, investigating the impact of different discretization schemes for the state space could potentially improve the agent's ability to make fine-grained treatment decisions.

The success of our model-free approach indicates that complex biological interactions between population and treatments can be effectively managed without explicit modeling of their underlying dynamics, opening possibilities for similar applications in other aquaculture and agricultural settings.

VIII. Contributions

In this project, Will focused on simulating lice levels in a model-free approach. This simulation allows us to train and evaluate our policies. Olivia worked on implementing the Q-learning model with a binary action space $A = \{\text{No treatment}, \text{ Mechanical only}\}$. The initial Q-learning model with a binary action space allowed us to explore and establish an appropriate reward function. She also implemented and explored the performance of Q-learning with value function approximation. Learning that the Q-learning performed worse with value function approximation, we shifted our focus to our exact Q-learning. Ali expanded this Q-learning model to include the complete action space with eight possible treatment combinations, and worked on evaluating the Q-learning agent against a random policy. These experiments allowed us to better understand which action spaces, reward models, and hyperparameters increased the performance of the Q-learning agent. All three members contributed to various components of the final paper, and each spent an extra 30 hours making sure the paper is conference-ready.

References

[1] 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.

[2] Abderrazak Chahid, Ibrahima N'Doye, John E Majoris, Michael L Berumen, and Taous-Meriem Laleg-Kirati. Fish growth trajectory tracking via reinforcement learning in precision aquaculture. *arXiv preprint arXiv:2103.07251*, 2021.

[3] Abderrazak Chahid, Ibrahima N'Doye, John E Majoris, Michael L Berumen, and Taous-Meriem Laleg-Kirati. Fish growth trajectory tracking using q-learning in precision aquaculture. *Aquaculture*, 550:737838, 2022.

[4] 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.

[5] Jason S. Link and Reg A. Watson. Global ecosystem overfishing: Clear delineation within real limits to production. *Science Advances*, 5(6):eaav0474, 2019.

[6] 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.

[7] Norwegian Ministry of Fisheries and Coastal Affairs. Regulation on salmon lice control - regulations relating to the control of salmon lice in aquaculture facilities (norway). Legislative Document LEX-FAOC144013, Food and Agriculture Organization of the United Nations, 2010.

[8] 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.

[9] Crawford W Revie, C Robbins, G Gettinby, L Kelly, and JW Treasurer. A mathematical model of the growth of sea lice, lepeophtheirus salmonis, populations on farmed atlantic salmon, salmo salar l., in scotland and its use in the assessment of treatment strategies. *Journal of Fish Diseases*, 28(10):603–613, 2005.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.