UiT The Arctic University of Norway

Norwegian College of Fishery Science

Identifying significant factors and optimal sites for commercial salmon farming in northern Norway.

An integrated GIS and machine learning approach using random forest.

Nikolai Aspen

Master's thesis resource management
FSK-3960 (60ECT). May 2022

Acknowledgements

This master´s thesis marks the end of five academic years at the Norwegian College of Fishery Science – UiT Norwegian Arctic University. I want to take this opportunity to thank some of you who have made this journey enjoyable and valuable.

First, I would like to thank my primary supervisor and teacher for three years, Keshav Paudel. Without you, I would have never started with GIS, machine learning and spatial analysis. Thus, this master´s thesis would never have happened.

I would also like to thank Cosmin Crăciun at Dualog for helping me with creating a script for data- cleaning and processing. Then I would like to thank Jon Albretsen at the Institute of Marine Research for guidance and for providing me with all the environmental data from NorKyst-800, SWAN and AROME. Without your great help, this master´s thesis would not be possible.

Thanks to all the aquaculture companies in northern Norway participating in this thesis with the sharing of location-specific environmental data, Kvarøy Fiskeoppdrett, Lerøy Aurora, Norway Royal Salmon, Wilsgård, Northern lights salmon, Salaks, Nova Sea and Flakstadvåg Laks.

Finally, I would like to thank everyone who participated in the voyage, fellow students, scientifically employees, family, and friends.

Abstract

This study presents a data-driven modelling approach to identify important factors influencing the growth- and mortality rate for farmed salmon in northern Norway. Furthermore, a model is trained to determine the best fish farming sites and identify optimal areas with the best geographical conditions.

Aquaculture site production and location data from 323 salmon farming sites (all licensed aquaculture sites) in northern Norway were obtained from the Directory of Fisheries. Two dependent variables, growth- and mortality rate, were calculated based on the monthly increase in biomass and mortality. These variables were combined with state-of-the-art environmental- and exploratory socio-economic data obtained from the institute of marine research (IMR), the Norwegian Meteorological Institute, Delft University of Technology, Norwegian Coastal Administration, and Statistics Norway.

Using random forest regression and recursive feature elimination, a data-driven ensemble approach identified significant variables. Prediction of optimal sites for salmon farming in northern Norway was done with a species distribution modelling approach using random forest classification.

The important factors affecting salmon growth were specific feeding rate, temperature, and total biomass. The important factors influencing salmon mortality were temperature and total biomass. The predicted optimal areas were inside Vefsnfjorden, Ranfjorden, Sørfjorden and Glomfjorden, small areas near the coast and around the small islands stretching from Gladstad to Narvik. Areas near the coast of Lofoten, Værøy, Røst, Vesterålen, Sortland and Senja. Further north, some dispersed regions were predicted as optimal outside Tromsø and Sørøya. Also large areas around Varangerhalvøya, Olderdalen/Kåfjorden, Lille Altafjorden and near the shore on both sides of Stjernøysundet.

The results clearly show that space is a scares resource and that there is an urge to evaluate the regulations and legislations concerning aquaculture in Norway. Especially the minimum distances between the fairways and aquaculture locations. The incorporation of machine learning approaches in GIS-based MCE analysis is suggested to help planners and decision-makers make informed and sustainable decisions about sea-area use.

Table of Contents

List of tables

List of figures

1 Introduction

Over the last 50 years, global food demand has approximately tripled and is expected to further increase in the coming years (Bodirsky et al., 2015). It is estimated that agricultural land contributes about 90% of food calories and 80% of protein and fats Viana, Freire, Abrantes, Rocha, and Pereira (2022). Currently, Land-based agriculture uses about 70% of the world's freshwater reservoir, occupies almost half of the earth's ice-free surface and accounts for 1/3 of all greenhouse gas emissions (Schubel & Thompson, 2019). Based on these facts, we can assume that current agricultural practices cannot scale to meet the world's growing demand for food. Therefore, an increased share of goods and services should come from the sea.

70% of the globe is covered by water, yet still, only two per cent of the world's food supply currently comes from the ocean; furthermore, the carbon footprint of commercial agriculture dramatically outweighs that of aquaculture. According to The High-Level panel for Sustainable Ocean Economy, the world can produce many times more seafood than today, and seafood can cover more than 2/3 of the need for animal protein, which we need to saturate a population of 10 billion by 2050 (Ministry of Trade Industry and Fisheries, 2021). These estimates are only realistic if the food is harvested and produced sustainably. The aquaculture industry is now the fastest-growing food production sector worldwide and is responsible for more than half of the world's seafood production (Falconer et al., 2020; Lymbery, 2002; Stevenson, 2007). It is expected that aquaculture will keep growing steadily, with an estimated 14.5% increase in production by 2030. It is predicted that 62% of all seafood production will be farm-raised (Morro et al., 2021).

Norway is the largest producer of salmonids in the world. In 2021 Norwegian aquaculture industry produced about 1.3 million tons of farmed salmonids. Norway has natural advantages for farming salmonids with deep fjords, good current conditions, and oxygen-rich waters with favourable temperatures (Ministry of Trade Industry and Fisheries, 2021). In 2014 the Norwegian government introduced plans to increase the national production of salmonids fivefold, i.e., a production expansion from 1 million tones to over 5 million tons of farmed salmon. The government aims to increase the production within a sustainable framework, i.e., the government will facilitate that the industry safeguards good fish- health and welfare and that the product has a low climate- and environmental footprint. However, to reach these goals sustainably, there is an urge to adopt more sea areas for salmon farming because of the maximum total biomass allowance at each location and the environment carrying capacity (Robertsen et al., 2020). Moreover, an industry expansion of this size is expected to 1) increase fish mortality because of transmission of pathogens and diseases (Johansen et al., 2011; Torrissen et al., 2013) and 2) increase conflicts in coastal and marine areas (user-user conflicts, or user-environmental conflicts) or further intensification of existing conflict (Bailey & Eggereide, 2020).

In Norway, the interest in the coastal areas has increased substantially over the recent years. Various competing users with divergent interests compete for the same sea areas (Robertsen et al., 2020). The consequences of this are increased total pressure on the marine ecosystems and loss of habitat and biodiversity (Douvere, 2008). Different approaches such as Integrated Coastal Zone Management (ICZM), ecosystem-based management, marine spatial planning (MSP) and simple suitability models have been developed to allocate human activities, minimize conflicts and at the same time maintain environmental, economic, and social sustainability. As a management tool, these approaches can help to allocate space for new activities such as aquaculture at sites with favourable geographical and operational characteristics and the minimum potential for conflicts. Choice of location is a crucial factor in any aquaculture operation. For sustainable development, the site should guarantee product quality, that the production does not adversely affect the ecosystems or other coastal zone users.

In Norway, the aquaculture industry reports that access to good sea areas with optimal environmental conditions is the most critical obstacle to further development and growth (Hersoug, Andreassen, Johnsen, & Robertsen, 2014; Hersoug, Mikkelsen, & Osmundsen, 2021; Longva & Elvenes, 2016). For management purposes, the Norwegian coast is separated into 13 different aquaculture production areas. These areas are part of a traffic-light system implied by the government to differentiate between healthy, regular, and “sick” zones, associated with green, yellow, and red colours. This system aims to regulate growth, where total biomass can increase in healthy areas whilst not allowing growth in other regions (Falconer et al., 2020). Furthermore, aquaculture is not the only user of the coastal zone in Norway. Many different users, such as fisheries, energy, transport, and tourism. The expansion of aquaculture would always be at the cost of other coastal zone uses (user-user conflicts) (Stelzenmüller, Gimpel, Gopnik, & Gee, 2017; Wever, Krause, & Buck, 2015).

Disregarding this and following the regulation within the traffic light system, the potential for further growth is in the two northernmost counties (Robertsen et al., 2020). The southern parts of Norway, where the density of salmon farms is highest, also have the most significant challenges with pathogens and diseases.

Planning ocean space and allocating space for aquaculture is a complex spatial decision problem. Methods such as MSP, ICZM, and suitability mapping have been developed to help planners and decision-makers make informed decisions. In Norway, to perform aquaculture, one must have a permit and a locality. Aquaculture companies can apply for areas assigned to aquaculture through a coastal planning process, but they can also apply for other areas not included in a plan. The former leaves the responsibility of impact- and suitability assessment with the planning authority. The latter goes the responsibility of suitability with the applicant. In both cases, appropriate suitability modelling or identification of optimal sites could help the industry's sustainable growth.

In recent years, the development and availability of high-quality temporal and spatial environmental and socioeconomic data along with fine-scale site-specific production data have made it possible to analyze yesterday's farming sites to improve the production of tomorrow's farming sites. Enhancement of computational power and machine learning approaches combined with geographical information systems (GIS) can provide an abundant opportunity to identify significant factors and identify optimal farming sites to assist planners, stakeholders, and decision-makers in choosing optimal farming sites. Planning with such tools could increase fish growth, reduce mortality, maximize profit while maintaining social and environmental sustainability and thus help Norway achieve its ambitious goals of aquaculture development.

Recently there have been some efforts to identify optimal sites for aquaculture (DEFRA, 2014; Falconer, Telfer, & Ross, 2016; Longva & Elvenes, 2016). For example, Longva & Elvenes (2016) tried to model optimal salmon farming sites in south Troms using MaxEnt with a species distribution modelling (SDM) technique. Telfer & Ross (2016) used Mahalanobis and MaxEnt with the same approach as Longva & Elvenes (2016) to identify optimal aquaculture sites in Vietnam. Despite Longva & Elvenes (2016) efforts, very few studies have used fine-scale empirical data from the industry and combined this with available high resolution environmental and socio-economic data to model optimal sites for salmon farming in Norway. There have been rapid advances in artificial intelligence and machine learning in recent years. These advances have made it possible to solve complex tasks that previously were impossible or required human expertise (Malde, Handegard, Eikvil, & Salberg, 2019). Random forest is such a model. The algorithm was developed by Breiman (2001) but, in later years, integrated with GIS and other software to help analysts solve complex spatio-temporal challenges using big data. Random forest is a highly acknowledged machine learning algorithm and performs exceptionally well in ecological predictions (Mi, Huettmann, Guo, Han, & Wen, 2017).

Experience from the industry indicates that some aquaculture sites can produce more robust, healthier fish faster than other sites in the same operational environment when the human factor linked to operating routines and regulations is disregarded (Longva & Elvenes, 2016). Suppose the performance of the aquaculture site is linked to external factors or governed by temporal and spatial site-specific environmental conditions. The random forest algorithm can help identify these factors and predict optimal salmon farming sites in the study area. This thesis is based on this fundamental concept.

This thesis aims to combine fine-scale production and location data from salmon farms in northern Norway in the period from 1. Jan 2018 to 31. Dec 2021 with environmental and socioeconomic data in a purely data-driven ensemble approach. The main objective is to develop a predictive model to identify optimal sites for salmon farming, defined by high salmon growth and low mortality, in northern Norway. The specific objectives are:

1. To identify significant variables for salmon growth and mortality in northern Norway based on fine-scale aquaculture site production and location data.
2. To identify and predict optimal areas for salmon farming in northern Norway.

1.1 Structure of thesis

This thesis is structured with a short introduction that starts with the relevance of this subject from a global perspective. Chapter two presents a brief literature review and the conceptual and methodological framework of the study. The chapter starts with salmon farming in Norway, continuing with the area challenges and then the governance of aquaculture farming in Norway. The chapter also briefly introduces MSP and the methodological framework, i.e., suitability modelling and the machine learning approach used in this study. In chapter 3, a short introduction to the study area and analysis methods are presented. Chapter 4 presents the main results. Chapter 5 discusses the results and describes the limitations of this study. Lastly, chapter 6 provides a concluding summary of the research findings and gives a short answer to how these research outcomes addressed the objectives of the thesis.

2 Literature review and conceptual framework

2.1 Salmon farming in Norway - area challenges

Aquaculture in Norway has developed from being a primitive experimental stage industry to a research-based, technologically redefined industry in the last four to five decades. Norway is currently the largest producer of farmed salmonids in the world. In 2021 production was approximately 1.3 million tons of Atlantic salmon and rainbow trout, of which 97% was exported (Hersoug et al., 2021; V. Oliveira, Dean, Qviller, Kirkeby, & Bang, 2021). At present, salmon farming constitutes 74% of total seafood export value from Norway, thus by far surpassing the traditional fisheries (Hersoug et al., 2021). Moreover, the Norwegian government aims to increase aquaculture production fivefold by 2050 (Ministry of Trade Industry and Fisheries 2014-15).

Today there are 1067 live fish farms cultivating salmonids (Norwegian Directorate of Fisheries, 2021). Access to productive areas is one of Norway’s most critical competitive advantages (Gullestad et al., 2011). Norway has a coastline of 103 000 km (including inlets and islands) and an exclusive economic zone (EEZ) of nearly one million km. It may seem paradoxical that aquaculture's lack of coastal space is a significant problem (Hersoug et al., 2021). However, when asked about the most critical challenges for further growth, the farmers highlighted lack of space as one of the biggest obstacles (Hersoug et al., 2014). In 2010 the aquaculture installations occupied a total area of 59 km^2 for the 900 licenses. However, in practical terms, the occupation is much more excellent. Transport should keep a distance of 25 m, and it is prohibited to fish closer than 100 m to the net pen. Furthermore, the anchoring area around the aquaculture farm often stretches up to 1000 m from the farm. If we use the anchoring area as the occupational area of the 900 licenses, the total area equals 420km^2, which is less than 0.5 % of the entire coastal area (Andreassen, Johnsen, & Hersoug, 2010).

So why the shortage of area? First, not all space is equally valuable. In modern salmon farming, fish farmers look for so-called “super localities” with favourable characteristics. The net-pen must be located in a sheltered area to ensure the safety of gear and operators (Hersoug et al., 2021). In addition, the locations must be located so that they can be part of an appropriate logistics system (transport of intermediate goods and personnel) and so that the companies can utilize any economies of scale (Hersoug et al., 2014). It is, however, not the sea are per se that most interests the fish farmers, but the water body beneath (Hersoug et al., 2021).

The Atlantic Salmon aquaculture is based primarily on open water sea cage production, where fish exposes to a complex of natural and artificial environments. The areas must have a suitable temperature, salinity, wave height, and other ecological factors. Current is essential so that faces and spilt feed are removed from the locations and that there is a continuous supply of oxygen-rich water, but not too strong so that the fish is exhausted from swimming. All these factors strongly reduce the available space for aquaculture operations. Secondly, aquaculture is only one of the many legitimate coastal zone users. Norway has a sizeable inshore fishing fleet that utilizes these areas. These fishing vessels play an essential role in the population of many of the small coastal communities along the coast. Usually, the coastal vessel is smaller boats that cannot go too far out in the open sea. Therefore, they are dependent on the fishing fields available inside fjords and along the coast. The Directory of fisheries has mapped out all these fishing areas and spawning grounds. Aquaculture can normally not be suited in these areas.

Further limitations to aquaculture include the sea transport in Norway, a large industry that carries nearly as much cargo as the roads on land (Hersoug et al., 2021). A system of fairways exists (in Norwegian: hoved- and biled). The Norwegian Coastal Administration is responsible for keeping these fairways free of intervention. In addition, the Norwegian Food and Safety Authority, as part of the general infection risk assessment, together with living environment requirements, is decisive for whether establishment applications are granted (Bjørnshol & Khan, 2020; Regulations on The Establishment and Expansion of Aquaculture Facilities etc., 2008). The Norwegian Food Safety Authority has, on a general basis, assessed the risk of infection to and from well boats to the extent that there is one recommended minimum distance of 1500 meters at sea to «important transport routes (fairway) for living farmed fish (Bjørnshol & Khan, 2020).

The energy sector is an emerging sector in the coastal areas. Production of oil and gas takes place offshore and is thus not in direct conflict with the coastal planning. However, large pipelines occupy a lot of the seabed from the installations. In recent years, marine wind power has been proposed to produce more sustainable energy (Heidenreich, 2016). Wind power production is often sought for shallow water and will thus have overlapping interests with the coastal fishing fleet and aquaculture interest. More recently, the interest in tourism and especially fishing in coastal areas have expanded. Fishing tourists usually want to come to untouched areas and regard salmon farms as a nuisance, as they limit the areas for fishing (Borch, 2009). There has been tremendous growth in fishing tourism, with 430 dedicated fish farms in 2014, while today, there are more than 1100 companies registered as fishing tourism companies (Hersoug et al., 2021). In addition, recreational kayaking, aurora lights tourism and whale safari are prominent actors in northern Norway.

The Norwegian military is also an actor worth mentioning in this section and claims large areas for training and exercise. In Troms, one of the most important areas for expanding the aquaculture industry, the navy has exclusive rights to ca 30% of the inshore coastal area, Rånes (2017) cited in (Hersoug et al., 2021). Further limiting all interest in the coastal regions is marine conservation which now holds 3% of the coastal zone as protected. The goal is to increase this to 7% (NOU 2004: 28, 2004).

In 2007 it was also decided that 53 national salmon rivers and 27 national salmon fjords should be protected from the further establishment of aquaculture farms due to concerns of transmission of pathogens and escapes that could affect these rivers.

Those mentioned above are significant stakeholders, limiting further expansion in the aquaculture industry. However, the most limiting factor to further growth is the industry itself. The Norwegian Food Safety Authority has decided that each farm should have at least 2,5 km to the next farm and a minimum of 5 km to a processing plant (Hersoug et al., 2021). This is due to protecting other farms from the transmission of diseases and pathogens. It could be argued that this distance could be reduced if the current and local conditions did not affect neighbouring farms. There have been efforts to model pathogens' transmission, especially salmon lice, using hydrographical models (Sanvik, Asplin, & Skardhamar, 2019), but more knowledge is required before a detailed system can be developed put in place (Hersoug et al., 2021).

2.1.1 Governing area access to aquaculture

To perform aquaculture in Norway, you need a license containing two independent licenses. First production license. These licenses were first allocated for free, object to strict regulations and regulated according to net-pen volume. This was later combined with feed quotas and permits sold at fixed prices due to production limitations following the EU´s threats of trade restrictions due to accusations of Norwegian dumping and the use of irregular subsidies (Hersoug, 2015). Throughout the 1990s, the licenses and quotas acted as production control measures to limit growth so that the industry would not be affected by any more sanctions. Later in 2005, feed quotas were replaced by the maximum allowed biomass (MAB) and licenses subject to public auctions. A standard permit used to be 780 MAB for the licenses in the south. While licenses far north were set at 945 MAB to compensate for lower temperatures and slower growth rates (Hersoug et al., 2014). Since 2017 the industry has been regulated by the traffic light system, determined by the production area's environmental status, as measured by one single indicator; the frequency of salmon lice (Osmundsen, Olsen, & Thorvaldsen, 2020). In green zones, total production can increase. In yellow zones, production must be stable, and in red zones, production must be reduced, measured by maximum allowable biomass (MAB). A reporting scheme was introduced in 1996 to control feed consumption. When the government changed from feed quotas to MAB, the scheme continued to control the standing biomass at facilities. The fish farming companies report to the Directorate of Fisheries monthly at the cage level; one report per locality, per company. The reporting scheme from 2006 has been modified and modernized over the years, and today appears as electronic reporting (Norwegian Directorate of Fisheries, 2021).

In addition to the production license, companies must have access to several locations to operate according to government requirements and newer principles for infection prevention and environmental considerations. It usually also involves co-location or cooperation with other licenses, so each locality usually has to be cleared for a larger production than the MAB for each license (Hersoug et al., 2014).

The aquaculture location is regulated according to eight sectors laws (Aquaculture Act, 2006; Outdoor Recreation Act, 1957; Planning and Construction Case Processing Act, 2008; Ports and Waterways Safety Act, 2019; Regulations on The Establishment and Expansion of Aquaculture Facilities etc., 2008; The Food and Safety Act, 2004; The Pollution Control Act, 1983; Water Resources Act, 2001). A locality can only be allocated if permissions are granted according to the different sector laws governed by, The Norwegian Coastal Authority, Norwegian Food Safety Authority, The county governor, and The Directory of Fisheries. This gives some sector directorates a de facto veto right to block an application for aquaculture space (Hersoug et al., 2021). The counties in Norway coordinate and grant the locality permit.

According to the Plan and Building Act, the locality must be allocated following the municipality's area plans.

By the Planning and Building Act, the municipality is required to prepare and adopt a municipal planning strategy at least once each election period (Planning and Construction Case Processing Act, 2008, § 10-1). This plan will discuss a long-term strategy for the municipality that applies community development and long-term land use. The municipality must also prepare an area plan for the entire municipality (Planning and Construction Case Processing Act, 2008, § 11-5). This plan shall provide a detailed overview of the municipality's land use and conditions for new measures. A part of this plan is coastal zone planning up to one nautical mile from the baseline. Planning of coastal areas (sea-area) is a voluntary activity, but almost all municipalities have either the sea area included as a part of the municipality plan (in Norwegian: kommunends arealdel) or as a separate coastal zone plan.

According to the Planning Construction Case Processing Act, all regional plans and municipal plans that may have significant effects on the environment and society must have an impact assessment on the plan's impact on the environment and society (Planning and Construction Case Processing Act, 2008, § 4-1, § 4-2). Planning, including coastal zone planning, involves balancing various sectoral interests. Together with risk and vulnerability analyses, impact assessments shall reveal possible consequences for the environment and society if the land use is changed or other measures are initiated. This is a part of the basis for the trade-offs made in coastal zone planning (Berg, Solås, Kvalvik, & Hersoug, 2017).

Coastal zone planning is driven by the increasing interest in the coastal zone of various interests, especially aquaculture. Before aquaculture became a major industry, sea transport at fishing dominated the nearby coastal area, and there was little need for detailed plans. Today we are in a situation where the number of uses has increased dramatically. We also have a lot more knowledge about ecosystems than we had a few years ago. Because of this, coastal zone planning has become a very complex task. Hersoug and Johnsen (2012) stated that lack of competence and financial resources were major bottlenecks in coastal zone planning. To make this any more complex, To, In addition, the ecosystem does not care about the municipality border, and it’s an advantage that the neighbouring municipalities coordinate their coastal zone plans. The advantage of cooperation between several municipalities is that they can combine their financial resources, knowledge, and planning capacity.

To create a successful coastal zone plan, components, processes in the ecosystems and human impacts must be translated symbolically (Osmundsen et al., 2020). This is the rationale for MSP, where areas managed are governance objects representing abiotic and biotic conditions and patterns of use and stakeholders' interests (Hersoug et al., 2021). There might not be any optimal areas for the different uses in the coastal zone. It’s especially true with aquaculture locations. Trade-offs would always be required. However, it is possible to make informed and well developed marine spatial- and coastal plans to reduce user-user conflicts and user-environmental conflicts whilst maximising the potential of the salmon farming locality. The following section aims to explore this.

2.2 MSP – aquaculture sites

The standard methodology for dealing with multiple interests and conflicts at sea is MSP. The concept was introduced in the 1990s and promoted to achieve more ecosystem-based marine management, focusing on holistically balancing multiple management objectives (Stelzenmüller et al., 2017). MSP analyses and allocates parts of three-dimensional marine spaces to specific uses. The overall goal is to achieve ecological, economic, and social objectives usually specified through a political process (Douvere, 2008). It is the preferred framework for an integrated approach that effectively deals with conflicts and is essential to sustainable ecosystem-based sea management. Douvere (2008) presented a base method in Marin Policy consisting of at least three phases (Figure 1), 1. planning and analysis, 2. implementation, and 3. monitoring and evaluation. The planning and analysis phase should generate and adopt an integrated spatial plan for the protection and sustainable use of the sea and its resources. This phase should be based on research and initiatives (including mapping) of human and environmental processes. The implementation

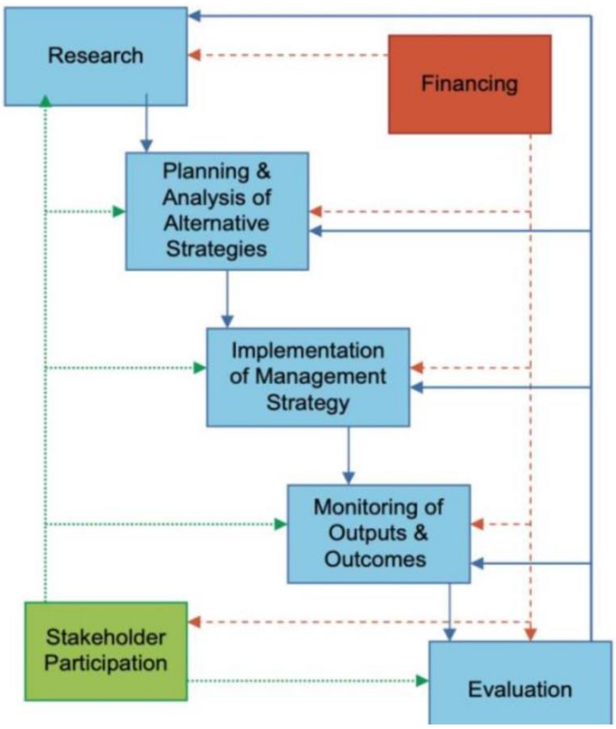


Figure 1: Marine spatial planning, as shown in Douvere (2008). It explains the concept of MSP in a three step process.

of the plan through the execution of programmed works or investments, encouraging improvement and through regulation, incentives, and enforcements of proposed changes and ongoing activities. The monitoring and evaluation phase involves assessing the effectiveness of the plans, their time scale, and their implementation mechanisms. Results of the assessments are fed back to all steps in the process (Douvere, 2008).

Previously, very few frameworks facilitated integrated strategic and comprehensive planning concerning all activities in the marine areas. The lack of such frameworks often translates to, but is not limited by, the spatial overlap of human activities and their objectives, causing conflicts between users and between users and the environment (Douvere, 2008). Because the users of the marine space have different goals, their uses are often not compatible with one another. Thus, users are competing for the same ocean space. For example, a study in Belgium focusing on ocean space interactions found that the negative interactions by far outreached the positive (Maes et al., 2005). Compared to conventional planning, MSP is a framework that provides an ecosystem-based approach for multi-sectoral decision-making to reduce socioeconomic conflicts and environmental impacts.

After some initial hesitation, fisheries and aquaculture stakeholders are now actively becoming engaged in MSP to secure the most suitable sites for their industry (Jentoft & Knol, 2014). MSP is also seen to help increase animal welfare and social acceptance, decreasing mortality and adverse environmental impact (Bryde, Krause, & Rosenthal, 2011). While it could be argued that other place-based approaches adopting areas of aquaculture would deliver similar benefits, statutory MSP brings certain strategic advantages (Stelzenmüller et al., 2017). MSP brings a more coordinated approach to widespread sea use, giving more accountability and transparency to decision-making by including a wide range of stakeholders (Stelzenmüller et al., 2017). It is also a cost-effective tool that increases the effectiveness of investment, reducing duplication of effort and speeding up decision-making (Stelzenmüller et al., 2017).

As a strategic tool, MSP can be performed with a Geographical Information System GIS to map and allocate space for aquaculture with favourable operational characteristics, also called suitability modelling. Identification of the most suitable sites for aquaculture should minimize environmental stress, maximize the potential for species growth, minimize production costs and avoid or minimize potential conflicts. Hence, site selection is a complex spatial decision problem that requires spatially explicit methods to assess both the environmental opportunities and risks and explore the relationship between multiple human pressures and the status of the ecosystem. GIS is often used to solve such complex spatial decision problems.

GIS is a system for collecting, organizing, storing, analyzing, and presenting geographically located information (Dick., 2015). To conduct efficient and good decisions, one must have a common platform to store all knowledge and information. Given the high spatial context of suitability modelling and MSP, GIS is the most commonly used decision support tool (Kapetsky, Aguilar, & Jenness, 2013). The recent development of computational power and GIS has fueled the development of GIS-based multi-criteria evaluation techniques (MCE), and spatial multi-criteria evaluation (SMCE). SMCE visualize opportunities and challenges such that decision-makers can make informed decisions easier (Alkema, 2007; Alkema, Boerboom, Ferlisi, & Cascini, 2014; Malczewski, 2006).

2.2.1 Suitability modelling

Sustainable site selection requires spatially explicit information about suitable areas and spatial assessments of overlap with other human activities (Stelzenmüller, Lee, South, Foden, & Rogers, 2013) and their combined impact on the marine environment (Douvere, 2008). Investigating the interplay and connection of phenomena and their underlying causes is essential for further developing salmon farming in open sea cages. Ensuring low mortality is vital for sustainable production since fewer resources are wasted when the animals are grown to slaughter (Hjeltnes, Bang, Bornø, Haukaas, & Walde, 2019). If we could explain the environmental factors associated with growth and mortality, marine spatial decision-makers could decide where to locate salmon farms through suitability modelling and/or MCE.

GIS-based suitability modelling is one of the most used applications to evaluate potential aquaculture sites and can support the further sustainable development of the industry. It is also expected to be significant for future MSP (Filgueira, Guyondet, Comeau, & Grant, 2014). Suitability modelling refers to the spatial overlay of geo-data layers to identify suitable aquaculture sites by determining, for instance, favourable environmental factors or constraints (Stelzenmüller et al., 2017). The development of suitability modelling techniques dates to the 1980s. The first application of these techniques is the siting of aquaculture and inland fisheries using GIS and remote sensing data conducted by FAO. Until the mid-1990, studies continued to target small-scaled data-rich areas (Stelzenmüller et al., 2017). The earliest efforts focused on simple siting models. At the beginning of 2000, the focus shifted to land-based and coastal aquaculture related to different species (W. L. Fisher, Fisher, & Rahel, 2004). Suitability modelling applied to larger areas was limited because of the lack of fine-scale data with the necessary temporal and spatial resolution (Stelzenmüller et al., 2017).

Later, when spatially fine-scaled spatial and temporal data became readily available, GIS-based spatial multi-criteria evaluation approaches were developed and used (Malczewski, 2006). MCE methods are primarily used in risk assessments and predicting, for example, the risk of floods and earthquakes (Alkema, 2007; Dang, Babel, & Luong, 2011; Rashed & Weeks, 2003), but now it is also used to identify the suitability of aquaculture sites often using SMCE approaches (Dapueto et al., 2015; Perez, Telfer, & Ross, 2005).

To identify optimal aquaculture sites in an SMCE process, one must first identify significant factors for aquaculture operations (Perez et al., 2005). These factors are often identified and ranked based on expert knowledge or through a participatory group discussion process between experts, decision-makers, stakeholders, and other interest groups (Malczewski, 2006), giving each factor a weight. Combining weights, for example, using the weighted linear combination method, creates an output map with different suitable classes (Dapueto et al., 2015). The challenge with such weighted ranking methods is that they are very subjective. Furthermore, complicated relationships between variables are simply translated to a weight. Thus, bias and prejudice will be central sources of error and could impact the result ("Weighted by ranking," 2013).

Further development has led to suitability analysis using a combination of GIS and dynamic models. For example, Silva et al. (2011) used a dynamic model to identify suitable sites for shellfish aquaculture. Recently, the ecosystem approach to aquaculture has been advanced by developing different GIS-based dynamics models and decision support systems (DSS), for example: - AquaSpace (Gimpel et al., 2018), Akvavis (Norwegian: dynamisk GIS-verktøy for lokalisering av oppdrettsanlegg for nye oppdrettsarter), Farm Aquaculture Resource Management (FARM; http://farmscale.org/), or management level eutrophication screening model (ASSETS) (Stelzenmüller et al., 2017). However, the main problem with such DSS tools is data availability. The data requirements of a suitability modelling expand with the scale of the aquaculture operation and the study area. Thus, it is challenging to use aquaculture suitability modelling in a broader spatial planning system, such as MSP, where a large ecosystem scale is covered (Stelzenmüller et al., 2017).

In Norway, state-of-the-art models have been developed to deal with data availability issues. Models such as the Norwegian Coastal model (NorKyst800, see: Albretsen et al. (2011)), Simulating Waves Nearshore (SWAN; http://swanmodel.sourceforge.net/), and Applications of Research to Operations at Mesoscale - Meteorological Cooperation on Operational Numerical Weather Prediction (AROME – MetCoOp; Müller et al. (2017)), developed by respectively, The institute of marine research, Delft University of Technology and the Norwegian Meteorological institute have given Norway an excellent opportunity to model spatial relationships and make predictions based on the best data available.

2.3 Modelling spatial relationships: approaches

Identifying the nature of spatial relationships between variables is crucial in modelling. The concept of a model is that it represents reality in a way that would be useful. It is a generalization of something that can be used to make predictions. The goal is to use available information to predict information that is not available. The development of modelling techniques dates back 200 years; in 1805, Legendre first published the method of least squares, a model that has played a central role in the statistical methodology used in physics, biological and social sciences (Eichler et al., 2012).

In general, with modelling spatial relationships, a principle exists called parsimony. The idea of parsimonious models stems from Occam's razor or the law of briefness – "plurality should not be posited without necessity" or "entities are not to be multiplied beyond necessity" (Blumer, Ehrenfeucht, Haussler, & Warmuth, 1987; Brian, 2021). There is a trade-off between the goodness of fit and parsimony. Low parsimony models tend to have a better fit than high parsimonious models. It is not usually a good approach. Adding more parameters usually results in a good model that fits the data at hand. The same model will likely be useless for predicting another dataset or other study areas. Occam's razor also applies to the choice of method; we want to use the most straightforward method possible and nothing simpler.

There are different techniques to estimate relationships between variables. Ordinary least squares (OLS) are one of the simplest methods. OLS is a global model that estimates the relationship between one or more independent variables and a dependent variable. The method estimates the relationship by minimizing the sum of squares in the difference between the observed and predicted values of the dependent variable. Such regression techniques are popular when modelling driving factors (Lugert, Thaller, Tetens, Schulz, & Krieter, 2016; Maceina, 1992). Usually, regression techniques are used with the goal of using the least amount of variables to give the most significant explanation of variability in the dependent variables (Graham, 2003). Multiple regression analysis has several ways to select the most relevant exploratory variables, following stepwise procedures based on Akaike's information criterion, F, or other measures (S. Oliveira, Oehler, San-Miguel-Ayanz, Camia, & Pereira, 2012).

When working with complex spatial relationships, it is crucial to understand the strength and weaknesses of the models. Especially with OLS, spatial autocorrelation tends to cause problems. Spatial autocorrelation describes the spatial dependencies or presence of systematic spatial variation in a variable. Tobler's first law of geography describes this concept: everything is related to everything, but near things are more connected than distant things (Tobler, 1970). Non-spatial statistics such as global OLS assume that all data is randomly distributed. Internal relationships between the data can cause non-negligible interferences in modelling behaviour and calculating essential statistics in the dataset (Luc, 2016). Spatial autocorrelation creates an overcount type of bias (Jenora, Alberto, & Flora, 2019).

Alternatively, the geographically weighted regression (GWR) model could be applied to a global model. GWR typically does not have the same problems related to spatial autocorrelation as the non-spatial global models. GWR is a local spatial regression method that evaluates the variable we try to understand by fitting a regression equation to every future dataset. The equation is calibrated based on the neighbouring type and neighbourhood selection method parameters. Several studies have shown that GWR performs better than any other global regression method when spatial autocorrelation is present (For example, Windle et al. (2010). Windle and co-authors explored GWR methods applied to fisheries data. They compared the performance of GWR with global logistics regression models of the distribution of northern cod based on environmental and biological variables. They found that in the presence of spatial autocorrelation, the GWR performs better than the other global methods. Furthermore, Cullen and Guida (2021) studied nonstationary environmental effects on black sea bass and scup distribution. They also found that the GWR models had higher goodness of fit and predictive accuracy than the non-spatial methods.

The problem with linear models such as those mentioned above is that they are unlikely to find a suitable model when working with complex data relationships when nonlinear relationships exist between variables. They are better at describing overall data relationships in the study area when those relationships are consistent. However, when the explanatory variables exhibit nonstationary relationships, such global models tend to fall apart (Vale, 2021). In addition, linear regression methods are susceptible to outlier effects and the problem of multicollinearity. One of the proposed methods for dealing with this type of complex data relationship across a wide study area is machine learning techniques. The interest in modelling through data-driven approaches and machine learning techniques has grown in popularity in the last few years. Especially models that evaluate species-habitat relationships have gained more attention with increasing interest in ecosystem management. However, developing models that can incorporate large dependent and independent variables is challenging (K. Miller, Huettmann, Norcross, & Lorenz, 2014). It is important to note that all machine learning techniques have a foundation in statistics, but it's not statistics. They will not show the same level of confidence or levels of statistical significance that one would get in a regression analysis such as OLS.

Several data-driven approaches have been developed to model species distributions (J. Miller, 2010). One widely used machine learning method in ecology is Maximum Entropy (MaxEnt). MaxEnt is a principle to find probability distribution, at which an event occurs with the most significant uncertainty while being subject to some constraints that the statistical moments of the distribution match with the sample moments of observations (Farrell et al., 2019). MaxEnt has become a benchmark in environmental niche modelling. MaxEnt analyzes the connection between observations and the environment in which they are made and indicate the probability that the model object will be found elsewhere in the modelled area (Phillips, Anderson, & Schapire, 2006; Phillips & Dudík, 2008). Longva and Elvenes (2016) attempted to conduct a modelling experiment with MaxEnt using the physical environmental data from the Aastafjord project linked to production data from the aquaculture industry (in their study, Ewos growth index (EGI) based on production data was used to rank or classify order of the aquaculture locations). The study aimed to demonstrate which physical environmental parameters were significant for aquaculture sites in south Troms, and identify new places with the best physical conditions for salmon farming. The idea was that a point dataset indicating optimal farming sites could be used similarly to a dataset with species observations. Suppose the physical conditions determine the quality of a locality, and the most important physical conditions are represented in the model. In that case, MaxEnt can point out which physical parameters are decisive and any other areas in the study area with the same physical parameters (Longva & Elvenes, 2016).

In addition to MaxEnt, random forest (RF), a machine learning technique based on an automatic combination of decision trees, is often seen to reach top predictive performance in building predictive habitat models for species distribution. RF has become one of the most popular machine learning algorithms for SDM (Mi et al., 2017; Valavi, Elith, Lahoz-Monfort, & Guillera-Arroita, 2021; Williams et al., 2009). RF is based on the algorithm provided by Breiman (2001) and is an ensemble of classification or regression trees (CART). RF fits many individual trees, usually several hundred and combines their predictions (Valavi et al., 2021). Each tree is trained on a bootstrap sample which is called in bag samples, and the remaining are called the out of bag (OOB) samples and are used to estimate model error. For example, if 2/3 of the records are used for the bootstrap sample, 1/3 are called OOB. Each split RF only evaluates a random subset of predictors to identify the best predictor (Degenhardt, Seifert, & Szymczak, 2017; Valavi et al., 2021). This produces decorrelated trees, making overprediction less likely (Breiman, 2001). The RF algorithm uses many random samples from the original data, fits classification and regression trees to each random selection, and then aggregates the votes over all the trees to make classifications or numeric predictions (Breiman, 2001). RF may achieve excellent performance for suitability predictions unmatched by other machine learning methods by minimizing the mode's variance and bias (Breiman, 2001; Farrell et al., 2019). RF can also estimate variable importance by rank ordering the predictive significance. Importance is calculated using Gini coefficients, which can be thought of as the number of times a variable is responsible for a split and the impact of that split divided by the number of trees. Splits are each decision within a decision tree (Grömping, 2009).

In this study, I extended the Longva and Elvenes (2016) approach to find the optimal aquaculture site using RF. The fundamental assumption is that if given a set of points indicating an optimal aquaculture area, RF would be able to predict other locations in the study area inhabiting similar attributes with top performance. As with any different modelling approach, efforts should be made to obtain parsimony with RF. However, this is not motivated by reducing the risk of over-fitting but rather driven by interpretability (Evans, Murphy, Holden, & Cushman, 2011). Another reason for seeking parsimony in RF is model performance – improving model fit and predictive performance (Evans et al., 2011). As spurious variables are removed, the trees become shallower. This reduces the size of the plurality vote matrix by reducing votes that account for noise, resulting in a higher signal to noise ratio and an overall reduction in error (Evans et al., 2011). Several different approaches are developed to reduce the number of exploratory variables and identify the most important variables based on ranking (See elaborated description in Degenhardt et al. (2017)). The recursive feature elimination method is by far the most popular, with over 671 citations (Degenhardt et al., 2017). The overall goal for this method is a minimal set of features (Díaz-Uriarte & Alvarez de Andrés, 2006). It removes a specific proportion of the least important variables.

2.4 Welfare needs of salmon and animal welfare indicators

Welfare indicators are widely used to measure the performance of the farm sites (for example, Longva and Elvenes (2016) used EGI). The welfare needs of Atlantic salmon can be directly linked to its available resources, water environment, health, and behavioural freedom. Fish welfare is a crucial issue in commercial salmon farming. It is central to all farmers when making decisions during their daily husbandry practices and longer-term production planning (Noble et al., 2018). Fish farmers should be very interested in keeping fish welfare at their attention.

Good fish welfare often results in a superior product with a low mortality rate. Fish welfare addresses fish's physical and mental health as an individual or part of a group. Excellent fish welfare is thus making sure that the fish is treated well, that they have a good quality of life, and that suffering is avoided. There are many factors influencing fish welfare. Fish farmers know this and have directly or indirectly tried to optimize fish welfare over the years. There is no consensus or universal definition of animal welfare, but most animal welfare scientists agree that animal welfare relates to the individual animal experience. Stien et al. (2013) defined animal welfare as the quality of life perceived by the animal itself (Noble et al., 2018). Nolbe et al. (2018) divided the welfare status of the fish into four main categories, resources, environment, health, and behaviour. An overview of this model is provided below (Figure 2)

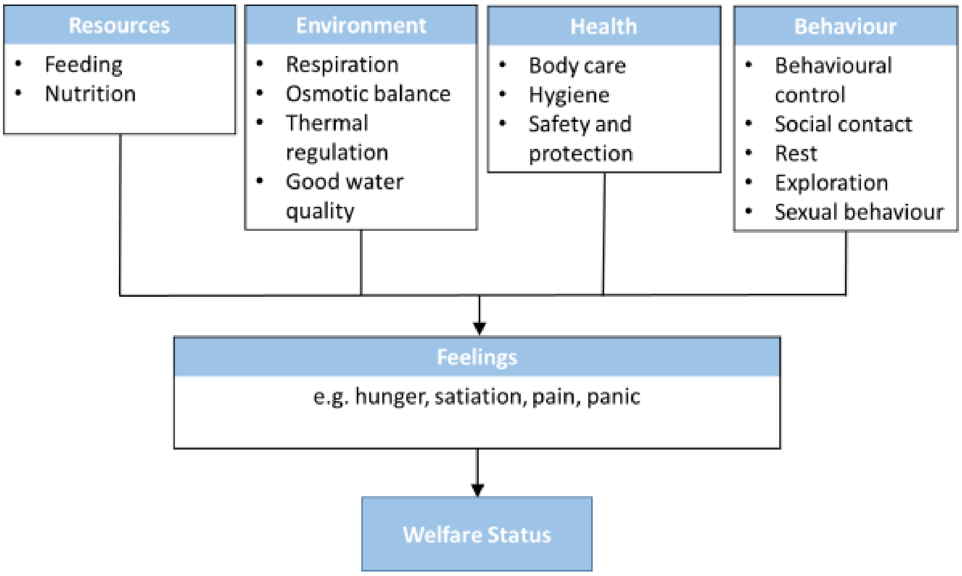


Figure 2: Welfare status of the fish. The welfare needs of salmon are categorized into available resources, environment, health, and behaviour. Figure adapted from "Mellor, D. J., Patterson-Kane, E. & Stafford, K. J. (2009). Noble et al. (2018).

The fulfilment of all of these would affect the fish's mental state and, thereby, their welfare status. The development of animal welfare indicators has made it easier for the farmers to know the physical and mental state of the fish. Noble et al. (2018) describe 22 welfare indicators; 7 of these are within the group category, including mortality rate, behaviour, surface activity, appetite, growth, scales or blood in the water and disease. In this thesis, I focus on the two categories: resources and environment. Over the years, a different set of welfare indicators (WI) have been developed to get information about the fish and the current state of its welfare. WI can be direct animal-based, the observation of attributes of the animal itself, or WI can be an indirect environmental-based indicator centred on the resources and environment the animals are subjected to, such as the level of dissolved oxygen or the temperature (Noble et al., 2018).

Animal-based welfare indicators are directly linked to the state of the fish. They can indicate prior welfare problems, e.g., previously poor nutrition, storms, high density inside the net pen or low oxygen. In addition, animal-based indicators can be behavioural responses from the fish. Low oxygen in the net-pen will cause higher ventilation rates and fish gasping for air at the surface. "Animal-based WI are sometimes also called outcome-based WI emphasizing that these Wi's measure the result of the treatment in the animals themselves" (Noble et al., 2018). Fulfilling the needs of the brain would release opioids to give pleasurable emotions and feelings, telling the fish that their actions were appropriate. When the states of one or more needs worsen, their punishment circuits release neurotransmitters that give unpleasant emotions and feelings" (Dawkins, 1990; Noble et al., 2018; Spruijt, Van den Bos, & Pijlman, 2001).

Noble et al. (2018) separated the welfare indicators into two main categories: group and individual. Group based welfare indicators are indicators that do not involve the handling of the fish. In comparison, individual welfare indicators require handling or examination of the fish. Table 1 shows the group-based and individual-based animal welfare indicators and the respective relationship with their needs. From the table, we can derive that mortality rate, behaviour, appetite, growth, and diseases have a relationship with all associated needs. Welfare indicators requiring observation of the fish are not applicable in this study. Therefore, behaviour and diseases were not included further in the analysis as dependent variables. It is possible to calculate appetite based on the amount of feed divided by the number of fish at the location. I explain this further in chapter 2.4.3. However, appetite was not used as a dependent variable in this study because it is highly correlated with growth.

Table 1: Animal-based welfare indicators. divided into group-based and individual-based indicators. Source: Noble et al. (2018).

	Needs	Environment				Health			Behaviour					Reso-urces	
	Welfare indicators	Respiration	Osmotic balance	Thermal reg.	Good water q.	Body care	Hygiene	Safety and prot.	Beh. control	Social contact	Rest	Exploration	Sexual beh.	Feeding	Nutrition
Group	Mortality rate	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Group	Behaviour	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Group	Surface activity					×	×		×			×			
Group	Appetite	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Group	Growth	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Group	Scales or blood in the water	×	×					×	×						
Group	Disease	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Individual	Gill beat rate	×			×			×	×						
Individual	Sea lice	×	×			×	×	×							
Individual	Gill bleaching and gill status	×	×				×				×				
Individual	Condition factor													×	×
Individual	Emaciation state		×				×							×	×
Individual	Sexual maturity stage		×										×		
Individual	Smoltification stage		×												
Individual	Vertebral deformation								×		×				
Individual	Fin damage (non-active)								×		×				
Individual	Fin status		×				×	×							
Individual	Scale loss and skin condition		×				×	×							
Individual	Eye damage and status						×	×	×					×	×
Individual	Deformed opercula	×													
Individual	Abdominal organs						×	×							×
Individual	Vaccine-related pathology													×	×
Individual – Blood	Cortisol		×					×	×	×		×		×	
Individual – Blood	Osmolality		×												
Individual – Blood	Ionic composition		×												
Individual – Blood	Glucose							×						×	×
Individual – Blood	Lactate							×	×		×				

2.4.1 Mortality

One of the most limiting factors for growth in the production of salmonids in Norway is mortality. In 2019 more than 50 million Atlantic Salmon died in the final production stage in marine cages. From 2015 to 2019, there has been a 27.8% increase in mortality in sea cages from 41.3 million to 52.8 million Atlantic salmon. Reducing mortality in salmon farming is crucial to ensure sustainable production and growth. Mortality represents a significant economic loss for producers and signals a need to improve fish welfare (V. Oliveira et al., 2021). Events such as algae blooms and infectious disease outbreaks can explain mass mortality, but little is known about environmental factors contributing to baseline mortality in salmon sea cages (Oliveira et al., 2021). Numerous factors have been suggested as contributors to these losses. Amongst them are management, diseases, parasites, and environmental factors (Moriarty et al., 2020). Oliveira et al. (2021) looked at 1627 Atlantic salmon cohorts put to sea between 2014 and 2019. They found that sea lice treatments were associated with salmon mortality, especially the increased mortality from chemical sea lice treatment to thermal delousing. They also found that salinity and temperature influenced Atlantic salmon mortality (Oliveira et al., 2021). Jensen et al. (2019), cited in Oliveria et al. (2021), used mortality data from 2014 to 2018 and found significant Spatio-temporal variations in mortality between the different production zones. The areas with the highest density of farmed salmon also had the most increased mortality. Oliveira et al. (2021) found that the first month of stocking and the number of months at sea were statistically significant determinants for mortality. Previous reports of national mortality have also indicated distinct patterns related to Spatio-temporal and cohort properties (Bleie & Skrudland, 2014).

However, variations in mortality can also be related to climate, particularly water temperatures (Thyholdt, 2014). Warming temperatures positively affect salmon farming, including a higher growth rate. However, the associated adverse effects include increased prevalence and severity of disease and parasitic diseases (H. Fisher et al., 2020; Wiltsey Stirman et al., 2012). For Norway, parasites such as sea lice can directly impact salmonid production growth if not strictly managed (Kristoffersen et al., 2014; Torrissen et al., 2013).

The mortality rate is one of the most used health-related welfare indicators. Farmers are very interested in minimizing the mortality rate to increase economic profit. Therefore, knowing which factors are associated with low base mortality is essential. Mortality as an indicator can be calculated as either long-term mortality or short-term mortality. Stien et al. (2016) cited in Noble et al. (2018) reported that the standard (median) mortality curve is highest during the weeks following sea cage transfer. Then the mortality curve gradually declines and stabilizes at around 0.2% per month (Figure 3).

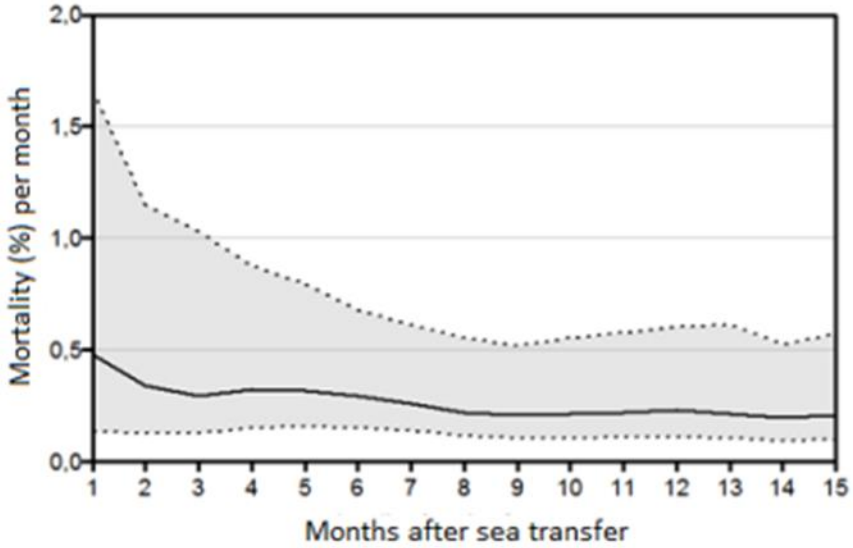


Figure 3: Standard mortality curve for the first 15 months of salmonids in cages based on reported data from all Norwegian farmers 2009-2015. The curve gives the monthly median mortality rate, in addition to 25 and 75 percentiles (Noble et al., 2018

2.4.2 Growth

Growth is one of the most critical metrics in fisheries and aquaculture. It has been used as a welfare indicator for animal production for a long time (DeLucas et al., 1986). Growth is tightly connected to feeding and the nutritional welfare needs of the fish. If these requirements are not met, the fish can exhibit poor growth performance (Noble et al., 2018). Jobling (1983) argues that growth rates are tightly connected to life stage, fish size and the nutritional content of the feed, diseases, social interactions, and water quality (Jobling, 1983). The growth rate can also be affected by internal cues in the fish of sexual mature salmon. Sexual mature Atlantic salmon can enter a state of anorexia with increasing temperatures from June to July (Kadri, Thorpe, & Metcalfe, 1997). In Northern Norway, there have also been situations where herring have entered fjords, and fjords have accordingly turned hypoxic. To reduce the oxygen consumption of the Atlantic salmon, farmers need to stop feeding; hence growth rates also decrease. If the fish have reduced growth over a long period, it could indicate that they face a welfare problem, and farmers need to investigate to identify the cause (Noble et al., 2018).

2.4.3 Appetite

Atlantic salmon are visual feeders, meaning they react to moving and visible feed, hence why commercially farmed salmon only are fed during daylight. The economic viability of any aquaculture operation depends on the farmer's ability to deliver a high-quality product at an acceptable cost for the consumer. Consequently, much focus has been directed to minimizing operational costs while maximizing production. Feeding is one of the significant cost drivers in commercial salmon farming. This is also one of the essential welfare factors to ensure stable growth and a healthy, high-quality product. Therefore, ensuring that feeding is delivered at the correct times during the day when the fish is motivated to feed is significant. According to Noble and co-authors (2018), appetite results from an array of factors, with the three predominant factors.

1. The nutritional status of the fish, including its energy reserves
2. The fullness of the stomach at the time of potential feeding
3. Seasonal adaptions and the fish's motivation to feed.

Once a fish decides to eat, the appetite can also be affected by other factors such as competition or the nutritional composition of the feed (Noble et al., 2018). Environmental factors can also affect and dictate the fish's appetite. Previous studies have identified temperature as one of the key drivers of increased appetite. Other factors, e.g., photoperiod, oxygen saturation, and health status, would also affect appetite (Noble et al., 2018). Since salmon are visual feeders, there is a highly positive correlation between photoperiod and growth, and numerous studies confirm this (Handeland, Berge, Björnsson, & Stefansson, 1998; Lundqvist, 1980; Saunders, Specker, & Komourdjian, 1989). This is probably due to the increased food intake as the day is long and the temperature increases accordingly (Jobling & Jørgensen, 1992). "The relationship between photoperiod and growth can also result from physiological interactions, entrained by photo stimulated neuro-endocrine activity" (Jobling & Jørgensen, 1992; Lundqvist, 1980). Management practices and exposure to stressors can profoundly affect appetite, e.g., handling and noise. Earlier in commercial salmon farming, it was common to use a growth chart to determine the amounts of feed to be supplied to the fish. Jobling (1995) argues that this rigid adherence to growth charts leads to sub-optimal growth on most days and may prevent the fish from reaching its maximum growth potential (M. Jobling, Arnesen, Baardvik, Christiansen, & Jørgensen, 1995). It is well established that individual and group appetite levels vary between days, even under stable environmental conditions with minimal disturbance. Therefore, feeding according to predetermined growth charts will often lead to overfeeding, which can lead to reduced water quality due to excess uneaten food (Johansen & Jobling, 1998)

Today the most common practice amongst commercial producers is to feed the fish based on behaviour and vertical movement in the cage. This way, it's possible to adjust feeding towards the fish's appetite. Jobling et al. (1998) found that fish fed to satiation ate more and grew faster than those provided with rations predicted to support high growth rates.

2.5 Deriving environmental factors for salmon farming

The environmental conditions under which salmon are grown are essential for their welfare and are reflected in physiological responses (Bowden, Smail, & Ellis, 2002; Johansson, Juell, Oppedal, Stiansen, & Ruohonen, 2007; Turnbull, Bell, Adams, Bron, & Huntingford, 2005). Water quality depends highly on local environmental characteristics, and the different characteristics effects both growth and mortality (Edwards & Edelsten, 1977; Hargrave, Duplisea, Pfeiffer, & Wildish, 1993; Stigebrandt, Aure, Ervik, & Hansen, 2004). Fish that experience a wide range of environmental variations might induce stress responses that incur a physiological cost. Changes in ecological conditions generally lead to a mismatch between physiological states and the environment, causing reduced maximum oxygen uptake rate and increased oxygen consumption. Torgersen et al. (2009) found that temperature variation induces an extra energetic cost measured as increased oxygen consumption for individuals, with 20-25% per day acclimation rates towards the new temperature (Torgersen et al., 2009). Conversely, a negative psychological and physiological impact occurs in salmon exposed to an acute temperature increase from 8—14 C (Folkedal, Torgersen, Nilsson, & Oppedal, 2010). Studies have shown that stressful rearing periods due to environmental stressors such as temperature and oxygen are highly correlated with increased susceptibility to disease and suppressed cytokine expression in fish (Fast, Hosoya, Johnson, & Afonso, 2008; Ndong, Chen, Lin, Vaseeharan, & Chen, 2007; F Oppedal, Dempster, & Stien, 2011; Pérez-Casanova et al., 2008). Furthermore, outbreaks of pancreas disease caused by the salmon alphavirus are stress-related (McLoughlin & Graham, 2007). The following environmental factors are considered essential for salmon farming in the literature.

2.5.1 Dissolved oxygen

Several authors, including Bowden et al. (2002); Kazakov and Khalyapina (1981); Kindschi and Koby Jr (1994), found that dissolved oxygen (DO) in the water is a crucial factor affecting fish health and thus fish metabolism and growth. Significant spatial and temporal variations in dissolved oxygen levels in salmon sea cages exist. "Strong vertical gradients in

DO typically coincide with the pycnocline, while fluctuating patterns occur over days to weeks" (F Oppedal et al., 2011). A study by Viken (2008) recorded severely hypoxic conditions in periods up to 1 hour in the centre of commercial cages, which correlated with low water flow (Vigen, 2008). If DO levels are at hypoxic levels or very low, it can be a tremendous significant threat to fish welfare and growth. Adequate DO levels are a vital requirement (Ellis et al., 2002; Kindschi & Koby Jr, 1994; Van Raaij, Pit, Balm, Steffens, & Van, 1996). For example, Atlantic salmon held in seawater at 16°C given fluctuating hypoxic saturation levels of 70% led to reduced appetite, feed conversion and growth. Fish are somewhat able to acclimatize to fluctuations in DO. However, if fish over substantial periods are introduced to low or fluctuating levels of DO, they would not recover between periods of unfavourable conditions. The fish will be chronically stressed, inducing higher mortality (Vigen, 2008). According to Johansson et al. (2007), few studies have looked at the more complex mechanism that controls oxygen levels in commercial sea cages. Hargrave et al. (1993) found that the photosynthesis capacity in the vicinity of a farm is not ordinarily sufficient to supply the oxygen demand of the total fish biomass in a marine fish farm (Hargrave et al., 1993). Therefore oxygen requirements must be met by physical transport such as tidal movement, current, or freshwater runoff (Johansson et al., 2007). The salmon itself also influences the fish farm's oxygen level and water quality (Furevik, Bjordal, Huse, & Fernö, 1993; Huse & Holm, 1993; Johansson et al., 2007; Juell & Fosseidengen, 2004).

2.5.2 Temperature

Temperature varies significantly across longitudinal degrees. The temperature within sea cages positioned in surface waters varies with depth, and vertical profiles differ considerably between seasons. The temperature is positively correlated with depth in the wintertime while negatively correlated with depth in the summer, with transition periods where the highest temperature is generally mid cage (Frode Oppedal, Juell, & Johansson, 2007). According to Oppedal et al. (2007), F Oppedal et al. (2011), Korsøen, Dempster, Fjelldal, Oppedal, and Kristiansen (2009), salmon were observed to position themselves vertically in correlation with seawater temperature within sea cages. This indicates that salmon prefer the highest available temperature or avoid colder temperatures (F Oppedal et al., 2011). Gamperl, Zrini, and Sandrelli (2021) studied Atlantic salmon behaviour under the New Foodland heatwave. Inhomogeneous water between 10 – 12°C, the fish mainly occupied the top 2-3 meters. With warming surface waters to 14-16°C, the fish frequented these depths despite having access to calmer waters. When maximum surface waters were recorded at 19.2 – 19.4°C, the fish moved slightly more profound in the cage, preferring temperatures between 14-16°C (Gamperl et al., 2021). Johansson et al. (2007) performed multivariate analysis and found that temperature most influenced vertical distribution. The preferred temperature range was 16-18°C within a1-20°C (Johansson et al., 2007; Frode Oppedal et al., 2007). Morisette et al. (2020) found similar results with wild Atlantic salmon. It decreased development time with increasing temperatures with an optimal temperature span of 16-20°, where maximum growth occurs between 16-18°C. They also found that growth rate, metabolism and swimming performance were not affected when exposed to cycling temperatures from 16-21°C—suggesting that within Atlantic salmon's thermal optimum range, temperature variation does not affect logical properties (Morissette, Swart, MacCormack, Currie, & Morash, 2020). Elliott et al. (1991) and Elliott & Elliott (2010) found that Atlantic salmon has an excellent threshold for temperature changes, but very low or very high temperatures can induce stress. Temperatures below 0°C or temperatures at 32°C are lethal within some minutes (Elliot, 1991; J. Elliott & Elliott, 2010).

2.5.3 Salinity

Atlantic salmon is a euryhaline anadrome fish that spends the first part of life in freshwater and migrates to higher salinities during adult life. "Many farming sites located near the coast, in fjords or near rivers are affected by freshwater runoff. Surface waters on these sites become less saline, developing a distinct halocline with a brackish layer of variable thickness and salinity on top and water with typical marine salinity below" (F Oppedal et al., 2011). However, salinity's effect on Atlantic salmon growth is not apparent in the literature. (Sutterlin & Stevens, 1992) suggested that salinity was one factor that regulated the swimming depth of fish in a sea change in stratified along with temperature and social aspects. (Emerman, 2016) found that salinity did not affect the growth rate during a period of 96 days. Thorarensen and Farllen (2011), cited in (Emerman, 2016), found that salinity affected growth rate. However, newly transferred salmon are affected by salinity and show a distinct preference for the halocline layer, independent of temperature for the first two months in the sea. Similar choices have been observed with wild salmon smolts migration from rivers toward the open ocean (Plantalech Manel-La et al., 2009). This strategy might be beneficial because osmoregulation in saltwater is physiologically costly for small salmon (Smith, 1982). Larger fish have much greater osmoregulatory ability than smaller salmon. This is due to their reduced relative water leakage due to their relatively more minor surface area to water ratio. Jeppesen et al. (2007); F Oppedal, Juell, Tarranger, and Hansen (2001) found accordingly that

the behaviour of salmon in sea cages older than three months was not affected by salinity. In contrast, Carvalho et al. (2020) found that fish distribution in sea cages was first determined by salinity, second by temperature and this dissolved oxygen.
Sea lice proliferation is strongly modulated by salinity. Adults do not survive during low salinities <12, and salinities below 30‰ partly prevent the development of nauplii onto the copepodite stage. Evidence exists that a salinity below 29‰ severely reduces survival of free-swimming stages (Bricknell, Dalesman, O'Shea, Pert, & Luntz, 2006), with "50% survival after 24 hours 29‰, 11hours at 26‰, 8 hours at 23‰, 6hours 19‰, 4hours 16‰ and >1 hour at 12‰, 9‰ and 5‰ (Morro et al., 2021). When given a choice, most copepods sit in salinities of 34‰ and actively avoid salinities below 27‰ (Bricknell et al. 2006). Since salmon lice is a major problem in Norway, nearshore farms close to rivers could benefit from the deleterious effects of low salinity on sea lice (Morro et al., 2021).

2.5.4 Currents

Atlantic salmon is an athletic, long-distance swimmer who can work with high intensity over an extended period. The high aerobic capacities make it perfect for long migrations. Wild salmon can pace themselves during migrations, take advantage of currents, and choose a vertical level that opposes less resistance to their advance. In captivity, these choices are limited. Fish must swim at speeds dictated by the farm environment. Strong currents can force the fish to swim against the current, exerting substantial effort. If the fish swim above their threshold, they will produce lactic acid and get stressed over time. Therefore strong currents can become a legitimate welfare concern in fish farms (Morro et al., 2021). High energy expenditure can decrease production and fish growth because a significant amount of energy is diverted to exercise. A necessary contrary to this, a moderate current velocity (0.36 – 0.63 body lengths per second, BL pers) has proven significantly beneficial to fish health and growth during their entire growth stage in the open sea (Morro et al., 2021).
Remen et al. (2016) found that higher current velocity (up to 2.5 BL per s) increases growth, muscle fibre size and several metabolic factors, but at the expense of fish welfare, where there was a higher incidence of inflammation and skin pelvic lesions. (Hvas, Folkedal, & Oppedal, 2021), using a swim tunnel, found that salmon post-smolts with lengths of 43cm and 850 grams were capable of swimming speeds of 97.2cm individually. When tested for endurance, only a fraction of the fish could sustain this speed for the whole period. When studying group behaviour, Remen et al. (2016) found that all fish coped with 78 cm/s for the entire four hours tested in a similar setup. Post smolt of around 29.2cm in this study achieved a critical

swimming speed (Ucrit) of 65.5cm/s. When tested in groups, there was a significant increase in performance due to a reduction in overall drag (Morro et al., 2021).
Groups of fish with lengths respectively 19.6cm, 29cm and 80.6cm achieved Ucrits of 80.6, 90.9 and 99,5 cm/s. Hvas M, Folkedal O, Solstorm D (2001) also showed that larger post-smolts of around 63.5cm (3.4kg) could withstand even higher current velocity but would become fatigued above 125 cm/s. Post smolts in Norway are generally transferred to open water sea cages weighing 250-300 grams. At this size, the maximum current velocity should not exceed 80.6 cm/s. It is suggested that the sustained swimming capacity over time is about 80% of this, indicating a current rate of about 65 cm/s that will increase with growth (Hvas et al., 2021; McKenzie et al., 2021; Morro et al., 2021; Remen et al., 2016) This Ucrit is uncertain because the test was done in swimming tunnels. However, in commercial cages, even lower Ucrit may affect growth and fish welfare. Performance and endurance are reduced for fully fed fish or fish held in high densities. If current reaches Ucrit in commercial sea cages and differential current speeds exist vertically, it is expected that salmon will modify their behaviour (F Oppedal et al., 2011). Therefore, when looking at current fish welfare and growth rates, it is essential to investigate several vertically distributed data points in depth.

2.5.5 Waves

Most of the fish farms in Norway are found in sheltered areas that are not primarily affected by waves. However, some fish farms are in exposed areas where waves exceed several meters. In such circumstances, waves can submerge the open net pens and cause salmon to collide with each other, and there is a risk that salmon get stuck in the net; this can cause injury and stress, leading to declining fish welfare. Over time stress can cause diseases and, in the worst cases, death (Kapetsky et al., 2013). During storm periods, it is expected that salmon may seek deeper in the sea cage to avoid waves and strong currents caused by wind transport because the waves decrease in the water column.
Contrary to this, Stockwell et al. (2021) found that whilst performing an acoustic telemetry study on fish movement during a storm. There was no change in behaviour during a storm event (Stockwell, Filgueira, & Grant, 2021). Atlantic salmon are phystostomes, which means they need to fill their air bladder to maintain buoyancy. In turbulent periods with waves, strong currents and fish that struggle to maintain buoyancy need to use more energy to maintain a vertical position. This thesis used temperature, current, and salinity on three different vertical levels, wind speed, and wave height, as exploratory variables. Efforts were

made to collect data for dissolved oxygen with minor success. Therefore, oxygen data were not included further in modelling.

3 Methodology

3.1 Study area

The study area covers the two northernmost counties in Norway: Nordland, Troms and Finnmark (Figure 4). The study area includes a total of 527 locations that are adopted for aquaculture purposes. They are located in fjords, bays, or are scattered amongst islands within archipelagos and some further out in open water across the entire coastline (Falconer et al., 2020). Farms located in the coastal areas typically have relatively homogeneous water and are likely to experience wind-driven upwelling events of colder water with lower oxygen saturation (F Oppedal et al., 2011). Farms located in fjords are less likely to experience upwelling events. Still, they experience more significant seasonal variation in environmental conditions with substantial stratification variations in salinity, temperature, oxygen and water currents (Johansson et al., 2007).

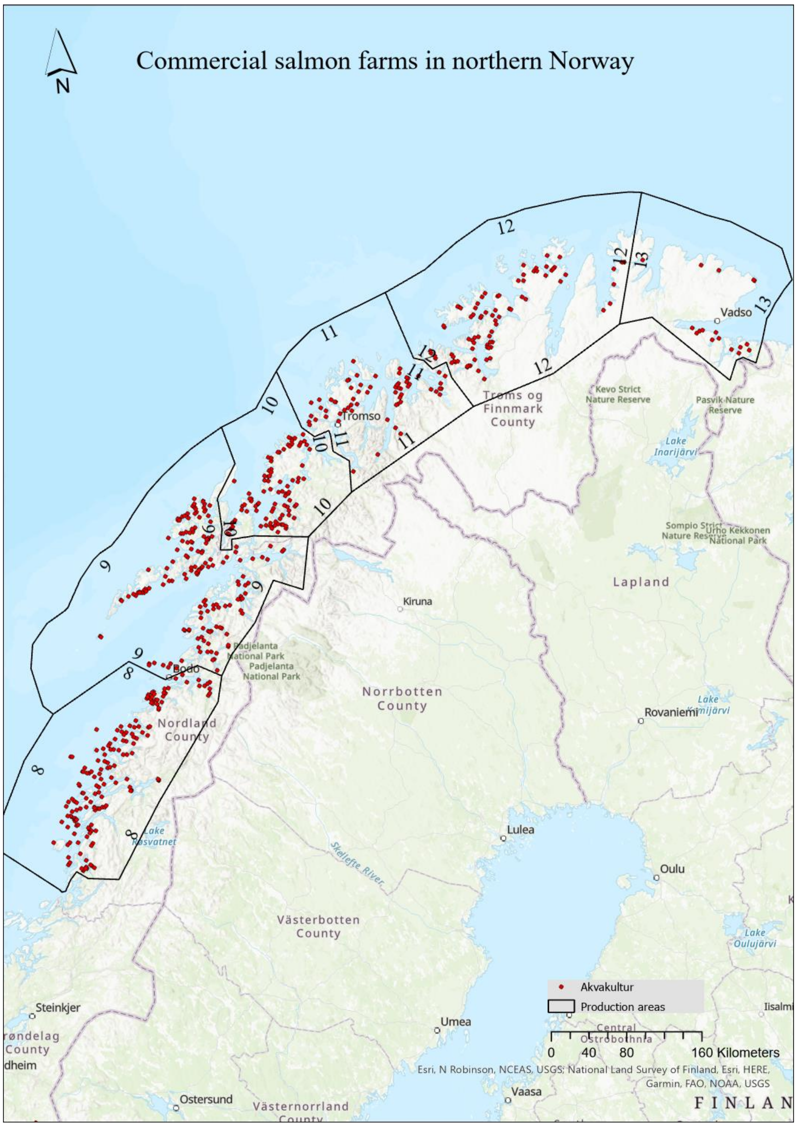


Figure 4: Study area map. The number indicates production areas: 8 – Helgeland to Loppa, 9 – Vestfjorden and Vesterålen, 10 – Andøya to Senja, 11 – Kvaløya to Loppa, 12 – Vest Finnmark, 13 – Øst Finnmark. Red dots indicate the 527 registered salmon farming locations

Only including the six northernmost production areas in this study is threefold. First, in these areas, the industry's most significant opportunities for expansion lie according to the traffic light system (Hersoug, Mikkelsen, Robertsen, & Osmundsen, 2020). Second, the environmental conditions, growth and mortality rate vary greatly at longitudinal degrees. This can highly affect the model we are trying to train, and therefore a minor homogeneous study area would be favourable regarding reliable results. Third, this reduces the amount of data processing to a manageable computational level.

3.2 Data and materials

3.2.1 Empirical production data salmon farming

Aquaculture site production data was obtained by the Directory of Fisheries (DOF). It was structured in three different excel sheets (2018, 2019, 2020) with monthly biomass per locality between 1. January 2018 – 31. December 2020. An overview of the data is attached in the appendix (Appendix A). The data hold information about the fish farm location, given biomass each month, and several other categories of fish losses, including mortality. The statistics provide important indicators for evaluating the government and industry goal of achievement and long-term utilization of living marine resources and contribute to the profitable and self-sustaining aquaculture industry. Because of the urge to increase the aquaculture industry's sustainability and lower mortality, the production data has grown increasingly important in the later years. The data is confidential at the locality levels but is shared with other authorities and research environments to help improve the industry's sustainable development. Confidentiality is maintained when processing this data and presenting it. Only aggregated results are presented as results and figures. There is not possible to track individual companies' performance and results in this thesis.

Only commercial farms producing Atlantic salmon were included in this study. Hence, research farms or other salmonids-producing farms were excluded from the analysis. After removing all locations that were not commercial salmon farms, only 364 sites were included in further data cleaning and processing. The raw data contained 24 799 unique rows with ID and dates for months when fish was present. However, there were records with no data and incomplete time series with empty rows for each locality. This made the monthly calculations difficult. Therefore, a script was made to complete the time series with null values for the no data records. The code snippet is attached in the appendix (Appendix B). After cleaning the data, conducting the calculations, and removing the null values. The remaining data included 323 unique aquaculture locations with 6048 rows.

3.3 Dependent variables

This thesis is based on the fundamental assumption that the suitability of a farming site is based on the specific growth- and mortality rate. Second that all salmon farms operate equally with the same technology. Growth rate and mortality rate were used as the dependent variables following the animal-based welfare indicators described by Noble et al. (2018). The following subsection describes this process in detail.

3.3.1 Mortality

The production data collected by the Norwegian directory of fisheries is monthly based and consists of four categories; (1) dead, (2) discarded, (3) escapes and (4) other. Only losses attributed to the "dead" category were used to calculate the mortality rate. Following the same assumption as Jensen et al. (2020), "for the number of fish in a farm, the number reported for a month constitutes the current number of fish, that is fish present at the end of the month. For the number of dead fish, the number of fish reported constitutes the number of fish that died during the month" Mortality rate (Mrate) for a month (*i*) is calculated as:

$$(1) \qquad Mrate = \frac{deaths}{(Fstart + Fend)/2}$$

Fstart is the number of fish at risk of death at the start of the month, Fend is the number of fish at the end of the month. The denominator calculates the average number of fish alive during the month by assuming that fish die or are added/removed uniformly (Jensen, Qviller, & Toft, 2020).

3.3.2 Growth

The individual growth rate is one of the most commonly calculated vital rates in aquaculture and fisheries management (Quist & Lsermann, 2017). Growth can be expressed in many ways, e.g., relative growth, instantaneous growth, and size-specific growth. The common factor with all of these is that they require knowledge of the size of the fish at two or more points in time, either from direct or indirect measurements (Crane, Ogle, & Shoup, 2020). In aquaculture farming in Norway, once every week, farmers randomly select a minimum of 20 fish from each cage and measure size and weight. This is used to calculate the total biomass in each pen following Crane et al. (2020). Crane et al. (2020) argued that the weight of fish increased exponentially over short periods and that it's possible to create an exponential function that models weight (w_2) at some future time (t_2) from weight (w_1) at time (t_1) with:

$$(2)\ w2 = w1e^{-g\Delta t}$$

In equation 2, *g* is the instantaneous growth rate and $\Delta t = t_2 - t_1$ is the elapsed time between t_1 and t_2 (Crane et al., 2020). *g* can further be obtained by algebraic rearrangement (Equation 3), which is considered a well-known equation in the fisheries literature (Crane et al., 2020; Ricker, 1975).

$$(3)\ g = \frac{log_e(w2) - log_e(w1)}{\Delta t}$$

Instantaneous growth rates are difficult to interpret because *g* represents the additive change in log weight per unit time"(J. M. Elliott & Hurley, 1995). Therefore, a more interpretable metric of growth can be obtained by rearranging equation 2 to:

$$(4)\ \left(\frac{w2}{w1}\right)^{\frac{1}{\Delta t}} = e^g$$

In equation 4, "e^g is the multiplicative change in weight per unit time. In Atlantic salmon farming usually $w_2 > w_1$ such that $e^g > 1$ and $e^g - 1$ gives the proportional increase in weight per unit time" (Crane et al., 2020). Multiplying this by 100 gives:

$$(5)\ G = 100(e^g - 1)$$

Equation 5 is the per cent increase in weight per unit of time, and thus G is called the specific growth rate (SGR) (Houde & Schekter, 1980). According to the recommendations given by Crane et al. (2020), the specific growth rate was used (equation 5) as the dependent variable.

3.4 Explanatory variables

The study area was of such a large extent, both in space and time. Different methods were needed to retrieve and extract exploratory variables for the various research questions. High-resolution data were required to identify the significant variables favour salmon growth and depict mortality. Therefore, exploratory variables were extracted (very high spatial resolution) at the farm location. Predicting optimal areas for salmon farming required exploratory data for the whole study area. Raster surface layers with a resolution of 160x160m for each exploratory variable were extracted from the environmental models described in the following chapter.

Seventy different exploratory variables related to ocean environmental conditions and other indirect variables affecting growth and mortality were extracted from different databases. All calculations and transformations were conducted with Excel, MATLAB, Power-Query and Python. Spatial analysis and modelling were performed in ArcGIS Pro. The projected coordinate system ETRS 1989 33N was applied. The following subsection provides a detailed description of the data used (i.e., salinity, temperature, and current data from NorKyst800, IMR. Wind speed from AROME MetCoOp (Meteorological Co-operation on Operational Numerical Weather Prediction) and Wave height from Simulating waves nearshore (SWAN).

3.4.1 Environmental data

Ocean current speed and hydrographical data were derived from the main hydrodynamical model system for the Norwegian coastal zone provided by the Institute of Marine Research. Five fjord models with a horizontal resolution of 160m x 160m covering northern Norway were run parallel with input along the open boundaries from the Norwegian coastal model NorKyst800 (Asplin, Albretsen, Johnsen, & Sandvik, 2020). Data was extracted at vertical layers: 0m, 10m, 20m, and 30m. A similar simulation of the 160m-model system is explained in Dalsøren, Albretsen, and Asplin (2020), where the open-source Regional Ocean Modeling System (ROMS) is used (Haidvogel et al., 2008) (see also http://myroms.org). ROMS is a state-of-the-art, three-dimensional, free-surface, hydrostatic, primitive equation ocean model that uses generalized terrain-following s-coordinates in the vertical. Significant wave height is retrieved from simulations conducted by the state-of-the-art, open-source wave model SWAN (http://swanmodel.sourceforge.net), developed at Delft University of Technology, Netherlands. Six model grids cover northern Norway, and all grids applied a 200m x 200m horizontal resolution. The application of similar wave height data is demonstrated in Van Son et al. (2020). Wind speed is arranged as seasonal averages based on data from 2018 to 2020 from the operational weather forecasting model at the Norwegian Meteorological Institute, AROME MetCoOp 2.5 km (Müller et al., 2017).

Since the empirical data is reported monthly, all the environmental data was extracted or transformed into monthly averages and matched with ID for location and date. The ocean current and hydrographical data were converted to monthly 10th percentile, mean and 90th percentile for the vertical layers. The wave height was derived from a representative period and included the mean and 90th percentile for each location at a given month. Wind speed is based on east-west and north-south vectors every third hour for each site and was transformed to monthly mean values. The location-specific current speed is based on daily averages.

Table 2: Overview of environmental, location-specific data. Ocean current speed and hydrographical data were derived from the main hydrodynamical model system for the Norwegian coastal zone run at the Institute of Marine Research. Significant wave height is retrieved from simulations conducted by the state-of-the-art, open-source wave model SWAN. Wind strength is arranged as seasonal averages based on data from 2018 to 2020 from the operational weather forecasting model at the Norwegian Meteorological Institute, AROME MetCoOp (Meteorological Co-operation on Operational Numerical Weather Prediction).

Variable type	Distribution	Vertical layer	Code	Source	Coordinate system	Format
Salinity	10th percentile	1	S1m_10p	Norkyst-800 (IMR)	WGS84	ASC
		10	S10m_10p			
		20	S20m_10p			
		30	S30m_10p			
	Mean values	1	S1m_avg	Norkyst-800 (IMR)	WGS84	ASC
		10	S10m_avg			
		20	S20m_avg			
		30	S30m_avg			
	90th percentile	1	S1m_90p	Norkyst-800 (IMR)	WGS84	ASC
		10	S10m_90p			
		20	S20m_90p			
		30	S30m_90p			
Temperature	10th percentile	1	T1m_10p	Norkyst-800 (IMR)	WGS84	ASC
		10	T10m_10p			

		20	T20m_10p			
		30	T30m_10p			
	Mean values	1	T1m_avg	Norkyst-800 (IMR)	WGS84	ASC
		10	T10m_avg			
		20	T20m_avg			
		30	T30m_avg			
	90th percentile	1	T1m_90p	Norkyst-800 (IMR)	WGS84	ASC
		10	T10m_90p			
		20	T20m_90p			
		30	T30m_90p			
Current	10th percentile	1	U1m_10p	Norkyst-800 (IMR)	WGS84	ASC
		10	U10m_10p			
		20	U20m_10p			
		30	U30m_10p			
	Mean values	1	U1m_avg	Norkyst-800 (IMR)	WGS84	ASC
		10	U10m_avg			
		20	U20m_avg			
		30	U30m_avg			
	90th percentile	1	U1m_90p	Norkyst-800 (IMR)	WGS84	ASC
		10	U10m_90p			
		20	U20m_90p			
		30	U30m_90p			
Wind speed	Mean values	Surface	Windspeed_monthly_avg	AROME MetCoOp	WGS84	ASC

Wave hight	90^{th} percentile	Surface	WH_90perc	SWAN	WGS84	ASC
	Median	Surface	WH_median	SWAN		

Table 3: Overview of area-wide environmental variables. Ocean current speed and hydrographical data were derived from the main hydrodynamical model system for the Norwegian coastal zone run at the Institute of Marine Research. Significant wave height is retrieved from simulations conducted by the state-of-the-art, open-source wave model SWAN. Wind strength is arranged as seasonal averages based on data from 2018-2020 from the operational weather forecasting model at the Norwegian Meteorological Institute, AROME MetCoOp (Meteorological Co-operation on Operational Numerical Weather Prediction).

Variables	Quarter	Area name	Spatial resolution	Coordinate system	Format	Source
Ocean- and hydrographical data (average)	1	A09	160x160m	WGS84	NetCDF	Norkyst-800
	2	A09	160x160m	WGS84	NetCDF	Norkyst-800
	3	A09	160x160m	WGS84	NetCDF	Norkyst-800
	4	A09	160x160m	WGS84	NetCDF	Norkyst-800
Ocean- and hydrographical data (average)	1	A10	160x160m	WGS84	NetCDF	Norkyst-800
	2	A10	160x160m	WGS84	NetCDF	Norkyst-800
	3	A10	160x160m	WGS84	NetCDF	Norkyst-800
	4	A10	160x160m	WGS84	NetCDF	Norkyst-800
Ocean- and hydrographical data (average)	1	A11	160x160m	WGS84	NetCDF	Norkyst-800
	2	A11	160x160m	WGS84	NetCDF	Norkyst-800
	3	A11	160x160m	WGS84	NetCDF	Norkyst-800
	4	A11	160x160m	WGS84	NetCDF	Norkyst-800
Ocean- and hydrographical data (average)	1	A12	160x160m	WGS84	NetCDF	Norkyst-800
	2	A12	160x160m	WGS84	NetCDF	Norkyst-800
	3	A12	160x160m	WGS84	NetCDF	Norkyst-800
	4	A12	160x160m	WGS84	NetCDF	Norkyst-800

Ocean- and hydrographical data (average)	1	A13	160x160m	WGS84	NetCDF	Norkyst-800
	2	A13	160x160m	WGS84	NetCDF	Norkyst-800
	3	A13	160x160m	WGS84	NetCDF	Norkyst-800
	4	A13	160x160m	WGS84	NetCDF	Norkyst-800
Wind Speed (average)	Na	A0601	200x200m	WGS84	NetCDF	AROME
	Na	A0602	200x200m	WGS84	NetCDF	AROME
	Na	A0701	200x200m	WGS84	NetCDF	AROME
	Na	A0702	200x200m	WGS84	NetCDF	AROME
	Na	A0801	200x200m	WGS84	NetCDF	AROME
	Na	A0802	200x200m	WGS84	NetCDF	AROME
Wave height (median)	Na	A0601	200x200m	WGS84	NetCDF	SWAN
	Na	A0602	200x200m	WGS84	NetCDF	SWAN
	Na	A0701	200x200m	WGS84	NetCDF	SWAN
	Na	A0702	200x200m	WGS84	NetCDF	SWAN
	Na	A0801	200x200m	WGS84	NetCDF	SWAN
	Na	A0802	200x200m	WGS84	NetCDF	SWAN

3.4.2 Socio-economic factors

Data-driven analysis of socio-economic factors included the specific feeding rate as the proxy indicator of economic factors (cost–feed) derived from the production data. Other socio-economic factors were derived based on distance from the aquaculture locations, distance to fairways (as a proxy measure of the effect of sea traffic) and distance from the urban settlements (as a proxy indicator of access to labour, infrastructure, market, and social environment).

3.4.2.1 Distance dependence variables

Distance calculations are fundamental in GIS. In its simplest form, distance is measured as a line between one object to another object. Three different distance dependence variables were included. 1. Distance to fairways as a measure for noise effect and transmission of pathogens.

2. Distance to urban settlements as a proxy for accessibility to infrastructure and labour. 3. Distance to nearest neighbouring aquaculture locations as a proxy of the mortality due to transmission of pathogens.

Fairways are included to see if traffic and noise can cause salmon to stress and lower overall welfare. There is also a theoretical chance that aquaculture locations close to high traffic areas (fairways) are more exposed to the transmission of pathogens from ballast water. Fairways were derived from GeoNorge and provided by the Norwegian Coastal Authority.

Urban settlements are defined by a collection of houses with at least 200 inhabitants where the distance between houses cannot exceed 50 meters. Statistics Norway provided the dataset. This was included as a socio-economic variable. Farmers often want to have localities close to urban settlements and infrastructure such that the daily husbandry is not overly costly. Moreover, it is hard to find labour in areas far away from urban settlements.

The last distance feature included was the internal distance between the locations themselves. The Food Safety Authority has prohibited localities from being established closer than 2,5 km because of the chance of transmission of pathogens. The hypothesis is that localities closer to each other's or cluster of localities have higher mortality than those more dispersed.

3.4.2.2 Specific-feed-ratio

Fish are among the most effective animals to produce as the feed conversion ratio is usually near 1. This indicator says something about the fish´s appetite and indicates how efficient the feed or the feeding strategy is during a life cycle. This thesis calculated the monthly specific feeding rate as a proxy for monthly economic investment into the locality. The specific feeding rate (SFR) is calculated as:

$$(6) \quad \frac{Feed}{Average\ Biomass} * 100$$

Feed is the monthly mean kg feed, and biomass is the average monthly biomass for the same month (Nilsen, Nielsen, & Bergheim, 2020).

3.4.3 Bathymetry and slope

The study area's bathymetry (160 m resolution raster) was created by interpolating depth points, shallows, coastal lines, and contours. These data were derived from GeoNorge and provided by the Norwegian Mapping and Cadastre Authority. Bathymetry was projected to ETRS UTM 33N. The slope was derived from bathymetry data using the planar method, measuring slope as the maximum rate of change in value from a cell to its immediate neighbours. The slope is computed as the rate of change of the surface horizontal (d/dx) and vertical (dz/dy) directions from the centre to each adjacent cell. The slope is commonly measured in units of degrees using equation 6. The slope was projected to ETRS UTM 33N with a cell size of 160x160m.

(6) $slope\ degrees\ =\ ATAN\ (\sqrt{}([dz/dx]2\ +\ [dz/dy]2)\ )\ *\ 57.29578$
(Esri, 2021) ¨

3.4.4 Solar radiation and duration of daylights

Solar radiation and day length were obtained from calculations in excel and ArcGIS Pro. Incoming solar radiation originates from the sun and is modified as it travels through the atmosphere. It is further affected by topography and the earth's surface. Three different radiation measurements were included; direct radiation (direct line from the sun), diffuse radiation (scattered by atmospheric constituents) and duration (the sum of the two prior). The solar radiation calculations are based on Rich et al. (Fu & Rich, 1999, 2002; Rich, Dubayah, Hetrick, & Saving, 1994). Since radiation is greatly affected by surface and topography, a DTM (Digital terrain model) over mainland Norway with heights in a grid of 10x10m and projection EUREF89 UTM zone 35 was downloaded from the Norwegian mapping authority and projected to ETRS 1989 33N. Solar radiation was first calculated for the specific coordinates for all the salmon farms included in this study. Calculations are based on the average latitude for all locations with monthly intervals for the year 2020 based on Julian days. The solar radiation calculations are computationally demanding and very time-consuming. In principle, a larger sky size would increase the accuracy of the calculations but dramatically increase the calculation time. Therefore, the sky size was set to 200, sufficient for the whole DEM with large day intervals.

Daylength (sunrise and sunset) for every dependent variable was calculated based on the corresponding date. The dependent variables are monthly averages. Day length is thus calculated based on a month's median day and not average daylight in the month. Calculations are based on the NOAA method for solar calculations based on Meeus (1991)`s equations from astronomical algorithms. Calculations are attached in the appendix (Appendix C).

3.5 Modelling and identifying driving factors

3.5.1 Linear regression techniques

Two different models were used to identify driving factors for salmon growth and mortality, with specific growth rate (SGR) and mortality rate (Mrate) as dependent variables. Exploratory data analysis revealed that both dependent variables had non-normal distribution. Different transformations were tested to obtain a normal distribution in the residuals as required for regression models. As a result, SGR was transformed using square-root transformation to reduce the influence of some extreme values. Similarly, Mrate was log-transformed. QQ plot of growth and mortality with and without transformations is attached in the appendix (Appendix- D and E). In line with the principle of Occam’s razor, OLS was the first modelling approach tested. OLS Equation:

$$(8) \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n + \in$$

Y is the dependent variable (monthly growth- and mortality rate). X is the explanatory variable. It is the variable we believe in causing or explaining Y. β is the coefficients and represents the strength and type of relationship that X has to Y. Coefficients can have a positive, negative or no relationship – the value for X is not correlated with Y. $\in$ is the residuals, these are the model over- and under predictions. $\in$ is the difference between the observed value and the predicted value. All parametric explanatory variables were input in the analysis, and several iterations were performed. To be included in the next iteration, the variable had to have a robust p-value of 0.05 and a VIF-value of 7.5. The final model was assessed for model bias using the Jarque-Bera statistic. Residuals were also checked for statistically significant high and/or low clustering.

3.5.2 Random forest regression

Two independent data processes were completed, first with SGR as the dependent variable and then with Mrate as the dependent variable. All parametric explanatory variables were input in the first RF training model. The recursive feature elimination method was used (Figure 5) to eliminate less important variables in an iterative process with the overall goal of a minimal set of important variables. The model was built with 100 trees but increased to 500 to obtain a more stable model. No data was left out for validation when training the model. After achieving a stable model with important variables, the performance evaluation was measured, excluding 10% of the training data.
The same methodology has been applied in ecological studies, for example, Cutler et al. (2007). Furthermore, Dietrich et al. (2016); Habermann et al. (2009) applied this method to analyze high-dimensional molecular data sets generated, e.g., transcriptomics and metabolomics experiments. Showing very high accuracy and the ability to model complex interactions between variables.

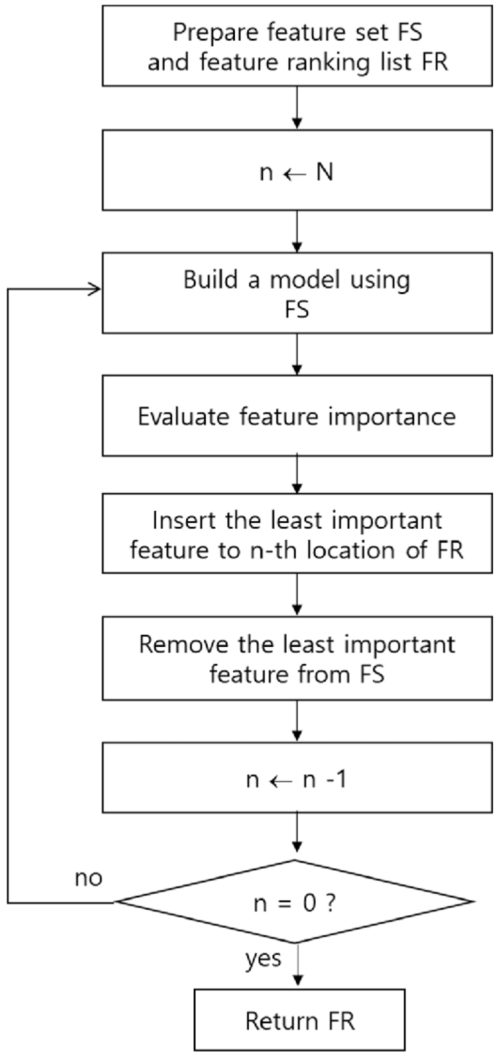


Figure 5: Feature importance based recursive feature elimination. Figure adapted from (Jeon & Oh, 2020)

3.6 Predicting with Random Forest

The requirement to predict a surface raster using RF classification is a point dataset with two binary classes: optimal/best site (having high growth and low mortality is considered as presence) and other sites (having low growth and high mortality is considered as absence). RF would perform classification instead of regression, considering votes from all decision trees to predict or classify the prediction raster for any unknown observation (Breiman, 2001; Valavi et al., 2021). Since classification requires presence and absence data, the dependent variables used in regression could not be applied in prediction. These values must be aggregated into average values for each aquaculture site. The following subsection explains this process in detail.

3.6.1 Presence (optimal) - and absence (others) data

The empirical data on hand is in monthly intervals. Since we are interested in predicting a surface raster, SGR and Mrate need to be aggregated and combined for all aquaculture sites. Prediction with RF does not work with multipoint data. 6048 rows were aggregated into average values for both SGR and Mrate. The results were 309 individual rows with frequency fields representing the count of rows aggregated. Aquaculture sites with less than five counts were removed from the analysis (15 aquaculture locations). Bivariate colours were used to visualize and identify aquaculture sites with high growth rates and low mortality rates based on natural breaks. All presence points obtained a value of 1, while the records that did not meet the minimum criteria (SGR > 34,61 and Mrate < 0,64 obtained 0).

Several authors have demonstrated poor predictive performance attributed to presence-background data (Bateman, VanDerWal, & Johnson, 2012; Shabani, Kumar, & Ahmadi, 2016). This is not a general SDM problem since Maxent and Boosted regression trees have performed strongly on these data types (Elith et al., 2006). This specific problem seems to be related to RF. This can result in biased classification accuracy due to the bootstrap over-representing the majority class, thus leading to underprediction of the minority class (Evans et al., 2011). According to Evans et al. (2011), this is less a problem with regression trees. The deal with imbalanced datasets in RF classification, Evan et al. (2011) developed a novel approach. In ArcGIS Pro, it is possible to compensate for sparse categories. This ensures that each category is represented in each tree to create a more balanced model. All predictions in this thesis were compensated for sparse categories. Presence points were tested for spatial autocorrelation with the Global Morans I index. I used inverse distance as a conceptualization for the spatial relationship, and the distance method was Euclidean distance with ETRS UTM 33N as the projected coordinate system. Test for spatial autocorrelation is attached in the appendix (Appendix F).

3.6.2 Processing of exploratory data

A RF classifier extracts the value from a raster surface (exploratory variable) and associates it with feature input data (presence-absence). If the feature input data is missing one of the exploratory raster data, it will be left out of the model. The environmental data were extracted as netCDF files from the corresponding models (NorKyst800, AROME, SWAN). The longitude and latitude arrays were used to locate every single grid point. The oceanographic variables were preprocessed in MATLAB before integrating them into ArcGIS Pro. Quarterly averages and only two vertical layers (0 m and 30m depth) were extracted to reduce the amount of computational power and geoprocessing time.

The standard cell size for all variables in the ocean and the hydrographical group was 160x160m. All further geoprocessing used this cell size to the same extent. A workflow was created in model builder to import, process, and batch project all netCDF files. A multidimensional raster layer was created three times for every netCDF file containing ocean- and hydrographical data (salinity, temperature, current speed). This was done by slicing data along with the defined variables and dimensions in the netCDF file. Wind speed and wave height only had one defined variable and were directly imported for all dimensions. The extent and cell size masked all raster layers according to Norkyst-800 resolution. Wind speed and wave height had a spatial resolution of 200x200m and were resampled using the nearest neighbour method.

3.6.3 Quarterly models

Four different models were built, first quarter (Q1), second quarter (Q2), third quarter (Q3) and fourth quarter (Q4). To maintain a similar procedure between the regression and classification models, the recursive feature elimination, as described in chapter 3.5.2, was also used as a method in classification. Every available exploratory variable was included in the first run. Note that all rasters are continuous, and RF classification will find a breakpoint and decide if a given value is present or absent. As described earlier, we compensate for sparse values because of imbalance. The number of trees was increased iteratively from 100 to 1500. The validation was set to 20% excluded data for two runs each build. The most relevant variables included in the final model were chosen based on variable importance. The final model was built with the more minor, most significant variables assessed with the importance table and out of bags error. Since all data was used in prediction, we need to determine the goodness of fit with OOB. Two final maps were calculated using map algebra and cell statistics using all layers (Q1, Q2, Q3, Q4). The results were one map with the sum of all rasters and another with the product of all four suitability layers (quarters).

4 Results

This section first describes salmon farms' spatial and temporal patterns with a focus on growth and mortality and significant environmental and socio-economic variables. The latter section presents the results from the suitability modelling of optimal aquaculture sites in Northern Norway.

4.1 The existing pattern of salmon farming in northern Norway

Figure 6 presents the existing pattern of coastal use and salmon farming in northern Norway. Looking at the map, one may conclude that there is not so much space available for the aquaculture industry along the coast – within one nautical mile from the baseline. The coastal areas are primarily used for inshore fishing, and there are several important spawning grounds. All farms are located within 39km of major urban settlements (minimum: 60.7m, maximum: 39km). The average distance to fairways from the salmon farms is 3 km (minimum: 104m, and maximum 19,1 km). In general, all the farms are located within 391 m from the shoreline (minimum: 101m, and maximum 1,1 km). The average specific feeding rate in northern Norway was 21.5, ranging from 2,33 to 239,6. It indicates a very high variation in specific feeding ratios among different farms, indicating a high variation of investment in the various locations.

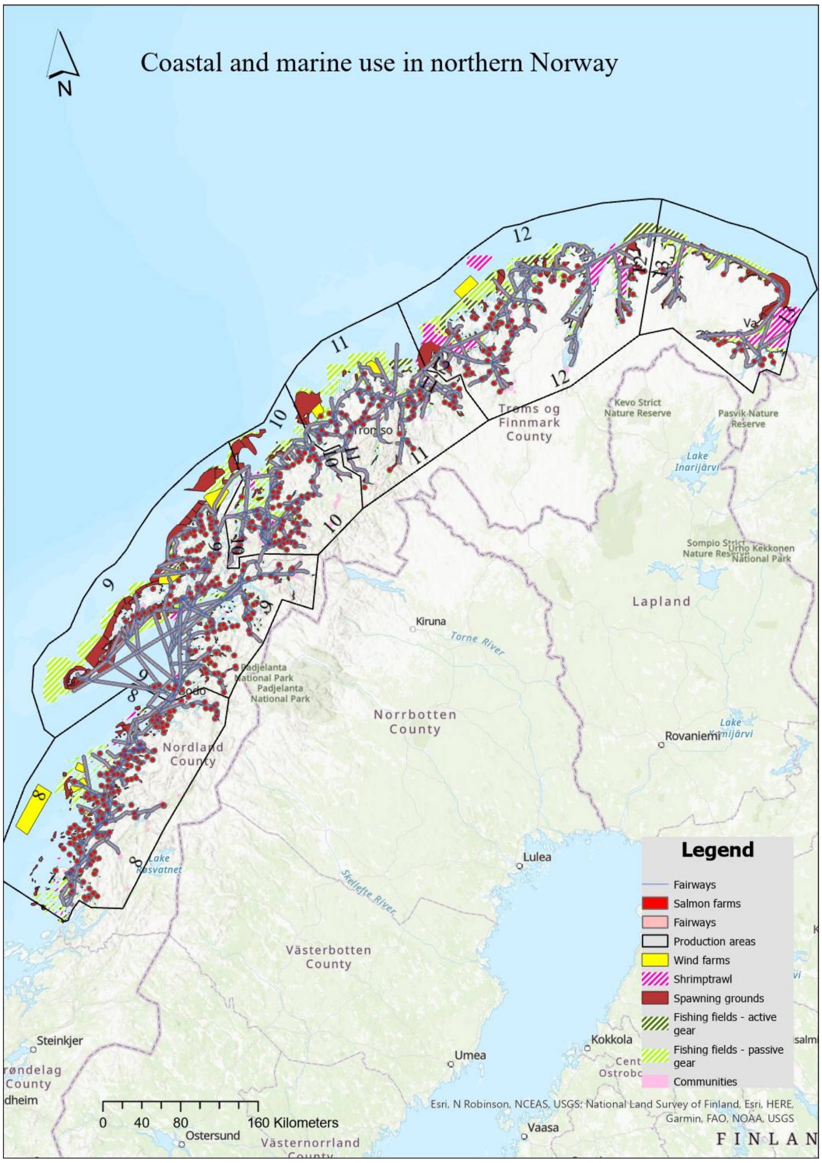


Figure 6: Existing pattern of coastal use in northern Norway.

4.1.1 Spatio-temporal variation environmental variables

The environmental variables varied greatly with longitudinal degrees and distance from land (Figure 7). Salinity increases with depth and distance from shore. The water inside fjords was less saline than the water at the coast or in open water. The average salinity was 32-33‰, with a minimum of 18.1‰ and a maximum of 34.5‰. Especially in the summer months, the water was significantly less saline inside fjords and along the coast.

The temperature varied considerably between the seasons and at different locations. Aquaculture locations had an average temperature of 6.9° C – 7.3° C. The highest and lowest temperatures were found at the surface (-0,1° C and 14,4). Water temperatures were lowest in the north (Finnmark area), especially in Q1 and Q2 water held a temperature around 0° C - 1° C. Cold temperatures were also found inside fjords, especially in the summer months due to freshwater runoff from snow and ice.

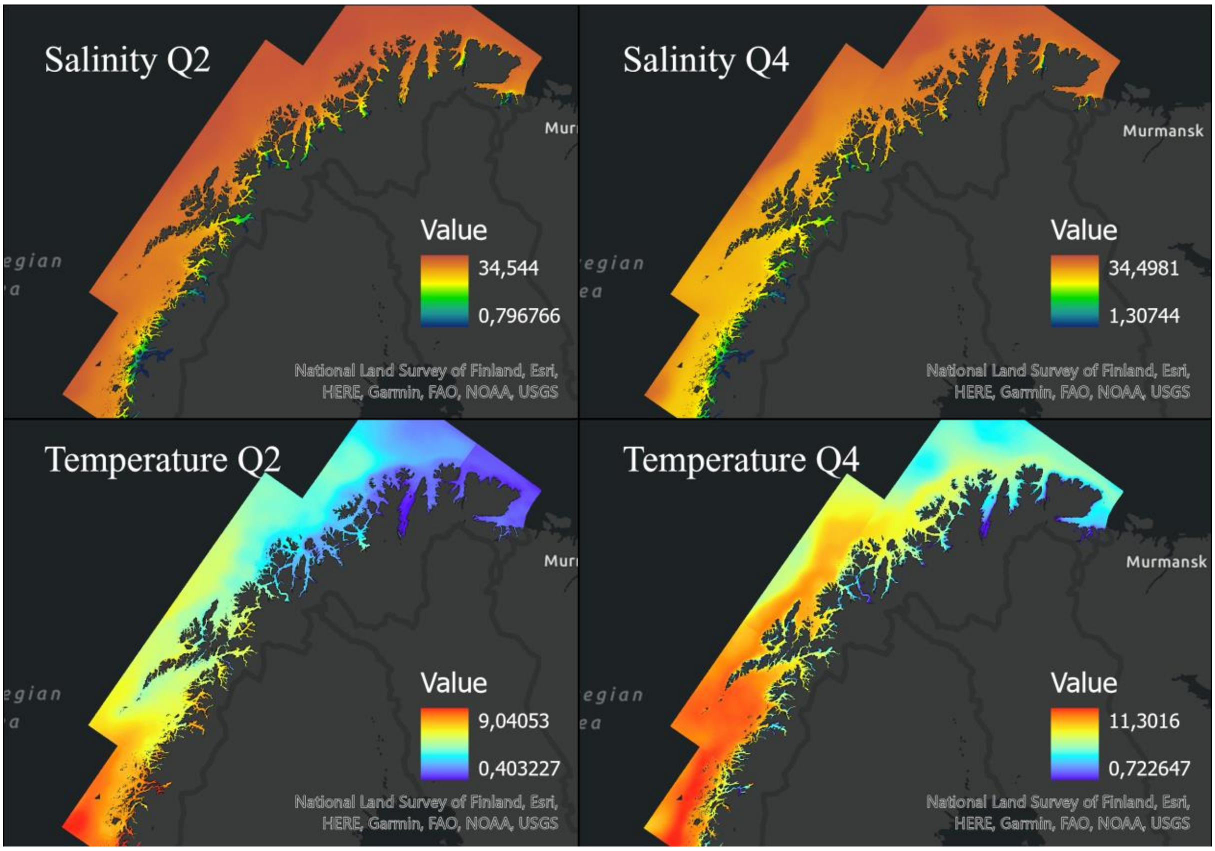


Figure 7: Spatio-temporal variation of salinity and temperature surface averages for Q2 and Q4.

The current was seemingly calm at the aquaculture locations with a minimum speed of 0m/s and a maximum speed ranging from 0,36 – 0,45m/s, strongest at the surface (Figure 8). Because of many null values, the average current speed was low, ranging from 0,038 – 0,079m/s. During the study period, the most substantial current is off the coast outside

Lofoten and Finnmark (Figure 8). However, the areas inside the baseline and nearshore had strong currents due to tidal water, reefs, islands, fjords, and shallow water, contributing to a significant variation in current. In addition, freshwater runoffs contribute to vertical circulation due to variation in temperature and salinity.

The fjord and sites near the coast are sheltered from the more significant wind speeds. At the aquaculture locations, the wind speed was 2,1 m/s, with a highest of 6,3 m/s and a lowest of 0,12 m/s. Windspeed is highest in the Barents Sea during the winter period, having strong wind from August to December in Nordland and smaller areas just outside the coast of Finnmark.

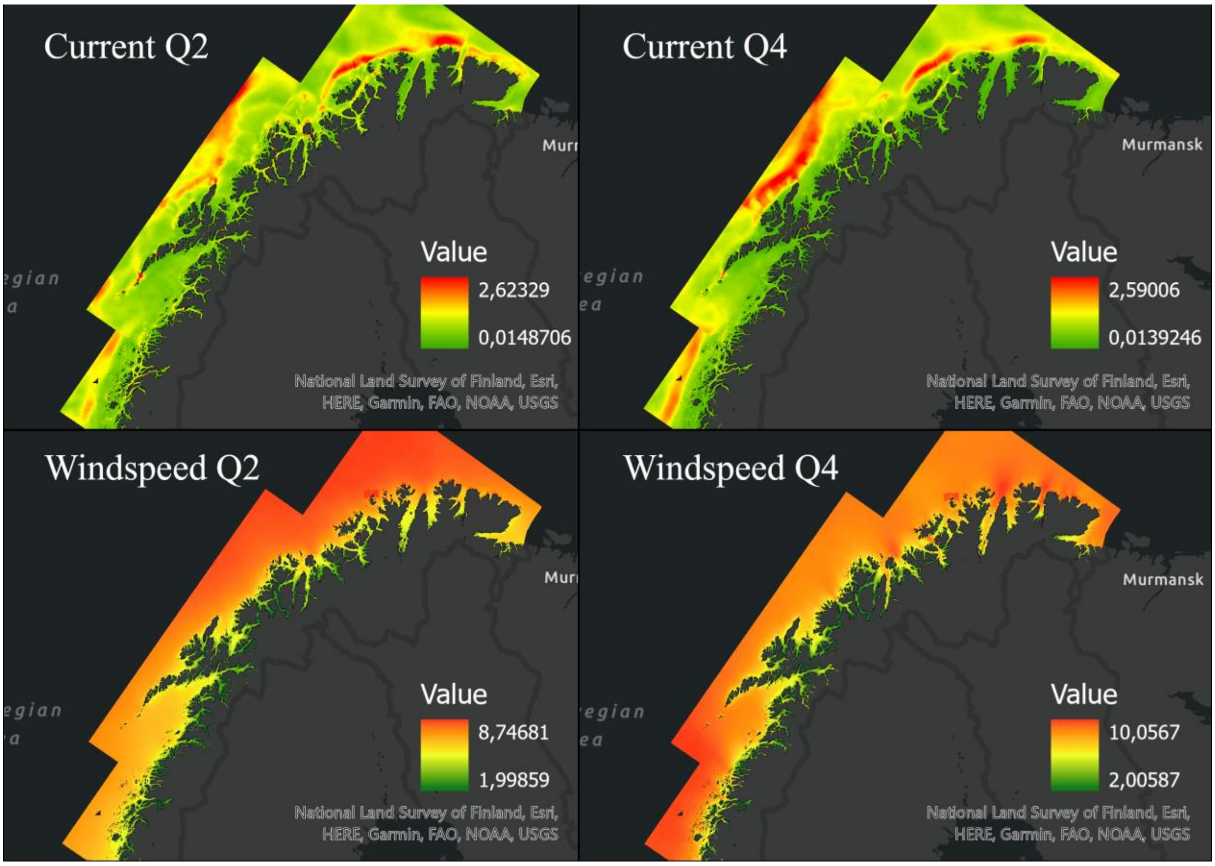


Figure 8: Spatio-temporal variation of current and windspeed surface averages for Q2 and Q4.

Day-length had a minimum value of 0 min and a maximum value of 1440 min. The average day length at the aquaculture locations was 754,9 min. Solar radiation varied greatly in the study area; direct sunlight was more present in areas south-facing (Figure 9). The average depth was 81m at the aquaculture locations with a slope of 10.7 degrees. Wave height had an average of 0,51 m, a minimum of 0,15 m and a maximum of 2,2 m. It is evident that the aquaculture locations are placed in sheltered areas inside fjords and are less affected by waves. Summery statistic of explanatory variables is attached in the appendix (Appendix G).

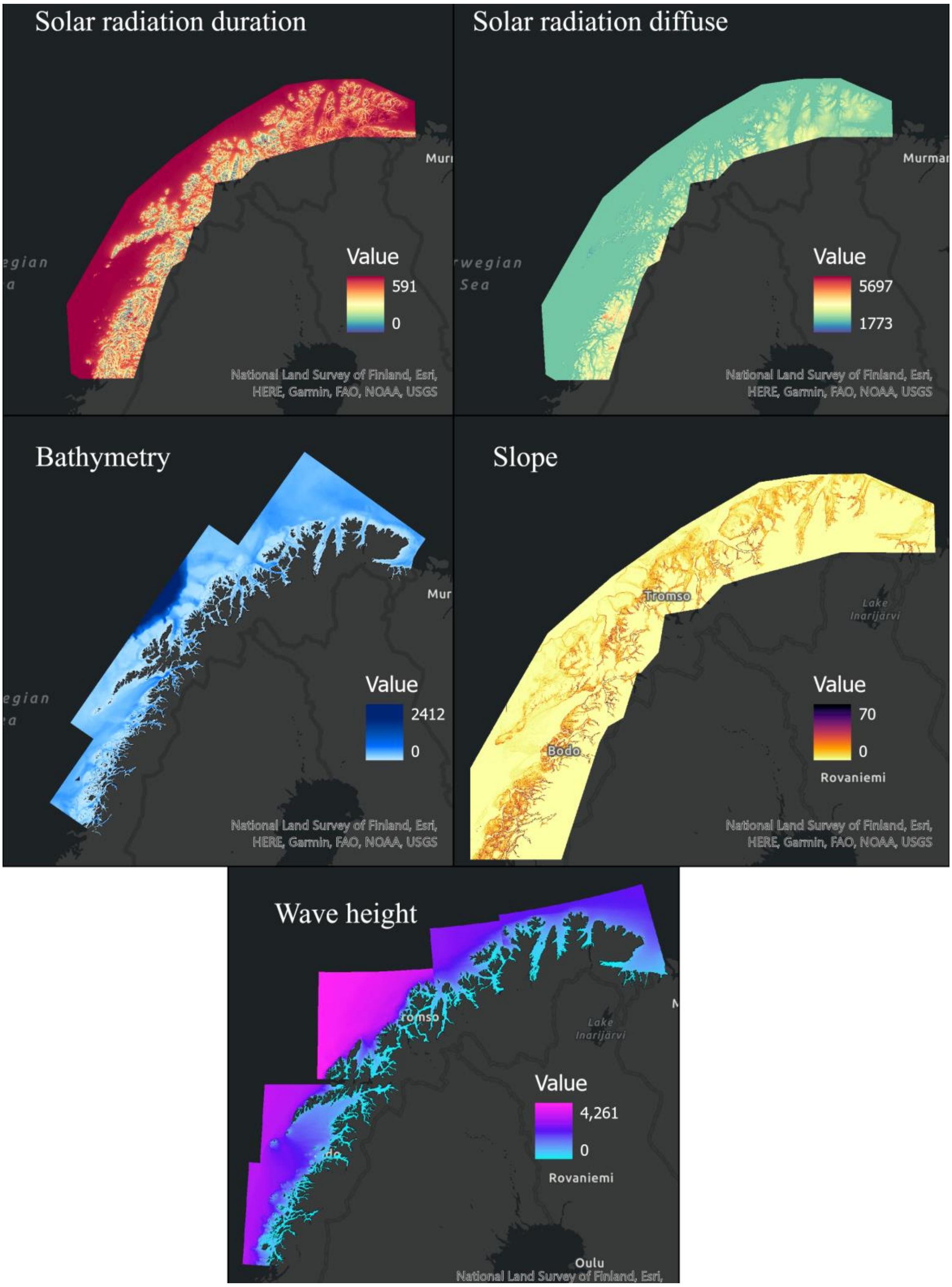


Figure 9: The average solar radiation, average wave height, bathymetry and slope in the study area.

4.1.2 Spatial and temporal variation of salmon- growth and mortality

There were some outliers in growth and mortality, as described in the method section., All data with SGR values below 0 and above 120 were removed. Similarly, all rows with mortality greater than or equal to 1 were removed. The results show that salmon farms in northern Norway had a mean SGR of 23,7 and a mean Mrate of 0,0093 during 2018 - 2020 (Table 4).

Table 4: Dependent variables statistics.

	Min	**Mean**	**Max**
Specific growth rate	0,035	23,7	120
Mortality rate	0,00008	0,0093	0.99

4.1.2.1 Temporal pattern of growth

The results show the unimodal (bell-shaped) pattern with maximum annual growth in August. The peak in August was slightly higher in 2020 than in 2019 and 2018. The slowest growth was between February and April every year, with the slowest growth in March, at the end of the month. In figure 10 below, SGR is aggregated to mean monthly values. The figure clearly shows the temporal differences in salmon growth.

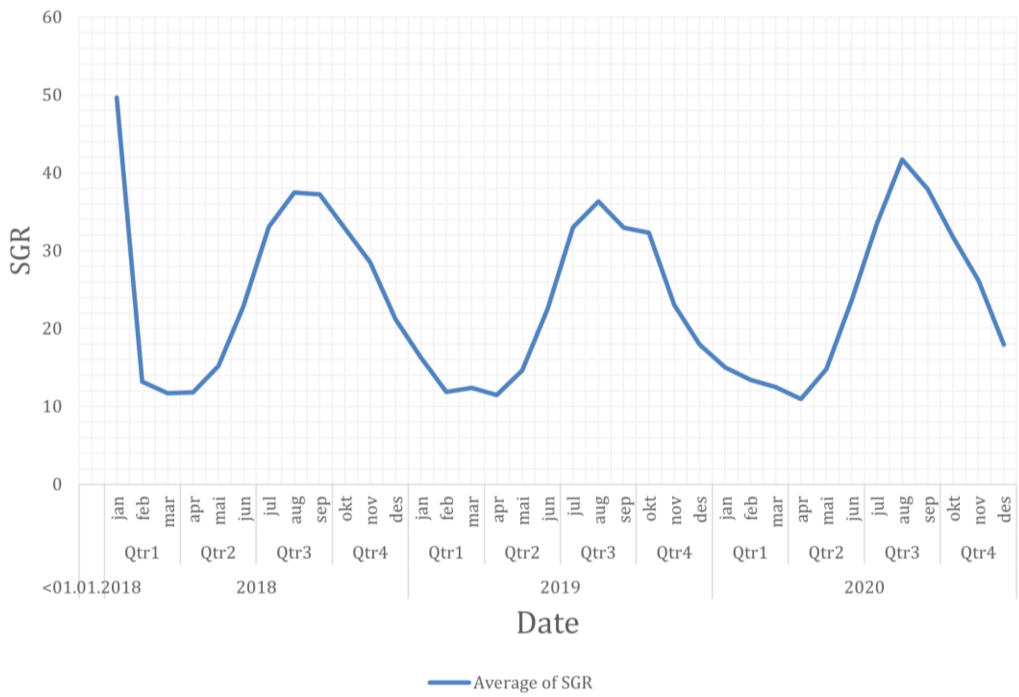


Figure 10: The temporal pattern of mean SGR between 2018-2021. The results show unimodal patterns with maximum growth in Q3 and the slowest growth in Q1.

The change in growth rate corresponds very well with the seasonal changes in water temperature. In figure 11 growth rate is plotted against temperature. The period with the highest temperature has the highest growth rates and vice versa. The average surface temperature was higher in 2020 compared to 2019 and 2018.

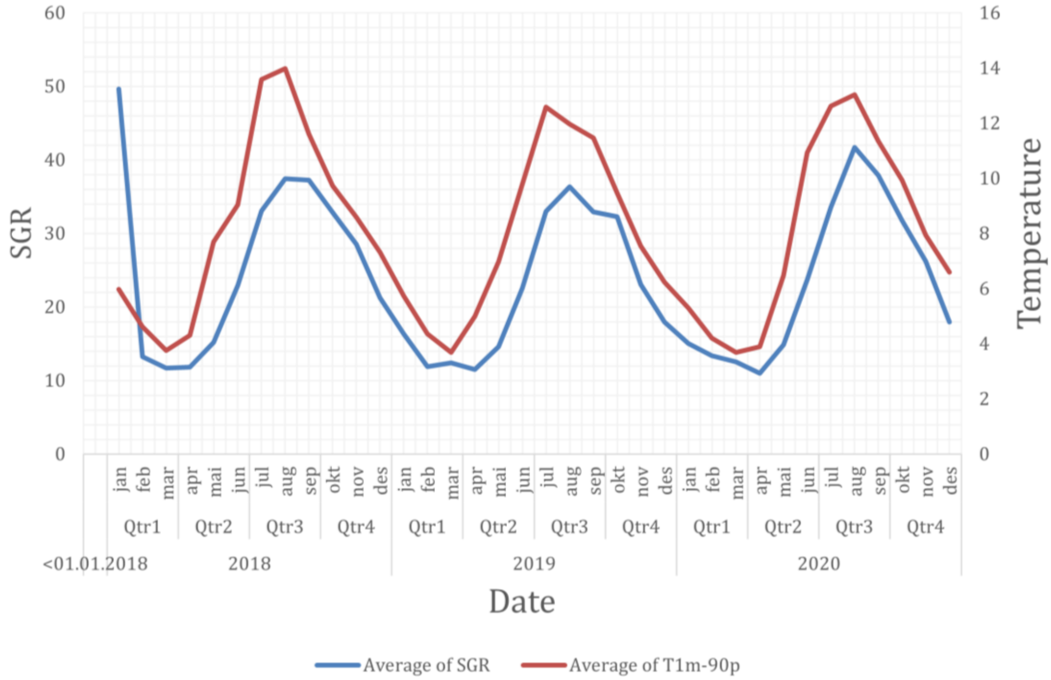


Figure 11: Comparison of change in mean SGR and average temperature.

4.1.2.2 Temporal pattern of mortality

The mortality rate was highest in May 2019 during the last three years (Figure 12). This corresponds with the harmful algae blooms of Chrysochromulina leadbeateri that killed more than 7.5millions farmed salmon at aquaculture sites in Nordland and south Troms (Hommedal, Lorentzen, & Hoddevik, 2019).

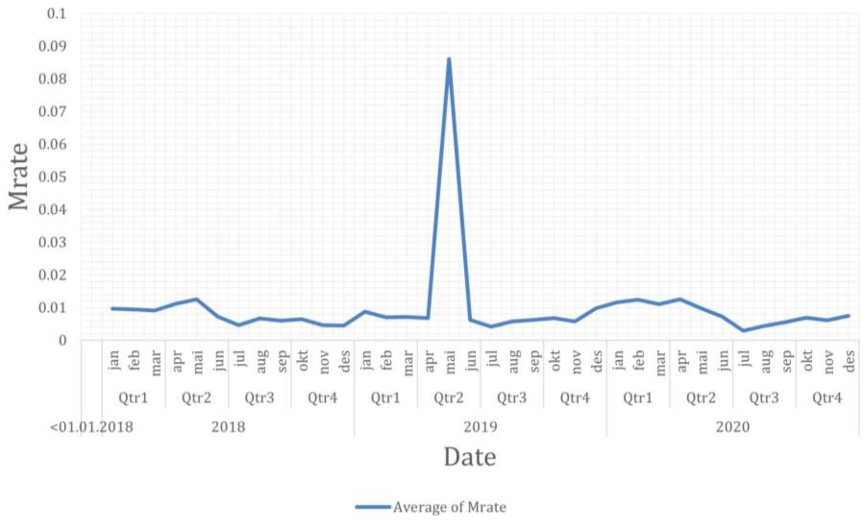


Figure 12: Temporal pattern of mortality rate. It is possible to spot a clear peak in Q2 2019 due to the algae bloom in south Troms.

Since such extreme values would interfere with the rest of the analysis, 46 rows were removed, obtaining a higher resolution of the mean mortality. The mortality rate did not show the same clear oscillations as the growth rate. However, mortality increased from March to April every year, having the highest mortality in 2020.

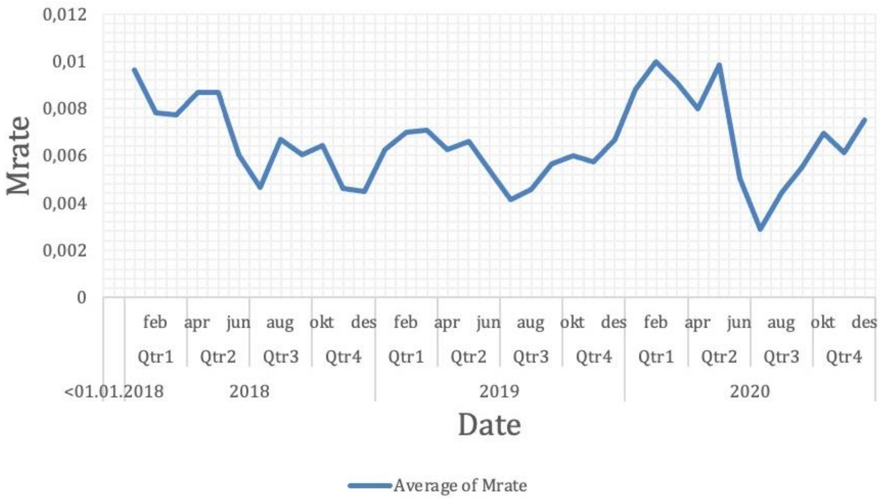


Figure 13: Temporal pattern of mortality rate. In this figure, 46 rows with excessive mortality rates were removed to get a clearer picture of temporal patterns for mortality rates.

4.1.2.3 Spatio-temporal pattern of growth and mortality

When aggregated growth and mortality data by production zone, spatial distribution patterns of both variables show somehow random distribution. However, it was possible to spot a trend with higher SGR on average in the southernmost production areas, with SGR declining at higher latitudinal degrees. Production areas 13 and 9 had significant variations in SGR between 2018 – 2021, while production areas 8, 10 and 12 had stable results. The mortality was dispersed, and it was difficult to see any spatial trend. The highest mortality was observed in 2019 for production areas 10 and 11.

Table 5: Summary statistics of spatio-temporal pattern of growth and mortality. Specific growth and mortality rate are average values grouped by production area.

Variable	Date	Production area 8	Production area 9	Production area 10	Production area 11	Production area 12	Production area 13
Specific growth rate	2018	25.229	27.465	25.162	23.131	23.719	20.931
	2019	24.262	22.331	23.119	19.793	19.769	27.492
	2020	24.495	24.845	23.243	24.797	22.388	21.084
Mortality rate	2018	0.007	0.007	0.007	0.006	0.007	0.009
	2019	0.006	0.009	0.017	0.021	0.009	0.006
	2020	0.006	0.007	0.007	0.008	0.011	0.009

Furthermore, as shown earlier, growth has a clear seasonal pattern in all production areas (Figure 14). Though, there are some spatial differences between the production areas. The northernmost production areas (11, 12 and 13) seem to have slightly lower SGR overall than production areas 8 and 9. Production area 13 has an unexpectedly high growth rate in the third quarter of 2019.

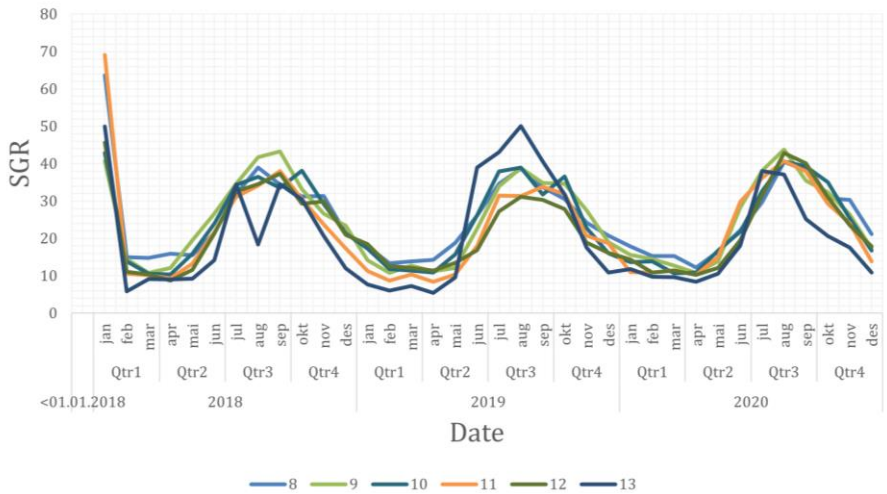


Figure 14 shows a specific growth rate plotted for the production area in a line chart. It is possible to derive that the southernmost areas have an overall slightly higher growth rate than those further north.

I did not observe any clear pattern of mortality when all data were used (Figure 15). However, when the extreme values were removed, very high mortality variability was observed though no clear trend or seasonal pattern was observed. Interestingly, the three northernmost production areas seem to have an overall higher mortality rate than the southernmost (Figure 16). Production area 13 also had a high peak of mortality in the first quarter of 2019.

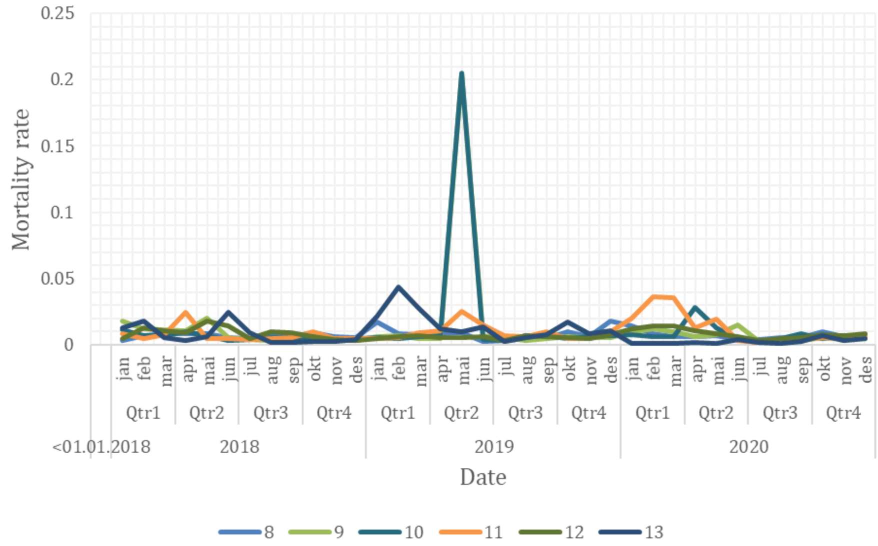


Figure 15: Mortality rate plotted for production area in a line chart. It is easy to spot the peak in May 2019 in production area 10.

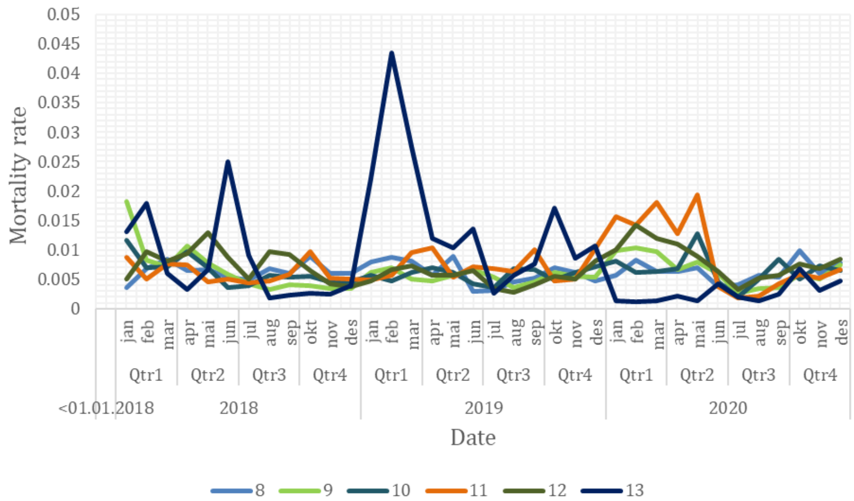


Figure 16: Mortality rate plotted for production area in a line chart where 46 outliers from algae bloom in production area 10 were removed.

4.2 Modelling

4.2.1 Linear regression OLS and GWR

Linear regression of the independent variables showed multicollinearity and internal redundancy (Appendix- H and I). The results show that only growth rate and temperature were correlated, indicating no linear relationship between growth rate and other socio-economic and environmental variables. Similarly, I did not find any significant correlation (and/or linear relationship) between morality and explanatory variables.

Ordinary least squares analysis revealed that the model was biased with a statistically significant Jarque-Bera value $p< 0.05$ and Global Moran's I p-value < 0.05. The results show that residuals were highly clustered (Appendix J). Further analysis also revealed non-linear trends between dependent and independent variables, indicating that ordinary parametric models like OLS cannot capture the relationship. Since the data is a time series between 2018 – 2021, data are highly clustered, with high multicollinearity, GWR cannot be applied.

4.2.2 Significant variables for salmon growth

RF classifier and regression identified the most significant variables for salmon growth using the recursive elimination method described in the methodology section. The five top variables were specific feed rate, temperature, biomass, day length and direct solar radiation. The specific feed rate and temperature were the most important variables. Both variables combined accounted for 70% of the importance in the model. Interestingly, current, wave height and distance features were the least important variables included in the final results.

The R-squared for both the training data and the validation was strong, with 0.968 in training and 0.819 for validation. The mean squared error was 0.494 for 50% of the trees, with 79.6% of the variation explained. Table 6 gives an overview of all independent variables included in the modelling results.

Table 6: The most significant variables for salmon growth. Results are derived from numerous RF iterations and recursive feature elimination.

Variable	Importance	%
Specific feed rate	5556.28	53
Temperature	1823.69	17
Biomass	941.82	9
Length of day	446.67	4
Direct solar radiation	303.02	3
Salinity	272.46	3
Windspeed	222.87	2
Distance to fairways	160.67	2

Slope	160.25	2
Bathymetry	158.50	1
Distance to aquaculture sites	153.40	1
Distance to communities	152.53	1
Wave height	152.35	1
Current	67.60	1

Since specific feeding rates had such high importance for growth, another model was built, excluding this variable. The result was a similar model with temperature as the most important variable. However, salinity was now included as one of the 5 top important variables. Temperature and biomass accounted for 54% of the importance (Table 7). The training data had a strong R-squared of 0.949. However, there was a decrease in validation data with an R-squared of 0.655.

Table 7: The most significant variables for salmon mortality. Results are derived from numerous RF iterations and recursive feature elimination.

Variable	**Importance**	**%**
Temperature	3507.32	33
Biomass	2228.96	21
Length of day	912.26	9
Direct solar radiation	772.11	7
Salinity	530.67	5
Windspeed	462.53	4
Bathymetry	351.07	3
Slope	325.87	3
Distance to aquaculture sites	319.76	3
Distance to fairways	315.65	3
Distance to communities	311.36	3
Wave height	287.29	3
Current	148.57	1

4.2.3 Significant variables for salmon mortality

Based on the RF model, salmon mortality was mainly explained by the five variables: the temperature at the surface, biomass, length of the day, the temperature at 10m depth and diffuse solar radiation. Temperature and biomass accounted for 54% of the importance of the model (Table 8). However, I observed more spread in the importance of the variables here rather than with the growth rate. Wind speed and wave height have importance of 8%. The model obtained a training R-squared at 0.948 and a validation R-squared at 0.905.

Table 8: The most significant variables for salmon mortality, excluding SFR. Results are derived from numerous RF iterations and recursive feature elimination.

Variable	**Importance**	**%**
Temperature at 1m depth (90 percentile)	2.35	23
Biomass	2.11	20
Length of day	1.30	12
Temperature at 10m depth (ave)	1.18	11
Diffuse solar radiation	0.96	9
Wind speed	0.83	8
Wave height	0.80	8
Salinity	0.50	5
Current	0.42	4

4.3 Prediction of optimal salmon farming sites with RF

4.3.1 Optimal sites (presence) and other sites (absence)

As mentioned in the method section, break points were growth rate > 34,61 and mortality rate < 0,64 obtained 0. This resulted in an imbalanced presence-absence dataset with 42 presence- and 252 absence points. An overview of the statistics is shown in table 9. Test for spatial autocorrelation indicated no spatial autocorrelation. Given these values, the pattern is not significantly different from random. An overview of the Global Morans I summary is attached in table 10.

.

Table 9: Summary statistics of presence- and absence points.

	Growth rate	**Mortality rate**
Count	294	294
Min	2,31	-7,62
Max	7,07	-3,86
Mean	4,57	-5,56
Standard Deviation	0,68	0,54

Table 10: Global Moran's I summary of presence points.

Global Moran´s I Summary	
Moran´s Index	0.0288
Variance	0.000521
z-score	1.4146
p-value	0.1571

4.4 Quarterly prediction

In the quarterly prediction, variable importance varied throughout the years, but the difference between variable importance is low. The temperature was the most significant variable for both vertical layers in the first quarter, followed by depth and current. Second-quarter, current at the surface was the most important, followed by temperature. Third-quarter, the current was the most important. While in the fourth quarter, current and wave height was most important. Interestingly, in the fourth quarter, the temperature at 30 m did not show any importance. As expected, solar duration was only significant in the second and third quarters. At the same time, windspeed was important in the first and fourth quarters. Table 11 provides an overview of the variable importance thorough the year.

Table 11: Most significant variables with RF classification method.

Variable	**Importance**			
	Q1	**Q2**	**Q3**	**Q4**
Temperature Surface	**14**	**12**	11	9
Temperature 30m	**13**	**12**	11	Na
Bathymetry	**13**	10	10	**11**
Current Surface	12	**13**	**13**	**12**
Current 30m	12	Na	**12**	**11**
Wave Height	12	11	11	**11**
Windspeed	12	Na	Na	10
Solar Diffuse	12	Na	11	9
Salinity Surface	Na	**12**	11	9
Solar Duration	Na	10	11	Na
Slope	Na	10	Na	9

The results for the prediction distribution model provided a good fit for all quarters. The OOB error ranges from 14 – 20% for absence points and <1% for presence. The highest OOB errors were found in the third quarter (Table 12). The model's sensitivity and accuracy proved very strong for all four quarters (Table 13).

Table 12: Model out of bag error shows OOB errors below 20% for absence points and 1% for presence points.

Quarter	**Number of trees**	**750**	**1500**
1	MSE	13.563	12.714
	0	15.716	14.737
	1	0.078	0.039
2	MSE	12.905	13.063
	0	14.998	15.168

	1	0.053	0.027
3	MSE	17.959	18.340
	0	20.823	21.276
	1	0.125	0.062
4	MSE	13.074	12.760
	0	15.114	14.768
	1	0.226	0.113

Table 13: Training data classification diagnostics with sensitivity and accuracy.

Quarter	Category	F1-Score	Sensitivity	Accuracy
1	0	0.93	0.87	0.89
	1	0.71	1	0.89
2	0	0.92	0.85	0.87
	1	0.69	1.00	0.87
3	0	0.88	0.79	0.82
	1	0.60	1.00	0.82
4	0	0.92	0.86	0.88
	1	0.69	1.00	0.88

Optimal areas for aquaculture varied greatly throughout the year. Third-quarter had the most area per square meter that resulted as optimal, while the first quarter had the least. In common for all the quarters, areas are optimal from Gladstad in the south to Vadsø in the north (Figure 17). There are areas along the whole coast that are predicted as optimal. In the first quarter, many fjords in Nordland were optimal, especially Vefsnfjorden, Ofotfjorden, Nordfolda (Figure 17). Large areas in Andfjorden around Rolla and Andørja. Areas around Lyngenhalvøya, Olderdalen, Loppa, south of Sørøya and east of Vadsø was also optimal (Figure 17). In the second quarter, this changed. The fjords in Nordland optimal in Q1 were no longer as significant. However, there is a tendency for areas deeper inside fjords to be optimal rather than the whole fjord. This is especially visible in Glomfjorden and Holandsfjorden. The same can be observed in and around Rolla and Andørja. Gratangen and Lavagen, which was not optimal in Q1, are optimal in Q2. The same is true for Sørfjorden, Lyngenfjorden, Olderdalen and Altafjorden. Interestingly in Q2, almost the whole Porsangerfjorden was predicted as optimal for salmon farming and large areas inside Varangerfjorden. In Q2 large and more dispersed coast areas outside Senja, Tromsø and Ringvassøya were also predicted as optimal. The third-quarter predictions overlapped with Q2, but the predictions were more profound (dense). Especially the fjords around Sandnessjøen, Ranfjorden, Vefsnfjorden, Sørfjorden and Melfjorden showed very significant predictions. The same goes for Skjerstadfjorden, Ofotfjorden, Malangsfjorden, Lyngsfjorden and Varangerfjorden. In Q3, Porsangerfjorden did not show nearly the same predictions as in Q2. The optimal sites were moved towards the fjord estuary. The most significant areas predicted optimal in Q3 were the areas off the coast outside Honningsvåg, Nordkapp, Hammerfest and Sørøya. The fourth quarter mostly overlapped with Q3. This was especially true for the fjords already mentioned in Nordland and the fjords around Lyngen in Troms. New predicted optimal areas were Stjernøysundet, Stjernsundet, Sørøysundet, just west of Alta. There was also a long and tiny bowel stretching from Sortland to Ringvassøya predicted as optimal, which was not predicted in any other quarters.

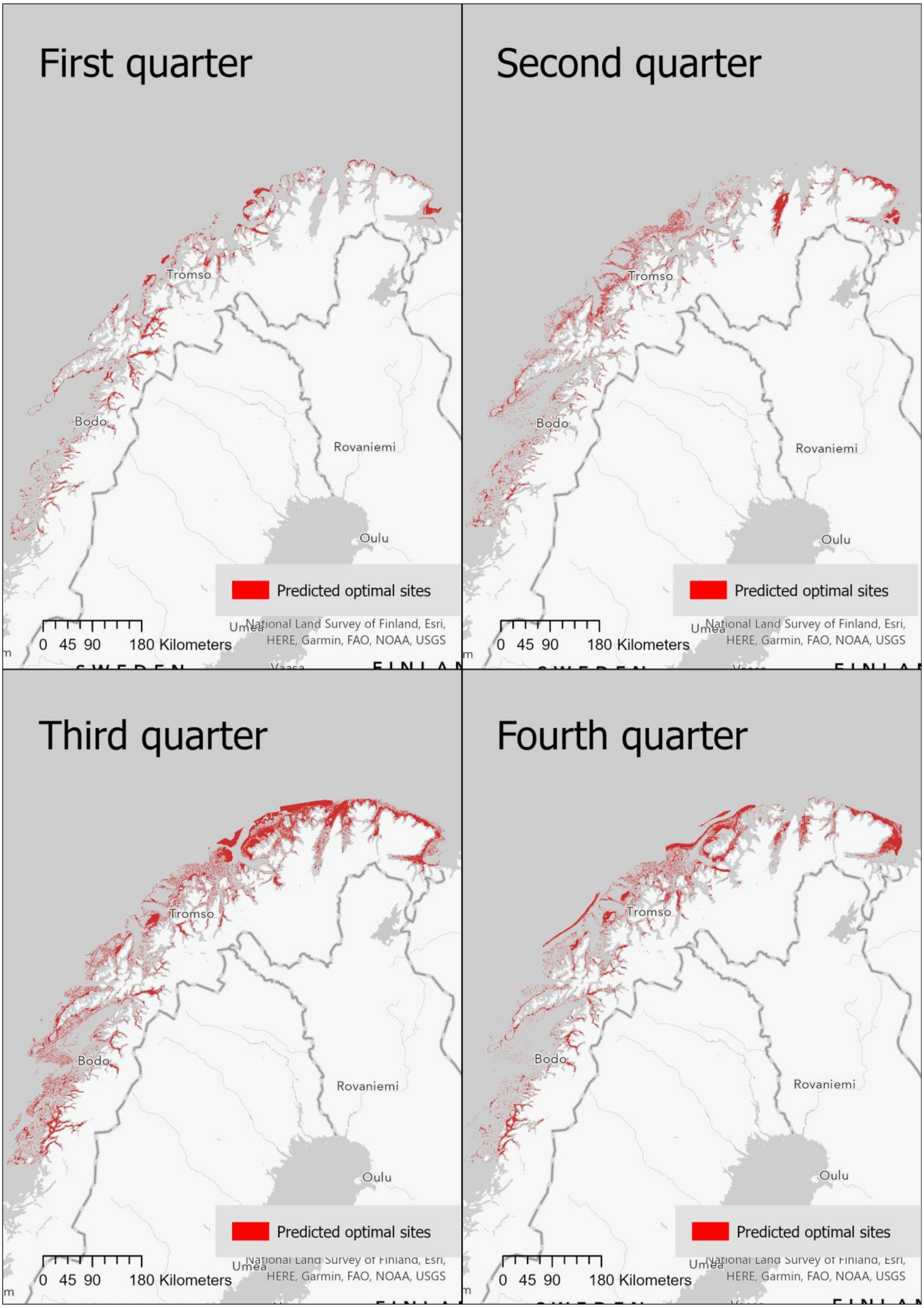


Figure 17: Quarterly predictions of optimal salmon farming sites in northern Norway.

4.5 Predicted optimal salmon farming sites

As mentioned in the methodology, all quarterly predictions were multiplied or summarized, creating two output maps of optimal sites. The product map (Figure 18) shows only areas that were predicted optimal for all four quarters. The summarized map (Figure 19) shows areas that were predicted as optimal for two, three or all four quarters (recommended, highly recommended, optimal sites).

The map displays the predicted most optimal areas for the whole year. Optimal areas are dispersed, ranging from inside fjords to the archipelago and open waters. Moreover, optimal areas were present in the entire study area from Sandnessjøen in the south to Vardø in the north. From the south, areas inside Vefsnfjorden, Ranfjorden, Sørfjorden and Glomfjorden, and small areas near the coast and around the small islands stretching from Gladstad to Narvik. Areas near the coast of Lofoten, Værøy, Røst, Vesterålen, Sortland and Senja was also predicted as optimal. Further north, some dispersed areas are predicted as optimal outside Tromsø and Sørøya. Large areas around Varangerhalvøya, and fjords such as Olderdalen/Kåfjorden, Lille Altafjorden/Burfjorden, and near the shore on both sides of Stjernøysundet.

Figure 18: Optimal sites for aquaculture. Product of all four quarterly predictions.

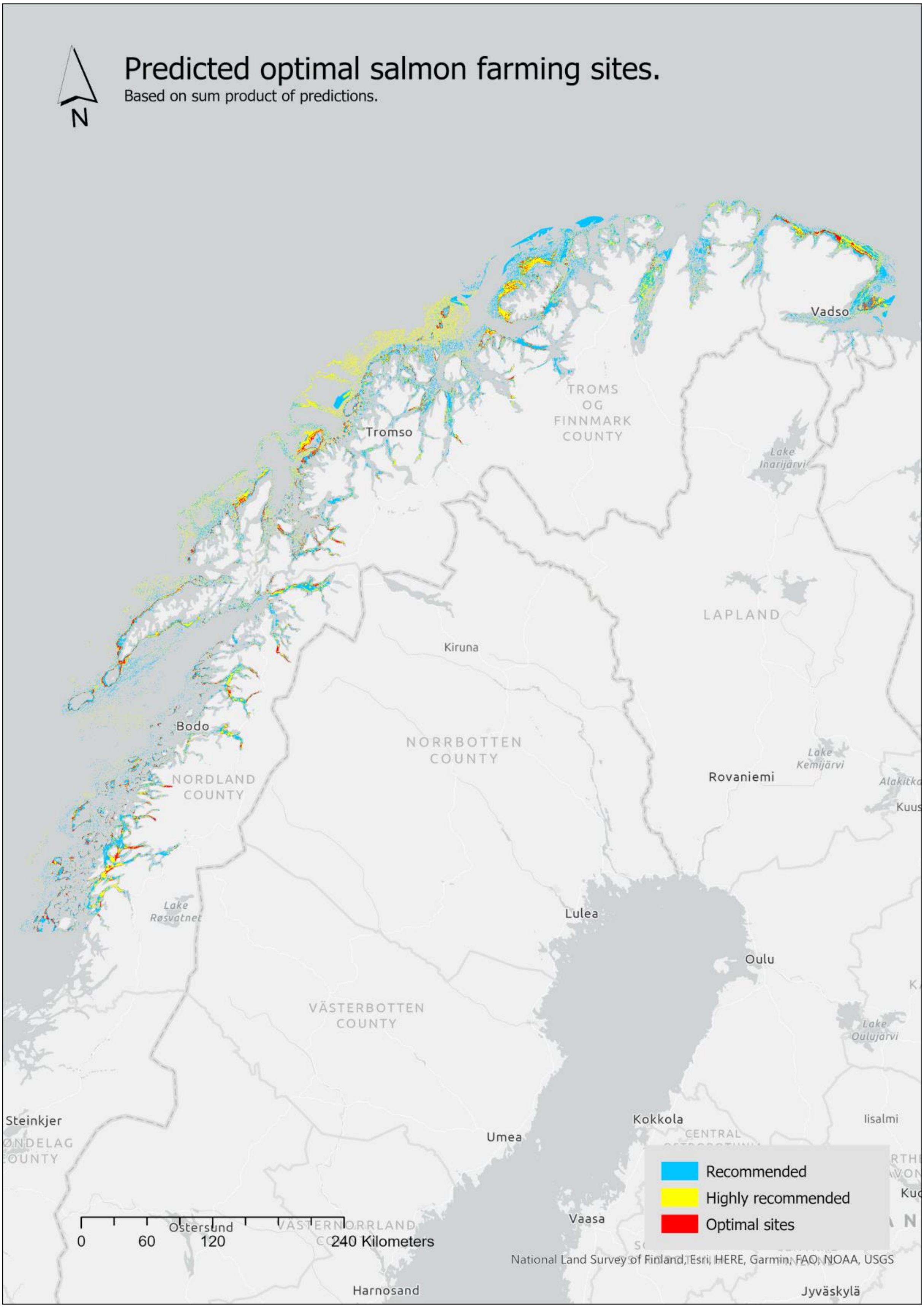


Figure 19: Optimal sites for aquaculture. Sum of all four quarterly predictions. Optimal areas are classified from 1-3 (1. Recommended, 2. Highly recommended, 3. Optimal sites).

Plotting the sum of the quarterly prediction in an overlay map with the aquaculture sites gives a clearer picture of the commercial salmon farm's spatial siting versus the predicted optimal areas. In Nordland, some of the farms in Vefsnfjorden, Ranfjorden and Sørfjorden are ideally located following the predictions. There are also some salmon farms situated in optimal areas in Sørfolda. In Troms, the aquaculture locations south and east of Andørja, Gratangen, and Kåfjorden are located optimally. Lastly, in Finnmark, one salmon farm located in Lille Altafjorden is optimal. There are many locations situated in recommended areas southwest of Hammerfest. However, regarding this study's results, many of the salmon farms in northern Norway are located in sub-optimal areas (Figure 20)

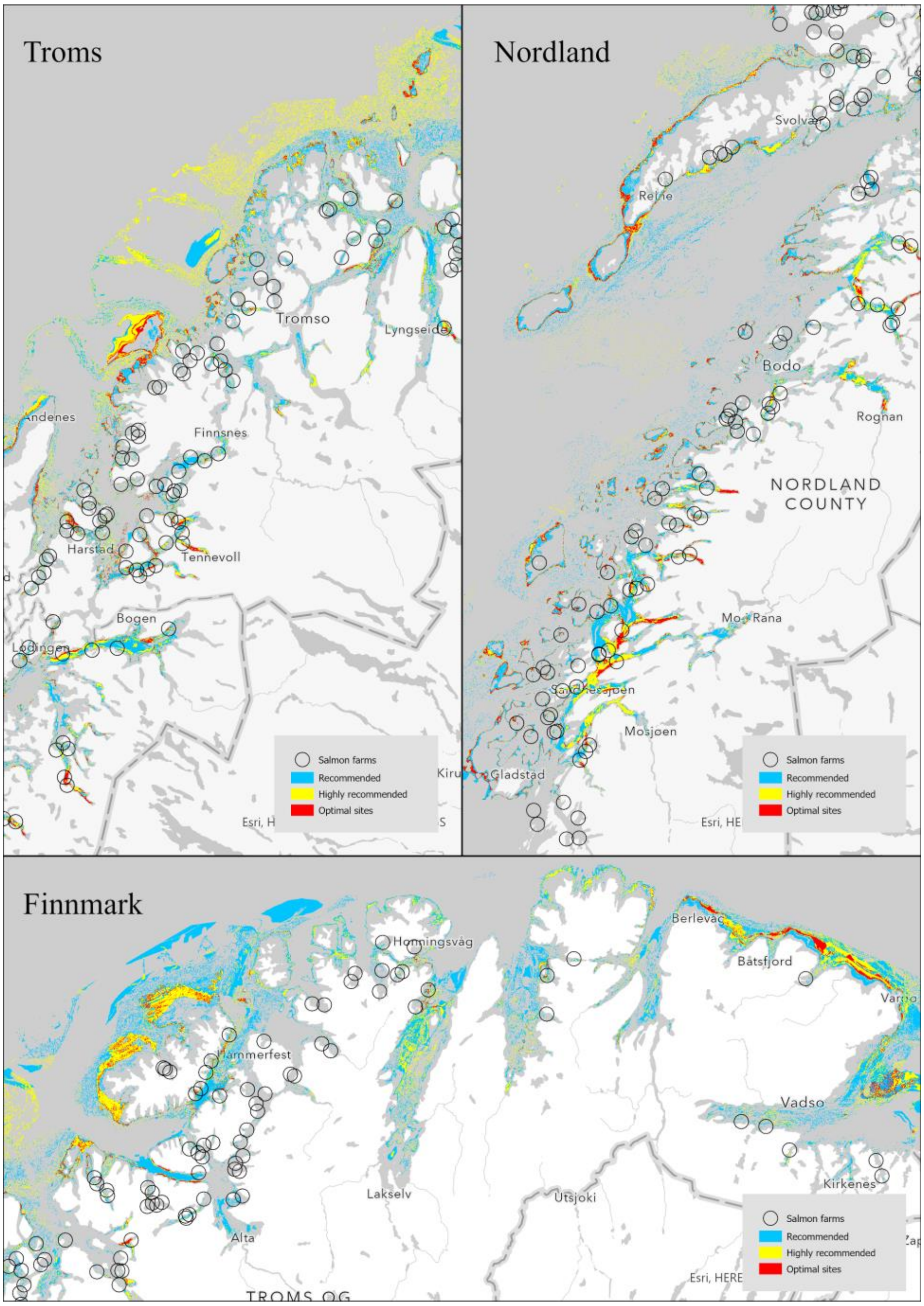


Figure 20: Sum of quarterly predictions and salmon farming sites. The map shows which salmon farms are in the predicted optimal areas.

5 Discussion

In this study, using production data from all salmon farms in northern Norway for the last three years, I attempted to identify the most significant environmental factors for salmon growth and mortality. Further, I have tried to predict the optimal areas for salmon farming in northern Norway using a data-driven RF algorithm. To the best of my knowledge, this is the first attempt to use a data-driven machine learning approach to model optimal sites for salmon farming based on fine-scale production data and moderately high resolution environmental and socio-economic data.

5.1 Integration of welfare indicators, environmental and socioeconomic factors

The growth and mortality rates were used as the only welfare indicators in this study. Both welfare indicators explain all the needs of farmed salmon defined by Noble et al. (2018), respiration, osmotic balance, thermal reg., good water quality, body care, hygiene, safety and protection, behavioural control, social contact, rest, exploration, sexual behaviour, feeding and nutrition. Appetite, behaviour, and diseases are group welfare indicators that could be applied to the study. However, this was not possible because this would have required physical presence or a daily log from the different salmon farms. This would also have put restrictions on both the study area, period, and the number of salmon farms in the study.

The fundamental assumption in this thesis is that if the suitability of the site is explained by the exploratory variables present in the study, then data indicating optimal areas could be used with a machine learning approach to identify the most significant contributing variables and predict other areas in the study area that inhabits the same environmental attributes. In this study, I used and explored the impact/contribution of the available ecological and oceanographic data – salinity, temperature, current, wind speed, wave height, benthic and topographic factors - slope, solar radiation, and the impact of socioeconomic factors - distance from the urban settlement, distance from shipping lane and distance from other farms.

This thesis is based on Longva & Elvenes (2016) principles to categorize optimal areas based on welfare indicators and to develop a predictive model. In their study, Longva & Elvenes (2016) used MaxEnt and an SDM method to find the optimal sites for salmon farming in south Troms using the EGI and environmental data from the Aastafjord-project. However, I extended their approach, using a solid statistical, empirical dataset in a purely data-driven machine learning approach using RF regression and classification.

The environmental variables were included based on the best theoretical knowledge from the authors mentioned in the theory chapter. In chapter 2, I have discussed in detail that the environmental conditions under which salmon are grown are essential for their welfare and are reflected in physiological responses (Bowden et al., 2002; Johansson et al., 2007; Turnbull et al., 2005), and that water quality depends highly on local environmental characteristics, and the different characteristics affect both growth and mortality (Edwards & Edelsten, 1977; Hargrave et al., 1993; Stigebrandt et al., 2004). It is primarily the local conditions that are interesting in terms of growth and mortality. In this study, however, large scale models such as NorKyst800, SWAN, AROME are used to model the environmental conditions. Thus, the local variations and conditions are not present in the analysis, which can significantly impact the results.

The solar radiation variables included the topography of northern Norway. This was included to capture the impact of direct sunlight on different salmon farms. Northern Norway has a topography of deep fjords and high mountains stretching from west to east. Farms located in these fjords, especially on the south side of the fjord, would experience significantly less direct solar radiation than those located north. Daylight and diffuse solar radiation were also added to capture the difference in total daylight between the farms situated above and below the arctic circle. Many of the farms located far north have many days without any sunlight. They also have many days with the sun above the horizon for 24 hours.

It is not only the environmental factors that could explain an optimal farming location. The socio-economic variables are implemented as proxies for optimal salmon farming sites. Many of the small towns and coastal communities in northern Norway suffers de-population. Labour is critical to salmon farming as it requires daily husbandry. Farmers are also interested in localities near their existing farms or any vital infrastructure. This is solely an economic driven interest. Distance to communities was calculated and included in the study to capture accessibility to the labour market, accessibility to the city centre and accessibility to essential infrastructure.

Today, the Norwegian Food Safety Authority, as mentioned earlier, uses a recommended minimum distance of 2.5 kilometres between aquaculture facilities. The distance to production facilities should be no less than 4km because of the "unacceptable risk of spreading infection". In addition, the distance to fairways should be not less than 1.5km. It is not the fairway itself that can transmit these pathogens, but the ballast water from well boats or any other ship that uses ballast. There is also more traffic, hence more noise and disturbance near fairways, disturbing the salmon. In this thesis, distance to aquaculture sites and distance to fairways is included to see if salmon farms close to these objects has higher or lower growth- and mortality rate.

Disregarded the many factors included in this analysis, many important variables are not included. Efforts were made to collect location-specific data. Several authors, including Bowden et al. (2002); Kazakov and Khalyapina (1981); Kindschi and Koby Jr (1994), found that dissolved oxygen (DO) in the water is a crucial factor affecting fish health and thus fish metabolism and growth. I could only collect this from some locations, but not near as much I needed to include in the analysis. With the few data points collected, I used a simple linear regression to see the correlation between growth and DO. The results were a positive relationship with a weak correlation coefficient. More data is needed to be able to model this relationship. Oppedal et al. (2006) found that geographic locations were the best predictor of variation in oxygen levels inside cages. The oxygen level was also better at coastal sites than at fjord sites, and that oxygen level was closely linked to salinity and current. If there is a strong relationship between oxygen levels and growth, salmon farms in coastal/open areas would be preferable.

5.2 Essential factors affecting salmon growth

The specific feeding rate, temperature and biomass were the variables that were identified as the most important factors affecting salmon growth (having an R squared value of 0.968 in training and 0.819 in the validation test). The results clearly show that about 82 % variance in the salmon growth is explained by specific feeding rate, temperature, biomass, length of the day, direct solar radiation, salinity, wind speed, distance to fairways, slope, bathymetry, distance to aquaculture sites, distance to communities, wave height and current. Since I wanted to create a model only explaining external factors, SFR was removed as a variable as this can internally be modified. It is essential to note the strong relationship between growth and specific feeding rate. The second model had temperature and biomass as the variables that were identified as most important (having an R squared value of 0.949 in training and 0.655 in validation). The validation data dropped significantly when SFR was removed. The results showed that the same variables above explain 62% variance in salmon growth without SFR.

Growth was shown to be slightly higher in the southernmost areas. This is probably tightly connected to the temperature difference between the southern- and northern parts. The model confirmed this assumption, and the temperature was one of the most critical contributors to salmon growth. The temperature was also the only exploratory variable with an obvious positive correlation with a high correlation coefficient. Specific feeding rate (SFR) was an essential contributor with 53% importance. The correlation between SGR and SFR was largely positive. When removing SFR, biomass increased significantly in importance. Large biomass will reduce the space inside the cages, and this can cause hypoxic conditions and increase competition, hence reducing fish welfare and growth. This can be one of the explanations for why biomass is an important factor.

Other significant contributors were the length of day and the direct solar radiation. Both explain the same thing. The length of the day is minutes of daylight, and direct solar radiation is kw/h direct sunlight on the farming site. Since salmon are visual eaters, they are not fed when it's dark outside. In addition, when it's dark outside, it's usually much colder, and the metabolic processes will run much slower, decreasing the need for feed. This can be the cause of the importance of the sunlight factors. Artificial light was disregarded in this analysis because it was impossible to retrieve data on this matter. The socio-economic variables, distance to aquaculture sites, distance to fairway and distance to communities contributed 3% each to the model. This could indicate some underlying socioeconomic factors contributing to salmon growth. The farm's location can be in an area near important infrastructure. If something happens to the feeding system, it’s easy for the farmers to react and fix the problem. It can be that locations further away from other aquaculture sites and fairways experience less transmission of pathogens and noise, which again contributes to good fish welfare, thus growth.

Interestingly, the less significant variables were windspeed, bathymetry and slope, which is contrary to the findings of Longva and Elvenes (2016), who found that surface- and bottom current and depth were the most significant factors. The literature also suggests that current is important. As mentioned previously, Hargrave et al. (1993) found that the photosynthesis capacity in the vicinity of a farm is not ordinarily sufficient to supply the oxygen demand of the total fish biomass in a marine fish farm (Hargrave et al., 1993). Therefore oxygen requirements must be met by physical transport such as tidal movement, current, or freshwater runoff (Johansson et al., 2007). Current is also essential to change the water, remove faces and feed away from the farm.

5.3 Essential factors affecting salmon mortality

The temperature and biomass were the variables that were identified as the most important factors affecting salmon mortality (having an R squared value of 0.948 in training and 0.905 in the validation test). The results clearly show that about 91% variance in the salmon mortality is explained by temperature, biomass, length of the day, diffuse solar radiation, wind speed, wave height, salinity and current. Temperature is also the most important contributor to mortality. From the mortality rate graphs (Figure 12 and Figure 13), it seems like the total highest mortality every year is the first quarter, disregarding the peak in May 2019 caused by the algae bloom. According to Oliveira et al. (2021), the temperature had a non-linear effect on mortality, and temperatures outside a range of 5-10°C were associated with increased mortality, temperatures below 2°C can be lethal. This can explain why there was a slightly higher mortality rate for the three northernmost production areas. Another top variable was the length of day and the diffuse solar radiation. These variables probably describe the same as temperature; as the days get longer, the sun is more hours above the horizon, and the temperature also increases. When the sun is not present, the temperature decreases and mortality rises when temperatures are below optimal. Wind speed, wave height, salinity and current were also significant contributors. As mentioned in the previous chapter, DO levels are highly affected by the current running through the cages and the salinity. Mortality as such can be a subject of low DO levels. Oliveira et al. (2021) also found that water with salinity close to 33 ‰ and lower would cause generally higher mortality and that multivariate regression confirmed that salinity was an important environmental determinant. The less important variables in this model were the socio-economic variables. There was nothing in this model suggesting that distance to fairways and aquaculture sites was an important determinant of mortality because of transmission of pathogens.

Mortality in the salmon farming industry is tightly connected to sea lice, cardiomyopathy syndrome, pancreatic disease, and ILA, which this model does not include. It can be the case that the regulations are effective and that transmission between aquaculture sites and fairways does not happen. However, I doubt this mainly because there have been large outbreaks of ILA around Ibestad in south Troms in the last years, which strengthens the assumption of pathogens travelling with the current. In addition to this, there are important driving factors for mortality that this model did not cover. Treatments against salmon lice using baths with H_2O_2 or medicinal compounds and non-medical treatments significantly impact salmon mortality (Oliveira et al., 2020).

5.4 Predictions

Prediction to surface rasters cannot use multipoint data. Therefore, all monthly data from each location were aggregated into the average growth- and mortality rate for each unique site. Some of the sites had significant variations in monthly growth and mortality. This can be caused by many different factors, as explained in the discussion about growth and mortality. The average growth- and mortality rate would, in this model, serve as the best estimate of the overall performance of the site. Based on bivariate colours and natural breaks, I classified the best (presence) and the worst (absence) sites. The presence points will indicate the optimal farms based on these threshold values. However, this is based on a more subjective assessment, and the analysis results would be largely affected if threshold values are modified. All models in the prediction gave an accurate and reliable result with low error margins. The out of bag errors were below 20% for all runs. During the analysis, I realized that some salmon farms were in areas with depths shallower than 30m. The salmon farms in these areas were excluded from the study due to the requirements of RF.

The RF model with 1500 trees proved very stable for all four quarters. There was little to no difference between the top variables in the models. However, the temperature was the most significant variable in the first quarter, then the importance declined in the following quarters and was replaced by both surface- and 30 m current. The different quarters gave very different predictions. This is the result of fixed independent variables and variation in exploratory variables. Note that the predictions of optimal sites did not use the same environmental variables as the identification of significant variables. Two different methods were used to extract data from the hydrographical models. Data points were extracted at the corresponding date and location at three vertical layers when identifying significant salmon growth and mortality factors. I converted netCDF files to multidimensional raster layers for only two vertical layers and quarterly averages for the prediction. I should have extracted the same variables as used in the regression as for the classification, but extracting, elaborating, and processing the data was time-consuming and demanding computationally.

The first quarter predicted optimal sites to be very dispersed. Large areas inside fjords near Lofoten, South Troms, Lyngen, Alta, and an area east of Vadsø were predicted as optimal areas. There were also a lot of smaller areas along the whole coast. In the second quarter, areas, especially inside fjords, that were indicated as optimal in Q1 did not display as optimal in Q2. The second quarter has more favourable environmental conditions in coastal waters. However, a large area of Porsagerfjorden was predicted as optimal. In the third quarter, we observed that current, surface and 30m depth were the most significant variable, followed by temperature, the most important in Q1 and Q2. Accordingly, predictions seem to be affected by this. Large areas outside the coast of Finnmark are predicted as optimal, the same goes for, Skjerstadfjorden, Vefsnfjorden, Ranafjorden, Saltfjorden, Sørfold, Nordfolda, Ofotfjorden, Nordfjorden and Balsfjorden. In the fourth quarter, the temperature seemed to have even more declining importance in favour of current, wave height and bathymetry. The predictions show similar optimal areas as in Q3, but they were denser. The large area outside the coast of Finnmark that was previously optimal had now been reduced in the total area and was moved south towards Sørøya.

The significant variations in optimal areas for aquaculture in the different quarters are likely due to large seasonal environmental variations of temperature and current in northern Norway. Following the optimal quarterly areas for salmon farming, one could quickly think that moving alive salmon between locations in different areas during the year would be beneficial. This could be true following the assumptions of this study. Nevertheless, this study does not consider the consequences and mortality of fish transportation or the financial implications that follow. Considering that salmon is stationary, companies and the authorities should plan aquaculture locations as such. This thesis shows the most optimal sites for salmon farming by multiplications of all quarterly predictions. The truly optimal sites are dispersed areas outside Helgeland, and in Ranfjorden, Sjona, Melfjorden, Glomfjorden, small areas inside Folda, Hellmofjorden, some areas inside Ofotfjorden and almost the whole coast of Lofoten following the archipelago in Nordland. In Troms and Finnmark areas outside Senja, Tromsøflaket, Norfjorden, Aastafjorden, Rolla, Andrøja, Lavangen, Kåfjorden, areas around

Sørøya and Stjernøya, north of Vardangerhalvøya and areas outside Vadsø was predicted as optimal for aquaculture.

5.5 New aquaculture sites: area challenge?

As mentioned in the introduction, the Norwegian government plan to increase the production of farmed salmon fivefold by 2050. To meet the requirements of good fish health and sustainability, the municipalities have to open for aquaculture in many new locations. As mentioned in the theory chapter, an industry expansion of this size will increase user-user conflicts and user-environmental conflicts. There are a lot of competing interests in the Norwegian coastal zone, including, e.g. fisheries, sea transport, the energy sector, military training and exercise, recreational tours, and fishing. Figure 6 in the results section shows the different competing industries and the space available in the three northernmost counties. The conclusion is that space is a scares resource.

The output map (Figure 18 and Figure 19) represents the modelled optimal sites for salmon farming.

An overlay map of these two is presented in figure 21.

Competing uses of the coastal zone are plotted with striped fill. Optimal areas for salmon farming are marked in red. If regulations are such as today, only a few optimal areas are available for aquaculture (yellow circles in Figure 21). The common dominator for the available areas is

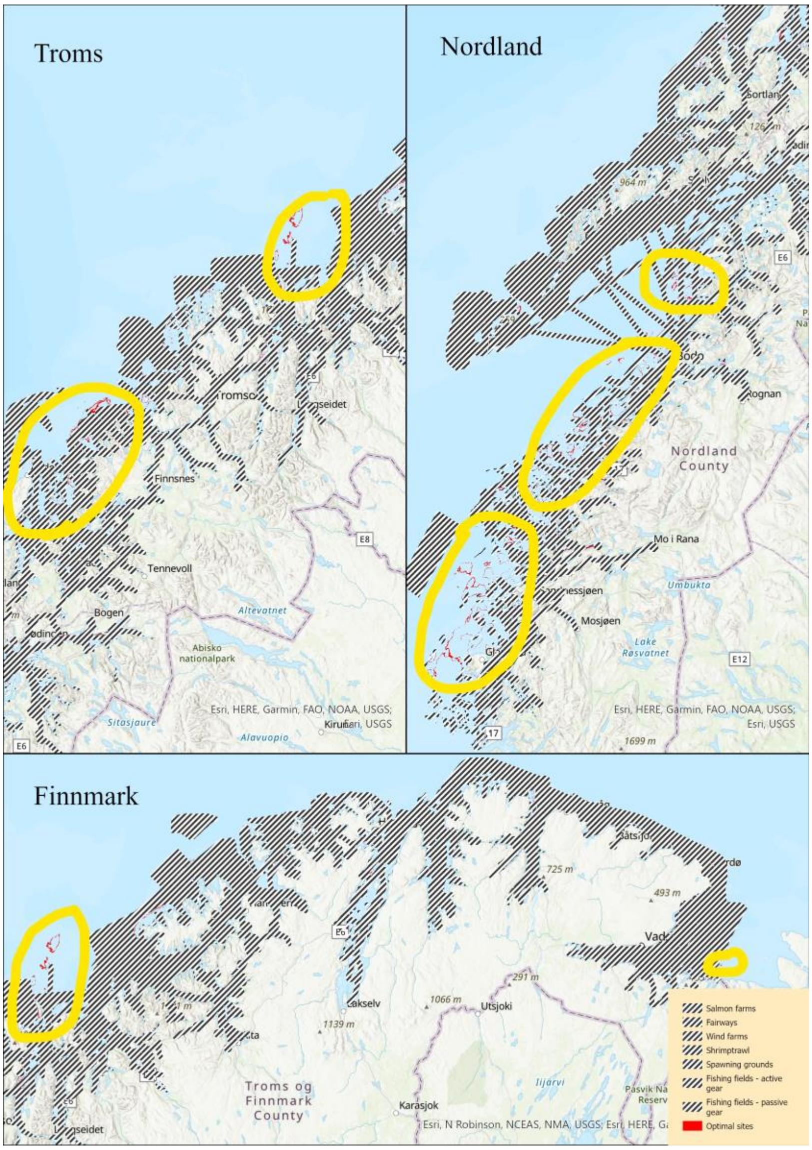


Figure 21: Map of study area showing the available space in accordance with the regulations and the predicted optimal areas.

that they are situated offshore. Some areas are available in sheltered areas and fjords located west of Lyngen, in Sørfjorden and Ofotfjorden in Nordland. To stress the area challenges, I have not included all the different uses of the coastal zone in this map. For example, salmon fjords, marine protected areas and the 4km buffer around production facilities are not included.

The fisheries are also a significant stakeholder in the coastal areas, where large regions marked as fishing fields or spawning grounds is unavailable for aquaculture. However, it should be said that the Directory of Fisheries does not have a de facto veto right to block an application for aquaculture space due to fisheries interest. This decision is made by the county governor, which in most cases agrees with DOF. It should be mentioned that the government is working on a new ERS regulation that enters into force on 1. July 2022 for all vessels below 15 m. One of the main objectives is to get more information regarding coastal fishing areas from the inshore fishing fleet. With this information, it is easier for DOF to accept an application for a farming site, which probably today conflicts with fishing areas.

The most limiting factor is the required distance between aquaculture farms and the distance from fairways. The reason attempts to assess effects on the spread of salmon lice, pancreatic diseases, and infectious salmon anaemia in production area 3 showed promising results. The analysis showed that a strategic relocation of biomass from the worst to the best sites could reduce the total infection between sites by 46% for salmon lice and 30% for viruses without decreasing the total production biomass (Huserbråten et al., 2020). These results strengthen the assumption that where local conditions allow for it, the distance between aquaculture locations can be reduced significantly without increasing the risk for infection pathogens, hence salmon mortality.

If efforts to relocate areas inside fjords do not work, there is also the possibility of moving farms offshore which we can see from the map (Figure 21) have a lot of space available. However, there are some bottlenecks. Morro et al. (2021) describe the different areas that need attention before offshore aquaculture can be a success. Firstly, the physical capabilities of the farmed fish species and infrastructure must be fully understood. Second, the oceanography of the different sites must be studied to confirm their capabilities. Thirdly, an economic plan considering operational costs and licensing limitations must be developed (Morro et al., 2021). RF predictive modelling can be used to identify suitable areas offshore.

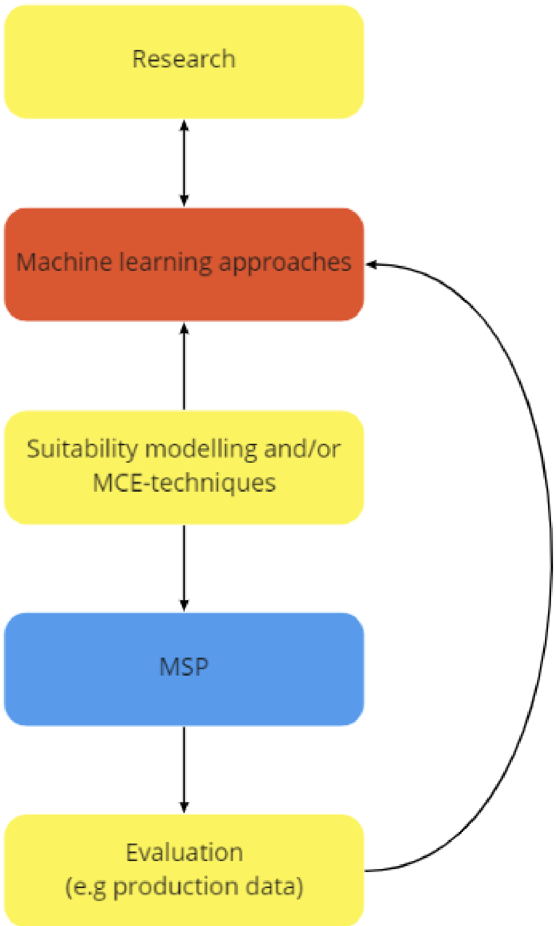
The results should be combined with MCE and include other factors, for example, economic cost, fish welfare, and new technology.

5.6 Implications for MSP

One challenge in Norway in coastal zone planning is that no standardized method exists for environmental impact assessment (Hersoug et al., 2019). The factors included in the impact assessment are often up to the planning authority to decide and not a defined minimum parameter based on research. Furthermore, there is a lack of knowledge about marine ecosystems, oceanography, and the user-environment effects. Lastly, many of the planning processes take a disproportionate amount of time. The result is a loss of initiative, community interest, and commitment erodes. The loss of initiative means that the municipalities, rather than planning for successful optimal farming. That can co-exist with other uses, instead has to allocate space via dispensation, or the farmers must wait pending a new coastal plan. One of the reasons that coastal planning is time-consuming is that the regulations and legislation are very complex. Some of the sectoral authorities express concern about possible double case management regarding aquaculture planning. Municipalities have, in some cases, submitted planning regulations concerning environmental conditions or other matters related to the operations of the salmon farm facility, which is the sector authority's area of responsibility (Hersoug et al., 2019). Blurred boundaries between rules and various legislations constitute a challenge for planning and development in the coastal areas. Therefore, it is an urgent need for clarification related to the planning authorities' area of responsibility and what is included in the sector administration's area of responsibility (Hersoug et al., 2019).

As mentioned in the theory section, MSP can be of great use for planners to optimize the locality structure in the municipality. Optimizing this can include the placement of salmon farming sites in areas with favourable conditions, such as the output map of this study. In addition, it would have been beneficial with a local model showing the transmission of pathogens. This can help reduce the distance between aquaculture sites and fairways, resulting in more available space. The output of this study can also be beneficial for farmers that want to apply for a salmon farming site that is not included in any coastal plan. Commercial companies use much time and financial resources on site-specific measurements of environmental conditions to find the "super localities". Using a machine learning approach to this problem could save much time and financial resources, as presented in this study.

A well-integrated plan gives the users of the coastal zone predictability and a particular long-term perspective to do investments and development. One of the significant problems in coastal planning is the knowledge gaps of the plan authorities. Machine learning approaches could be incorporated into a MSP process to help cover some of these knowledge gaps. As explained earlier, the standard methodology today in coastal planning is an impact- and risk assessment. This can be done in multiple ways, but a GIS overlay or multi-criteria analysis is often used. I suggest here an extension of the MSP process that Douvere (2008) presented in Marine Policy. Using the output of this thesis and incorporating other socioeconomic- and constraints variables in GIS-based MCE could be a better method for coastal zone planning for aquaculture locations. This extension can be used for different planning processes extending beyond aquaculture sitings. Figure 22 on the right explains this. The planning and mapping should always be based on the best available data and research. If site-specific data is available, it is the preferred data rather than the models used in this thesis. Since there is very difficult to see the linear relationships with the environmental data and optimal sites, a machine learning approach as suitability mapping is suggested. This can be done using an SDM approach with MaxEnt, RF, or other machine learning approaches. A spatial multi-criteria evaluation should be conducted in a GIS displaying all the different interests and stakeholders in the coastal zones. If the result of this analysis is incorporated into the MSP process, there is an excellent chance that salmon welfare could be increased. As in all MSP processes and machine learning techniques, evaluating the process and the results is significant for further success. Returning evaluation data, e.g., production data to the model, and additional training would increase the performance and accuracy over the years.



Figur 22: Flow-chart suggested MSP process. As an extension and enhancement of Douvere (2008).

Models such as the ones presented in this thesis can predict areas with favourable environmental conditions such that the maximum allowed biomass could be increased sustainably on sites with favourable conditions. Note that to estimate the holding capacity of a site, it is crucial to understand the physical, chemical, and biological processes that affect the water quality. Johansson et al. (2006) found that especially the strength of the pycnocline and the cage resistance, which affects the water flow, was significant factors. Therefor site, specific measurements should always be conducted at the identified optimal locations.

5.7 Limitations

There are many reasons to be critical of these results. The underlying assumption of this study is that the optimal site is explained by two variables: growth and mortality. When choosing the farming site, many other factors are valued from a farmer's perspective. The final decision is most likely economic. This means that it is beneficial to have farming sites near existing farming sites or close to infrastructure, such that the cost of transportation is not disproportional high. Since optimal sites per se are given solely from monthly specific growth- and mortality rate calculations – all salmon farms are treated equally. This is not true of farms in northern Norway. Different companies have different investments in the farms, e.g., lice skirts, lasers, feed types, net types, and several other factors. Moreover, they have different strategies such as feeding, delousing and movement of fish which can significantly affect growth and mortality. Optimally this model should have captured all these different factors. Further work on this should include companies, feeding strategies and different technology as exploratory variables. There are also several external factors that were not included in the analysis. The most important is the transmission of pathogens and the number of sea lice or other parasites in the net pen. This can dramatically reduce the fish welfare and, in most cases, be lethal.

The raw data from the Directory of Fisheries contained many outliers (very extreme values), several no data values for the number of fish added and the number of dead fish. This can be typos or that fish was added/removed to a location in the middle of a period. This could have caused the calculations of mortality and growth to be very high or adverse. The dataset was cleaned to the best of my ability, and the outliers were removed before analysis. However, errors could have been included in the analysis.

Since farmers move fish between sites, I experienced inconsistent time series throughout the years. The consequence is that it is not possible to follow the salmon from start to end. Many of the salmon in the study could have been to several locations, thus being exposed to environmental conditions in several areas without the model capturing this. Juvenile fish are more sensitive to ecological needs than adult salmon, and the quality of the juvenile fish is crucial for further growth. Therefore, it would have been helpful to know where the eggs and the juvenile fish were produced. Different companies use different providers, and many are now producing their own juvenile fish. Growth and mortality can be largely affected by this. In addition, the time of year juvenile fish is put into the sea can also largely affect the fish's performance. If put into the sea in the winter months, growth can be very low while mortality is high. All these scenarios are our model not able to capture. Further improvement should include producers and use size, age, and time of year as exploratory variables.

Because of the data material and the chosen methods, it was impossible to use the results from RF regression to predict new areas. The data material used in the regression was 6048 rows with overlapping data in time at each location, while the data used to predict was average values with 294 rows. Since it was impossible to obtain all the environmental data extracted as points in the regression for prediction, some data was left out. The identified important variables using the RF regression model are the most reliable. The predicted raster surface does not include all the variables used in regression and should therefore be interpreted with this in mind. Since the optimal sites are explained by growth and mortality, one could argue that the model in this study presents areas that favour fish health and is not directly the most "optimal area". The RF algorithm would, either with regression or classification, identify variables and predict areas based on the exploratory variables in the input model. The input environmental exploratory variables are based on state-of-the-art models and are as close as we get to actual data. However, we are still working with models. An interesting observation for the regression versus the classification was that currently did not seem to contribute to the regression model, but the classification included current as important. This is a strange observation since the literature suggested that current and change of water at the farm site is one important contributor to good fish health. The data extraction from NorKyst-800 can probably explain this phenomenon. Because of the massive data material, a shortcut was used such that current is based on the daily average current. The best would have been to have the current based on hourly values to include the tidal movement. This means that the current is only an estimate of the mean current at each locality and will not include that the tide causes the water to go back and forth with significant strength. Inside fjords, the tidal movement is primarily responsible for the current. Efforts were made to get actual data from all aquaculture companies in northern Norway, but it was impossible to obtain data from all locations in the requested periods. Data request was current, oxygen, salinity, and temperature. The level of data each company had varied, and some only collected temperature. However, data collected was exchanged with the model data and used further in the analysis. I managed to collect oxygen data from 6 different locations, but it was not enough to include in the study.

The prediction of optimal sites was based on the classification of "presence and absence" points done with the help of bivariate colours and natural breaks. There were significant differences in performance from month to month on many of the aquaculture sites included in the analysis. If some of the locations are hit by acute mortality or illnesses in the analysis period, such as the algae bloom in 2019 in south-Troms, they would probably not be included as presence points because mortality was above the threshold. Predictions done with SDM-modelling depend on the presence and absence points, and the exploratory raster's provided. Since 30 m depth was used as an exploratory raster, some locations were not included in the model because those were in areas where the depth was < 30m. Further work should use 20m as maximal depth for the vertical layers.

The selection method of presence and absence points gave an imbalanced dataset, which was compensated for, and all trees were given a subset of presence and absence data. However, because of the considerable variation in aquaculture performance, we cannot be entirely sure that the areas that are predicted as optimal are the best performing locations in northern Norway. A more solid statistical analysis should assess performance over time rather than the average values. A model result will never be better than the data the model is based on. Although this study's RF prediction and modelling have produced results that can be visualized on maps, I would be careful using these results as anything more than examples of how machine learning and RF can be used to model optimal sites for salmon farming.

6 Conclusion

This thesis aimed to identify the significant variables and predict optimal areas for salmon farming in northern Norway based on fine-scale production data from the industry and state of the art environmental data using the data-driven machine learning approach, RF regression and classification.

The important factors affecting salmon growth were specific feeding rate, temperature, and biomass. Other important factors were solar radiation, salinity, wind speed, bathymetry, slope, distance to aquaculture sites, fairways, and urban settlements. In addition, wave height and current showed some importance. The model gave strong results with an R squared value of 0.968 in training and 0.819 in validation. The important factors influencing salmon mortality were temperature and biomass. Length of the day, wind speed, wave height, salinity and current were also driving factors. Interestingly non of the socio-economic factors did not influence salmon mortality.

Prediction of optimal sites was performed quarterly. A final map of optimal areas for salmon farming was produced by multiplying the four different quarterly results. Optimal locations are dispersed, ranging from inside fjords to the archipelago and open waters and present in the whole study area. The predicted optimal sites were inside Vefsnfjorden, Ranfjorden, Sørfjorden and Glomfjorden, and small areas near the coast and around the small islands stretching from Gladstad to Narvik. Areas near the coast of Lofoten, Værøy, Røst, Vesterålen, Sortland and Senja. Further north, some dispersed areas were predicted as optimal outside Tromsø and Sørøya. Large areas around Varangerhalvøya, and fjords such as Olderdalen/Kåfjorden, Lille Altafjorden/Burfjorden, and near the shore on both sides of Stjernøysundet.

When plotting identified optimal- and available coastal areas on the same map, it is evident that coastal areas are a scares resource. There is a need to rethink the regulations and legislation, especially minimum distances to aquaculture locations and fairways. Regarding the implications to MSP or coastal zone planning, I have suggested incorporating machine learning approaches in GIS-based MCE analysis to help planners and decision-makers make informed and sustainable decisions about sea-area use.

7 References

Albretsen, J., Sperrevik, A., Staalstrøm, A., Sandvik, A., Vikebø, F., & Asplin, L. (2011). *NorKyst-800 Rapport nr. 1: User Manual and technical descriptions.* Retrieved from Tromsø: Institute of Marine Research.

Alkema, D. (2007). *Simulating floods: on the application of a 2D-hydraulic model for flood hazard and risk assessment*: Utrecht University.

Alkema, D., Boerboom, L., Ferlisi, S., & Cascini, L. (2014). 6.4 Spatial Multi-Criteria Evaluation. *Caribbean Handbook on Risk Information Management, Edited by ACP–EU Natural Disaster Risk Reduction Program*.

Andreassen, O., Johnsen, J. P., & Hersoug, B. (2010). The Aquaculture industry's minimal area use is weakly growing and strongly changing. *Norsk Fiskeoppdrett* (35), 48-51.

Aquaculture Act. (2006). *Lov om akvakultur* Retrieved from https://lovdata.no/dokument/NL/lov/2005-06-17-79

Asplin, L., Albretsen, J., Johnsen, I. A., & Sandvik, A. D. (2020). The hydrodynamic foundation for salmon lice dispersion modelling along the Norwegian coast. *Ocean Dynamics, 70*(8), 1151-1167.

Bailey, J. L., & Eggereide, S. S. (2020). Indicating sustainable salmon farming: The case of the new Norwegian aquaculture management scheme. *Marine Policy, 117*, 103925. doi:https://doi.org/10.1016/j.marpol.2020.103925

Bateman, B., VanDerWal, J., & Johnson, C. (2012). Nice weather for bettongs: using weather events, not climate means, in species distribution models. *Ecography, 35*(4), 306-314.

Berg, P., Solås, A.-M., Kvalvik, I., & Hersoug, B. (2017). *Hvordan planlegges kystsonen?* (1890-579X). Retrieved from Tromsø: https://www.regjeringen.no/contentassets/adc94a1dbb034519983b5090334beea0/nofima_planlegging_kystsonen.pdf

Bjørnshol, B., & Khan, s. (2020). *Hvordan påvirkes risikoen for smittespredning av de nye kravene til behandling av transportvannet fra brønnbåter?* : The Food Security Authority Retrieved from: https://www.mattilsynet.no/fisk_og_akvakultur/akvakultur/bronnbat/bestilling_fra_mattilsynet_til_veterinaerinstituttet__avstander_og_smittespredning_fra_bronnbaater_200520.39446/binary/Bestilling%20fra%20Mattilsynet%20til%20Veterin%C3%A6rinstituttet%20-%20Avstander%20og%20smittespredning%20fra%20br%C3%B8nnb%C3%A5ter%2020.05.20

Bleie, H., & Skrudland, A. (2014). Tap av Laksefisk i Sjø. *Rapport fra Mattilsynet*. Retrieved from https://www.mattilsynet.no/fisk_og_akvakultur/ fiskevelferd/tap_av_laksefisk_i_sjo_rapport.15430/binary/Tap%20 av%20laksefisk%20i%20sj%C3%B8%20rapport

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. (1987). Occam's razor. *Information Processing Letters, 24*(6), 377-380.

Bodirsky, B. L., Rolinski, S., Biewald, A., Weindl, I., Popp, A., & Lotze-Campen, H. (2015). Global Food Demand Scenarios for the 21st Century. *PLOS One, 10*(11), e0139201. doi:10.1371/journal.pone.0139201

Borch, T. (2009). Contested coastal commercialization: marine fishing tourism in Norway. *Maritime Studies, 8*(1), 33-51.

Bowden, T., Smail, D., & Ellis, A. (2002). Development of a reproducible infectious pancreatic necrosis virus challenge model for Atlantic salmon, Salmo salar L. *Journal of Fish Diseases, 25*(9), 555-563.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.
Brian, D. (2021). Occam's razor. In. Encyclopedia Britannica.
Bricknell, I. R., Dalesman, S. J., O'Shea, B., Pert, C. C., & Luntz, A. J. M. (2006). Effect of environmental salinity on sea lice Lepeophtheirus salmonis settlement success. *Diseases of Aquatic Organisms, 71*(3), 201-212.
Bryde, B., Krause, G., & Rosenthal, H. (2011). Marine Spatial Planning - A Norwegian Perspective. *Presentation at the Conference on Competetive and Sustainable Aquaculture*. Helsinki.
Carvalho, L. A., Whyte, S. K., Braden, L. M., Purcell, S. L., Manning, A. J., Muckle, A., & Fast, M. D. (2020). Impact of co‐infection with Lepeophtheirus salmonis and Moritella viscosa on inflammatory and immune responses of Atlantic salmon (Salmo salar). *Journal of Fish Diseases, 43*(4), 459-473.
Crane, D., Ogle, D., & Shoup, D. (2020). Use and misuse of a common growth metric: guidance for appropriately calculating and reporting specific growth rate. *Reviews in Aquaculture, 12*(3), 1542-1547.
Cullen, D., & Guida, V. (2021). Use of geographically weighted regression to investigate spatial non-stationary environmental effects on the distributions of black sea bass (Centropristis striata) and scup (Stenotomus chrysops) in the Mid-Atlantic Bight, USA. *Fisheries Research, 234*, 105795. doi:https://doi.org/10.1016/j.fishres.2020.105795
Cutler, R., Edwards, T., Beard, K., Cutler, A., Hess, K., Gibson, J., & Lawler, J. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783-2792.
Dalsøren, S., Albretsen, J., & Asplin, L. (2020). New validation method for hydrodynamic fjord models applied in the Hardangerfjord, Norway. *Estuarine, Coastal and Shelf Science, 246*, 107028.
Dang, N. M., Babel, M. S., & Luong, H. T. (2011). Evaluation of food risk parameters in the day river flood diversion area, Red River delta, Vietnam. *Natural Hazards, 56*(1), 169-194.
Dapueto, G., Massa, F., Costa, S., Cimoli, L., Olivari, E., Chiantore, M., . . . Povero, P. (2015). A spatial multi-criteria evaluation for site selection of offshore marine fish farm in the Ligurian Sea, Italy. *Ocean & Coastal Management, 116*, 64-77.
Dawkins, M. (1990). From an animal's point of view: Motivation, fitness, and animal welfare. *Behavioural and Brain Sciences, 13*, 1-&. doi:10.1017/S0140525X00077104
DEFRA. (2014). *East Inshore and East Offshore Marine Plans*. London
Degenhardt, F., Seifert, S., & Szymczak, S. (2017). Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics, 20*(2), 492-503. doi:10.1093/bib/bbx124
DeLucas, L. J., Suddath, F., Snyder, R., Naumann, R., Broom, M. B., Pusey, M., . . . Nelson, B. (1986). Preliminary investigations of protein crystal growth using the space shuttle. *Journal of Crystal Growth, 76*(3), 681-693.
Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. In. BMC Bioinformatics 7.
Dietrich, S., Floegel, A., Weikert, C., Prehn, C., Adamski, J., Pischon, T., . . . Drogan, D. (2016). Identification of serum metabolites associated with incident hypertension in the European Prospective Investigation into Cancer and Nutrition–Potsdam Study. *Hypertension, 68*(2), 471-477.
Douvere, F. (2008). The importance of marine spatial planning in advancing ecosystem-based sea use management. *Marine Policy, 32*(5), 762-771.

Edwards, A., & Edelsten, D. (1977). Deep water renewal of Loch Etive: A three basin Scottish fjord. *Estuarine and Coastal Marine Science, 5*(5), 575-595. doi:https://doi.org/10.1016/0302-3524(77)90085-8

Eichler, S., Kirischuk, S., Juettner, R., Schaefermeier, P., Legendre, P., Lehmann, T., . . . Meier, J. (2012). Glycinergic tonic inhibition of hippocampal neurons with depolarizing GABAergic transmission elicits histopathological signs of temporal lobe epilepsy. *Journal of Cellular and Molecular Medicine, 16*(4), 959-959.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., . . . Lehmann, A. (2006). Novel methods improve prediction of species distributions from occurrence data. *Ecography, 29*, 129-151.

Elliot, J. (1991). Tolerance and resistance to thermal stress in juvenile Atlantic salmon, Salmo salar. *Freshwater Biology, 25*(1), 61-70. doi:https://doi.org/10.1111/j.1365-2427.1991.tb00473.x

Elliott, J., & Elliott, J. (2010). Temperature requirements of Atlantic salmon Salmo salar, brown trout Salmo trutta and Arctic charr Salvelinus alpinus: predicting the effects of climate change. *Journal of Fish Biology, 77*(8), 1793-1817.

Elliott, J. M., & Hurley, M. A. (1995). The functional relationship between body size and growth rate in fish. *Functional Ecology*, 625-627.

Ellis, T., North, B., Scott, A., Bromage, N., Porter, M., & Gadd, D. (2002). The relationships between stocking density and welfare in farmed rainbow trout. *Journal of Fish Biology, 61*(3), 493-531.

Emerman, J. D. (2016). *Establishing the optimal salinity for rearing salmon in recirculating aquaculture systems.* University of British Columbia,

Evans, J., Murphy, M., Holden, Z., & Cushman, S. (2011). Modelling Species Distribution and Change Using Random Forest. *Predictive Species and Habitat Modeling in Landscape Ecology, 139-159, Springer, New York, NY. doi: https://doi.org/10.1007/978-1-4419-7390-0_8*

Falconer, L., Hjøllo, S. S., Telfer, T. C., McAdam, B. J., Hermansen, Ø., & Ytteborg, E. (2020). The importance of calibrating climate change projections to local conditions at aquaculture sites. *Aquaculture, 514*, 734487.

Falconer, L., Telfer, T. C., & Ross, L. G. (2016). Investigation of a novel approach for aquaculture site selection. *Journal of Environmental Management, 181*, 791-804. doi:https://doi.org/10.1016/j.jenvman.2016.07.018

Fast, M. D., Hosoya, S., Johnson, S. C., & Afonso, L. O. (2008). Cortisol response and immune-related effects of Atlantic salmon (Salmo salar Linnaeus) subjected to short- and long-term stress. *Fish & Shellfish Immunology, 24*(2), 194-204.

Filgueira, R., Guyondet, T., Comeau, L., & Grant, J. (2014). A fully-spatial ecosystem-DEB model of oyster (Crassostrea virginica) carrying capacity in the Richibucto Estuary, Eastern Canada. *Journal of Marine Systems, 136*, 42-54.

Fisher, H., Collins, S., Hanson, C., Mason, B., Colombo, S., & Anderson, D. (2020). Black soldier fly larvae meal as a protein source in low fish meal diets for Atlantic salmon (Salmo salar). *Aquaculture, 521*, 734978.

Fisher, W. L., Fisher, W. L., & Rahel, F. J. (2004). *Geographic information systems in fisheries*: American Fisheries Society.

Folkedal, O., Torgersen, T., Nilsson, J., & Oppedal, F. (2010). Habituation rate and capacity of Atlantic salmon (Salmo salar) parr to sudden transitions from darkness to light. *Aquaculture, 307*(1-2), 170-172.

Fu, P., & Rich, P. M. (1999). *Design and implementation of the Solar Analyst: an ArcView extension for modelling solar radiation at landscape scales.* Paper presented at the Proceedings of the nineteenth annual ESRI user conference.

Fu, P., & Rich, P. M. (2002). A geometric solar radiation model with applications in agriculture and forestry. *Computers and Electronics in Agriculture, 37*(1-3), 25-35.
Furevik, D. M., Bjordal, Å., Huse, I., & Fernö, A. (1993). Surface activity of Atlantic salmon (Salmo salar L.) in net pens. *Aquaculture, 110*(2), 119-128.
Gamperl, A. K., Zrini, Z. A., & Sandrelli, R. M. (2021). Atlantic Salmon (Salmo salar) Cage-Site Distribution, Behavior, and Physiology During a Newfoundland Heat Wave. *Frontiers in Physiology, 12*(1269). doi:10.3389/fphys.2021.719594
Gimpel, A., Stelzenmüller, V., Töpsch, S., Galparsoro, I., Gubbins, M., Miller, D., . . . Watret, R. (2018). A GIS-based tool for an integrated assessment of spatial planning trade-offs with aquaculture. *Science of The Total Environment, 627*, 1644-1655. doi:https://doi.org/10.1016/j.scitotenv.2018.01.133
Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology, 84*(11), 2809-2815.
Grömping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *The American Statistician, 63*(4), 308-319. doi:10.1198/tast.2009.08199
Gullestad, P., Bjørgo, S., Eithun, I., Ervik, A., Gudding, R., Hansen, H., . . . Røsvik, I. (2011). Effektiv og bærekraftig arealbruk i havbruksnæringen–areal til begjær. *Rapport fra et ekspertutvalg oppnevnt av Fiskeri-og kystdepartementet. ISBN*, 978-982.
Habermann, J. K., Doering, J., Hautaniemi, S., Roblick, U. J., Bündgen, N. K., Nicorici, D., . . . Ma, Y. (2009). The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. *International Journal of Cancer, 124*(7), 1552-1564.
Haidvogel, D. B., Arango, H., Budgell, W. P., Cornuelle, B. D., Curchitser, E., Di Lorenzo, E., . . . Lanerolle, L. (2008). Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. *Journal of Computational Physics, 227*(7), 3595-3624.
Handeland, S., Berge, Å., Björnsson, B. T., & Stefansson, S. (1998). Effects of temperature and salinity on osmoregulation and growth of Atlantic salmon (Salmo salar L.) smolts in seawater. *Aquaculture, 168*(1-4), 289-302.
Hargrave, B., Duplisea, D., Pfeiffer, E., & Wildish, D. (1993). Seasonal changes in benthic fluxes of dissolved oxygen and ammonium associated with marine cultured Atlantic salmon. *Marine Ecology Progress Series*, 249-257.
Heidenreich, S. (2016). Out of sight, out of mind? Controversy over offshore wind energy in Norway's news media. *Science as Culture, 25*(4), 449-472.
Hersoug, B. (2015). The greening of Norwegian salmon production. *Maritime Studies, 14*(1), 16. doi:10.1186/s40152-015-0034-9
Hersoug, B., Andreassen, O., Johnsen, J. P., & Robertsen, R. (2014). *Hva begrenser tilgangen på sjøareal til havbruksnæringen?* Retrieved from Tromsø: http://hdl.handle.net/11250/283227
Hersoug, B., Armstrong, C., Brattland, C., Eythórsson, E., Holmgaard, S., Johnsen, J.-P., . . . Solås, A.-M. (2019). Når det blåser i fra ØST. *Om bruken av økosystemtjeneste-perspektivet i kystsoneplanleggingen. Nofima Rapport*. Retrived from: https://nofimaas.sharepoint.com/sites/public/Cristin/Rapport%2002-2019.pdf?ga=1
Hersoug, B., & Johnsen, J. P. (2012). Kampen om plass på kysten. *Interesser og utviklingstrekk i kystsoneplanleggingen. Oslo: Universitetsforlaget*.
Hersoug, B., Mikkelsen, E., & Osmundsen, T. C. (2021). What's the clue; better planning, new technology or just more money? - The area challenge in Norwegian salmon farming. *Ocean & Coastal Management, 199*, 105415. doi:https://doi.org/10.1016/j.ocecoaman.2020.105415

Hersoug, B., Mikkelsen, E., Robertsen, R., & Osmundsen, T. (2020). Arealforvaltning av norsk havbruksnæring. Retrieved from https://nofimaas.sharepoint.com/sites/public/Cristin/Rapport%2009-2020.pdf
Hjeltnes, B., Bang, J., Bornø, G., Haukaas, A., & Walde, C. (2019). Fish health report 2018 [in Norwegian: Fiskehelserapporten].
Hommedal, S., Lorentzen, E., & Hoddevik, B. (2019). Skadelig oppblomstring av Chrysochromulina leadbeateri har forårsaket laksedød i Nordland og Troms. Algen er vanlig langs norskekysten. Retrieved from https://www.hi.no/hi/nyheter/2019/mai/dette-vet-vi-om-den-sakalte-dodsalgen-i-nord-norge
Houde, E. D., & Schekter, R. C. (1980). Feeding by marine fish larvae: developmental and functional responses. *Environmental Biology of Fishes, 5*(4), 315-334.
Huse, I., & Holm, J. (1993). Vertical distribution of Atlantic salmon (Salmo salar) as a function of illumination. *Journal of Fish Biology, 43*, 147-156.
Huserbråten, M., Ådlandsvik, B., Bergh, Ø., Grove, S., Karlsen, Ø., Taranger, G. L., . . . Johnsen, I. A. (2020). Endret lokalitetsstruktur i produksjonsområde 3 — vurdert virkning på spredning av lakselus, pankreassykdom og infektiøs lakseanemi. *Rapport fra havforskningen 2020-12*. Retrived from: https://www.hi.no/hi/nettrapporter/rapport-fra-havforskningen-2020-12
Hvas, M., Folkedal, O., & Oppedal, F. (2021). What is the limit of sustained swimming in Atlantic salmon post smolts? *Aquaculture Environment Interactions, 13*, 189-198.
Jenora, D., Alberto, N., & Flora, V. (2019). Beyond where: Modeling Spatial Relationships and Making Predictions. Retrieved from: https://www.esri.com/videos/watch?videoid=ob18SuuAJGI&title=beyond-where-modeling-spatial-relationships-and-making-predictions
Jensen, B., Qviller, L., & Toft, N. (2020). Spatio-temporal variations in mortality during the seawater production phase of Atlantic salmon (Salmo salar) in Norway. *Journal of Fish Diseases, 43*(4), 445-457. doi:https://doi.org/10.1111/jfd.13142
Jentoft, S., & Knol, M. (2014). Marine spatial planning: risk or opportunity for fisheries in the North Sea? *Maritime Studies, 12*(1), 1-16. Retrieved from: https://www.researchgate.net/publication/259828090_Marine_spatial_planning_risk_or_opportunity_for_fisheries_in_the_North_Sea
Jeppesen, E., Søndergaard, M., Pedersen, A. R., Jürgens, K., Strzelczak, A., Lauridsen, T. L., & Johansson, L. S. (2007). Salinity induced regime shift in shallow brackish lagoons. *Ecosystems, 10*(1), 48-58.
Jobling, M. (1983). Growth studies with fish—overcoming the problems of size variation. *Journal of Fish Biology, 22*(2), 153-157.
Jobling, M., & Jørgensen, E. H. (1992). Feeding behaviour and effect of feeding regime on growth of Atlantic salmon, Salmo salar. *Aquaculture, 101*, 135-146. doi:10.1016/0044-8486(92)90238-G
Johansen, L., Jensen, I., Mikkelsen, H., Bjørn, P., Jansen, P., & Bergh, Ø. (2011). Disease interaction and pathogens exchange between wild and farmed fish populations with special reference to Norway. *Aquaculture, 315*(3-4), 167-186. doi:https://doi.org/10.1016/j.aquaculture.2011.02.014
Johansson, D., Juell, J., Oppedal, F., Stiansen, J., & Ruohonen, K. (2007). The influence of the pycnocline and cage resistance on current flow, oxygen flux and swimming behaviour of Atlantic salmon (Salmo salar L.) in production cages. *Aquaculture, 265*(1-4), 271-287. doi:https://doi.org/10.1016/j.aquaculture.2006.12.047

Juell, J., & Fosseidengen, J. (2004). Use of artificial light to control swimming depth and fish density of Atlantic salmon (Salmo salar) in production cages. *Aquaculture, 233*(1-4), 269-282. doi:https://doi.org/10.1016/j.aquaculture.2003.10.026

Kadri, S., Thorpe, J. E., & Metcalfe, N. B. (1997). Anorexia in one-sea-winter Atlantic salmon (Salmo solar) during summer, associated with sexual maturation. *Aquaculture, 151*(1), 405-409. doi:https://doi.org/10.1016/S0044-8486(96)01495-0

Kapetsky, J. M., Aguilar, J., & Jenness, J. (2013). *A global assessment of offshore mariculture potential from a spatial perspective*. (2070-7010). Retrieved from https://www.fao.org/3/i3100e/i3100e.pdf

Kazakov, R., & Khalyapina, L. (1981). Oxygen consumption of adult Atlantic salmon (Salmo salar L.) males and females in fish culture. *Aquaculture, 25*(2-3), 289-292.

Kindschi, G. A., & Koby Jr, R. F. (1994). Performance and oxygen consumption of Snake River cutthroat trout reared at four densities with supplemental oxygen. *The Progressive Fish‐Culturist, 56*(1), 13-18.

Korsøen, Ø., Dempster, T., Fjelldal, P., Oppedal, F., & Kristiansen, T. (2009). Long-term culture of Atlantic salmon (Salmo salar L.) in submerged cages during winter affects behaviour, growth and condition. *Aquaculture, 296*, 373-381. doi:10.1016/j.aquaculture.2009.08.036

Kristoffersen, A. B., Jimenez, D., Viljugrein, H., Grøntvedt, R., Stien, A., & Jansen, P. A. (2014). Large scale modelling of salmon lice (Lepeophtheirus salmonis) infection pressure based on lice monitoring data from Norwegian salmonid farms. *Epidemics, 9*, 31-39.

Longva, O., & Elvenes, S. (2016). *Modellering av optimale oppdrettslokaliteter* (2016-005). Retrieved from Trondheim: https://www.ngu.no/upload/Publikasjoner/Rapporter/2016/2016_005.pdf

Luc, A. (2016). Spatial Autocorrelation. *Spatial Analysis*: GeoDA.

Lugert, V., Thaller, G., Tetens, J., Schulz, C., & Krieter, J. (2016). A review on fish growth calculation: multiple functions in fish production and their specific application. *Reviews in Aquaculture, 8*(1), 30-42.

Lundqvist, H. (1980). Influence of photoperiod on growth in Baltic salmon parr (Salmo salar) with special reference to the effect of precocious sexual maturation. *Canadian Journal of Zoology, 58*(5), 940-944.

Lymbery, P. (2002). *In too deep: the welfare of intensively farmed fish*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.180.6550&rep=rep1&type=pdf

Maceina, M. (1992). A simple regression model to assess environmental effects on fish growth. *Journal of Fish Biology, 41*(4), 557-565.

Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., . . . Vanden Abeele, P. (2005). GAUFRE: towards a spatial structure plan for the sustainable management of the Belgian part of the North Sea. *Science and Sustainable Management of The North Sea* 177-249. Retrieved from http://hdl.handle.net/1854/LU-374922

Malczewski, J. (2006). GIS‐based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science, 20*(7), 703-726.

Malde, K., Handegard, N. O., Eikvil, L., & Salberg, A.-B. (2019). Machine intelligence and the data-driven future of marine science. *ICES Journal of Marine Science, 77*(4), 1274-1285. doi:10.1093/icesjms/fsz057

McKenzie, D. J., Palstra, A. P., Planas, J., MacKenzie, S., Bégout, M. L., Thorarensen, H., . . . De Boeck, G. (2021). Aerobic swimming in intensive finfish aquaculture:

applications for production, mitigation and selection. *Reviews in Aquaculture, 13*(1), 138-155.
McLoughlin, M., & Graham, D. (2007). Alphavirus infections in salmonids–a review. *Journal of Fish Diseases, 30*(9), 511-531.
Meeus, J. (1991). *Astronomical Algorithms*. Richmond: Willmann-Bell Inc.
Mi, C., Huettmann, F., Guo, Y., Han, X., & Wen, L. (2017). Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ, 5*(6), doi:10.7717/peerj.2849
Miller, J. (2010). Species Distribution Modeling. *Geography Compass, 4*(6), 490-509. doi:https://doi.org/10.1111/j.1749-8198.2010.00351.x
Miller, K., Huettmann, F., Norcross, B., & Lorenz, M. (2014). Multivariate random forest models of estuarine-associated fish and invertebrate communities. *Marine Ecology Progress Series, 500*, 159-174. doi:10.3354/meps10659
Ministry of Trade Industry and Fisheries. (2014-15). *Forutsigbar og miljømessig bærekrafitg vekst i norske lakse og ørretoppdrett*. (Meld. St. 16 (2014-2015)).
Ministry of Trade Industry and Fisheries. (2021). *Et hav av muligheter*. Retrieved from https://www.regjeringen.no/no/dokumenter/havbruksstrategien-et-hav-av-muligheter/id2864482/
Moriarty, M., Murray, A. G., B.Berx, Christie, A. J., Munro, L. A., & Wallace, I. S. (2020). Modelling temperature and fish biomass data to predict annual Scottish farmed salmon, salmo salar L., losses: Development of an early warning too. . *Preventive Veterinary Medicine, 178*. Retrieved from https://doi.org/10.1016/j.prevetmed.2020.104985
Morissette, J., Swart, S., MacCormack, T. J., Currie, S., & Morash, A. J. (2020). Thermal variation near the thermal optimum does not affect the growth, metabolism or swimming performance in wild Atlantic salmon (Salmo salar). *Journal of Fish Biology, 8*(6), 1585-1589. doi:https://doi.org/10.1111/jfb.14348
Morro, B., Davidson, K., Adams, T., Falconer, L., Holloway, M., Dale, A., . . . Rey Planellas, S. (2021). Offshore Aquaculture of Finfish: Big expectations at sea. *Reviews in Aquaculture, 14*(2), 791-815. doi:https://doi.org/10.1111/raq.12625
Müller, M., Homleid, M., Ivarsson, K.-I., Køltzow, M. A., Lindskog, M., Midtbø, K. H., . . . Bjørge, D. (2017). AROME-MetCoOp: A Nordic convective-scale operational weather prediction model. *Weather and Forecasting, 32*(2), 609-627.
Ndong, D., Chen, Y.-Y., Lin, Y.-H., Vaseeharan, B., & Chen, J.-C. (2007). The immune response of tilapia Oreochromis mossambicus and its susceptibility to Streptococcus iniae under stress in low and high temperatures. *Fish & Shellfish Immunology, 22*(6), 686-694.
Nilsen, A., Nielsen, K. V., & Bergheim, A. (2020). A closer look at closed cages: Growth and mortality rates during production of post-smolt Atlantic salmon in marine closed confinement systems. *Aquacultural Engineering, 91*, 102124. doi:https://doi.org/10.1016/j.aquaeng.2020.102124
Noble, C., Gismervik, K., Iversen, M. H., Kolarevic, J., Nilsson, J., Stien, L. H., & Turnbull, J. F. (2018). Welfare Indicators for farmed Atlantic salmon: tools for assessing fish welfare. In: Nofima. Retrieved from Tromsø: http://hdl.handle.net/11250/2575780
Norwegian Directorate of Fisheries. (2021). Biomassestatistikk. Retrieved from: https://www.fiskeridir.no/Akvakultur/Tall-og-analyse/Biomassestatistikk
NOU 2004: 28. (2004). *Lov om bevaring av natur, landskap og biologisk mangfold—(Naturmangfoldloven)*. Retrieved from https://www.regjeringen.no/no/dokumenter/nou-2004-28/id388846/?ch=1

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012). Modelling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management, 275*, 117-129. doi:https://doi.org/10.1016/j.foreco.2012.03.003

Oliveira, V., Dean, K., Qviller, L., Kirkeby, C., & Bang, B. (2021). Factors associated with baseline mortality in Norwegian Atlantic salmon farming. *Scientific Reports, 11*(1), 14702. doi:10.1038/s41598-021-93874-6

Oppedal, F., Dempster, T., & Stien, L. (2011). Environmental drivers of Atlantic salmon behaviour in sea-cages: A review. *Aquaculture, 311*(1), 1-18. doi:https://doi.org/10.1016/j.aquaculture.2010.11.020

Oppedal, F., Juell, J.-E., & Johansson, D. (2007). Thermo- and photoregulatory swimming behaviour of caged Atlantic salmon: Implications for photoperiod management and fish welfare. *Aquaculture, 265*(1), 70-81. doi:https://doi.org/10.1016/j.aquaculture.2007.01.050

Oppedal, F., Juell, J. E., Tarranger, G., & Hansen, T. (2001). Artificial light and season affects vertical distribution and swimming behaviour of post‐smolt Atlantic salmon in sea cages. *Journal of Fish Biology, 58*(6), 1570-1584.

Osmundsen, T. C., Olsen, M. S., & Thorvaldsen, T. (2020). The making of a louse-Constructing governmental technology for sustainable aquaculture. *Environmental Science & Policy, 104*, 121-128.

Outdoor Recration Act. (1957). *Lov om friluftslivet*. Retrieved from https://lovdata.no/dokument/NL/lov/1957-06-28-16

Perez, O. M., Telfer, T. C., & Ross, L. G. (2005). Geographical information systems‐based models for offshore floating marine fish cage aquaculture site selection in Tenerife, Canary Islands. *Aquaculture Research, 36*(10), 946-961.

Pérez-Casanova, J., Rise, M., Dixon, B., Afonso, L., Hall, J., Johnson, S., & Gamperl, A. (2008). The immune and stress responses of Atlantic cod to long-term increases in water temperature. *Fish & Shellfish Immunology, 24*(5), 600-609.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling, 190*(3-4), 231-259.

Phillips, S. J., & Dudík, M. (2008). Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography, 31*(2), 161-175.

Planning and Construction Case Processing Act. (2008). *Lov om planlegging og byggesaksbehandling*. Retrieved from: https://lovdata.no/dokument/NL/lov/2008-06-27-71

Plantalech Manel‐La, N., Thorstad, E., Davidsen, J., Økland, F., Sivertsgård, R., McKinley, R., & Finstad, B. (2009). Vertical movements of Atlantic salmon post‐smolts relative to measures of salinity and water temperature during the first phase of the marine migration. *Fisheries Management and Ecology, 16*(2), 147-154.

Ports and Waterways Safety Act. (2019). *Lov om havner og farvann*. Retrieved from: https://lovdata.no/dokument/NL/lov/2019-06-21-70

Quist, M., & Lsermann, D. (2017). Age and Growth of Fishes: Principles and Techniques. *American Fisheries Society*. doi: https://doi.org/10.47886/9781934874486

Rashed, T., & Weeks, J. (2003). Assessing vulnerability to earthquake hazards through spatial multicriteria analysis of urban areas. *International Journal of Geographical Information Science, 17*(6), 547-576.

Regulations on The Establishment and Expansion of Aquaculture Facilities etc. (2008). *Forskrift om utvidelse av akvakulturanlegg mv*. Retrieved from: https://lovdata.no/dokument/SF/forskrift/2008-06-17-823

Remen, M., Solstorm, F., Bui, S., Klebert, P., Vågseth, T., Solstorm, D., . . . Oppedal, F. (2016). Critical swimming speed in groups of Atlantic salmon Salmo salar. *Aquaculture Environment Interactions, 8*, 659-664.
Rich, P., Dubayah, R., Hetrick, W., & Saving, S. (1994). *Using viewshed models to calculate intercepted solar radiation: applications in ecology. American Society for Photogrammetry and Remote Sensing Technical Papers.* Paper presented at the American Society of Photogrammetry and Remote Sensing.
Ricker, W. E. (1975). Computation and interpretation of biological statistics of fish populations. *Bulletin of the Fisheries Research Board of Canada, 191*, 1-382.
Robertsen, R., Mikkelsen, E. I., Karlsen, K. M., Solås, A.-M., Hersoug, B., Tveterås, R., . . . Sørgård, B. (2020). *Havbruksforvaltning mot 2030–faglig sluttrapport*. (Nofima Report 46/2020). Retrieved from https://hdl.handle.net/11250/2724971
Rånes, S. (2017). Kystplan Tromsøregionen. Retrived from https://www.tffk.no/_f/p1/i34afc7fa-5488-45c4-bc09-7158d4c70403/utkast-til-planprogram_kystsoneplan-for-tromso-regionen_versjon_07042020.pdf
Sanvik, A., Asplin, L., & Skardhamar, J. (2019). *Modellering av smittsomme lakseluslarveer bakgrunnsdata for Havforskningsinstituttets modellprodukt til Trafikklyssystemet*. Retrieved from https://hdl.handle.net/11250/2725014
Saunders, R. L., Specker, J. L., & Komourdjian, M. P. (1989). Effects of photoperiod on growth and smolting in juvenile Atlantic salmon (Salmo salar). *Aquaculture, 82*(1-4), 103-117.
Schubel, J. R., & Thompson, K. (2019). Farming the sea: the only way to meet humanity's future food needs. *GeoHealth, 3*(9), 238-244.
Shabani, F., Kumar, L., & Ahmadi, M. (2016). A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecology and volution, 6*(16), 5973-5986.
Silva, C., Ferreira, J., Bricker, S., DelValls, T., Martín-Díaz, M., & Yáñez, E. (2011). Site selection for shellfish aquaculture by means of GIS and farm-scale models, with an emphasis on data-poor environments. *Aquaculture, 318*(3-4), 444-457.
Smith, L. S. (1982). Decreased swimming performance as a necessary component of the smolt migration in salmon in the Columbia River. *Aquaculture, 28*(1-2), 153-161.
Spruijt, B. M., Van den Bos, R., & Pijlman, F. T. (2001). A concept of welfare based on reward evaluating mechanisms in the brain: anticipatory behaviour as an indicator for the state of reward systems. *Applied Animal Behaviour Science, 72*(2), 145-171.
Stelzenmüller, V., Gimpel, A., Gopnik, M., & Gee, K. (2017). Aquaculture Site-Selection and Marine Spatial Planning: The Roles of GIS-Based Tools and Models. In B. H. Buck & R. Langan (Eds.), *Aquaculture Perspective of Multi-Use Sites in the Open Ocean: The Untapped Potential for Marine Resources in the Anthropocene* (pp. 131-148). Cham, Switzerland: Springer International Publishing.
Stelzenmüller, V., Lee, J., South, A., Foden, J., & Rogers, S. I. (2013). Practical tools to support marine spatial planning: a review and some prototype tools. *Marine Policy, 38*, 214-227.
Stevenson, P. (2007). *Closed Waters: The Welfare of Farmed Atlantic Salmon, Rainbow Trout, Atlantic Cod and Atlantic Halibut: a Report*: Compassion in World Farming. Retrived from https://www.ciwf.org.uk/media/3818650/closed-waters-welfare-of-farmed-atlantic-salmon.pdf
Stien, L., Bracke, M. B. M., Folkedal, O., Nilsson, J., Oppedal, F., Torgersen, T., . . . Kristiansen, T. (2013). Salmon Welfare Index Model (SWIM 1.0): A semantic model for overall welfare assessment of caged Atlantic salmon: Review of the selected

welfare indicators and model presentation. *Reviews in Aquaculture, 5*, 33-57. doi:10.1111/j.1753-5131.2012.01083.x
Stigebrandt, A., Aure, J., Ervik, A., & Hansen, P. K. (2004). Regulating the local environmental impact of intensive marine fish farming: III. A model for estimation of the holding capacity in the Modelling–Ongrowing fish farm–Monitoring system. *Aquaculture, 234*(1-4), 239-261.
Stockwell, C. L., Filgueira, R., & Grant, J. (2021). Determining the effects of environmental events on cultured Atlantic salmon behaviour using 3-dimensional acoustic telemetry. *Frontiers in Animal Science, 2*, 26.
Sutterlin, A., & Stevens, E. (1992). Thermal behaviour of rainbow trout and Arctic char in cages moored in stratified water. *Aquaculture, 102*(1-2), 65-75.
The Food and Safety Act. (2004). *Lov om matproduksjon og mattrygghet mv.* Retrieved from https://lovdata.no/dokument/NL/lov/2003-12-19-124
The Pollution Control Act. (1983). *Lov om vern mot forurensninger og om avfall*. Retrieved from https://lovdata.no/dokument/NL/lov/1981-03-13-6
Thyholdt, S. B. (2014). The importance of temperature in farmed salmon growth: regional growth functions for Norwegian farmed salmon. *Aquaculture Economics & Management, 18*(2), 189-204.
Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography, 46*(1), 234-240.
Torgersen, T., Stien, L., Kvamme, B., Remen, M., Folkedal, O., & Kristiansen, T. (2009). How does salmon cope with fluctuating sea cage environments. *Kyst og Havbruk*, 164-166.
Torrissen, O., Jones, S., Asche, F., Guttormsen, A., Skilbrei, O. T., Nilsen, F., . . . Jackson, D. (2013). Salmon lice–impact on wild salmonids and salmon aquaculture. *Journal of Fish Diseases, 36*(3), 171-194.
Turnbull, J., Bell, A., Adams, C., Bron, J., & Huntingford, F. (2005). Stocking density and welfare of cage farmed Atlantic salmon: application of a multivariate analysis. *Aquaculture, 243*(1-4), 121-132.
Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography, 44*(12), 1731-1742. doi:https://doi.org/10.1111/ecog.05615
Vale, F. (2021). Spatial Autocorrelation. Retrieved from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/regression-analysis-basics.htm
Van Raaij, M., Pit, D. S., Balm, P. H., Steffens, A., & Van, G. (1996). Behavioural strategy and the physiological stress response in rainbow trout exposed to severe hypoxia. *Hormones and Behavior, 30*(1), 85-92.
Van Son, T. C., Nikolioudakis, N., Steen, H., Albretsen, J., Furevik, B. R., Elvenes, S., . . . Norderhaug, K. M. (2020). Achieving reliable estimates of the spatial distribution of kelp biomass. *Frontiers in Marine Science*, 107.
Viana, C. M., Freire, D., Abrantes, P., Rocha, J., & Pereira, P. (2022). Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. *Science of The Total Environment, 806*, 150718. doi:https://doi.org/10.1016/j.scitotenv.2021.150718
Vigen, J. (2008). *Oxygen variation within a seacage.* The University of Bergen. Retrieved from http://hdl.handle.net/1956/6128
Water Resources Act. (2001). *Lov om vassdrag og grunnvann*. Retrieved from https://lovdata.no/dokument/NL/lov/2000-11-24-82
Weighted by ranking. (2013). Retrieved from http://www.gitta.info/Suitability/en/html/Normalisatio_learningObject1.html

Wever, L., Krause, G., & Buck, B. H. (2015). Lessons from stakeholder dialogues on marine aquaculture in offshore wind farms: Perceived potentials, constraints and research gaps. *Marine Policy, 51*, 251-259.

Williams, J. N., Seo, C., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M., & Schwartz, M. W. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions, 15*(4), 565-576.

Wiltsey Stirman, S., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research. *Implementation Science, 7*(1), 1-19.

Appendix A

Overview of the fields in the production data. The information is collected for control purposes, and the individual breeder is responsible for reported information.

Date	Date
LOKNR	Location ID
LOKNAVN	Location Name
O_DESIMALGRADER_X	Desimal degrees East
N_DESIMIALGRADER_Y	Desimal degrees north
UTSATT_SMOLT_STK	The number of smolts put into the sea. Smolt is defined as fish with a net weight of less than 250g.
FISKEBEHOLDNING_ANTALL	Number of fish in the net pen at the end of the month.
BIOMASSE_KG	Biomass at the end of the month.
FORFORBRUK_KG	Total feed usage in one month.
TAP_DØDFISK_STK	Mortality. The number of fish dead in one month.
TAP_RØMMING_STK	Escape. The number of fish escapes during a month.
UTTAK_SLAKT_STK	The number of fish extracted from the sea and slaughtered.
UTTAK_SLAKT_KG	Uttak av fisk til slakt, målt kg (WFE). Ved omregning fra sløydvekt og sløyd hodekappet er omregningsfaktorer i henhold til NS 9417:2012 benyttet
UTTAK_LEVENDE_STK	Fish extracted and moved to another location.

Appendix B

The program was used to complete the time series 2018 - 2021 of all aquaculture locations in the analysis. In total, 36 rows were created for each unique ID.

```
using System.Collections.Generic;
using System.Linq;
using System.IO;
using CsvHelper;
using CsvHelper.Configuration;
using System;

namespace Nikolai_par_balai
{
    public class Program
    {
        public static int Main(string[] args)
        {
            if(!args[0].ToLowerInvariant().EndsWith(".csv")){
                Console.WriteLine("Csv file expected");
                return 0;
            }

            var list = new List<EntryRecord>();
            using (var reader = new StreamReader(args[0]))
            using (var csv = new CsvReader(reader))
            {
                csv.Configuration.HasHeaderRecord = true;
                csv.Configuration.Delimiter = ";";
                csv.Configuration.AllowComments = true;
                csv.Configuration.TrimOptions = TrimOptions.Trim;
                csv.Configuration.MissingFieldFound = null;
                var records = csv.GetRecords<EntryRecord>();

                foreach (var record in records)
                {
                    list.Add(record);
                }
            }

            var ids = list.Select(l => l.ID).Distinct().OrderBy(id => id).ToList();

            var newRecords = new List<EntryRecord>();


            foreach (var id in ids)
            {
                var entries = list.Where(l => l.ID == id).ToList();
                if (entries.Count < 6)
                {
                    continue;
                }

                var end = new DateTime(2021, 01, 01);

                for(
                var current = new DateTime(2018, 01, 01);
                    current < end;
                    current = current.AddMonths(1))
                {
                    var monthEntry = entries.FirstOrDefault(e => Convert.ToDateTime(e.Date).Date ==
current.Date);

                    if(monthEntry == null)
                    {
```

Appendix C

Solar calculations based on NOAA. An excel online file with complete calculations is attached here in this file: DayLengthNOAA.xls. Fields and formulas are attached in the table below.

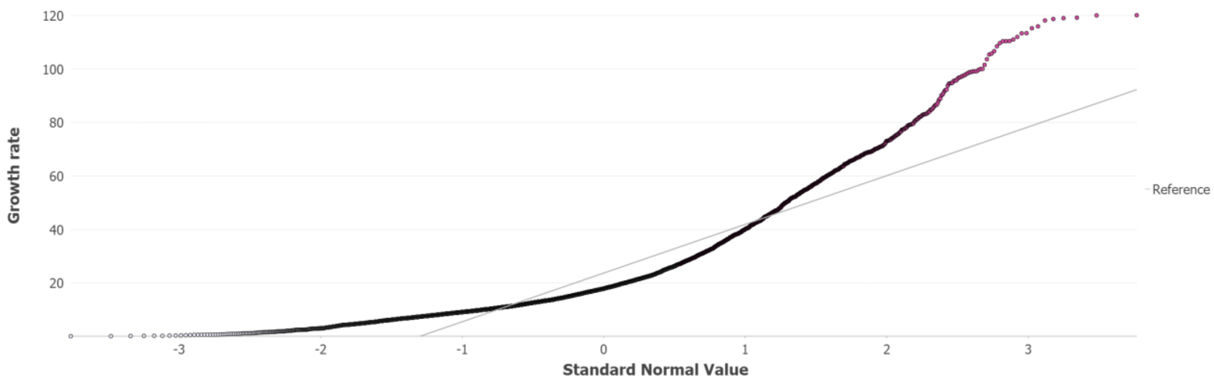
Appendix C.1: Solar calculation formulas derived from NOAA.

Fields	**Formulas/Calculations**
Date	1st in every month
Time (hrs past local midnight)	12:00hrs
Julian Day	Date+2415018,5+Time-1/24
Julian Century	(Julian Day-2451545)/36525
Geom Mean Long Sun (deg)	REST(280,46646+G2*(36000,76983 + Julian Century*0,0003032);360)
Geom Mean Anom Sun (deg)	357,52911+ Julian Century *(35999,05029 - 0,0001537* Julian Century)
Eccent Earth Orbit	0,016708634-Julian Century*(0,000042037+0,0000001267* Julian Century)
Sun Eq of Ctr	SIN(RADIANER(Geom Mean Anom Sun (deg)))*(1,914602-Julian Century *(0,004817+0,000014*Geom Mean Anom Sun (deg))+SIN(RADIANER(2* Geom Mean Anom Sun (deg)))*(0,019993-0,000101*Julian Century)+SIN(RADIANER(3*Geom Mean Anom Sun (deg)))*0,000289
Sun True Long (deg)	SIN(RADIANER(Geom Mean Anom Sun (deg))*(1,914602- Julian Century *(0,004817+0,000014* Julian Century))+SIN(RADIANER(2*Geom Mean Anom Sun (deg))*(0,019993-0,000101* Julian Century)+SIN(RADIANER(3*Geom Mean Anom Sun (deg))*0,000289
Sun True Anom (deg)	Geom Mean Long Sun (deg) + Sun Eq of Ctr
Sun Rad Vector (AUs)	Geom Mean Anom Sun (deg) + Sun Eq of Ctr
Sun App Long (deg)	(1,000001018*(1- Eccent Earth Orbit * Eccent Earth Orbit))/(1+ Eccent Earth Orbit *COS(RADIANER(Sun True Anom (deg))))

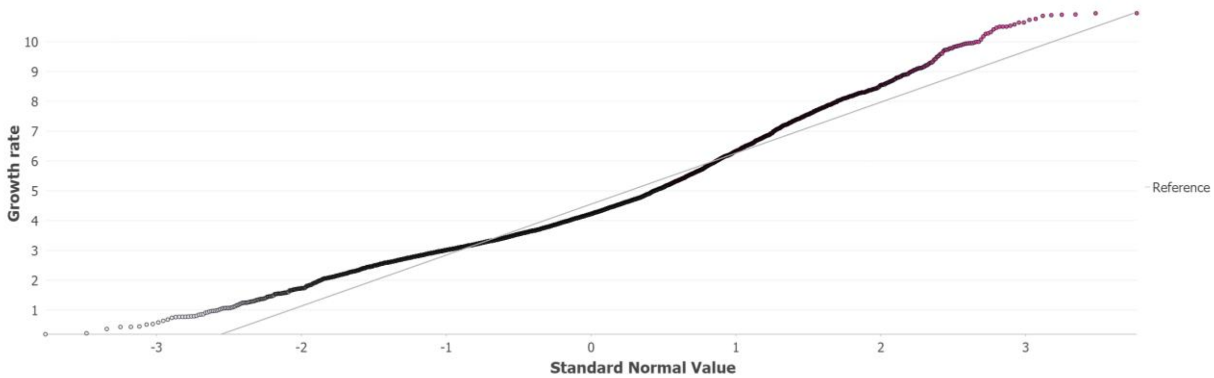
Mean Obliq Ecliptic (deg)	23+(26+((21,448- Julian Century *(46,815+ Julian Century *(0,00059- Julian Century *0,001813))))/60)/60
Obliq Corr (deg)	Mean Obliq Ecliptic (deg) +0,00256*COS(RADIANER(125,04-1934,136* Julian Century))
Sun Rt Ascen (deg)	GRADER(ARCTAN2(COS(RADIANER(Sun App Long (deg)));COS(RADIANER(Obliq Corr (deg)))*SIN(RADIANER(Sun App Long (deg)))))
Sun Declin (deg)	GRADER(ARCSIN(SIN(RADIANER(Obliq Corr (deg)))*SIN(RADIANER(Sun App Long (deg)))))
var y	TAN(RADIANER(Obliq Corr (deg) /2))*TAN(RADIANER(Obliq Corr (deg) /2))
Eq of Time (minutes)	4*GRADER(var y *SIN(2*RADIANER(I2))-2* Eccent Earth Orbit *SIN(RADIANER(Geom Mean Anom Sun (deg)))+4* Eccent Earth Orbit * var y z*SIN(RADIANER(Geom Mean Anom Sun (deg)))*COS(2*RADIANER(I2))-0,5* var y * var y *SIN(4*RADIANER(I2))-1,25* Eccent Earth Orbit * Eccent Earth Orbit *SIN(2*RADIANER(Geom Mean Anom Sun (deg))))
HA Sunrise (deg)	GRADER(ARCCOS(COS(RADIANER(90,833))/(COS(RADIANER(Lon))*COS(RADIANER(Sun Declin (deg))))-TAN(RADIANER(Lon))*TAN(RADIANER(Sun Declin (deg)))))
Solar Noon (LST)	(720-4*Lon-V2+Lat*60)/1440
Sunrise Time (LST)	(Solar Noon (LST) *1440- HA Sunrise (deg) *4)/1440
Sunset Time (LST)	(Solar Noon (LST) *1440+ HA Sunrise (deg) *4)/1440
Sunlight Duration (minutes)	8* HA Sunrise (deg)
True Solar Time (min)	REST(Eq of Time (minutes)*1440+Eq of Time (minutes)+4*Lat-60*1;1440)

Appendix D

QQ-plot of growth rate compared to normal distributions. Appendix D.1 is without transformations. Appendix D.2 is with square root transformations.



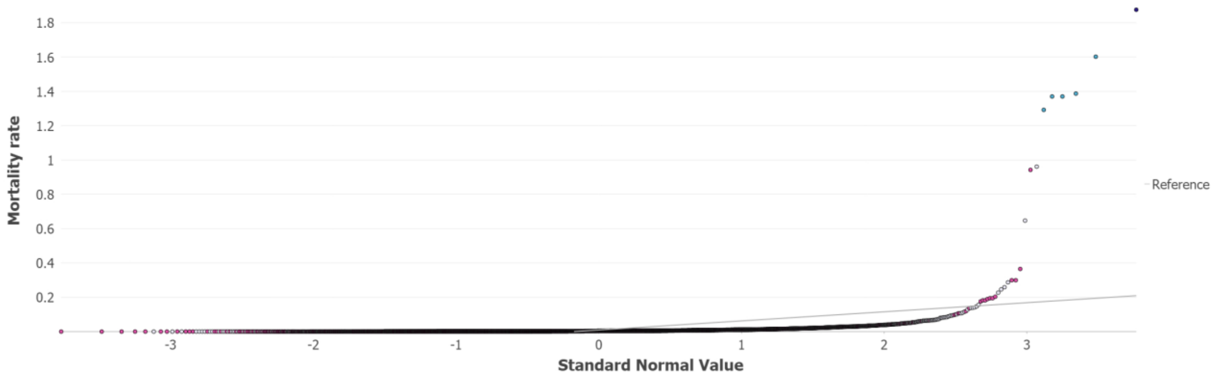
Appendix D.1: QQ plot comparison of growth rate and normal distributions.



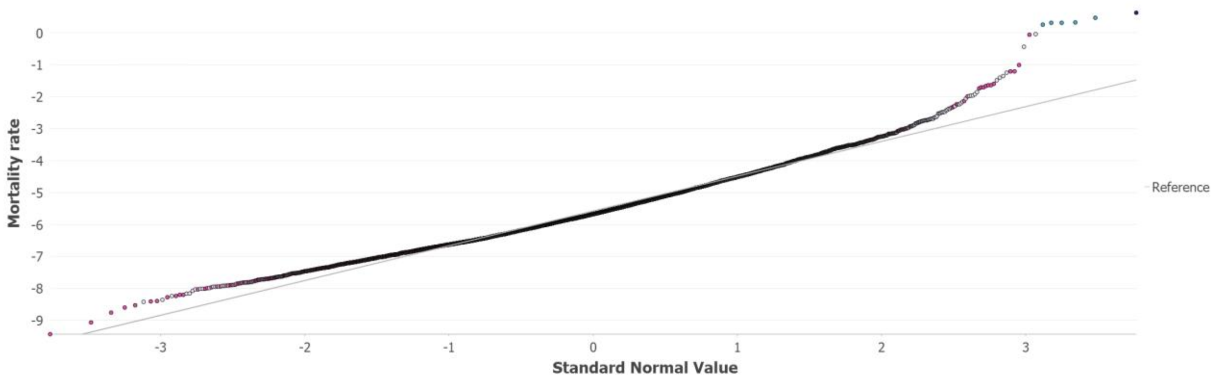
Appendix D.2: QQ plot comparison of growth rate and normal distributions with square root transformation

Appendix E

QQ-plot of mortality rate compared to normal distributions. Appendix E.1 is without transformations. Appendix E.2 is with logarithmic transformations.



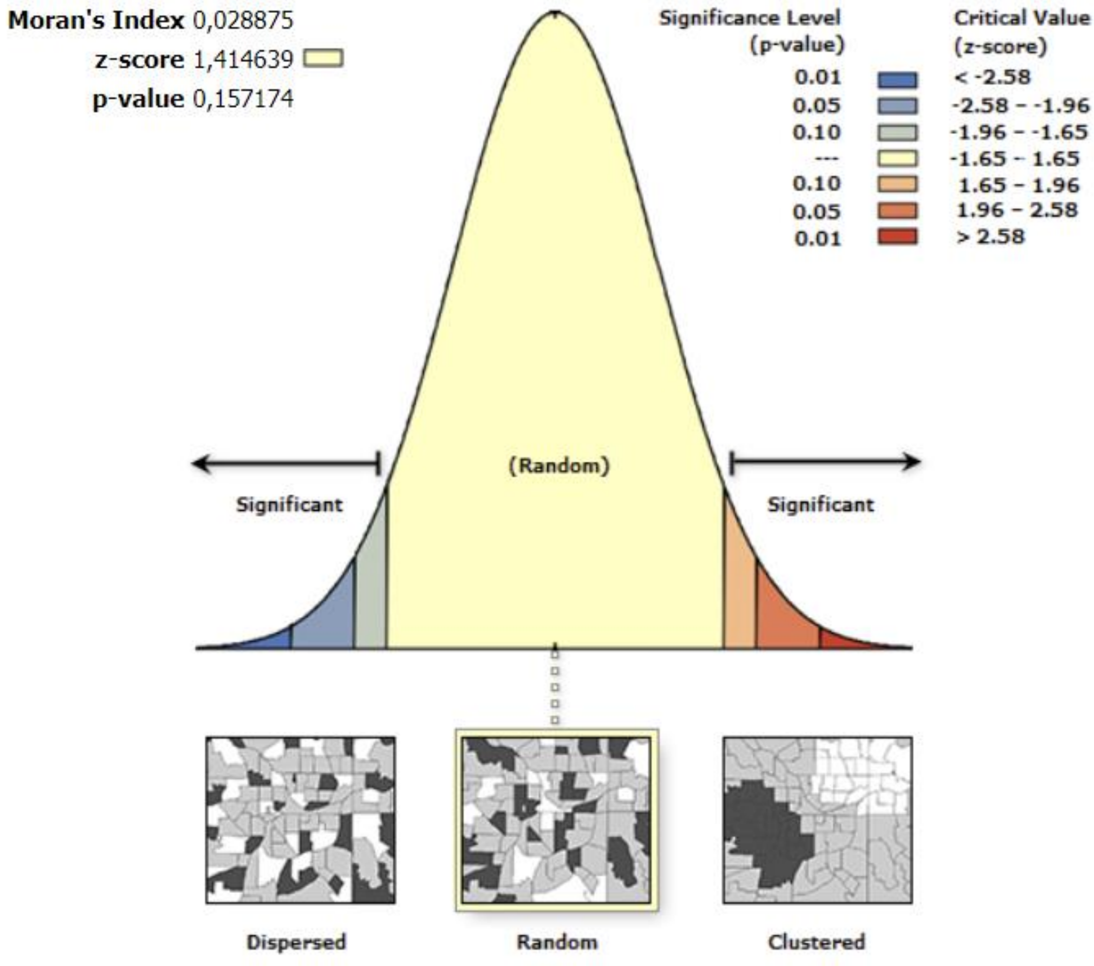
Appendix E.1: QQ plot comparison of mortality rate and normal distributions.



Appendix E.2: QQ plot comparison of mortality rate and normal distributions with logarithmic transformation

Appendix F

Test for spatial autocorrelation. Moran's Index of 0,028875 and p-value of 0,0157 indicated no spatial autocorrelation.



Appendix F.1: Test for spatial autocorrelation Globals Moran's I.

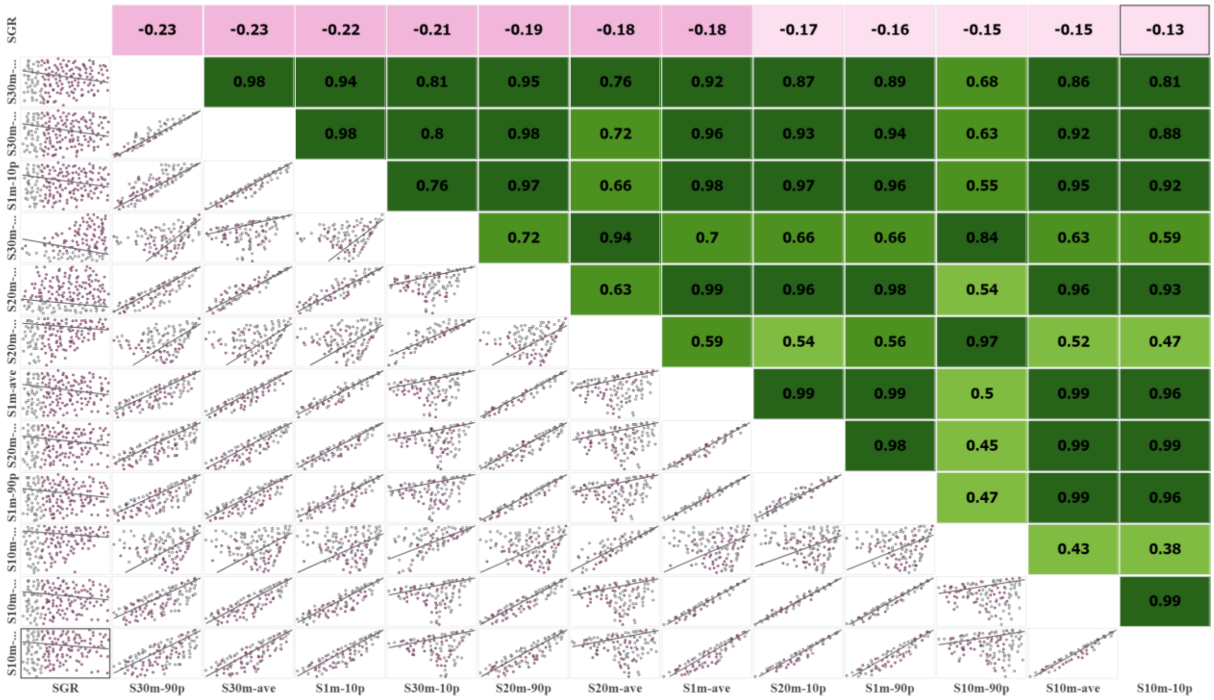
Appendix G

Appendix G.1: Summary statistics of all exploratory variables. The environmental variables are aggregated average values from monthly mean values to the average for the period 2018 – 2021. All other values are also transformed into averages. Note the difference in count ranging from 6048 to 278. Biomass, Slope, Depth, and distance dependence variables were calculated based on the presence- and absence point dataset. All calculations were done using ArcGIS pro.

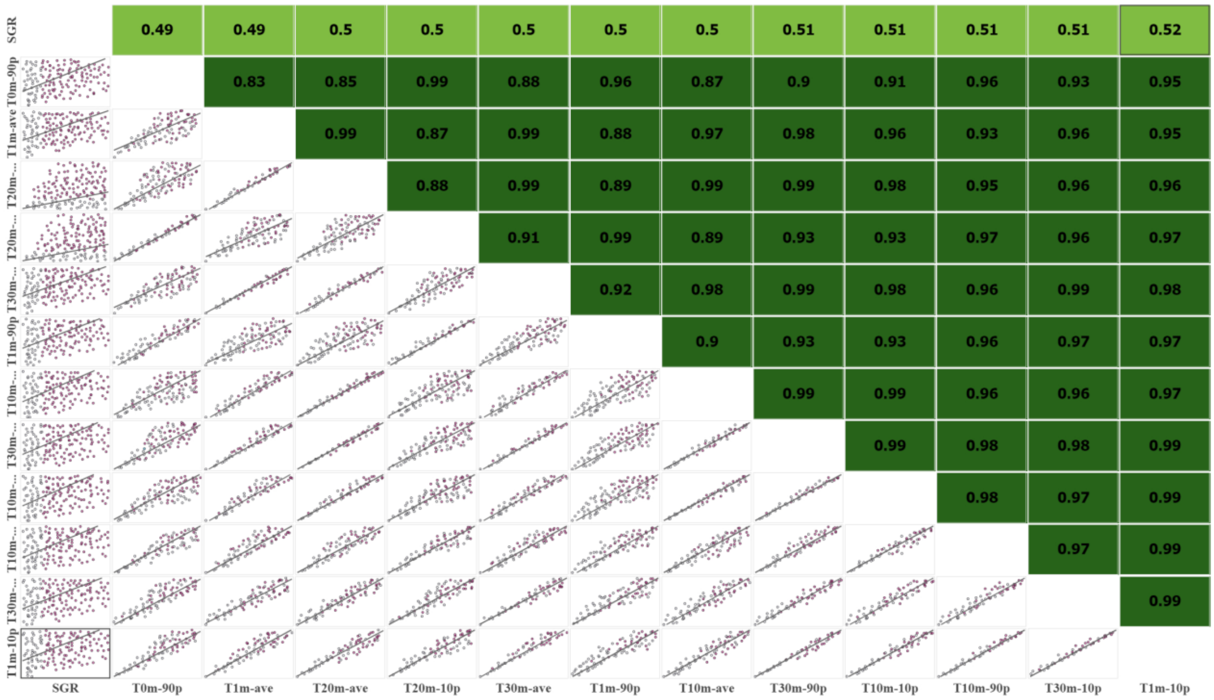
Variables	Min	Max	Mean	Count
Salinity 1m ‰	18.1	34.5	32	6048
Salinity 10m ‰	29.4	34.5	32.8	6021
Salinity 20m ‰	30.3	34.5	33	5762
Salinity 30m ‰	30.5	34.5	33.2	5546
Temperature 1m ° C	-0.1	14.4	7.3	6048
Temperature 10m ° C	0.33	14	7.2	6021
Temperature 20m ° C	0.79	13.2	7.0	5762
Temperature 30m ° C	0.98	13.0	6.9	5546
Current 1m m/s	0.01	0.45	0.079	6048
Current 10m m/s	0	0.41	0.05	6021
Current 20m m/s	0	0.39	0.042	5762
Current 30m m/s	0	0.36	0.038	5546
Windspeed m/s	0.12	6.3	2.1	6025
Wave height m	0.15	2.2	0.51	5953
Day length min	0	1 440	754.9	6025
Solar Duration	0	720	281	6048
Solar Diffuse	0	34 997.8	12.429.7	6048
Solar Direct	9	93 821.5	29 803.1	6048
SFR	2.33	239.6	21.5	6048
Biomass	2 110.5	7 393 067	1 549 874	6048
Slope	0	45.2	10.7	278
Depth	12.4	344.5	81.4	285
Nearest community	60.7	39 578	13 166	363
Nearest fairway	104	19 132	3 263	363
Nearest salmon farm	312	65 944	4885	363
Distance to shore	101	1 156	391	285

Appendix H

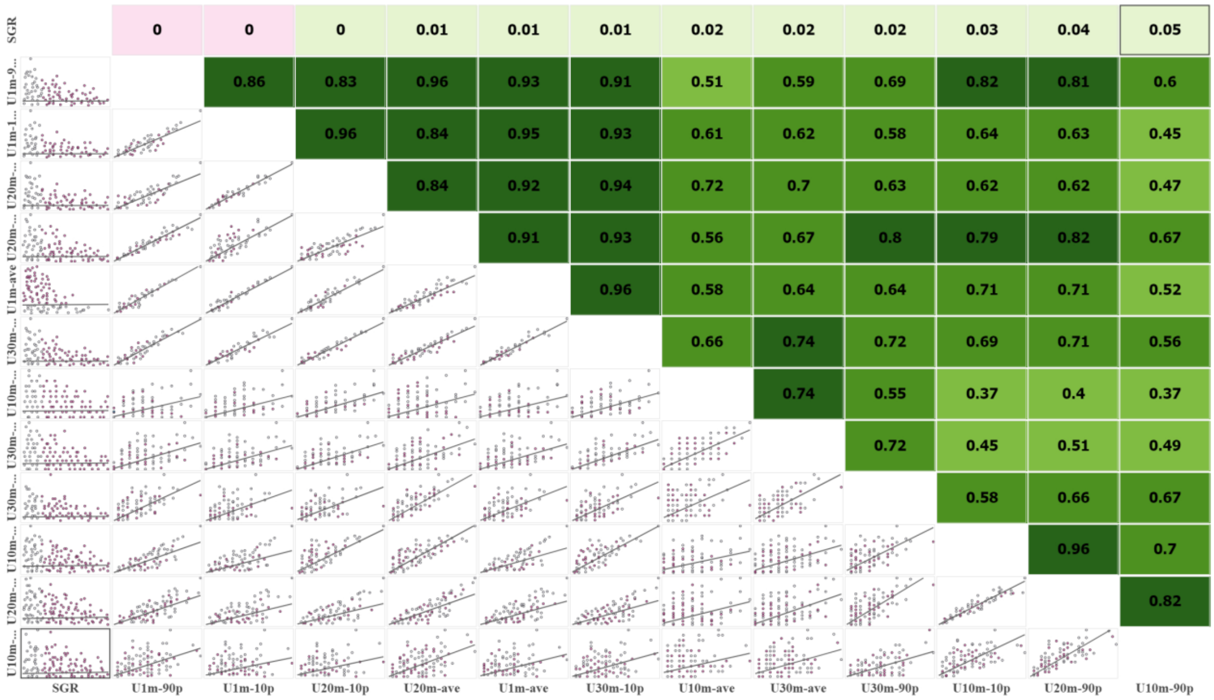
Linear regression of growth rate and independent variables displayed with Pearson's correlation coefficient.



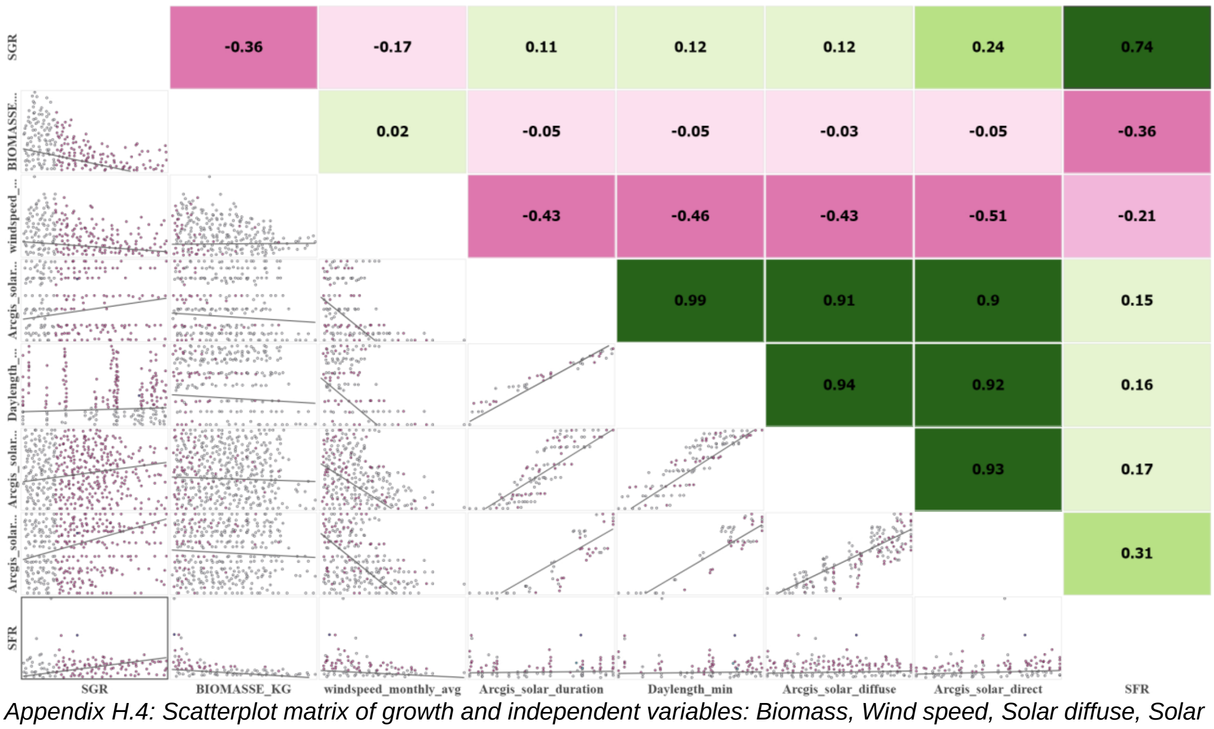
Apenndix H.1: Scatterplot matrix of growth and salinity sorted ascending by Pearson's r.



Appendix H.2: Scatterplot matrix of growth and temperature sorted ascending by Pearson's r.



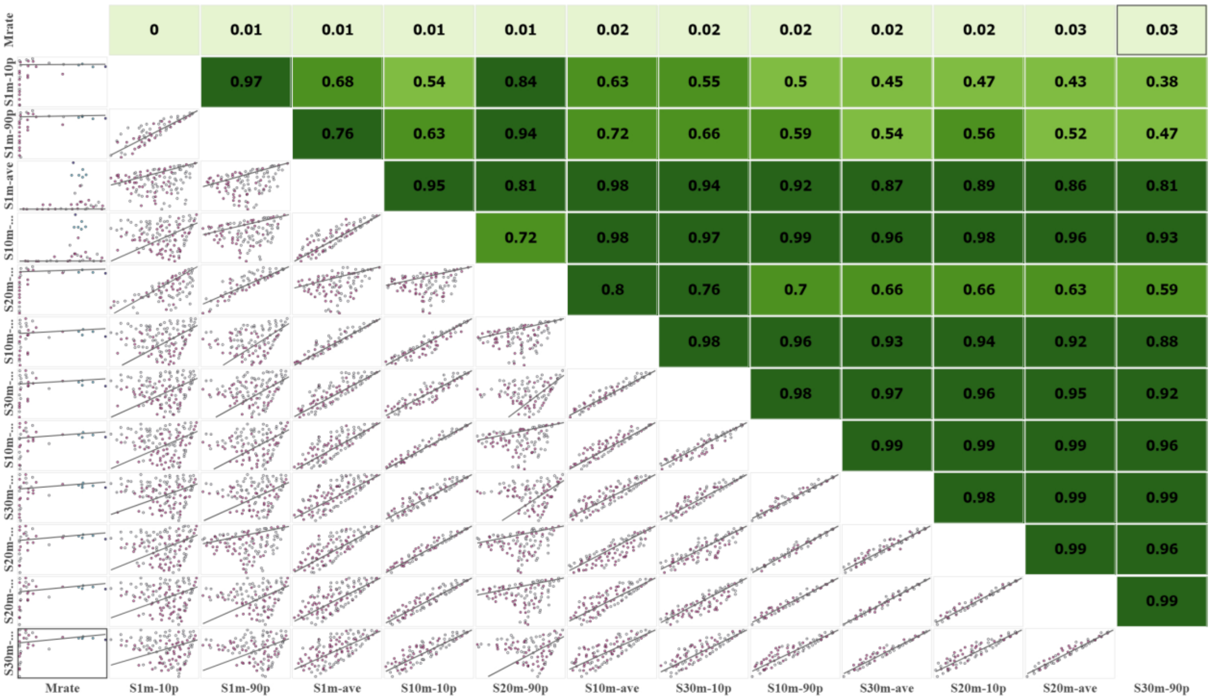
Appendix H.3: Scatterplot matrix of growth and current sorted ascending by Pearson's r.



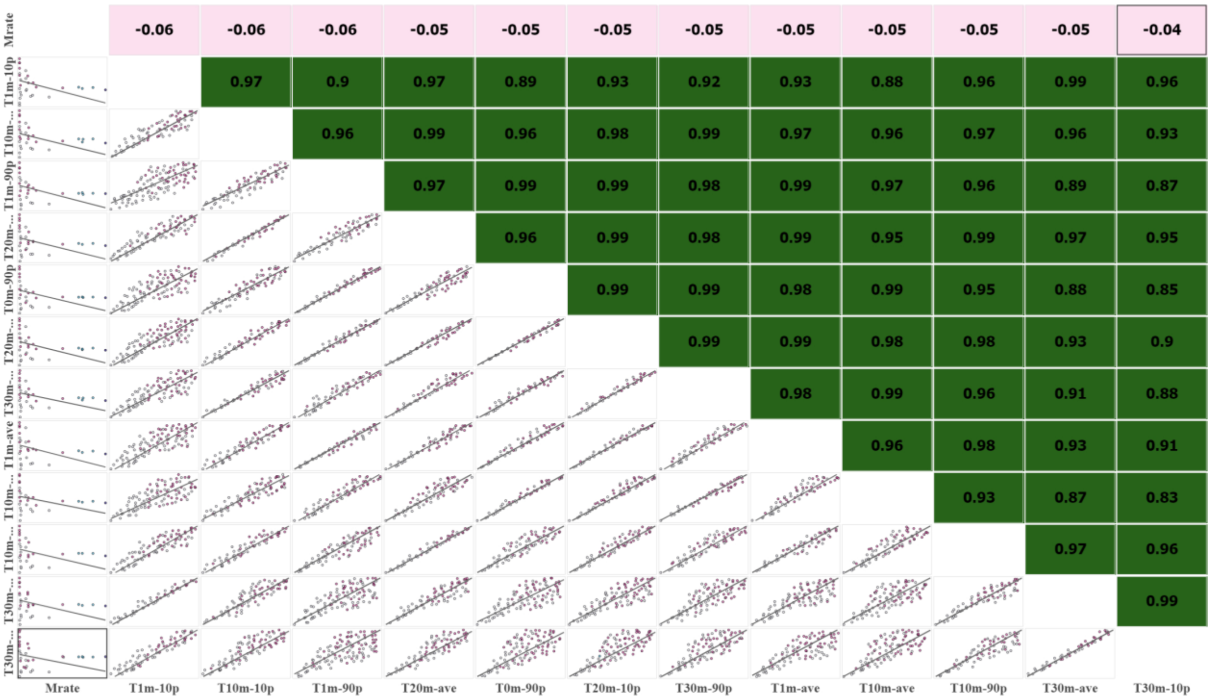
Appendix H.4: Scatterplot matrix of growth and independent variables: Biomass, Wind speed, Solar diffuse, Solar direct, Solar duration, Day length, and specific feeding rate.

Appendix I

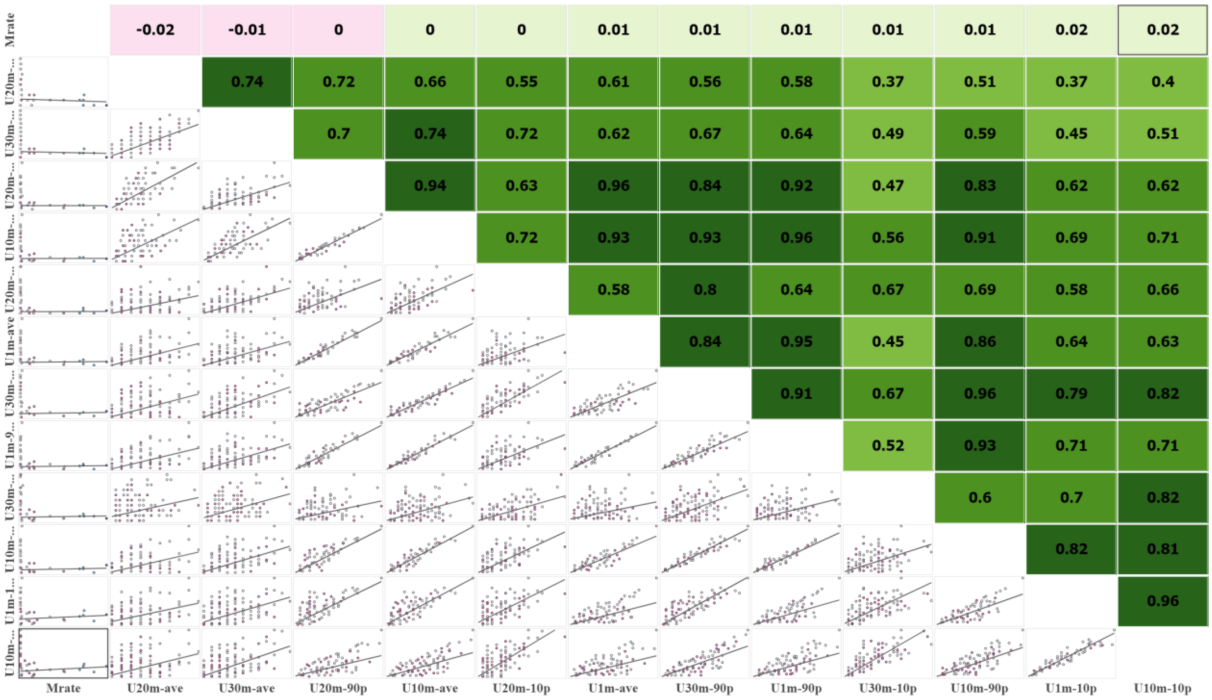
Linear regression of mortality rate and independent variables displayed with Pearson's correlation coefficient.



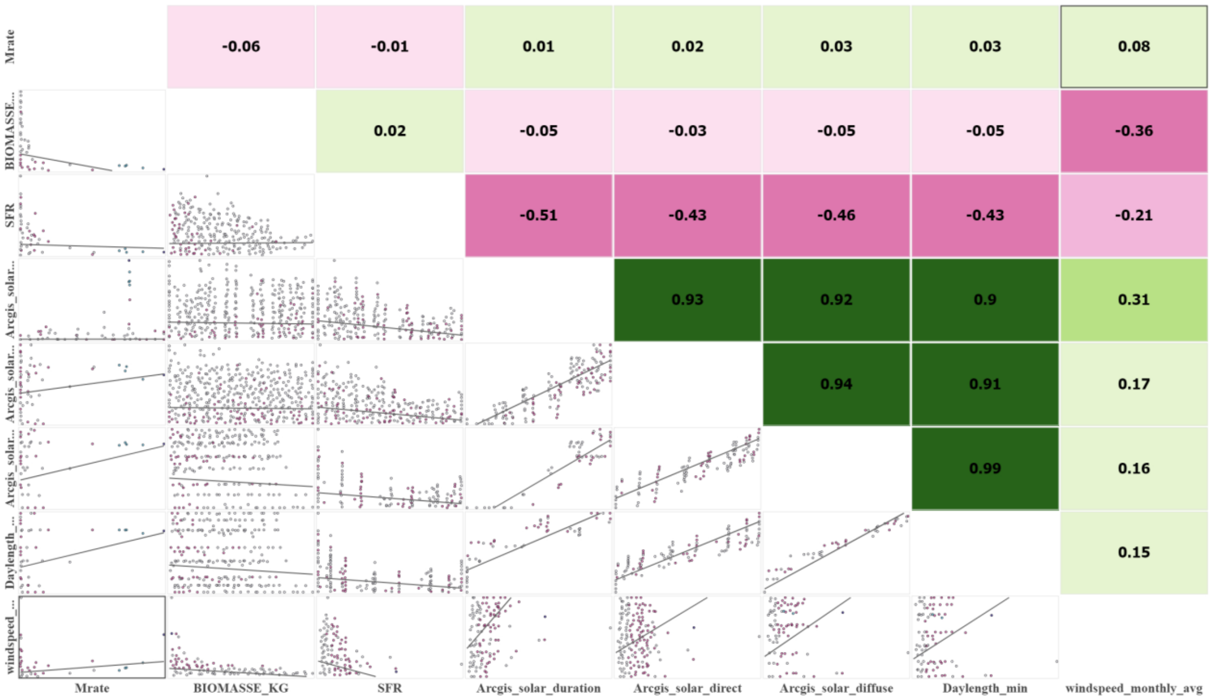
Appendix I.1: Scatterplot matrix of mortality and salinity sorted ascending by Pearson's r.



Appendix I.2: Scatterplot matrix of mortality and temperature sorted ascending by Pearson's r.



Appendix I.3: Scatterplot matrix of mortality and current sorted ascending by Pearson's r.



Appendix I.4: Scatterplot matrix of mortality and independent variables: Biomass, Wind speed, Solar diffuse, Solar direct, Solar duration, Day length, and specific feeding rate.

Appendix J

Ordinary least squares results.

Appendix J.1: Linear regression with OLS, dependent variable SGR with square root transformation.

Input feature	**SGR square root transformation**		
Number of observations	6048	Akaike’s information criterion (AICc)	11056.4
Multiple R Squared	0.788	Adjusted R squared	0.788
Joint F-statistics	2094.42	Prob(>F), (9.5047) degrees of freedom	0.00*
Joint Wald Statistics	14871.19	Prob(<chi-squared), (9) degrees of freedom	0.00*
Koenker (BP) Statistics	306.78	Prob(<chi-squared), (9) degrees of freedom	0.00*
Jarque-Bera Statistics	4340.89	Prob(<chi-squared), (2) degrees of freedom	0.00*

Appendix J.2: Linear regression with OLS, dependent variable Mrate with log-transformation.

Input feature	**Mrate log transformation**		
Number of observations	6048	Akaike’s information criterion (AICc)	6867.51
Multiple R Squared	0.058	Adjusted R squared	0.057
Joint F-statistics	56.52	Prob(>F), (9.5047) degrees of freedom	0.00*
Joint Wald Statistics	343.633	Prob(<chi-squared), (9) degrees of freedom	0.00*
Koenker (BP) Statistics	158.030	Prob(<chi-squared), (9) degrees of freedom	0.00*
Jarque-Bera Statistics	846..665	Prob(<chi-squared), (2) degrees of freedom	0.00*