Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Kaia Arnøy Høyheim

Particle Filtering Approaches for Atlantic Salmon Migration Based on Acoustic Telemetry Data

Master's thesis in Applied Physics and Mathematics
Supervisor: Jo Eidsvik
June 2020

NTNU
Norwegian University of Science and Technology

Kaia Arnøy Høyheim

Particle Filtering Approaches for Atlantic Salmon Migration Based on Acoustic Telemetry Data

Master's thesis in Applied Physics and Mathematics
Supervisor: Jo Eidsvik
June 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Summary

When young salmon, called smolts, migrate from the river to the ocean, the survival rate can be reduced because of sea lice infection. Sea lice infection can make the migrating smolts more vulnerable to predators, and they can more easily get other infections. Understanding where the smolts are at different times is important for initiating targeted actions to reduce the infection risk. To do this, it is necessary to obtain reliable estimates of the movement pattern of the salmon.

Data containing information about the position of salmon in Nordfjord during the migration period in 2017 were collected using acoustic telemetry. 118 salmon were equipped with acoustic transmitters and 66 stationary receivers were placed in the river and fjord system. The telemetry data gathered from this experiment were indicating the presence and absence of the salmon in the vicinity of the receivers.

A state-space model was used to model the movement of each salmon independently, and a particle filter was applied to estimate the movement pattern of the salmon based on the acoustic telemetry data. In addition, a sequential fixed-lag smoother adjustment was added to the particle filter.

The performance of the particle filter was compared to that of a sequential fixed-lag smoother adjustment to the particle filter with different lags. An approximation of the root mean square error and the effective sample size of the different models were compared. By applying the particle filter without a smoother adjustment to the data, we got a low root mean square error and a high effective sample size. This indicated that the predictions were both precise and robust. The particle filter gave a total approximated root mean square error of 5802 meters for all the salmon considered. With a sequential fixed-lag smoother adjustment to the particle filter, the root mean squared error increased. Overall, the particle filter algorithm seemed to work well on the data. The effective sample size was high, which should ensure robust results.

The state-space model did not consider environmental variables like temperature and salinity explicitly. Instead, the results of the particle filter were compared with data on environmental variables. These factors are expected to have an effect on the behavior of salmon. The correlations between the particle filter estimates and environmental variables from the fjord were generally found to be weak and not significant.

Sammendrag

Overlevelsesraten til migrerende laksesmolt kan reduseres på grunn av lakselussmitte. Den kan gjøre smolten mer sårbar overfor rovdyr og øke risikoen for å bli smittet av andre sykdommer. Innsikt i hvor smolten befinner seg til ulike tider er en viktig faktor for å iverksette målrettede tiltak for å redusere risikoen for lakselussmitte. For å gjøre dette er det viktig å ha tilgang på pålitelige og gode estimater av bevegelsesmønsteret til smolten.

Posisjonsdata for smolten i løpet av migrasjonsperioden i Nordfjord i 2017 ble samlet ved hjelp av akustisk telemetri. 118 laks ble utstyrt med akustiske merker som sendte ut signaler, og 66 stasjonære lyttebøyer som fanget opp slike signaler ble plassert i elv- og fjordsystemet. Telemetridataen som ble hentet inn indikerer om laksen oppholdt seg i nærheten av lyttebøyene til gitte tider. Fravær av deteksjoner på en lyttebøye tilsier at laksen ikke var i nærheten av den angitte lyttebøyen på det aktuelle tidspunktet.

En tilstandsmodell ble benyttet for å modellere bevegelsen til laks, og et partikkelfilter ble anvendt for å estimere bevegelsesmønsteret til laksen basert på telemetridataen. I tillegg ble det lagt til en justering til partikkelfilteret i form av en sekvensiell glatter med en fast forsinkelse.

Prestasjonen til partikkelfilteret ble sammenlignet med prestasjonen til glatteren med ulik forsinkelse. Dette ble gjort ved å sammenligne en tilnærming av den gjennomsnittlige kvadratiske feilen og den effektive utvalgsstørrelsen for de ulike modellene. Ved å benytte partikkelfilteret uten en glatter ble det oppnådd en lav kvadratisk feil og høy effektiv utvalgsstørrelse, noe som sikret både presise og robuste resultat. Ved å se på resultatet til alle smoltene var den totale approksimerte kvadratiske feilen 5802 meter. Ved å legge til en glatter til partikkelfilteret økte verdien av den gjennomsnittlige kvadratiske feilen. Totalt sett ga modellen gode resultater for den aktuelle dataen. Den effektive utvalgsstørrelsen var høy, noe som sikret robuste resultater.

Tilstandsmodellen som ble benyttet tok ikke eksplisitt hensyn til miljøvariabler som temperatur og salinitet. Resultatet fra partikkelfilteret ble i stedet sammenlignet med data som inneholdt informasjon om ulike miljøvariabler. Det er forventet at disse variablene kan ha en innvirkning på laksens oppførsel. Korrelasjonene mellom estimatene fra partikkelfilteret og miljøvariablene fra fjorden var generelt sett svake og ikke signifikante.

Preface

This thesis is the final part of the Master of Science in Applied Physics and Mathematics at the Norwegian University of Science and Technology (NTNU) and it adds up to my master's thesis. The work that is presented was executed during the spring and early summer of 2020.

I would like to thank my supervisor, Jo Eidsvik, for excellent feedback and support throughout the semester. I would also like to thank Jo Arve Alfredsen and Henning Urke for providing data and information about the problem, and for helping me with the thesis through discussions and valuable feedback. Lastly, I want to thank Helge Bostwick Bjerck for sharing cleaned and processed data.

Kaia Arnøy Høyheim
Trondheim, June 2020

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Atlantic salmon are anadromous fish that begin their life in freshwater rivers. When the salmon are between one and six years old and have a length of approximately 12-20 cm, they are called smolts. At this stage of the life cycle, the salmon are ready to migrate from freshwater to saltwater (Vøllestad, 2018). The average age for when the smolts migrate to sea is three years (Jensen & Johnsen, 1989). Prior to the migration, the juvenile salmon undergo physiological and behavioral transformations (Urke et al., 2013a). The physiological changes include the development of an increased salinity tolerance. Juvenile salmon actively swim against the water current, but the smolts develop a preference for moving downstream, and hence out of the river system. The changes occur to enable the smolts to enter the ocean at an appropriate time, avoid predation in the estuary and reduce the risk of predation in the open waters. Migration takes place during the spring and may be initiated by temperature and water flow (Høgåsen, 1998; Aas et al., 2011). The smolts migrate to the ocean to grow and mature before they return to the river to spawn (Hoar, 1988; McCormick et al., 1998; Stefansson et al., 2012).

In 2017, a study of Atlantic salmon smolts in Nordfjord was performed by Urke et al. (2018). One of the goals of the experiment was to study the migration timing and behavior of Atlantic salmon and brown trout smolts by collecting acoustic telemetry data. Acoustic telemetry is a popular tool for understanding and tracking migration of aquatic animals (Kraus et al., 2018). The animals are marked with acoustic transmitters, and the tracking can either be active or passive. In active tracking, the researcher usually uses a mobile receiver to follow the tagged fish. When using passive tracking, the receivers are placed in fixed locations in the environment (Brownscombe et al., 2019). By using this technique, individual fish can be identified and placed in time and space. In this thesis, the focus will be on evaluating the migration pattern of the salmon based on acoustic telemetry data collected using passive receivers in the study performed by Urke et al. (2018).

When the salmon migrate through the middle and outer fjord areas, they can be exposed to sea lice from the aquaculture industry in the area. The sea lice infection pressure on the wild salmon is connected to the size of the aquaculture industry. To ensure that the growth in the aquaculture industry is predictable and sustainable, especially with respect

to the wild salmon population, the so-called traffic light system has been introduced. If an area receives a green light, it means that the allowed production capacity of the aquaculture industry can be increased. A yellow light indicates that the production capacity in the area can remain unchanged, while a red light means that it must be decreased. The production capacity is a measure of the upper limit of the amount of salmon a farmer can have in the ocean at any time (Ministry of Trade, Industry and Fisheries, 2019). Indicators for the sustainability of the wild salmon should determine the allowed production in the aquaculture. To lay the grounds for determining the lights in the different production areas in Norway, Vollset et al. (2019) have evaluated the mortality of wild fish induced by sea lice infection.

Sea lice infection is a problem among wild salmon, and it can reduce the survival rate of the migrating smolts. Impaired fish can be subject to increased risk of predation and have reduced growth and grazing ability (Thorstad et al., 2012; Miller et al., 2014). The study by Vollset et al. (2019) concludes that the mortality of wild fish induced by sea lice infection is high in the area from Nordhordaland to Stadt, where Nordfjord is located. As a result, the area has received a red light, and the production capacity is to be reduced.

Studies have been performed to evaluate the risk of contagious diseases spreading in the Nordfjord areas and to determine the infection risk in the different areas (Daae et al., 2011). When the salmon reside in the middle and outer fjord areas, the sea lice infection pressure from the aquaculture is presumed to be highest (Urke et al., 2018; Vollset et al., 2019). Important factors for infection risk are the time spent in different areas and the time of the migration. Today, the knowledge about the time and duration of the smolts' stay in the different parts of the fjord is limited (Haugen et al., 2014; Urke et al., 2018).

If the knowledge about the migration pattern of the salmon is increased, it enables a more certain evaluation and possibly a better management of the salmon population in Nordfjord. As a result of the red light, the aquaculture industry in Nordfjord must be reduced. In addition, targeted actions can be initiated from the remaining aquaculture industry in the periods where the salmon are in the areas considered to have the highest infection risk. If there is knowledge about when and where the salmon will be, one can create an improved surveillance program to monitor the condition of the salmon (Ugedal et al., 2014; Urke et al., 2018).

The objective of this thesis is to study the migration and estimate the movement pattern of Atlantic salmon using acoustic telemetry data. This is done by using a state-space model for individual fish movement and by applying a sequential importance resampling algorithm, a particle filter, to the data obtained in the experiment performed in Nordfjord in 2017 (Urke et al., 2018). If we can create a robust and accurate prediction model, it is possible to create more effective and targeted protection and surveillance programs of the salmon. Today, passive receivers are used to detect the salmon. There is an ongoing project to develop techniques to enable autonomous tracking of the salmon by placing a receiver on a moving vessel. We can then perform adaptive sampling by using a prediction model to determine how the vessel should move. This could ensure that we get more detections during the migration, which can make the measurements and predictions more accurate. A second objective of the thesis is to investigate if there is a correlation between the estimated movement pattern and environmental data from the fjord.

In a related study performed by Johnson et al. (2008), a continuous-time correlated

random walk for animal telemetry data is put into a state-space model, and a Kalman filter is used for computing maximum likelihood estimates of the movement parameters. A Kalman filter is also used to estimate errors, movement parameters and most probable tracks for data collected on bigeye tuna near Hawaii (Siebert et al., 2003). Xydes et al. (2013) present a particle filter method for estimating position and speed based on data from acoustically tagged fish. A random walk is used to propagate the particles during the prediction step, and detections of the fish were used for the correction step. In this thesis, a particle filter is used to track the salmon. We actively use the presence data for the importance distribution, and then the particle weights are updated using both the presence and absence data. This approach seems to be well suited for the movement situation in Nordfjord. In addition, a sequential fixed-lag smoother will be applied to the particle filter estimates.

In Chapter 2, the case is more thoroughly described. The data are analyzed, prepared and cleaned before a particle filter and sequential fixed-lag smoother are applied. The theory behind sequential filtering and smoothing is presented in Chapter 3. A state-space model and a particle filter for the salmon tracking problem are presented in Chapter 4. The results when applying the state-space model to the data are presented in Chapter 5, with a subsequent discussion in Chapter 6.

Chapter 2

Case Description and Data Analysis

2.1 Description of the Case

Nordfjord is a drainage basin situated in Vestland and Innlandet, Norway, see **Figure 2.1** (Norgeskart, 2019). According to the Norwegian Ministry of Climate and Environment, Nordfjord is a national salmon watercourse (Ministry of Climate and Environment, 2019). National watercourses are important for rebuilding the size and composition of the salmon population, such that this in turn will secure the diversity and reproduction ability of the species.

Figure 2.1: Map of Norway. The position of Nordfjord is marked with a black rectangle.

In the experiment performed by Urke et al. (2018), acoustic telemetry was used to register the migration of the fish. Acoustic telemetry is a popular tool for understanding and tracking migration of aquatic animals (Kraus et al., 2018). Salmon were marked with acoustic transmitters, and stationary receivers that capture signals from these transmitters were placed in different parts of Nordfjord. By using this technique, individual salmon was identified and placed in time and space. Acoustic telemetry works both in freshwater and seawater, and it was therefore a well suited method for mapping the migration pattern of salmon both in the river and in the fjord.

118 pre-smolt salmon were caught using electrical fishing and marked with acoustic transmitters after standard methods (Urke et al., 2013b) in the period April 24 to 29, 2017. The salmon were marked with two different types of transmitters. There were in total 53 salmon equipped with transmitters with a depth sensor. These tags emitted the ID of the fish, in addition to information about the depth of the fish with a resolution of 0.2 meters. The remaining 65 salmon were equipped with tags that only emitted the ID. In addition, all transmitters emitted a check sum that was used to verify that the signals detected at the receivers were from one of the transmitters used in the experiment.

66 stationary receivers (VR2W, Vemco: `http://vemco.com/products/vr2w-69khz/`) were placed in the fjord before the salmon were released. The detection range of a receiver depends on various factors. For example, areas with clear water, flat bottoms and low current usually have the highest ranges, while areas with muddied water, rocky bottoms and high current exhibit low ranges. In addition, extreme weather can reduce the range significantly (Brownscombe et al., 2019; Vemco, 2019). The transmitter range will also depend on the conditions of the water around the receivers, and it can vary between a couple of meters up to 200-500 meters. The tags were programmed to emit information with a periodic interval between 30 and 90 seconds (Urke et al., 2018). Due to the detection range of the receivers, and the transmitter range of the tags, not all fish were registered when they passed a receiver.

As presented in **Figure 2.2**, the receivers were placed in nine zones. UTM coordinates, zone 32, are used. The zones can further be divided into three main parts: the inner, middle and outer part of Nordfjord. The inner part consisted of the zones River, Estuary, Inner and Utvik. Lote, Krokneset and Isane made up the middle part of the fjord while the zones Maurstad and Outer made up the outer part.

2.2 Data Preparation

In addition to the data collected from the receivers during the experiment period, there are also data about the receivers that contain the name, ID, coordinates and the zone of the receivers. For analysis and plotting purposes, the receivers are sorted from the inner to the outer parts of Nordfjord. In general, this is from east to west, but in some areas, we sorted the receivers manually to account for turns in the fjord. By doing this, each receiver gets assigned a numbered position. The IDs and the position of the receivers in Nordfjord, with the coordinates given in UTM coordinates, can be seen in **Figure 2.3** to **2.5**. See Appendix A.1 for a list of all the receivers with name, ID, coordinates, zone and position number. In addition, information about all the fish in the study with at least one detection is presented in Appendix A.2.

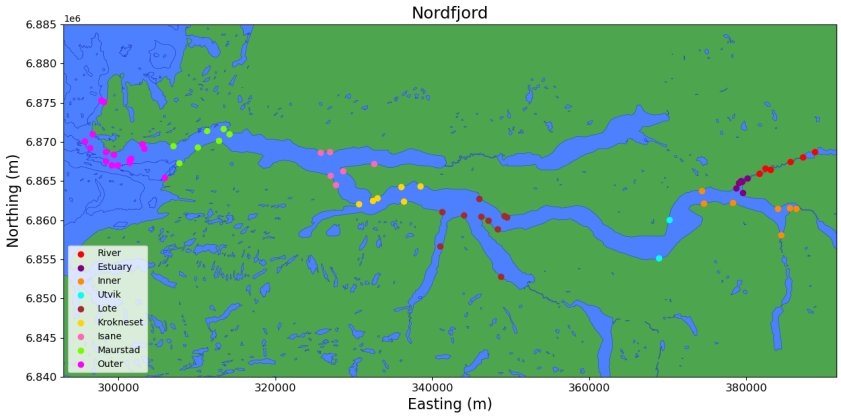


Figure 2.2: Nordfjord with receivers. The receivers are placed in nine different zones. UTM coordinates, zone 32, are used.

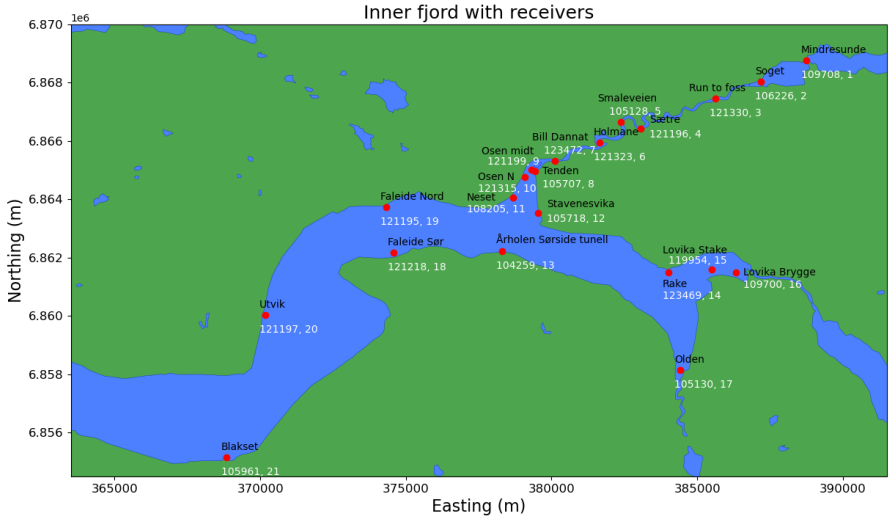


Figure 2.3: The inner part of Nordfjord with receivers. The name of the receivers are given in black, while the ID and the position of the receivers are given in white. UTM coordinates, zone 32, are used.

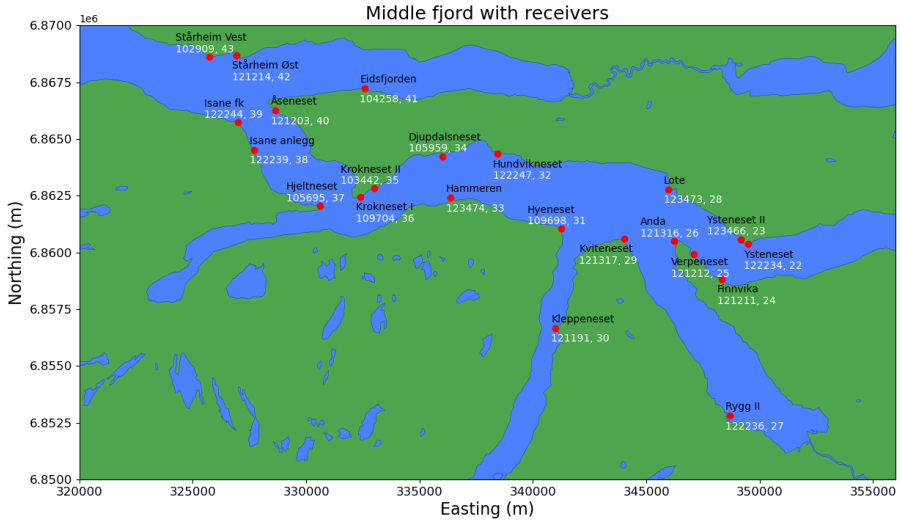


Figure 2.4: The middle part of Nordfjord with receivers. The name of the receivers are given in black, while the ID and the position of the receivers are given in white. UTM coordinates, zone 32, are used.

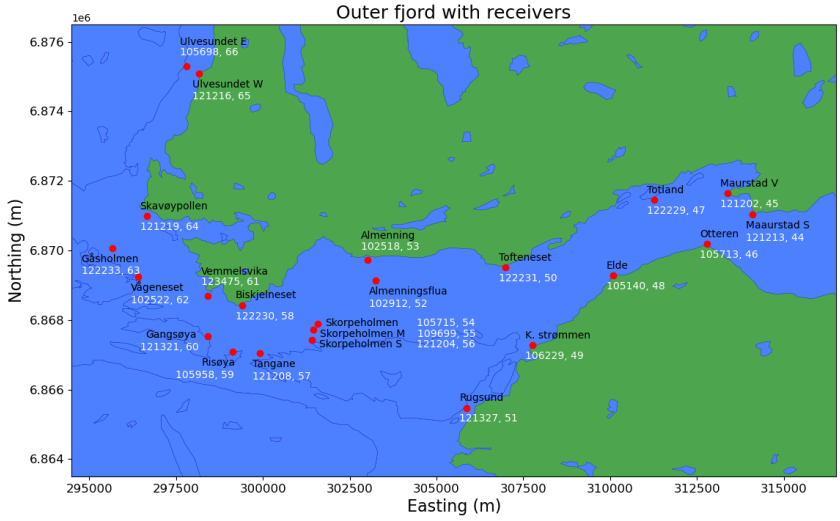


Figure 2.5: The outer part of Nordfjord with receivers. The name of the receivers are given in black, while the ID and the position of the receivers are given in white. UTM coordinates, zone 32, are used.

After combining the detection data with the information about the receivers, the data contain information about the name, ID and coordinates of each receiver, the ID of the fish registered and the date and time for each detection. A sample of the data is presented in **Table 2.1**. The columns "Name" and "Receiver" refer to the name and ID of the receiver. "Northing" gives the north coordinate of the receiver, and "Easting" gives the east coordinate. UTM coordinates, zone 32, are used. "Transmitter" is the transmitter ID of the detected fish and "Date & Time" gives the date and time of the detection. The transmitters emit information with an interval between 30 and 90 seconds, but the interval between each detection varies. Even though a fish is detected at the same receiver several times, the time between registrations may be higher than 90 seconds because all transmitter signals are not necessarily registered by the receivers. In addition to the fields presented in the table, the data also contain information about the zone of the receivers, the sensor value of the registrations and the length and weight of the fish.

Name	Receiver	Easting	Northing	Transmitter	Date & Time
Run to foss	121330	385615	6867449	2935	2017-04-29 02:28:11
Run to foss	121330	385615	6867449	2935	2017-04-29 02:33:13
Run to foss	121330	385615	6867449	2935	2017-04-29 02:35:07
Run to foss	121330	385615	6867449	2935	2017-04-29 02:35:46
...	...	...	...	...	...
Sætre	121196	383056	6866420	100	2017-04-30 00:28:01
Sætre	121196	383056	6866420	100	2017-04-30 00:28:47
Smaleveien	105128	382368	6866651	100	2017-04-30 01:24:08
Smaleveien	105128	382368	6866651	100	2017-04-30 01:27:06
...	...	...	...	...	...

Table 2.1: A sample of the data.

2.3 Data Analysis

We next conduct data analysis for the study described in Chapter 2.2. Analysis of detections at the different receivers are presented, together with analysis of the movement pattern of the salmon on both a general basis and on an individual basis.

Detections at the different receivers

Of the 66 receivers placed in Nordfjord, 57 of them have at least one detection of a salmon during 2017. The receivers placed in the river and estuary have registered most unique salmon. At Sætre, 69 unique salmon are detected and 68 salmon are detected at Smaleveien and Bill Dannat. The receivers Mindresunde, Soget, Ysteneset II, Kleppeneset, Maurstad V, Rugsund, Vemmelsvika, Ulvesundet W and Ulvesundet E do not have any detections during the period. **Figure 2.6** shows the ratio of unique fish out of the total number of fish detected at least once at each receiver. The receivers are numbered based on their position in Nordfjord, as described in Chapter 2.2.

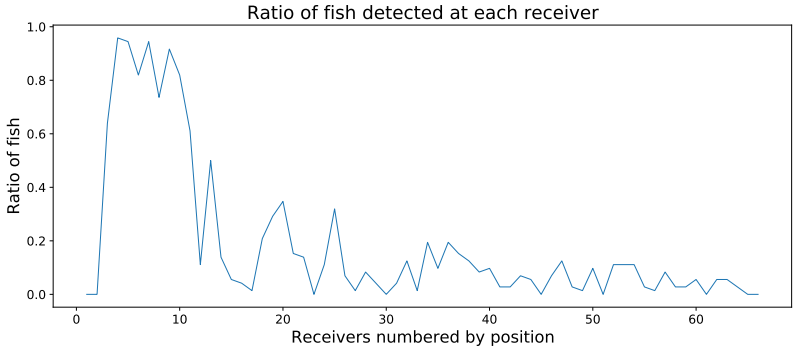


Figure 2.6: Ratio of salmon detected at each receiver. The receivers are numbered by position.

Detection pattern of the migrating salmon

The cleaned data contain detections of 72 of the 118 tagged salmon. 49 salmon are detected in the middle part, and 27 salmon are detected in the outer part of Nordfjord, as presented in **Table 2.2**. The mean time between the first and last registration is found to be approximately 10.5 days, while the median time is approximately 7.6 days. For the 27 salmon that are registered in the outer part of Nordfjord, the mean time between the first registration and the last registration is 11.1 days and the median time is 7.9 days. This is somewhat higher than the times found for all salmon. The results are presented in **Table 2.3**.

	Number	Percentage
Detected in inner fjord	72	61 %
Detected in middle fjord	49	42 %
Detected in outer fjord	27	23 %

Table 2.2: Number and percentage of fish detected in the inner, middle and outer parts of the fjord.

	All fish	Outer
Mean time	10.5 days	11.1 days
Median time	7.6 days	7.9 days

Table 2.3: The mean and median time between the first and last detection. The first column presents the results for all fish. The last column presents the results when only the fish that are registered in the outer parts of Nordfjord are considered.

The migration is triggered by an increase of water discharge (Urke et al., 2018). **Figure 2.7** shows which dates the salmon are detected for the first time. The salmon mainly start their migration in two periods, which can be seen from the figure. There is a small increase in migrations on May 5 and 7, and a large increase in migrations on May 16. In **Figure 2.8**, the dates of the first detections are presented when the salmon are divided into two groups

based on whether or not they are registered in the outer part of the fjord. A higher ratio of the salmon detected in the outer part of Nordfjord migrate before May 16 compared to the salmon not detected in the outer part.

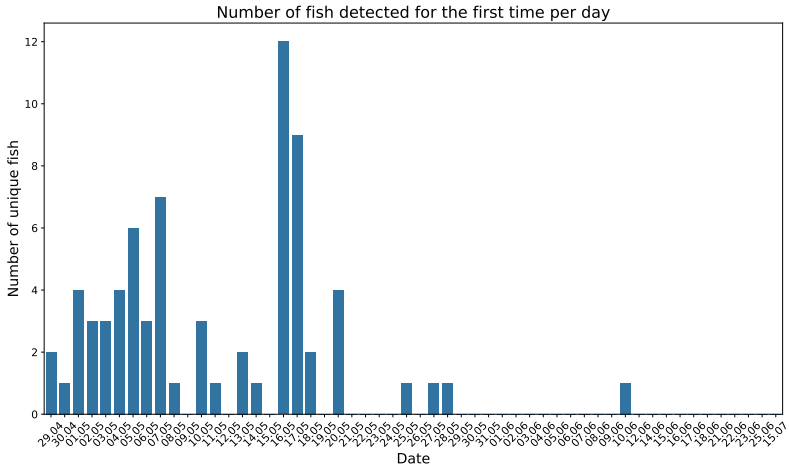


Figure 2.7: Number of salmon registered for the first time at different dates.

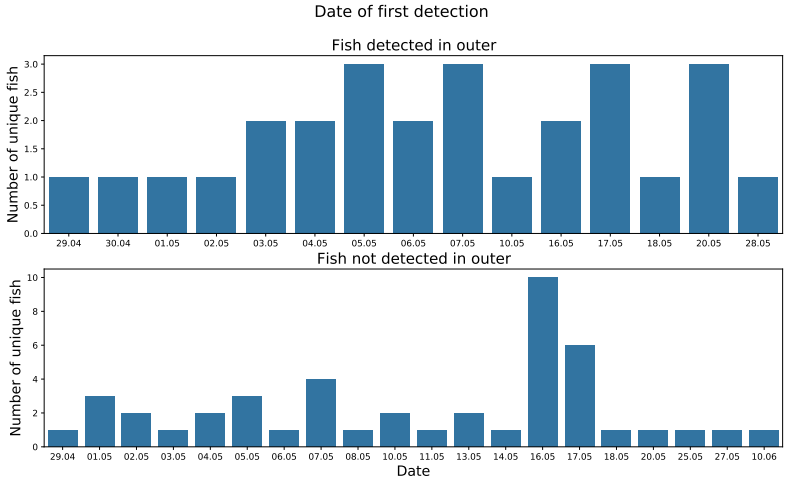


Figure 2.8: Number of salmon registered for the first time at different dates. The salmon are divided into two groups, the ones that are detected in the outer part of the fjord, and the ones that are not.

The number of salmon detected for the last time at each date are presented in **Figure 2.9**. The dates for the when the salmon are last detected when dividing the salmon based on whether or not they are registered in the outer part of the fjord are presented in **Figure 2.10**. There is a more unclear pattern to the display of the last detection dates compared to the dates of the first detections, presented in **Figure 2.7** and **2.8**. Most salmon are only registered in May, and only 8 of the 72 salmon are registered in June.

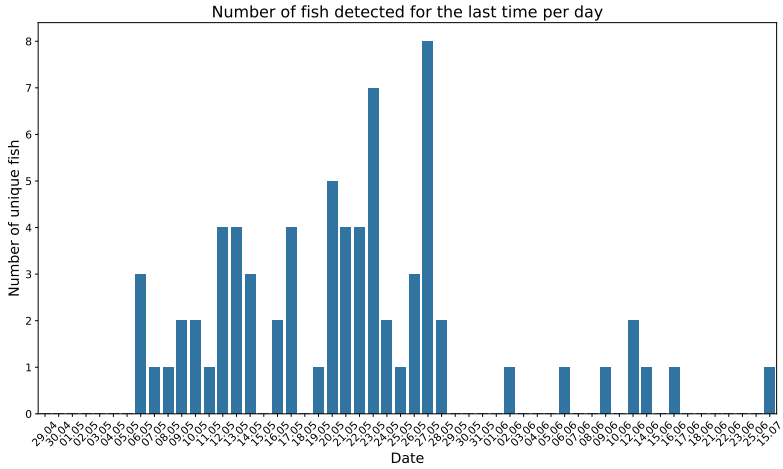


Figure 2.9: Number of salmon registered for the last time at different dates.

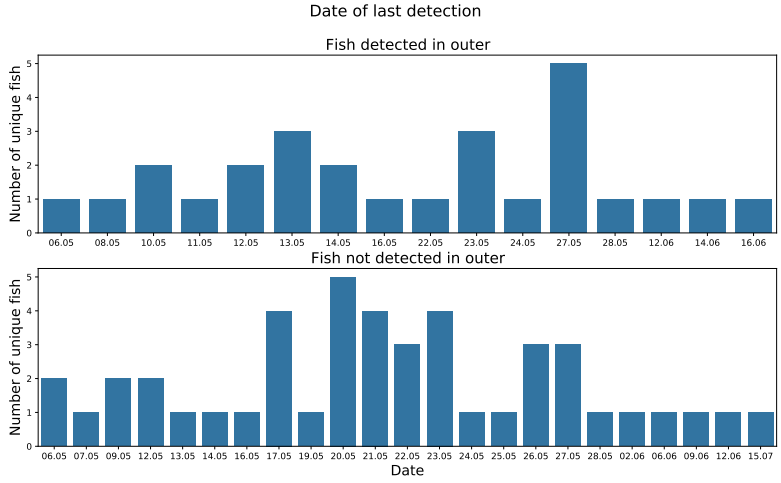


Figure 2.10: Number of salmon registered for the last time at different dates. The salmon are divided into two groups, the ones that are detected in the outer part of the fjord, and the ones that are not.

The total number of unique fish detected per day is presented in **Figure 2.11**. The main part of the detections happens in May. After June, there are few salmon registered. After June 16, we only have detections of one unique salmon. From **Figure 2.10**, we see that this fish is among the salmon not detected in the outer part of Nordfjord.

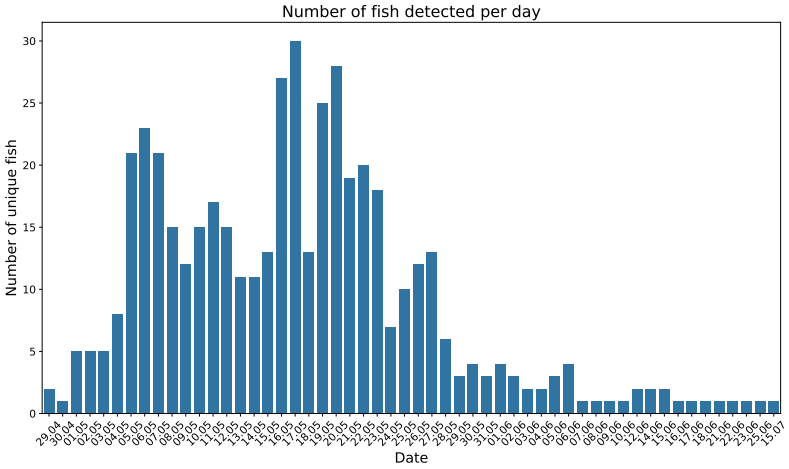


Figure 2.11: Number of unique salmon registered on each day in the migration period.

Sætre is the receiver where most salmon are registered, and this is where most of the salmon are registered for the first time. **Figure 2.12** and **2.13** compare the detections at the receivers Sætre and Osen N. Sætre is placed in the beginning of the river, while Osen N is in the estuary. The Euclidean distance between the receivers is approximately 4.3 km. **Figure 2.12** presents the number of unique fish registered at the two receivers at different dates. At Sætre there are clear peaks for detections on May 5 and May 16, which coincide with the dates for the first detections presented in **Figure 2.7**. At Osen N, the pattern is similar, with peaks on May 6 and 7 and on May 17. This indicates that the salmon can reach Osen N from Sætre in one day. By comparing **Figure 2.7** and **2.12**, we can see that there are a higher number of unique fish registered at Sætre compared to the number of fish first registered at the different dates. This is natural, as the first figure only counts the fish that have not been registered at an earlier time. When counting the number of fish registered at a specific receiver at a given date, all registrations that day are considered.

Figure 2.13 presents the total number of registrations at Sætre and Osen N. The difference from **Figure 2.12** is that each fish can be counted several times. The number of detections at Sætre is higher than at Osen N, which indicates that the salmon have a tendency to spend more time in the beginning of the river compared to the estuary. Salmon can be registered by a receiver before it starts migrating, and this can lead to a high number of registrations in the upper parts of the river.

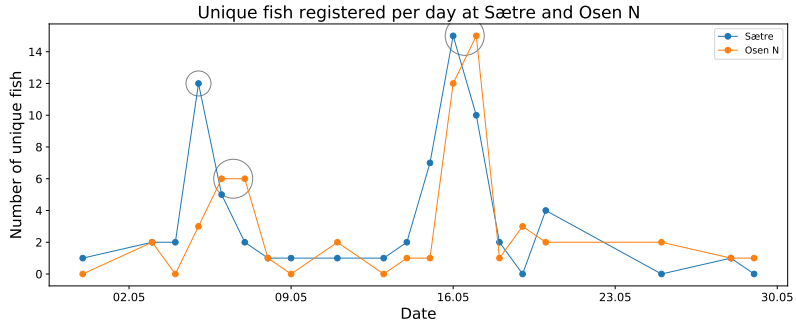


Figure 2.12: Number of unique fish registered at Sætre and Osen N. The grey circles mark the two detection peaks at the receivers.

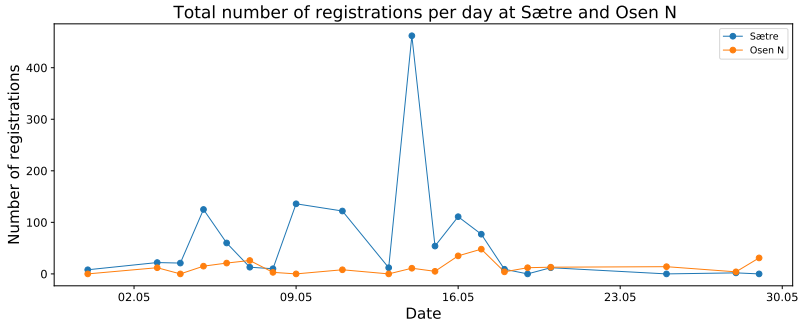


Figure 2.13: Total number of registrations at Sætre and Osen N.

Krokneset I is placed in the middle of the fjord, approximately 50.8 km from Sætre when considering the Euclidean distance. From **Figure 2.14** it is clear that there are fewer salmon registered at this receiver compared to Sætre, and that the peaks for the most registered fish have the same pattern as at Sætre, but at a later date. The dates of the first top are on May 5 and May 11 at Sætre and Krokneset I, respectively. The second top are on May 16 at Sætre and on May 23 at Krokneset I. This can indicate that the salmon use approximately one week to get from Sætre to Krokneset I.

Detection pattern of some specific fish

Figure 2.15 presents the detections of four unique fish. Fish ID 100 is first registered at Sætre on April 30. It is detected in the outer part of the estuary on May 3, on the receiver Neset. The last registration is on May 6 on the receiver Almenningsflua, in the outer part of the fjord. Fish ID 42 has a similar pattern to fish ID 100, but it stays longer in the river before it is detected in the estuary for the first time. The last detection is on Almenning, on May 22, which is also among the outermost receivers. Both Fish ID 100 and 42 are first detected in the beginning of May.

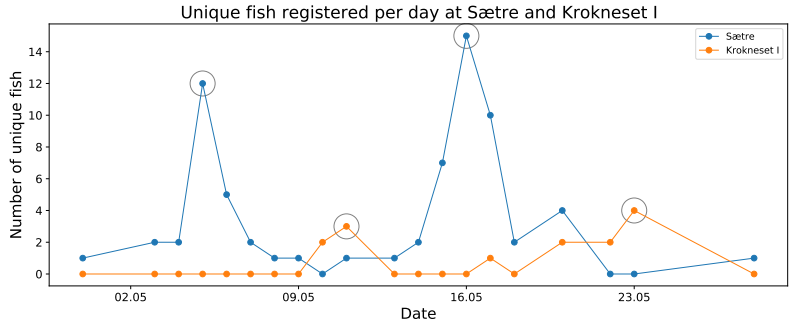


Figure 2.14: Number of unique fish registered at Sætre and Krokneset I. The grey circles mark the two detection peaks at the receivers.

From the data presented in Appendix A.2 it can be seen that the salmon with ID 157 is detected at 24 different receivers. This is the highest number of receivers for any of the salmon. We see that fish ID 157 migrates in the second group and that it turns around when it reaches Gåsholmen on May 27. It is last registered on Toftneset on June 6. Fish ID 160 is also in the second migration group and it starts from the same position as the three other salmon. It does not go further than Hundvikneset for the entire period.

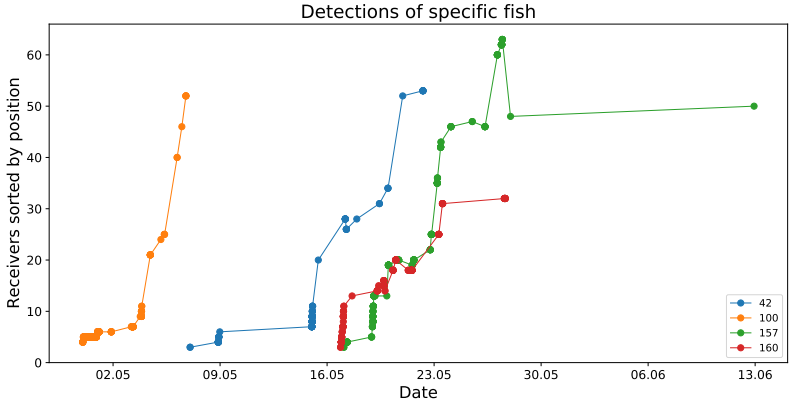
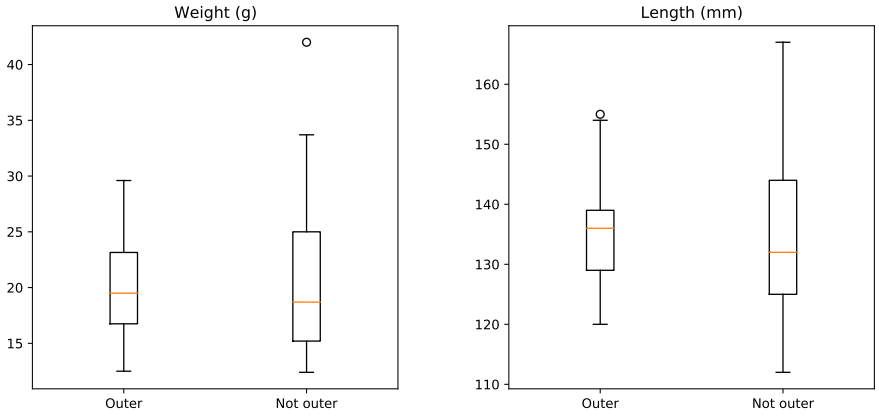


Figure 2.15: Detections of fish IDs 42, 100, 157 and 160.

Physical properties of the salmon

When the salmon are captured and tagged, their weight and length are recorded. The data of all the salmon detected at least once during the migration are presented in Appendix A.2. The mean weight of the 72 salmon is 20.2 g, with a standard error of 5.9 g. The mean length is 135.1 mm, with a standard error of 11.5 mm. A total of 27 salmon are

registered in the outer part of the fjord. The mean weight and length of these are 20.2 g and 136.1 mm, respectively. The mean weight and length of the 45 salmon not detected in the outer part are 20.2 g and 134.5 mm, respectively. Box plots of the weight and length of the two groups are presented in **Figure 2.16**. By performing a Welch's t-test (Welch, 1947; Heumann & Shalabh, 2016) with null hypothesis of the two means being identical, the p-value is found to be 0.95 and 0.53 for the weight and length respectively. Hence, we do not reject the null hypothesis, and we conclude that there are no significant difference in the means of either weight or length.



(a) Weight.

(b) Length.

Figure 2.16: Box plots of the weight and length of the salmon. The salmon are divided into two groups, the ones that are detected in the outer part of the fjord, and the ones that are not.

Chapter 3

Sequential Filtering and Smoothing Methods

This chapter covers the theory behind Bayesian filtering and smoothing equations. This includes the theory behind the Kalman filter and smoother, the particle filter and a sequential fixed-lag smoother. Bayesian filtering and smoothing equations are used to compute solutions to linear Gaussian and non-linear/non-Gaussian state-space models (Särkkä, 2013).

3.1 Probabilistic State-Space Models

The states $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots\}$ of a time-varying system can be estimated by filtering and smoothing methods when the states are indirectly observed through noisy measurements $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots\}$. The Bayesian formulation of optimal filtering is called Bayesian filtering, in which the state of the system at a given time is estimated given the measurements up to this time. Bayesian smoothing is a class of methods often considered to be a part of the Bayesian filtering methods, where the current state of the system is estimated using both previous and future measurements (Särkkä, 2013). In Bayesian filtering and smoothing, a collection of dynamic variables that fully describes the system makes up the state of the system.

The case described in Chapter 2 can be considered a target tracking problem on which we can use Bayesian filtering and smoothing to predict the state. We have a set of passive receivers that detects signals from transmitters attached to the salmon. The state we want to predict contains the position and velocity of the salmon. The available measurements are only that of presence or absence in the vicinity of a receiver. Hence, the measurement vector contains the coordinates of the receiver that the salmon have been in the vicinity of, and not the actual position of the salmon. If we only consider the east and north direction, not the depth, and we consider a specific fish with a measurement at time t, the state vector $\boldsymbol{x}_t$ and the measurement vector $\boldsymbol{z}_t$ can be written as

$$\begin{aligned} \boldsymbol{x}_t &= \begin{pmatrix} E_t & N_t & v_{E,t} & v_{N,t} \end{pmatrix}^\mathsf{T}, \\ \boldsymbol{z}_t &= \begin{pmatrix} z_{E,t} & z_{N,t} \end{pmatrix}^\mathsf{T}. \end{aligned} \tag{3.1}$$

A probabilistic state-space model describes the conditional dependence of the state of the system, $\boldsymbol{x}_t \in \mathbb{R}^n$, and the observed measurement, $\boldsymbol{z}_t \in \mathbb{R}^m$. This behavior is illustrated in **Figure 3.1**. We start from a prior distribution $p(\boldsymbol{x}_0)$. At time step t, the hidden layer is $\boldsymbol{x}_t$ and the observed layer is $\boldsymbol{z}_t$. The main goal is to estimate the hidden layers $\boldsymbol{x}_{0:T} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots \boldsymbol{x}_T\}$ given the observed measurements $\boldsymbol{z}_{1:T} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots \boldsymbol{z}_T\}$. In Bayesian filtering, the goal at each time step t is to estimate the hidden layer $\boldsymbol{x}_t$ based on the observed measurements $\boldsymbol{z}_{1:t}$. In Bayesian smoothing, the hidden layer $\boldsymbol{x}_t$ is estimated based on the observed measurements up to a time step $T > t$, $\boldsymbol{z}_{1:T}$.

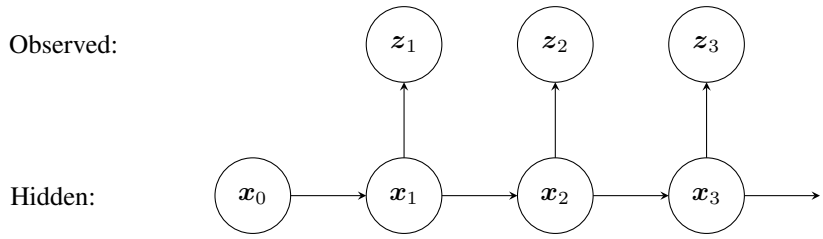


Figure 3.1: Illustration of a hidden Markov model, where $\boldsymbol{z}_t$ refers to the observed layer and $\boldsymbol{x}_t$ is the hidden layer at time step t. The hidden layer, $\boldsymbol{x}_t$, is observed indirectly through the noisy measurement, $\boldsymbol{z}_t$.

The Markov property of the sequence of states $\{\boldsymbol{x}_t : t = 0, 1, 2, \dots\}$ says that $p(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1})$, that is, the conditional density of $\boldsymbol{x}_t$ at the present time step t given $\boldsymbol{x}_{0:t-1}$, only depends on the state at the previous time step $t-1$. Hence, it is conditionally independent of what takes place at all time steps before $t-1$:

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1}) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}). \tag{3.2}$$

In **Figure 3.1**, this relation is indicated by the horizontal edges. The Markov property also states that the past is independent of the future, given the present,

$$p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_{t:T}) = p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t), \tag{3.3}$$

where $T > t$. If a sequence of states satisfies the Markov property stated in Equation (3.2) and (3.3), the states form a Markov sequence.

The measurement $\boldsymbol{z}_t$ at time step t, given the state $\boldsymbol{x}_t$, is conditionally independent of the measurement and state histories if

$$p(\boldsymbol{z}_t|\boldsymbol{x}_{0:t}, \boldsymbol{z}_{1:t-1}) = p(\boldsymbol{z}_t|\boldsymbol{x}_t). \tag{3.4}$$

This behavior is indicated by the vertical edges in **Figure 3.1**.

With conditionally independent data, this also holds for the conditional density of $\boldsymbol{x}_t$ at time step t given $\boldsymbol{x}_{0:t-1}$ and $\boldsymbol{z}_{1:t-1}$:

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1}, \boldsymbol{z}_{1:t-1}) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}), \tag{3.5}$$

and for

$$p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_{t:T}, \boldsymbol{z}_{t:T}) = p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t). \tag{3.6}$$

Assuming the states form a Markov sequence and that the measurements are conditionally independent of the measurement and state histories, a probabilistic state-space model can be written as a sequence of conditional probability distributions where $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ is the dynamic model of the state and $p(\boldsymbol{z}_t|\boldsymbol{x}_t)$ is the measurement model, for the time steps $t = 1, 2, \ldots$.

3.2 Bayesian Filtering Equations

In Bayesian filtering, the goal is to compute the filtering distribution, that is the marginal posterior distribution, of the state $\boldsymbol{x}_t$ at each time step t given all the measurements up to t:

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}), \tag{3.7}$$

where the observation vector from time 1 to t, including both 1 and t, is denoted by $\boldsymbol{z}_{1:t}$.

Assuming that the marginal posterior distribution of the previous time step is known, $p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1})$, the joint distribution of $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$ given $\boldsymbol{z}_{1:t-1}$ can be computed as

$$\begin{aligned} p(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1}) &= p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{z}_{1:t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1}) \\ &= p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1}), \end{aligned} \tag{3.8}$$

where the last equality comes from the Markov assumption of the states defined in Equation (3.5). Integrating over $\boldsymbol{x}_{t-1}$ gives

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}) = \int p(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1})d\boldsymbol{x}_{t-1}. \tag{3.9}$$

By inserting the result from Equation (3.8) into Equation (3.9), we get the following expression for the predicted distribution of the state $\boldsymbol{x}_t$:

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1})\mathrm{d}\boldsymbol{x}_{t-1}. \tag{3.10}$$

The filtering distribution of the state $\boldsymbol{x}_t$ at time step t, given the measurements $\boldsymbol{z}_{1:t}$, can be computed using Bayes' rule

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) &= \frac{p(\boldsymbol{z}_t|\boldsymbol{x}_t, \boldsymbol{z}_{1:t-1})p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1})}{p(\boldsymbol{z}_t|\boldsymbol{z}_{1:t-1})} \\ &= \frac{p(\boldsymbol{z}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1})}{\int p(\boldsymbol{z}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1})d\boldsymbol{x}_t} \\ &= \frac{1}{Z_t}p(\boldsymbol{z}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}), \end{aligned} \tag{3.11}$$

where the second equality comes from the conditional independence of $\boldsymbol{z}_t$ of the measurement history, as defined in Equation (3.4). In the above equation, Z_t is defined as

$$Z_t = \int p(\boldsymbol{z}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1})\mathrm{d}\boldsymbol{x}_t. \tag{3.12}$$

If the state contains some discrete components, the corresponding integrals are replaced with summations.

Equation (3.8) to (3.12) give the Bayesian filtering equations and can be summarized as follows:

Initialization: The recursion starts from the prior distribution $p(\boldsymbol{x}_0)$.

Prediction step: The predicted distribution of the state $\boldsymbol{x}_t$ at time step t is given by

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1})\mathrm{d}\boldsymbol{x}_{t-1}. \tag{3.13}$$

Update step: The filtering distribution of the state $\boldsymbol{x}_t$ at time step t is given by

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) = \frac{1}{Z_t}p(\boldsymbol{z}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}), \tag{3.14}$$

where Z_t is the following normalization constant

$$Z_t = \int p(\boldsymbol{z}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1})\mathrm{d}\boldsymbol{x}_t. \tag{3.15}$$

3.3 Bayesian Smoothing Equations

In Bayesian filtering, the measurements before and at the current time step are used. In smoothing, all the measurements up to a certain time step are used, including future measurements. The goal of Bayesian smoothing is to compute the marginal posterior distribution of the state $\boldsymbol{x}_t$ at time step t, given all the measurements up to a time step T, where $T > t$:

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:T}), \tag{3.16}$$

which is also called the smoothed distribution of time step t.

Instead of conditioning on the measurements up to a fixed time step T, a fixed-lag smoother is an alternative smoothing approach. This method uses a fixed delay L, called a lag, between the measurement and the state estimation. The goal is then to estimate the state $\boldsymbol{x}_t$ given the smoothing density at time step $t + L$, $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t+L})$. With this approach it is not necessary to have all measurements up to a fixed time step T. The states can be estimated in real time, only with a fixed delay. The theory behind the fixed-lag smoother is presented in Chapter 3.7.

The distribution of $\boldsymbol{x}_{t+1}$ given $\boldsymbol{z}_{1:t}$ can be expressed as

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t}) = \int p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})d\boldsymbol{x}_t. \tag{3.17}$$

Based on the model illustrated in **Figure 3.1**, the distribution of $\boldsymbol{x}_t$ given $\boldsymbol{x}_{t+1}$ and $\boldsymbol{z}_{1:T}$ is

given by

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:T}) &= p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:t}) \\ &= \frac{p(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t})}{p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t})} \\ &= \frac{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{z}_{t:1})p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})}{p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t})} \\ &= \frac{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})}{p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t})}, \end{aligned} \tag{3.18}$$

using Bayes's rule and the Markov property of state $\boldsymbol{x}_t$, that is, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:T}) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:t})$. Hence, the joint distribution of $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$ given $\boldsymbol{z}_{1:T}$ is

$$\begin{aligned} p(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:T}) &= p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:T})p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:T}) \\ &= p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:t})p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:T}) \\ &= \frac{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:T})}{p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t})}. \end{aligned} \tag{3.19}$$

The marginal distribution of $\boldsymbol{x}_t$ given $\boldsymbol{z}_{1:T}$ is found by integrating over $\boldsymbol{x}_{t+1}$ in Equation (3.19):

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:T}) = p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) \int \left[\frac{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:T})}{p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t})} \right] d\boldsymbol{x}_{t+1}. \tag{3.20}$$

Here, $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})$ is the filtering distribution defined in Equation (3.11) and $p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)$ is the model equation given in Equation (3.2). The smoothing distribution from the subsequent step is denoted by $p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:T})$ and $p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t})$ is the one-step prediction found from Equation (3.17). The calculations are done in a backward recursion.

3.4 Kalman Filter

When the dynamic model and the measurement model of a system are linear Gaussian, the closed form solution to the Bayesian filtering equations for the filtering model is the Kalman filter. In probabilistic terms we have that

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) &= \mathrm{N}(\boldsymbol{x}_t|\boldsymbol{A}_{t-1}\boldsymbol{x}_{t-1}, \boldsymbol{Q}_{t-1}), \\ p(\boldsymbol{z}_t|\boldsymbol{x}_t) &= \mathrm{N}(\boldsymbol{z}_t|\boldsymbol{B}_t\boldsymbol{x}_t, \boldsymbol{R}_t), \end{aligned} \tag{3.21}$$

where $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2)$ denotes the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}^2$. This can be written using random variables directly:

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{A}_{t-1}\boldsymbol{x}_{t-1} + \boldsymbol{q}_{t-1}, \\ \boldsymbol{z}_t &= \boldsymbol{B}_t\boldsymbol{x}_t + \boldsymbol{r}_t. \end{aligned} \tag{3.22}$$

In Equation (3.22), the state of the system is denoted by $\boldsymbol{x}_t \in \mathbb{R}^n$ and the measurement by $\boldsymbol{z}_t \in \mathbb{R}^m$. Moreover, $\boldsymbol{A}_{t-1}$ is the transition matrix of the dynamic model, $\boldsymbol{B}_t$ is the

measurement model matrix, $\boldsymbol{q}_{t-1} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_{t-i})$ is the process noise and $\boldsymbol{r}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{R}_t)$ is the measurement noise. The prior distribution of the state is Gaussian, $\boldsymbol{x}_0 \sim \mathrm{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$.

The Bayesian filtering equations for the linear filtering model in Equation (3.22) can be evaluated on closed form and in the following paragraphs we will show that the resulting distributions are Gaussian. If we have

$$p\left(\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{pmatrix}\right) = \mathrm{N}\left(\begin{pmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_z\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xz}\\ \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_z\end{pmatrix}\right), \tag{3.23}$$

then $p(\boldsymbol{x}) = \mathrm{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$, $p(\boldsymbol{z}) = \mathrm{N}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$ and the conditional distributions of $\boldsymbol{x}$ given $\boldsymbol{z}$ and $\boldsymbol{z}$ given $\boldsymbol{x}$ are

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{z}) &= \mathrm{N}\left(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_z), \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\Sigma}_{zx}\right),\\ p(\boldsymbol{z}|\boldsymbol{x}) &= \mathrm{N}\left(\boldsymbol{\mu}_z + \boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x), \boldsymbol{\Sigma}_z - \boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xz}\right). \end{aligned} \tag{3.24}$$

If $p(\boldsymbol{x}) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $p(\boldsymbol{z}|\boldsymbol{x}) = \mathrm{N}(\boldsymbol{B}\boldsymbol{x} + \boldsymbol{u}, \boldsymbol{R})$, then we know that the joint distribution of $\boldsymbol{x}$ and $\boldsymbol{z}$ is

$$p\left(\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{pmatrix}\right) = \mathrm{N}\left(\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{B}\boldsymbol{\mu} + \boldsymbol{u}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{B}^\intercal\\ \boldsymbol{B}\boldsymbol{\Sigma} & \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\intercal + \boldsymbol{R}\end{pmatrix}\right). \tag{3.25}$$

At time step t, the distribution of $p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1})$ is known from the previous time step. Then the marginal distribution of $\boldsymbol{x}_t$ given $\boldsymbol{z}_{1:t-1}$ is

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}) &= p(\boldsymbol{A}_{t-1}\boldsymbol{x}_{t-1} + \boldsymbol{q}_{t-1}|\boldsymbol{z}_{1:t-1})\\ &= \mathrm{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_t^-, \boldsymbol{\Sigma}_t^-), \end{aligned} \tag{3.26}$$

where

$$\begin{aligned} \boldsymbol{\mu}_t^- &= \boldsymbol{A}_{t-1}\boldsymbol{\mu}_{t-1},\\ \boldsymbol{\Sigma}_t^- &= \boldsymbol{A}_{t-1}\boldsymbol{\Sigma}_{t-1}\boldsymbol{A}_{t-1}^\intercal + Q_{t-1}. \end{aligned} \tag{3.27}$$

To find the marginal distribution of $\boldsymbol{x}_t$ given $\boldsymbol{z}_{1:t}$, we first find the joint distribution of $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ given $\boldsymbol{z}_{1:t-1}$:

$$\begin{aligned} p(\boldsymbol{x}_t, \boldsymbol{z}_t|\boldsymbol{z}_{1:t-1}) &= p(\boldsymbol{z}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1})\\ &= p(\boldsymbol{B}_t\boldsymbol{x}_t + \boldsymbol{r}_t|\boldsymbol{x}_t)p(\boldsymbol{A}_{t-1}\boldsymbol{x}_{t-1} + \boldsymbol{q}_{t-1}|\boldsymbol{z}_{1:t-1})\\ &= \mathrm{N}(\boldsymbol{z}_t|\boldsymbol{B}_t\boldsymbol{x}_t, \boldsymbol{R}_t)\mathrm{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_t^-, \boldsymbol{\Sigma}_t^-)\\ &= \mathrm{N}\left(\begin{pmatrix}\boldsymbol{x}_t\\ \boldsymbol{z}_t\end{pmatrix}|\boldsymbol{\mu}', \boldsymbol{\Sigma}'\right), \end{aligned} \tag{3.28}$$

where

$$\begin{aligned} \boldsymbol{\mu}' &= \begin{pmatrix}\boldsymbol{\mu}_t^-\\ \boldsymbol{B}_t\boldsymbol{\mu}_t^-\end{pmatrix},\\ \boldsymbol{\Sigma}' &= \begin{pmatrix}\boldsymbol{\Sigma}_t^- & \boldsymbol{\Sigma}_t^-\boldsymbol{B}_t^\intercal\\ \boldsymbol{B}_t\boldsymbol{\Sigma}_t^- & \boldsymbol{B}_t\boldsymbol{\Sigma}_t^-\boldsymbol{B}_t^\intercal + \boldsymbol{R}_t\end{pmatrix}. \end{aligned} \tag{3.29}$$

Combining this result with Equation (3.24) we find the marginal distribution of $\boldsymbol{x}_t$ given $\boldsymbol{z}_{1:t}$

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{z}_t, \boldsymbol{z}_{1:t-1}) &= p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) \\ &= \mathrm{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \end{aligned} \tag{3.30}$$

where

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_t^- + \boldsymbol{\Sigma}_t^- \boldsymbol{B}_t^\intercal (\boldsymbol{B}_t \boldsymbol{\Sigma}_t^- \boldsymbol{B}_t^\intercal + \boldsymbol{R}_t)^{-1} (\boldsymbol{z}_t - \boldsymbol{B}_t \boldsymbol{\mu}_t^-) \tag{3.31}$$

and

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^- - \boldsymbol{\Sigma}_t^- \boldsymbol{B}_t^\intercal (\boldsymbol{B}_t \boldsymbol{\Sigma}_t^- \boldsymbol{B}_t^\intercal + \boldsymbol{R}_t)^{-1} \boldsymbol{B}_t \boldsymbol{\Sigma}_t^- . \tag{3.32}$$

Equation (3.26) to (3.32) can be summarized in the following way:

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}) &= \mathrm{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_t^-, \boldsymbol{\Sigma}_t^-), \\ p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) &= \mathrm{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \\ p(\boldsymbol{z}_t|\boldsymbol{z}_{1:t-1}) &= \mathrm{N}(\boldsymbol{z}_t|\boldsymbol{B}_t \boldsymbol{\mu}_t^-, \boldsymbol{S}_t), \end{aligned} \tag{3.33}$$

where the parameters can be calculated in the following Kalman filter steps:

Prediction step:

$$\begin{aligned} \boldsymbol{\mu}_t^- &= \boldsymbol{A}_{t-1} \boldsymbol{\mu}_{t-1}, \\ \boldsymbol{\Sigma}_t^- &= \boldsymbol{A}_{t-1} \boldsymbol{\Sigma}_{t-1} \boldsymbol{A}_{t-1}^\intercal + \boldsymbol{Q}_{t-1}. \end{aligned} \tag{3.34}$$

Update step:

$$\begin{aligned} \boldsymbol{v}_t &= \boldsymbol{z}_t - \boldsymbol{B}_t \boldsymbol{\mu}_t^-, \\ \boldsymbol{S}_t &= \boldsymbol{B}_t \boldsymbol{\Sigma}_t^- \boldsymbol{B}_t^\intercal + \boldsymbol{R}_t, \\ \boldsymbol{K}_t &= \boldsymbol{\Sigma}_t^- \boldsymbol{B}_t^\intercal \boldsymbol{S}_t^{-1}, \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_t^- + \boldsymbol{K}_t \boldsymbol{v}_t, \\ \boldsymbol{\Sigma}_t &= \boldsymbol{\Sigma}_t^- - \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{K}_t^\intercal. \end{aligned} \tag{3.35}$$

The recursion is started from the prior mean $\boldsymbol{\mu}_0$ and the prior covariance $\boldsymbol{\Sigma}_0$.

3.5 Kalman Smoother

A method for finding the closed form smoothing solution,

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:T}) = \mathrm{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_t^s, \boldsymbol{\Sigma}_t^s), \tag{3.36}$$

to the linear filtering model presented in Equation (3.22) is the Kalman smoother, also called the Rauch-Tung-Striebel smoother (Särkkä, 2013). This is a backward recursion started from $p(\boldsymbol{x}_T|\boldsymbol{z}_{1:T}) = \mathrm{N}(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T)$, where $\boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T$ are the mean and covariance from the last time step computed by the Kalman filter in Equation (3.35).

By Equation (3.25), the joint distribution of $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$ given $\boldsymbol{z}_{1:t}$ is

$$\begin{aligned} p(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:t}) &= p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) \\ &= p(\boldsymbol{A}_t\boldsymbol{x}_t + \boldsymbol{q}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) \\ &= \mathrm{N}(\boldsymbol{x}_{t+1}|\boldsymbol{A}_t\boldsymbol{x}_t, \boldsymbol{Q}_t)\mathrm{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \\ &= N\left(\begin{pmatrix}\boldsymbol{x}_t \\ \boldsymbol{x}_{t+1}\end{pmatrix}|\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\Sigma}}_1\right), \end{aligned} \tag{3.37}$$

where

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_1 &= \begin{pmatrix}\boldsymbol{\mu}_t \\ \boldsymbol{A}_t\boldsymbol{\mu}_t\end{pmatrix}, \\ \tilde{\boldsymbol{\Sigma}}_1 &= \begin{pmatrix}\boldsymbol{\Sigma}_t & \boldsymbol{\Sigma}_t\boldsymbol{A}_t^\intercal \\ \boldsymbol{A}_t\boldsymbol{\Sigma}_t & \boldsymbol{A}_t\boldsymbol{\Sigma}_t\boldsymbol{A}_t^\intercal + \boldsymbol{Q}_t\end{pmatrix}, \end{aligned} \tag{3.38}$$

and $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ are the mean and covariance for time step t computed by the Kalman filter in Equation (3.35). By the Markov property of the states and Equation (3.24) we have

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:T}) &= p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:t}) \\ &= \mathrm{N}(\boldsymbol{x}_t|\tilde{\boldsymbol{\mu}}_2, \tilde{\boldsymbol{\Sigma}}_2), \end{aligned} \tag{3.39}$$

where

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_2 &= \boldsymbol{\mu}_t + \boldsymbol{G}_t(\boldsymbol{x}_{t+1} - \boldsymbol{A}_t\boldsymbol{\mu}_t), \\ \tilde{\boldsymbol{\Sigma}}_2 &= \boldsymbol{\Sigma}_t - \boldsymbol{G}_t(\boldsymbol{A}_t\boldsymbol{\Sigma}_t\boldsymbol{A}_t^\intercal + \boldsymbol{Q}_t)\boldsymbol{G}_t^\intercal, \\ \boldsymbol{G}_t &= \boldsymbol{\Sigma}_t\boldsymbol{A}_t^\intercal(\boldsymbol{A}_t\boldsymbol{\Sigma}_t\boldsymbol{A}_t^\intercal + \boldsymbol{Q}_t)^{-1}. \end{aligned} \tag{3.40}$$

Hence, the joint distribution of $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$ given all the data up to time step T, $\boldsymbol{z}_{1:T}$, is

$$\begin{aligned} p(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:T}) &= p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{z}_{1:T})p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{1:T}) \\ &= \mathrm{N}(\boldsymbol{x}_t|\tilde{\boldsymbol{\mu}}_2, \tilde{\boldsymbol{\Sigma}}_2)\mathrm{N}(\boldsymbol{x}_{t+1}|\boldsymbol{\mu}_{t+1}^s, \boldsymbol{\Sigma}_{t+1}^s) \\ &= N\left(\begin{pmatrix}\boldsymbol{x}_{t+1} \\ \boldsymbol{x}_t\end{pmatrix}|\tilde{\boldsymbol{\mu}}_3, \tilde{\boldsymbol{\Sigma}}_3\right), \end{aligned} \tag{3.41}$$

where

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_3 &= \begin{pmatrix}\boldsymbol{\mu}_{t+1}^s \\ \boldsymbol{\mu}_t + \boldsymbol{G}_t(\boldsymbol{\mu}_{t+1}^s - \boldsymbol{A}_t\boldsymbol{\mu}_t)\end{pmatrix}, \\ \tilde{\boldsymbol{\Sigma}}_3 &= \begin{pmatrix}\boldsymbol{\Sigma}_{t+1}^s & \boldsymbol{\Sigma}_{t+1}^s\boldsymbol{G}_t^\intercal \\ \boldsymbol{G}_t\boldsymbol{\Sigma}_{t+1}^s & \boldsymbol{G}_t\boldsymbol{\Sigma}_{t+1}^s\boldsymbol{G}_t^\intercal + \tilde{\boldsymbol{\Sigma}}_2\end{pmatrix}. \end{aligned} \tag{3.42}$$

Finally, by Equation (3.24), the marginal distribution of $\boldsymbol{x}_t$ is given as

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:T}) = \mathrm{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_t^s, \boldsymbol{\Sigma}_t^s), \tag{3.43}$$

where

$$\begin{aligned}\boldsymbol{\mu}_t^s &= \boldsymbol{\mu}_t + \boldsymbol{G}_t(\boldsymbol{\mu}_{t+1}^s - \boldsymbol{A}_t\boldsymbol{\mu}_t),\\ \boldsymbol{\Sigma}_t^s &= \boldsymbol{\Sigma}_t + \boldsymbol{G}_t(\boldsymbol{\Sigma}_{t+1}^s - \boldsymbol{A}_t\boldsymbol{\Sigma}_t\boldsymbol{A}_t^\intercal - \boldsymbol{Q}_t)\boldsymbol{G}_t^\intercal.\end{aligned} \tag{3.44}$$

Equation (3.37) to (3.44) can be summarized in the following backward recursion equations:

Kalman smoother:

$$\begin{aligned}\boldsymbol{\mu}_{t+1}^- &= \boldsymbol{A}_t\boldsymbol{\mu}_t,\\ \boldsymbol{\Sigma}_{t+1}^- &= \boldsymbol{A}_t\boldsymbol{\Sigma}_t\boldsymbol{A}_t^\intercal + \boldsymbol{Q}_t,\\ \boldsymbol{G}_t &= \boldsymbol{\Sigma}_t\boldsymbol{A}_t^\intercal\left(\boldsymbol{\Sigma}_{t+1}^-\right)^{-1},\\ \boldsymbol{\mu}_t^s &= \boldsymbol{\mu}_t + \boldsymbol{G}_t\left(\boldsymbol{\mu}_{t+1}^s - \boldsymbol{\mu}_{t+1}^-\right),\\ \boldsymbol{\Sigma}_t^s &= \boldsymbol{\Sigma}_t + \boldsymbol{G}_t\left(\boldsymbol{\Sigma}_{t+1}^s - \boldsymbol{\Sigma}_{t+1}^-\right)\boldsymbol{G}_t^\intercal.\end{aligned} \tag{3.45}$$

The recursion is started from the last time step T, with $\boldsymbol{\mu}_T^s = \boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T^s = \boldsymbol{\Sigma}_T$.

3.6 Particle Filter

Methods based on sequential importance resampling can be a better alternative than Kalman filter and smoother approaches, if, for example, the problem is non-linear or if some of the state components are discrete (Särkkä, 2013). This is the case for presence-absence data. For state-space models that are non-linear and non-Gaussian, the performance of Monte Carlo methods have been shown to be better than the performance of classical sequential techniques (Clapp & Godsill, 1999). Particle filter methods are one example of such a method, and they are a class of Monte Carlo approximations to the solutions of Bayesian filtering equations (Doucet & Johansen, 2008; Särkkä, 2013).

3.6.1 Monte Carlo Approximations in Bayesian Inference

Monte Carlo is a class of methods where statistical quantities are approximated by drawing samples form the distribution and estimating the quantities by sample averages. Monte Carlo methods provide a numerical method for calculating integrals of the form

$$\mathbb{E}[\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{z}_{1:T}] = \int \boldsymbol{g}(\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{z}_{1:T})d\boldsymbol{x}, \tag{3.46}$$

where $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ is an arbitrary function and $p(\boldsymbol{x}|\boldsymbol{z}_{1:T})$ is the posterior probability density of the state $\boldsymbol{x}$ given the measurements $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$. Here, $\boldsymbol{x}$ is treated as a continuous random variable, but it is possible to formulate analogous results for discrete random variables.

In Bayesian inference, including Bayesian filtering, the main inference problem can often be reduced into computation of expectations of the form in Equation (3.46). In Monte Carlo methods, the target distribution $p(\boldsymbol{x}|\boldsymbol{z}_{1:T})$ is approximated by a set of samples that

are distributed according to this distribution. We draw N_p independent random samples $\boldsymbol{x}^{(p)} \sim p(\boldsymbol{x}|\boldsymbol{z}_{1:T}), p = 1, \ldots, N_p$, and the expectation is estimated as

$$\mathbb{E}[\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{z}_{1:T}] \approx \frac{1}{N_p} \sum_{p=1}^{N_p} \boldsymbol{g}(\boldsymbol{x}^{(p)}). \tag{3.47}$$

3.6.2 Importance Sampling

In many cases, the distribution $p(\boldsymbol{x}|\boldsymbol{z}_{1:T})$ is complex and high-dimensional, and thus it is not possible to obtain samples directly from it. A solution to this is to use the importance sampling method, which relies on the introduction of an importance distribution, $\hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T})$. This is an approximate distribution to the target distribution from which we can easily draw samples. The support of the importance density $\hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T})$ is required to be greater than or equal to the support of $p(\boldsymbol{x}|\boldsymbol{z}_{1:T})$, that is,

$$p(\boldsymbol{x}|\boldsymbol{z}_{1:T}) > 0 \implies \hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T}) > 0. \tag{3.48}$$

The expectation over the posterior probability density $p(\boldsymbol{x}|\boldsymbol{z}_{1:T})$ can be decomposed as

$$\int \boldsymbol{g}(\boldsymbol{x}) p(\boldsymbol{x}|\boldsymbol{z}_{1:T}) d\boldsymbol{x} = \int \left[\boldsymbol{g}(\boldsymbol{x}) \frac{p(\boldsymbol{x}|\boldsymbol{z}_{1:T})}{\hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1_T})} \right] \hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T}) d\boldsymbol{x}. \tag{3.49}$$

We can draw a Monte Carlo approximation to the expression in the brackets in Equation (3.49) by drawing N_p samples from the importance distribution $\boldsymbol{x}^{(p)} \sim \hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T}), p = 1, \ldots, N_p$. Because we sample from an approximation to the target distribution, we need to correct the approximations by associating a weight with each of the samples. Hence, the Monte Carlo approximation can be formed as

$$\begin{aligned} \mathbb{E}[\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{z}_{1:T}] &\approx \frac{1}{N_p} \sum_{p=1}^{N_p} \frac{p(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T})}{\hat{p}(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T})} \boldsymbol{g}(\boldsymbol{x}^{(p)}) \\ &= \sum_{p=1}^{N_p} \tilde{w}^{(p)} \boldsymbol{g}(\boldsymbol{x}^{(p)}), \end{aligned} \tag{3.50}$$

where the weights are defined as

$$\tilde{w}^{(p)} = \frac{1}{N_p} \frac{p(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T})}{\hat{p}(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T})}. \tag{3.51}$$

This is a direct importance sampling method which requires us to be able to evaluate $p(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T})$. Bayes' rule gives that

$$p(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T}) = \frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x}^{(p)}) p(\boldsymbol{x}^{(p)})}{\int p(\boldsymbol{z}_{1:T}|\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}}. \tag{3.52}$$

In many cases, the normalization constant $\int p(\boldsymbol{z}_{1:T}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$ cannot be evaluated directly. By also approximating the normalization constant by importance sampling, this problem can be overcome:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{z}_{1:T}] &= \int \boldsymbol{g}(\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{z}_{1:T})d\boldsymbol{x} \\
&= \frac{\int \boldsymbol{g}(\boldsymbol{x})p(\boldsymbol{z}_{1:T}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}}{\int p(\boldsymbol{z}_{1:T}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}} \\
&= \frac{\int \left[\frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x})p(\boldsymbol{x})}{\hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T})}\boldsymbol{g}(\boldsymbol{x})\right]\hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T})d\boldsymbol{x}}{\int \left[\frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x})p(\boldsymbol{x})}{\hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T})}\right]\hat{p}(\boldsymbol{x}|\boldsymbol{z}_{1:T})d\boldsymbol{x}} \\
&\approx \frac{\frac{1}{N_p}\sum_{p=1}^{N_p}\frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x}^{(p)})p(\boldsymbol{x}^{(p)})}{\hat{p}(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T})}\boldsymbol{g}(\boldsymbol{x}^{(p)})}{\frac{1}{N_p}\sum_{j=1}^{N_p}\frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x}^{(j)})p(\boldsymbol{x}^{(j)})}{\hat{p}(\boldsymbol{x}^{(j)}|\boldsymbol{z}_{1:T})}} \\
&= \sum_{p=1}^{N_p}\left[\frac{\frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x}^{(p)})p(\boldsymbol{x}^{(p)})}{\hat{p}(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T})}}{\sum_{j=1}^{N_p}\frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x}^{(j)})p(\boldsymbol{x}^{(j)})}{\hat{p}(\boldsymbol{x}^{(j)}|\boldsymbol{z}_{1:T})}}\right]\boldsymbol{g}(\boldsymbol{x}^{(p)}) \\
&= \sum_{p=1}^{N_p} w^{(p)}\boldsymbol{g}(\boldsymbol{x}^{(p)}).
\end{aligned} \tag{3.53}$$

The normalized weights can be written as

$$w^{(p)} = \frac{w^{*(p)}}{\sum_{j=1}^{N_p} w^{*(j)}}, \tag{3.54}$$

where the unnormalized weights are given by

$$w^{*(p)} = \frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x}^{(p)})p(\boldsymbol{x}^{(p)})}{\hat{p}(\boldsymbol{x}^{(p)}|\boldsymbol{z}_{1:T})}. \tag{3.55}$$

Hence, an approximation to the posterior probability density can formally be written as

$$p(\boldsymbol{x}|\boldsymbol{z}_{1:t}) \approx \sum_{p=1}^{N_p} w^{(p)}\delta(\boldsymbol{x} - \boldsymbol{x}^{(p)}), \tag{3.56}$$

where $\delta(\cdot)$ is the Dirac delta function (Arfken & Weber, 2005).

3.6.3 Sequential Importance Sampling

The importance sampling method can be modified such that it becomes possible to compute an estimate of $p(\boldsymbol{x}_{0:t}|\boldsymbol{z}_{1:t})$ without modifying the previously simulated trajectories $\{\boldsymbol{x}_{0:t-1}^{(p)}, p = 1, \ldots, N_p\}$. This is called sequential importance sampling (Doucet et al., 2001). Importance sampling approximations to filtering distributions of the form described in Chapter 3.1 can be found using sequential importance sampling.

The sequential importance sampling algorithm uses a weighted set of particles $\{(w_t^{(p)}, \boldsymbol{x}_t^{(p)}) : p = 1, \ldots, N_p\}$ for representing the filtering distribution $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})$. The set of particles are samples from an importance distribution and their corresponding weights. Hence, at every time step t, the approximation of an arbitrary function $\boldsymbol{g}(\cdot)$ can be calculated as the weighted sample average

$$\mathbb{E}[\boldsymbol{g}(\boldsymbol{x}_t)|\boldsymbol{z}_{1:t}] \approx \sum_{p=1}^{N_p} w_t^{(p)} \boldsymbol{g}(\boldsymbol{x}_t^{(p)}). \tag{3.57}$$

The sequential importance sampling method can equivalently be interpreted as forming the following approximation to the filtering distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) \approx \sum_{p=1}^{N_p} w_t^{(p)} \delta(\boldsymbol{x}_t - \boldsymbol{x}_t^{(p)}). \tag{3.58}$$

The posterior distribution of the states $\boldsymbol{x}_{0:t}$ given the measurements $\boldsymbol{z}_{1:t}$ can be expressed as

$$\begin{aligned} p(\boldsymbol{x}_{0:t}|\boldsymbol{z}_{1:t}) &\propto p(\boldsymbol{z}_t|\boldsymbol{x}_{0:t}, \boldsymbol{z}_{1:t-1}) p(\boldsymbol{x}_{0:t}|\boldsymbol{z}_{1:t-1}) \\ &= p(\boldsymbol{z}_t|\boldsymbol{x}_t) p(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1}, \boldsymbol{z}_{0:t-1}) p(\boldsymbol{x}_{0:t-1}|\boldsymbol{z}_{1:t-1}) \\ &= p(\boldsymbol{z}_t|\boldsymbol{x}_t) p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) p(\boldsymbol{x}_{0:t-1}|\boldsymbol{z}_{1:t-1}) \end{aligned} \tag{3.59}$$

by using the Markov properties of the model. As in Chapter 3.6.2, we draw samples from a given importance distribution $\boldsymbol{x}_{0:t}^{(p)} \sim \hat{p}(\boldsymbol{x}_{0:t}|\boldsymbol{z}_{1:t})$. It is convenient to select the importance distribution to be Markovian in the sense that

$$\hat{p}(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1}, \boldsymbol{z}_{1:t}) = \hat{p}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{z}_{1:t}). \tag{3.60}$$

Then it is only necessary to store the current states $\boldsymbol{x}_t^{(p)}$, and not the whole histories $\boldsymbol{x}_{0:t}^{(p)}$. If we form the importance distribution for the states $\boldsymbol{x}_t$ recursively as

$$\hat{p}(\boldsymbol{x}_{0:t}|\boldsymbol{z}_{1:t}) = \hat{p}(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1}, \boldsymbol{z}_{1:t}) \hat{p}(\boldsymbol{x}_{0:t-1}|\boldsymbol{z}_{1:t-1}), \tag{3.61}$$

and use the result in Equation (3.59), the importance weights can be computed as

$$w_t^{(p)} \propto \frac{p(\boldsymbol{z}_t|\boldsymbol{x}_t^{(p)}) p(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{t-1}^{(p)})}{\hat{p}(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{0:t-1}^{(p)}, \boldsymbol{z}_{1:t})} \frac{p(\boldsymbol{x}_{0:t-1}^{(p)}|\boldsymbol{z}_{1:t-1})}{\hat{p}(\boldsymbol{x}_{0:t-1}^{(p)}|\boldsymbol{z}_{1:t-1})}. \tag{3.62}$$

If we assume that the samples $\boldsymbol{x}_{0:t-1}^{(p)}$ have already been drawn from the importance distribution $\hat{p}(\boldsymbol{x}_{0:t-1}|\boldsymbol{z}_{1:t-1})$ and that the importance weights $w_{t-1}^{(p)}$ have been calculated, samples $\boldsymbol{x}_{0:t}^{(p)}$ can be drawn from the importance distribution $\hat{p}(\boldsymbol{x}_{0:t}|\boldsymbol{z}_{1:t})$ by drawing state samples from time step t as $\boldsymbol{x}_t^{(p)} \sim \hat{p}(\boldsymbol{x}_t|\boldsymbol{x}_{0:t-1}^{(p)}, \boldsymbol{z}_{1:t})$. The importance weights from the previous time step are proportional to the last term in Equation (3.62):

$$w_{t-1}^{(p)} \propto \frac{p(\boldsymbol{x}_{0:t-1}^{(p)}|\boldsymbol{z}_{1:t-1})}{\hat{p}(\boldsymbol{x}_{0:t-1}^{(p)}|\boldsymbol{z}_{1:t-1})}, \tag{3.63}$$

and thus the weights satisfy the recursion

$$w_t^{(p)} \propto w_{t-1}^{(p)} \frac{p(\boldsymbol{z}_t|\boldsymbol{x}_t^{(p)})p(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{t-1}^{(p)})}{\hat{p}(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{0:t-1}^{(p)}, \boldsymbol{z}_{1:t})}. \tag{3.64}$$

3.6.4 Sequential Importance Resampling

A problem in the sequential importance sampling algorithm described in Chapter 3.6.3 is that it is common to encounter the situation where almost all the particles have zero or nearly zero weights. This is called the degeneracy problem (Doucet & Johansen, 2008). This problem can partially be solved by using a resampling procedure where we sample from $\{\boldsymbol{x}_t^{(p)}, p = 1, \ldots, N_p\}$, which have already been sampled from the importance distribution. If we assume that the weights have been normalized, we select $\boldsymbol{x}_t^{(p)}$ with probability $w_t^{(p)}$. The old set of N_p samples is then replaced with this new set, and the weights are set equal to $1/N_p$. This allows us to remove particles with very small weights, and duplicate particles with large weights.

When we add a resampling step to the sequential importance sampling algorithm, we obtain sequential importance resampling. This algorithm is usually referred to as the particle filter (Särkkä, 2013). Resampling introduces additional variance. Therefore it is most sensible to only resample when it is actually needed. One way to implement this is to do resampling on every nth step, where n is some predefined constant. This method is unbiased. An alternative method is called adaptive resampling. Here, we only perform resampling when the variance of the normalized weights are higher than some pre-determined threshold. The effective sample size (ESS), also called the effective number of particles,

$$\text{ESS} \approx \left(\sum_{p=1}^{N_p} \left(w_t^{(p)} \right)^2 \right)^{-1}, \tag{3.65}$$

can be used to assess the variability of the weights. In the expression, $w_t^{(p)}$ is the normalized weights of particle p at time step t. Resampling is performed when the number of effective particles is significantly less than the total number of particles, for example, $\text{ESS} < N_p/10$ or $\text{ESS} < N_p/2$ where N_p is the total number of particles (Doucet & Johansen, 2008; Särkkä, 2013).

The sequential importance resampling algorithm, hereby called the particle filter algorithm, can be summarized as follows:

At time t = 0: Draw N_p samples from the prior distribution

$$\boldsymbol{x}_0^{(p)} \sim p(\boldsymbol{x}_0) \tag{3.66}$$

and give all particles equal weights, $w_0^{(p)} = 1/N_p$ for $p = 1, \ldots, N_p$.

At time $t = 1, \ldots, T$:

1. Draw samples $\boldsymbol{x}_t^{(p)}$ from the importance distributions

$$\boldsymbol{x}_t^{(p)} \sim \hat{p}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}^{(p)}, \boldsymbol{z}_{1:t}), \qquad p = 1, \ldots, N_p. \tag{3.67}$$

2. Calculate new weights according to

$$w_t^{(p)} \propto w_{t-1}^{(p)} \frac{p(\boldsymbol{z}_t | \boldsymbol{x}_t^{(p)}) p(\boldsymbol{x}_t^{(p)} | \boldsymbol{x}_{t-1}^{(p)})}{\hat{p}(\boldsymbol{x}_t^{(p)} | \boldsymbol{x}_{0:t-1}^{(p)}, \boldsymbol{z}_{1:t})} \tag{3.68}$$

   and normalize them to sum to unity.

3. If the ESS in Equation (3.65) is too low, perform resampling.

For time step t, samples are drawn according to Equation (3.67). This equation propagates the particles forward in time. Based on the weights of the particles, the ESS is calculated, and the particles may be resampled. This behavior is illustrated in **Figure 3.2**. Black points represent particles at the current time step. Gray points are the particles at the next time step, after they have been propagated one step. Light points with a line around the points are particles that were not resampled at the current time step. In the first window, we have five particles that have been propagated forward in time according to Equation (3.67). The arrows illustrate the descendants of each particle. In the second window, the gray particles from the first window are now the particles at the current time step, and hence they are black. Resampling has been performed, and only three of the five particles have been resampled. Two of the three particles have been resampled twice. The particles are propagated forward in time, and the resulting particles are presented as gray points. The remaining two particles are presented as light points, and they do not longer contribute to the state estimation. In the final window, resampling has been performed for the last time step.

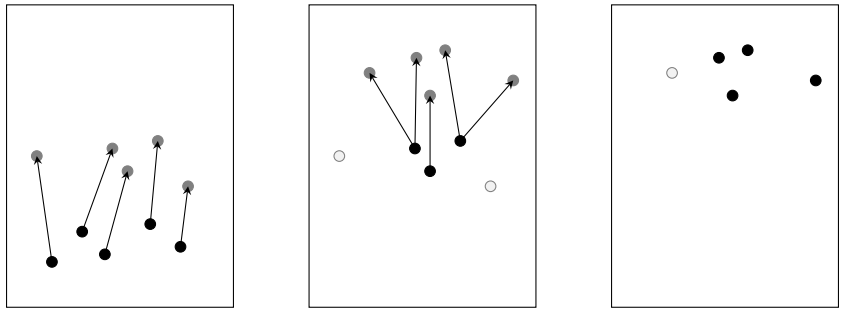


Figure 3.2: Illustration of the propagation and resampling of particles in the particle filter for three time steps. Black points represent particles at the current time step. Gray points are the particles at the next time step, after they have been propagated one step. The lightest points illustrate the particles that were not resampled.

3.7 Sequential Fixed-Lag Smoother for Particle Filters

A smoother adjustment to the particle filter is relevant in the current situation. In smoothers, more measurements are used to produce an estimate of the state $\boldsymbol{x}_t$, and hence we expect smoothers to perform better than filters, that is, produce more accurate estimates of the states (Anderson & Moore, 1979). For both the Kalman filter algorithm and the particle filter algorithm, described in Chapter 3.4 and Chapter 3.6 respectively, the aim is to assess the state $\boldsymbol{x}_t$ from the filtering density $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})$, as given in Equation (3.7).

If we allow a small delay between a measurement and the estimation of the corresponding state, it will be profitable to perform fixed-lag smoothing rather than filtering (Clapp & Godsill, 1999). For some fixed lag L, the fixed-lag smoother provides an estimate of a state $\boldsymbol{x}_t$ based on noisy measurements $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, z_{t+L}$. Hence, the goal of the fixed-lag smoother is to simulate the state $\boldsymbol{x}_{t|t+L}$ from the smoothing density at lag L, $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t+L})$.

There are several fixed-lag smoother algorithms. The following sequential fixed-lag smoother algorithm is presented by Nulsen et al. (2015), and it is used in the current situation. If $\boldsymbol{x}_t^{(i)}$ in the particle filter algorithm presented in Chapter 3.6.4 is obtained by progressing the particle $\boldsymbol{x}_{t-1}^{(p)}$ forward according to Equation (3.67), $\boldsymbol{x}_t^{(i)}$ is said to be descended from $\boldsymbol{x}_{t-1}^{(p)}$. The particle $\boldsymbol{x}_t^{(i)}$ is also said to be descended from $\boldsymbol{x}_{t-1}^{(p)}$ if it was generated by drawing the state occupied by $\boldsymbol{x}_{t-1}^{(p)}$ in resampling.

From Equation (3.68), the importance weights $w_t^{(p)}$ are proportional to $p(\boldsymbol{z}_t|\boldsymbol{x}_t^{(p)})$. For the sequential fixed-lag smoother, we want to express the updated weights $w_{t|t+L}^{(p)}$ in terms of $w_{t+L}^{(i)}$. The Chapman-Kolmogorov equations (Särkkä, 2013; Nulsen et al., 2015) give

$$\begin{aligned} w_{t|t+L}^{(p)} &\propto p(\boldsymbol{z}_{t+L}|\boldsymbol{x}_t^{(p)}) \\ &= \int p(\boldsymbol{z}_{t+L}|\boldsymbol{x}_{t+L}) p(\boldsymbol{x}_{t+L}|\boldsymbol{x}_{t+L-1}) \cdots p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^{(p)}) d\boldsymbol{x}_{t+1} \cdots d\boldsymbol{x}_{t+L}, \end{aligned} \tag{3.69}$$

where $p(\boldsymbol{z}_{t+L}|\boldsymbol{x}_t^{(p)})$ is the probability of observing $\boldsymbol{z}_{t+L}$ at time $t+L$ given the state at time t. Denote $i \leftarrow p$ as the particle i is descended from p. By using the following operator approximation

$$\int p(\boldsymbol{x}_{t+L}|\boldsymbol{x}_{t+L-1}) \cdots p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^{(p)}) d\boldsymbol{x}_{t+1} \cdots d\boldsymbol{x}_{t+L} \approx \sum_{i \leftarrow p} \tag{3.70}$$

and substituting it into Equation (3.69), we get

$$\begin{aligned} p(\boldsymbol{z}_{t+L}|\boldsymbol{x}_t^{(p)}) &\approx \sum_{i \leftarrow p} p(\boldsymbol{z}_{t+L}|\boldsymbol{x}_{t+L}^{(i)}) \\ &\propto \sum_{i \leftarrow p} w_{t+L}^{(i)}. \end{aligned} \tag{3.71}$$

Hence, the weight of a particle p at time step t is updated when the particle filter reaches time step $t+L$ according to

$$w_{t|t+L}^{(p)} = \sum_{i \leftarrow p} w_{t+L}^{(i)}. \tag{3.72}$$

This means that the weight of a particle at time t is determined by the weights of its descendants at time $t+L$. In particular, this means that the weights of the particles at time t without any descendants at time $t+L$ are set to zero.

Chapter 4

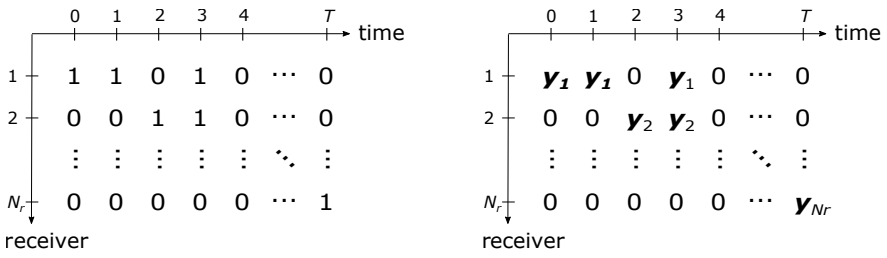
Model Formulation for Salmon Tracking

In the case described in Chapter 2, acoustic telemetry is used to register the presence of fish in the vicinity of stationary receivers placed in Nordfjord. If we consider T aggregated time intervals with equal length dt, we look at the time steps $t = 0, \ldots, T$. To estimate the position and velocity of a salmon given the detections, we use an importance resampling algorithm, a particle filter, with N_p particles. This is presented in Chapter 4.1. In Chapter 4.2, a sequential fixed-lag smoother adjustment to the particle filter is presented. Let $\boldsymbol{x}_t^{(p)}$ be the state vector for particle p at time t defined by

$$\boldsymbol{x}_t^{(p)} = \begin{pmatrix} E_t^{(p)} & N_t^{(p)} & v_{E,t}^{(p)} & v_{N,t}^{(p)} \end{pmatrix}^\intercal, \tag{4.1}$$

where $E_t^{(p)}$ and $N_t^{(p)}$ are the east and north coordinate respectively and $v_{E,t}^{(p)}$ and $v_{N,t}^{(p)}$ are the velocity in the east and north direction for particle p. Collecting the particles, the state vector at time t is represented by $\boldsymbol{x}_t = \begin{pmatrix} \boldsymbol{x}_t^{(1)} & \boldsymbol{x}_t^{(2)} & \cdots & \boldsymbol{x}_t^{(N_p)} \end{pmatrix}$, with dimensions $4 \times N_p$.

Let the number of time intervals with detections be denoted by N_z. Hence, we have that $N_z \leq T$. A detection, also called a registration or a measurement, is referring to a presence detection of a salmon at a receiver. The total number of receivers is denoted by N_r, and $\boldsymbol{y}_j$ are the coordinates of receiver j, for $j = 1, \ldots, N_r$. Let $Z_{t,j}$ be 1 if we have a detection at receiver j at time t and 0 if not. For all receivers at time t, we define the detection vector $\boldsymbol{Z}_t = (Z_{t,1}, \ldots, Z_{t,N_r})^\intercal$. This is visualized in **Figure 4.1a**. In the example presented in the figure, we have $\boldsymbol{Z}_0 = \boldsymbol{Z}_1 = (1, 0, 0, \ldots, 0)^\intercal$. This means that we only have detections at the first receiver for the first two time steps. At $t = 3$, we have detections at two receivers, and $\boldsymbol{Z}_3 = (1, 1, 0, \ldots, 0)^\intercal$. There can also be time steps without any detections, then all entries in the vector will be equal to 0.



(a) Visualization of $\boldsymbol{Z}_t$, for $t = 0, \ldots, T$.

(b) Visualization of $\boldsymbol{z}_t$, for $t = 0, \ldots, T$.

Figure 4.1: Visualization of the detection vectors $\boldsymbol{Z}_t$ and $\boldsymbol{z}_t$.

We define $\boldsymbol{z}_t$ to be a vector containing the east and north coordinates of all receivers with at least one detection in the interval $(t - dt, t]$. Hence, we have the relation $\boldsymbol{z}_t = \{\boldsymbol{y}_j \quad \text{if} \quad Z_{t,j} = 1, j = 1, \ldots, N_r\}$. In general, this can be written as

$$\boldsymbol{z}_t = \begin{pmatrix} z_{E,t,1} & z_{N,t,1} & \cdots & z_{E,t,N_{z,t}} & z_{N,t,N_{z,t}} \end{pmatrix}^\mathsf{T}, \tag{4.2}$$

where $N_{z,t}$ is the number of receivers with a detection at time step t. In **Figure 4.1b**, $\boldsymbol{z}_t$ is visualized. The example presented in **Figure 4.1b** corresponds with the one in **Figure 4.1a**. For the first time step, we only have detections at the first receiver. Hence, $N_{z,0} = 1$, and $\boldsymbol{z}_0 = (z_{E,0,1}, z_{N,0,1})^\mathsf{T}$. Here, $z_{E,0,1}$ is the east coordinate of receiver 1, and $z_{N,0,1}$ is the north coordinate. For $t = 3$, we have $N_{z,3} = 2$ and $\boldsymbol{z}_3 = (z_{E,3,1}, z_{N,3,1}, z_{E,3,2}, z_{N,3,2})^\mathsf{T}$. We have that $z_{E,3,1}$ and $z_{N,3,1}$ are the east and north coordinate, respectively, for the first receiver, and that $z_{E,3,2}$ and $z_{N,3,2}$ are the east and north coordinate, respectively, for the second receiver.

With the measurement vector as defined in Equation (4.2), the measurement model of the system can be expressed as follows:

$$\boldsymbol{z}_t = \tilde{\boldsymbol{B}}\boldsymbol{x}_t^{(p)} + \boldsymbol{r}_t^{(p)}, \tag{4.3}$$

which is the same as defined in Equation (3.22). Here, $\boldsymbol{r}_t^{(p)} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{R})$ is the measurement noise which is independent of the state, with

$$\boldsymbol{R} = \begin{pmatrix} \sigma_z^2 & 0 \\ 0 & \sigma_z^2 \end{pmatrix}, \tag{4.4}$$

where σ_z^2 is the variance in the measurement. Hence, the variance is the same for the measurement in the east and north direction. The dimensions of $\boldsymbol{r}_t$ are $2N_{z,t} \times 1$. The measurement model matrix, $\tilde{\boldsymbol{B}}$, is defined as

$$\tilde{\boldsymbol{B}} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{B} & \cdots & \boldsymbol{B} \end{pmatrix}^\mathsf{T} \tag{4.5}$$

with

$$\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{4.6}$$

Hence, $\tilde{\boldsymbol{B}}$ has dimensions $2N_{z,t} \times 4$.

4.1 Properties of the Target and Importance Distributions in the Particle Filter

For the particle filter algorithm, we need to have an expression for the target distribution $p(\cdot)$ and the importance distribution $\hat{p}(\cdot)$. When a salmon is in the vicinity of a receiver, we have a certain detection probability. This means that the absence of a detection at a receiver indicates that the salmon most likely is not close to this receiver. The target distribution of the system should take both detections and the absence of detections into account. The probability of having a detection at receiver j at time step t, given the predicted state at this time, can be modelled as an exponentially decreasing function:

$$\alpha_{t,j}(\boldsymbol{y}_j|\boldsymbol{x}_t^{(p)}) = p_{\max} \exp\left(-\frac{3}{\phi}\|\boldsymbol{y}_j - \boldsymbol{B}\boldsymbol{x}_t^{(p)}\|\right), \tag{4.7}$$

for each particle $p = 1, \ldots, N_p$. The variable $p_{\max}$ is a measure of the probability of detecting a salmon at a certain receiver when the distance between the receiver and the fish is equal to zero. Because of the experimental nature of the data collection, this probability is smaller than one. The variable ϕ is a measure of the detection range of the receiver. We know that $\exp(-3)$ is approximately equal to 0.05. This means that the factor $-3/\phi$ ensures that if a salmon is closer than ϕ to a receiver, the exponential function has a value greater than 0.05.

For the time steps t where at least one $Z_{t,j} = 1$, the target distribution can be expressed as

$$p(\boldsymbol{Z}_t|\boldsymbol{x}_t^{(p)}) = \prod_{j=1}^{N_r} \alpha_{t,j}(\boldsymbol{y}_j|\boldsymbol{x}_t^{(p)})^{I(Z_{t,j}=1)}(1 - \alpha_{t,j}(\boldsymbol{y}_j|\boldsymbol{x}_t^{(p)}))^{I(Z_{t,j}=0)}, \tag{4.8}$$

which is the probability of having the detections $\boldsymbol{Z}_t$ given the state $\boldsymbol{x}_t^{(p)}$ for particle p. The importance distribution is an approximate form of the target distribution. By using the Markov properties of the model, it can be written as

$$\hat{p}(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{0:t-1}^{(p)}, \boldsymbol{z}_{1:t}) \propto \hat{p}(\boldsymbol{z}_t|\boldsymbol{x}_t^{(p)})p(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{t-1}^{(p)}). \tag{4.9}$$

For each time step t with a detection, we define

$$\hat{p}(\boldsymbol{z}_t|\boldsymbol{x}_t^{(p)}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{z}_t - \tilde{\boldsymbol{B}}\boldsymbol{x}_t^{(p)})^\intercal \tilde{\boldsymbol{R}}^{-1}(\boldsymbol{z}_t - \tilde{\boldsymbol{B}}\boldsymbol{x}_t^{(p)})\right), \tag{4.10}$$

where

$$\tilde{\boldsymbol{R}} = \begin{pmatrix} \boldsymbol{R} & 0 & \cdots & 0 \\ 0 & \boldsymbol{R} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{R} \end{pmatrix} \tag{4.11}$$

with dimensions $2N_{z,t} \times 2N_{z,t}$ and $\tilde{\boldsymbol{B}}$ is as defined in Equation (4.5).

We can sample from $\hat{p}(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{t-1}^{(p)}, \boldsymbol{z}_{1:t})$ by linearly adjusting an unconditional sample. This will be explained in the following paragraphs. We let $\boldsymbol{x}_{u,t}$ denote the unconditional sample at time t. We draw the initial state from an initial distribution $\boldsymbol{x}_{u,0} \sim$

$\mathbb{N}(\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_0)$,where $\boldsymbol{\mu_0}$ and $\boldsymbol{\Sigma}_0$ are predefined. For $t = 1, \ldots, T$, predictions of the state are sampled according to the dynamic model of the system defined in Equation (3.22):

$$\boldsymbol{x}_{u,t} = \boldsymbol{A}\boldsymbol{x}_{u,t-1} + \boldsymbol{q}_{t-1}, \tag{4.12}$$

where the transition matrix of the system is

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.13}$$

and $\boldsymbol{q}_t \sim \mathbb{N}(0, \boldsymbol{Q})$, with

$$\boldsymbol{Q} = \begin{pmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & \sigma_x^2 & 0 & 0 \\ 0 & 0 & \sigma_v^2 & 0 \\ 0 & 0 & 0 & \sigma_v^2 \end{pmatrix}. \tag{4.14}$$

Here, σ_x and σ_v and are the process noise for the position and velocity, respectively. The predictions $\boldsymbol{x}_{u,t}$ are independent of all measurements, and they will be referred to as the unconditional predictions, or samples, at time t. We define $\boldsymbol{x}_{u,0:t}$ to be the unconditional state predictions for all particles at the time steps $0, 1, \ldots, t$, with dimensions $4(t+1) \times N_p$.

The next step is to correct the unconditional predictions based on the detections at the time steps t, where $Z_{t,j} = 1$ for at least one receiver j. We denote the previous time step with at least one detection t^- and we consider the interval $(t^-, t]$. The states before t^- should not be updated, as we have already conditioned on all the measurements up to, and including, t^-. If we assume that we only have one detection, that is, $N_{z,t} = 1$, then the covariance matrix for the predicted, unconditional states and a given detection at time t can be written as

$$Cov(\boldsymbol{x}^{(p)}_{u,t^-+1:t}, \boldsymbol{z}_t | \boldsymbol{Z}_{0:t^-}) = \begin{pmatrix} \boldsymbol{\Sigma}_{t^-+1}\boldsymbol{A}^\intercal \cdots \boldsymbol{A}^\intercal\boldsymbol{B}^\intercal \\ \vdots \\ \boldsymbol{\Sigma}_{t-2}\boldsymbol{A}^\intercal\boldsymbol{A}^\intercal\boldsymbol{B}^\intercal \\ \boldsymbol{\Sigma}_{t-1}\boldsymbol{A}^\intercal\boldsymbol{B}^\intercal \\ \boldsymbol{\Sigma}_t\boldsymbol{B}^\intercal \end{pmatrix}, \tag{4.15}$$

with dimensions $4(t - t^-) \times 2$. For each time step $k, k = t^- + 1, t^- + 2, \ldots, t$, the covariance matrices are given by $\boldsymbol{\Sigma}_k = \boldsymbol{A}\boldsymbol{\Sigma}_{k-1}\boldsymbol{A}^\intercal + \boldsymbol{Q}$. Neither $\boldsymbol{\Sigma}_k$, $\boldsymbol{A}$ nor $\boldsymbol{B}$ are dependent on the value of the detection $\boldsymbol{z}_t$. Hence, if we have several detections in the same interval, the calculated covariances will be the same for all measurements. Hence, for $N_{z,t} \geq 1$, the covariance matrix is given by

$$Cov(\boldsymbol{x}^{(p)}_{u,t^-+1:t}, \boldsymbol{z}_t | \boldsymbol{Z}_{0:t^-}) = \begin{pmatrix} \boldsymbol{\Sigma}_{t^-+1}\boldsymbol{A}^\intercal \cdots \boldsymbol{A}^\intercal\boldsymbol{B}^\intercal & \cdots & \boldsymbol{\Sigma}_{t^-+1}\boldsymbol{A}^\intercal \cdots \boldsymbol{A}^\intercal\boldsymbol{B}^\intercal \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{t-2}\boldsymbol{A}^\intercal\boldsymbol{A}^\intercal\boldsymbol{B}^\intercal & \cdots & \boldsymbol{\Sigma}_{t-2}\boldsymbol{A}^\intercal\boldsymbol{A}^\intercal\boldsymbol{B}^\intercal \\ \boldsymbol{\Sigma}_{t-1}\boldsymbol{A}^\intercal\boldsymbol{B}^\intercal & \cdots & \boldsymbol{\Sigma}_{t-1}\boldsymbol{A}^\intercal\boldsymbol{B}^\intercal \\ \boldsymbol{\Sigma}_t\boldsymbol{B}^\intercal & \cdots & \boldsymbol{\Sigma}_t\boldsymbol{B}^\intercal \end{pmatrix}, \tag{4.16}$$

which has dimensions $4(t - t^-) \times 2N_{z,t}$. The variance matrix for the detections at time step t is defined as

$$Var(\boldsymbol{z}_t|\boldsymbol{Z}_{0:t^-}) = \begin{pmatrix} \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} + \boldsymbol{R} & \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} & \cdots & \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} \\ \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} & \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} + \boldsymbol{R} & \cdots & \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} & \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} & \cdots & \boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B} + \boldsymbol{R} \end{pmatrix}. \tag{4.17}$$

We know that $\boldsymbol{B}^\intercal\boldsymbol{\Sigma}_t\boldsymbol{B}$ has dimensions 2×2, and hence the variance matrix has dimensions $2N_{z,t} \times 2N_{z,t}$.

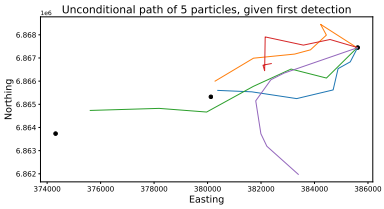
We start the iteration by drawing the initial state from an initial distribution $\boldsymbol{x}_{u,0} \sim \mathbb{N}(\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\mu_0}$ and $\boldsymbol{\Sigma}_0$ are predefined. The covariance is found from $Cov(\boldsymbol{x}_{u,0}^{(p)}, \boldsymbol{z}_0) = \begin{pmatrix} \boldsymbol{\Sigma}_0\boldsymbol{B}^\intercal & \cdots & \boldsymbol{\Sigma}_0\boldsymbol{B}^\intercal \end{pmatrix}$ and the variance $Var(\boldsymbol{z}_0)$ is found from Equation (4.17). The conditional state at the time step $t = 0$ is found by the following relation

$$\boldsymbol{x}_0^{(p)} = \boldsymbol{x}_{u,0}^{(p)} + Cov(\boldsymbol{x}_{u,0}^{(p)}, \boldsymbol{z}_0)Var(\boldsymbol{z}_0)^{-1}(\boldsymbol{z}_0 - \tilde{\boldsymbol{B}}\boldsymbol{x}_{u,0}^{(p)} + \boldsymbol{r}_0^{(p)}). \tag{4.18}$$

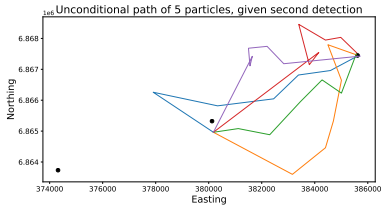
For the following time steps up to T, we consider the time steps t with a detection, that is, the time steps where $Z_{t,j} = 1$ for at least one receiver j. The previous detection was at time step t^-. Unconditional predictions of the states are found according to the dynamic model of the system defined in Equation (4.12) for the time steps $(t^-, t]$, that is $\boldsymbol{x}_{u,t^-+1:t}$. To find the conditional state at time step t, we condition on the detection at this time step, $\boldsymbol{z}_t$:

$$\begin{aligned} \boldsymbol{x}_{t^-+1:t}^{(p)} &= \boldsymbol{x}_{u,t^-+1:t}^{(p)} \\ &\quad + Cov(\boldsymbol{x}_{u,t^-+1:t}^{(p)}, \boldsymbol{z}_t|\boldsymbol{Z}_{0:t^-})Var(\boldsymbol{z}_t|\boldsymbol{Z}_{0:t^-})^{-1}(\boldsymbol{z}_t - \tilde{\boldsymbol{B}}\boldsymbol{x}_{u,t}^{(p)} + \boldsymbol{r}_t^{(p)}). \end{aligned} \tag{4.19}$$

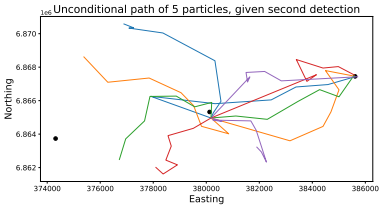
This relation is found by combining the properties stated above with the Kalman filter prediction and update steps, summarized in Equation (3.34) and Equation (3.35). The unconditional state found for time step t is replaced by the conditional state for time step t. In the next iteration, t from the previous iteration is now set to t^-, and hence the last predicted conditional state is used as the basis when we draw unconditional samples for the time steps $(t^-, t]$. This is illustrated in **Figure 4.2**. We have three detections, one at each receiver, marked as a black point. The path of five particles are plotted before and after conditioning on the given detections. The first step is to condition on the first detection. This can be seen from **Figure 4.2a**, as all particles have the same starting point. Then, unconditional particles are sampled according to Equation (4.12) for all time steps up to, and including, the second detection. In **Figure 4.2b**, we have conditioned on the second detection. The resulting conditional prediction for the time step of the second detection is the starting point for the next draw. In **Figure 4.2c**, unconditional particles are sampled for all time steps from the second detection, up to, and including, the third detection. Finally, as presented in **Figure 4.2d**, we condition on the third detection and update the last unconditional state of all particles.



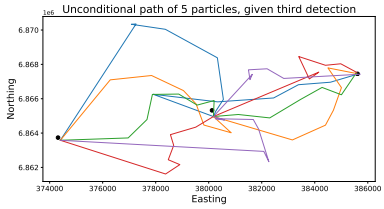
(a) Unconditional states of 5 particles are drawn after conditioning on the first detection.



(b) The unconditional state of the time step with the second detection is updated after conditioning on this detection.



(c) Unconditional states of 5 particles are drawn after conditioning on the second detection.



(d) The unconditional state of the time step with the third detection is updated after conditioning on this detection.

Figure 4.2: Illustration of the unconditional states of 5 particles.

4.2 Particle Filter and Sequential Fixed-Lag Smoother

After drawing samples from the importance distribution, the next step of the particle filter algorithm is to calculate the weights for each particle p according to Equation (3.68). At a time step t with a detection, we consider all time steps from the previous detection up to this, that is $(t^-, t]$. Hence, the weights can be calculated as follows

$$\begin{aligned} w_t^{(p)} &\propto w_{t^-}^{(p)} \frac{p(\boldsymbol{Z}_t|\boldsymbol{x}_t^{(p)})p(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{t-1}^{(p)})}{\hat{p}(\boldsymbol{x}_t^{(p)}|\boldsymbol{x}_{0:t-1}^{(p)}, \boldsymbol{z}_{1:t})} \\ &\propto w_{t^-}^{(p)} \frac{p(\boldsymbol{Z}_t|\boldsymbol{x}_t^{(p)})}{\hat{p}(\boldsymbol{z}_t|\boldsymbol{x}_t^{(p)})} \\ &= w_{t^-}^{(p)} \frac{\prod_{j=1}^{N_r} \prod_{k=t^-+1}^{t} \alpha_{k,j}(\boldsymbol{y}_j|\boldsymbol{x}_k^{(p)})^{I(Z_{k,j}=1)}(1-\alpha_{k,j}(\boldsymbol{y}_j|\boldsymbol{x}_k^{(p)}))^{I(Z_{k,j}=0)}}{\exp\left(-\frac{1}{2}(\boldsymbol{z}_t - \tilde{\boldsymbol{B}}\boldsymbol{x}_t^{(p)})^\intercal \tilde{\boldsymbol{R}}^{-1}(\boldsymbol{z}_t - \tilde{\boldsymbol{B}}\boldsymbol{x}_t^{(p)})\right)}. \end{aligned} \tag{4.20}$$

Finally, the weights are normalized to sum to unity and resampling is performed when the criterion for the effective sample size is met.

Finally, a sequential fixed-lag smoother as described in Chapter 3.7 is applied. A fixed lag of L time steps with detection is chosen. That is, we only count the time steps with at least one $Z_{t,j} = 1$, which are the time steps where we calculate weights. Hence, if $L = 1$,

the weights for the time step t^- is updated when we have calculated the weights for time step t. In general, we let t^* be the time step L time steps with detection before t. Based on Equation (3.72), the weights w_{t^*} are updated according to

$$w_{t^*|t}^{(p)} = \sum_{i \leftarrow p} w_t^{(i)}. \tag{4.21}$$

In this case, all particles are propagated forward in time according to the dynamic model of the system, as defined in Equation (4.12). If the effective sample size is too low, resampling is performed. If we perform resampling, a particle i is said to be a descendant of particle p if it was obtained by resampling particle p. If resampling is not performed, i is a descendant of p if it was obtained by propagating p forward in time. This is illustrated in **Figure 4.3**. The edges illustrates descending particles. If we choose $L = 2$, the updated weights can be calculated as follows:

$$\begin{aligned} w_{t^*}^{(1)} &= w_t^{(4)}, \\ w_{t^*}^{(2)} &= w_t^{(1)} + w_t^{(2)} + w_t^{(3)}, \\ w_{t^*}^{(3)} &= 0, \\ w_{t^*}^{(4)} &= 0. \end{aligned} \tag{4.22}$$

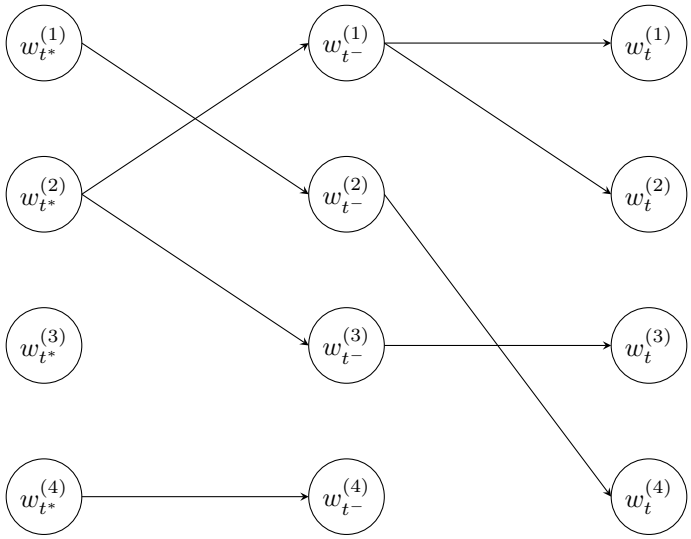


Figure 4.3: Illustration of the fixed-lag smoother.

4.2.1 Algorithm

With the covariance and variance as defined in Equation (4.16) and Equation (4.17) respectively, and the detection probability $\alpha_{t,j}(\cdot)$ as defined in Equation (4.7), the particle filter algorithm described in this chapter can be summarized as follows:

At time t = 0:

1. Draw N_p unconditional samples from the prior distribution

$$\boldsymbol{x}_{u,0} \sim \mathbb{N}(\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_0). \tag{4.23}$$

2. Condition on the detections $\boldsymbol{z}_0$

$$\boldsymbol{x}_0^{(p)} = \boldsymbol{x}_{u,0}^{(p)} + Cov(\boldsymbol{x}_{u,0}^{(p)}, \boldsymbol{z}_0) Var(\boldsymbol{z}_0)^{-1} (\boldsymbol{z}_0 - \tilde{\boldsymbol{B}} \boldsymbol{x}_{u,0}^{(p)} + \boldsymbol{r}_0). \tag{4.24}$$

   for $p = 1, 2, \ldots, N_p$.

3. Calculate weights according to

$$w_0^{(p)} \propto \frac{\prod_{j=1}^{N_r} \alpha_{0,j}(\boldsymbol{y}_j | \boldsymbol{x}_0^{(p)})^{I(Z_{0,j}=1)} (1 - \alpha_{0,j}(\boldsymbol{y}_j | \boldsymbol{x}_0^{(p)}))^{I(Z_{0,j}=0)}}{\exp\left(-\frac{1}{2}(\boldsymbol{z}_0 - \tilde{\boldsymbol{B}} \boldsymbol{x}_0^{(p)})^\intercal \tilde{\boldsymbol{R}}^{-1} (\boldsymbol{z}_0 - \tilde{\boldsymbol{B}} \boldsymbol{x}_0^{(p)})\right)} \tag{4.25}$$

   for $p = 1, 2, \ldots, N_p$ and normalize them to sum to unity.

4. If the effective sample size in Equation (3.65) is too low, perform resampling as described in Chapter 3.6.4

At time $t = 1, \ldots, T$: Draw unconditional samples $\boldsymbol{x}_{u,t}$ according to the dynamic model of the system:

$$\boldsymbol{x}_{u,t} = \boldsymbol{A} \boldsymbol{x}_{u,t-1} + \boldsymbol{q}_{t-1}. \tag{4.26}$$

If $Z_{t,j} = 1$ for at least one receiver j, for $j = 1, \ldots, N_r$, we define t^- as the previous time step where we had at least one detection, $Z_{t^-,j} = 1$, and perform the following:

1. Condition on the detections at time step t to find the conditional states for the time steps $(t^-, t]$:

$$\begin{aligned} \boldsymbol{x}_{t^-+1:t}^{(p)} &= \boldsymbol{x}_{u,t^-+1:t}^{(p)} \\ &\quad + Cov(\boldsymbol{x}_{u,t^-+1:t}^{(p)}, \boldsymbol{z}_t | \boldsymbol{Z}_{0:t^-}) Var(\boldsymbol{z}_t | \boldsymbol{Z}_{0:t^-})^{-1} (\boldsymbol{z}_t - \tilde{\boldsymbol{B}} \boldsymbol{x}_{u,t}^{(p)} + \boldsymbol{r}_t) \end{aligned} \tag{4.27}$$

   for $p = 1, 2, \ldots, N_p$.

2. Calculate new weights according to

$$w_t^{(p)} = w_{t^-}^{(p)} \frac{\prod_{j=1}^{N_r} \prod_{k=t^-+1}^{t} \alpha_{k,j}(\boldsymbol{y}_j | \boldsymbol{x}_k^{(p)})^{I(Z_{k,j}=1)} (1 - \alpha_{k,j}(\boldsymbol{y}_j | \boldsymbol{x}_k^{(p)}))^{I(Z_{k,j}=0)}}{\exp\left(-\frac{1}{2}(\boldsymbol{z}_t - \tilde{\boldsymbol{B}} \boldsymbol{x}_t^{(p)})^\intercal \tilde{\boldsymbol{R}}^{-1} (\boldsymbol{z}_t - \tilde{\boldsymbol{B}} \boldsymbol{x}_t^{(p)})\right)} \tag{4.28}$$

   for $p = 1, 2, \ldots, N_p$ and normalize them to sum to unity.

3. If the effective sample size in Equation (3.65) is too low, perform resampling as described in Chapter 3.6.4

4. Sequential fixed-lag smoother with lag L: Update the weights w_{t^*} by looking at the descendants at time t:

$$w_{t^*|t}^{(p)} = \sum_{i \leftarrow p} w_t^{(i)}, \tag{4.29}$$

   where t^* is the time step L time steps before t, when we only count time steps with detections. Perform resampling based on the updated weights.

Chapter 5

Results

5.1 Choice of Parameters

The goal of the state-space model described in Chapter 4 is to estimate the position and velocity of the fish. The parameters σ_x, σ_v, σ_z, ϕ and $p_{\max}$ must be specified. So must the prior distribution given by $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$. The mean of the initial state is set equal to

$$\boldsymbol{\mu}_0 = \begin{pmatrix} 385615 & 6867449 & -0.05 & 0.0 \end{pmatrix}^\intercal, \tag{5.1}$$

which is the coordinates to the receiver Run to foss. This is the receiver closest to the release point. The initial velocity is set to -0.05 m/s in the east direction, and 0 m/s in the north direction. This means that the salmon initially are expected to have a small drift toward west. The initial variance matrix is

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 500^2 & 0 & 0 & 0 \\ 0 & 500^2 & 0 & 0 \\ 0 & 0 & 0.03^2 & 0 \\ 0 & 0 & 0 & 0.03^2 \end{pmatrix}. \tag{5.2}$$

A good method should have high effective sample size (ESS) and low root mean squared error (RMSE). The effective sample size is defined in Equation (3.65). This is a measure of the variability of the weights. If the ESS is high, a high proportion of the particles contributes to approximating the target distribution. This gives more robust results. An approximation of the mean squared error of the prediction for a specific fish i is

$$e_i = \frac{1}{N_{z,i}} \sum_{t=0}^{N_{z,i}-1} \left((\bar{z}_{E,i,t} - \bar{x}_{E,i,t})^2 + (\bar{z}_{N,i,t} - \bar{x}_{N,i,t})^2 \right), \tag{5.3}$$

where $N_{z,i}$ is the total number of time intervals with detections for fish i. We let $\bar{z}_{E,i,t}$ and $\bar{z}_{N,i,t}$ denote the mean easting and northing position, respectively, of all receivers with a detection of fish i at time t. Because we do not know the true position of the salmon,

this mean position is used as an approximation. The mean easting and northing position, respectively, for all particles for fish i at time step t are denoted by $\bar{x}_{E,i,t}$ and $\bar{x}_{N,i,t}$. The estimate of the total mean squared error is

$$e = \sum_{i=1}^{N} e_i, \tag{5.4}$$

where N is the total number of fish. The approximated root mean squared error is then given by

$$\text{RMSE} = \sqrt{e}. \tag{5.5}$$

The models are run with a time interval of six hours and the parameter $p_{\max}$ is set to 0.95. The number of particles is 1000 and resampling is performed if ESS < $N_p/2$. The parameter ϕ is a measure of the detection range of a receiver and the transmitter range of the acoustic tags. As described in Chapter 2.1, the range is highly dependent of the water conditions, and can vary between a couple of meters up to 200-500 meters. The value of ϕ affects the value of $\alpha_{t,j}(\cdot)$, which is the probability of having a detection at receiver j at time step t, as defined in Equation (4.7). The value of α as a function of the distance from the receiver is plotted in **Figure 5.1** for different values of ϕ. The value of ϕ is set to 3000. This value is based on the sensor quality, and it is adjusted for the sampling interval. In addition, the value of ϕ should be high to ensure more robust estimates. A low value of ϕ will lead to a low detection range for most particles, and the ESS will be low.

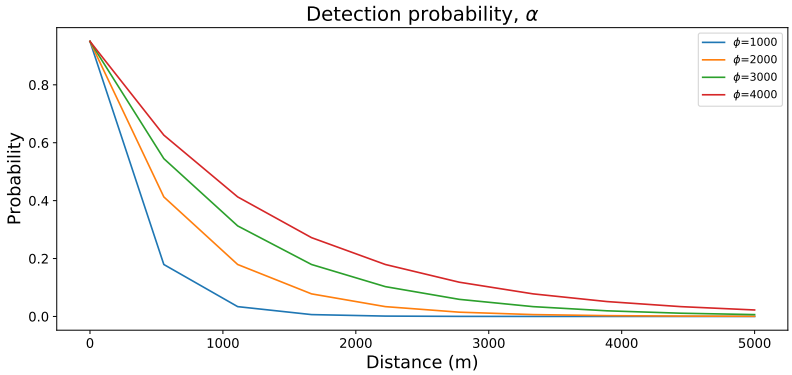


Figure 5.1: Detection probability with different parameter values for ϕ.

In **Table 5.1** and **5.2**, the total approximated RMSE and the mean ESS are presented for different values of σ_x, σ_v and σ_z when the models are run on a random selection of the salmon. The particle filter algorithm without a sequential fixed-lag smoother is denoted by $L = 0$, while $L = 1$ denotes the model where the sequential fixed-lag smoother with lag 1 is applied to the particle filter and so on. In general for all models, we see that the RMSE is lowest when σ_z is low. This is natural, as the value of the variance in the detection, σ_z^2, will affect how important a detection is for the estimation of the state. The ESS, on the other hand, is higher with higher values of σ_z.

For $L = 0, 1, 2, 3$, the RMSE is lowest for $\sigma_x = 1000$, $\sigma_v = 0.03$ and $\sigma_z = 500$. For these parameter values, the ESS is between 41 and 54, which means that the model estimates will not be robust. For $L = 4$, the RMSE is lowest for for $\sigma_x = 1000$, $\sigma_v = 0.01$ and $\sigma_z = 500$, which give an ESS of 54.

With $\sigma_x = 1000$, $\sigma_v = 0.03$ and $\sigma_z = 2000$, the RMSE is relatively low, and the ESS is high for $L = 0, 3, 4$. For $L = 1$ and $L = 2$, the best alternatives are $\sigma_x = 500$, $\sigma_v = 0.03, \sigma_z = 2000$ and $\sigma_x = 1000$, $\sigma_v = 0.01, \sigma_z = 2000$, respectively. These parameter values give the lowest RMSE when we only consider the values that give an ESS above 300. For all values of L, the ESS is relatively similar, varying between 321 and 328. The RMSE is lowest for $L = 4$ and second lowest for $L = 0$, with a value of 4254 and 4266, respectively. However, the ESS is 4 higher for $L = 0$ compared to $L = 4$.

	σ_x:	500	500	500	1000	1000	1000
	σ_z:	500	1000	2000	500	1000	2000
$L = 0$	RMSE	4182	7276	7088	3505	5807	4596
	ESS	60	151	330	58	167	340
$L = 1$	RMSE	3763	7110	7200	3375	6474	4710
	ESS	59	154	338	53	152	328
$L = 2$	RMSE	3940	7410	7317	3478	5534	4990
	ESS	63	146	339	47	156	328
$L = 3$	RMSE	3837	7086	7053	3311	5794	7052
	ESS	52	161	335	58	151	331
$L = 4$	RMSE	3824	7455	7248	3217	5758	5240
	ESS	67	153	333	54	160	326

Table 5.1: RMSE and ESS for different values of σ_x, σ_z and L, with $\sigma_v = 0.01$ fixed.

	σ_x:	500	500	500	1000	1000	1000
	σ_z:	500	1000	2000	500	1000	2000
$L = 0$	RMSE	3687	5711	4889	3381	5407	4266
	ESS	56	159	334	50	155	326
$L = 1$	RMSE	3595	6412	4404	3206	5606	5162
	ESS	61	160	327	41	142	329
$L = 2$	RMSE	3247	5594	5693	3202	6005	5702
	ESS	60	154	333	54	143	320
$L = 3$	RMSE	3467	6151	5282	3060	5799	4418
	ESS	60	154	335	53	141	323
$L = 4$	RMSE	3791	5917	4908	3230	5704	4254
	ESS	57	170	338	50	152	322

Table 5.2: RMSE and ESS for different values of σ_x, σ_z and L, with $\sigma_v = 0.03$ fixed.

5.2 Estimated Movement Pattern

The goal of the particle filter and the sequential fixed-lag smoother is to estimate the position and velocity of the salmon. The results obtained for fish ID 100 with $\sigma_x = 1000$, $\sigma_v = 0.03$ and $\sigma_z = 2000$ for $L = 0$ and $L = 4$ are presented in the following chapter.

Estimated position

Fish ID 100 is first registered on April 30. The last detection is on May 6. During the first 15 time steps, that is the first 90 hours of the migration, fish ID 100 is registered in the river and in the estuary. In the first time step, it is detected at Sætre and Smaleveien. It remains in the upper part of the river, and is registered at Sætre at time step 2, 3 and 4. In the 4th time step, it is also detected at Holmane. At time step 5 and 8 it continues to be detected at Holmane, before it is registered in the estuary for the first time at time step 13 . At the 16th time step, it is detected both at Osen midt, Osen N and Neset. These are the last detections in the estuary. After this, it is detected at Blakset at time step 18. At time step 21 and 22, it is detected in the zone Lote. There are no detections in the zone Krokneset. The next detection is in the zone Isane, at the receiver Åseneset, at time step 25. In the following time step it is detected at Otteren, in the zone Maurstad. The last detection is at time step 27. This is at Almenningsflua, in the outer part of the fjord.

The mean position of all particles with $L = 0$, that is, the particle filter, is presented in **Figure 5.2** together with the standard error of the particle estimates. The mean position is also presented on a map of Nordfjord in **Figure 5.3**. The equivalent results using a lag of $L = 4$ are presented in **Figure 5.4** and **Figure 5.5**. There are not any big, visible differences in the results.

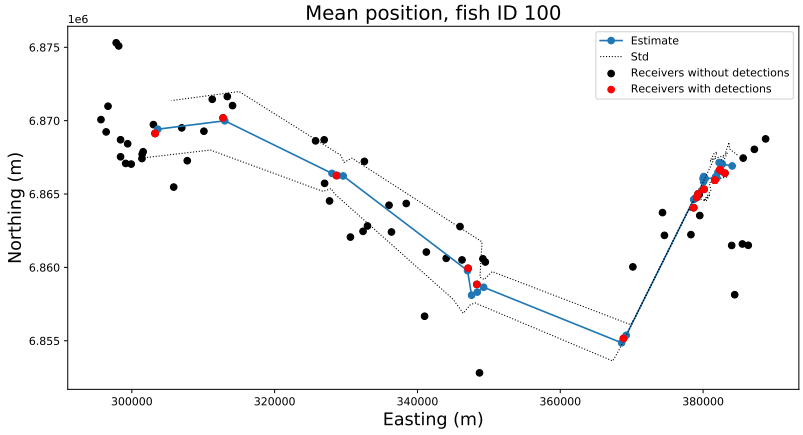


Figure 5.2: Estimated position for fish ID 100 with the standard error of the particle estimates added and subtracted to the mean with $L = 0$. Black and red dots illustrate receivers without and with any detections, respectively. UTM coordinates, zone 32, are used.

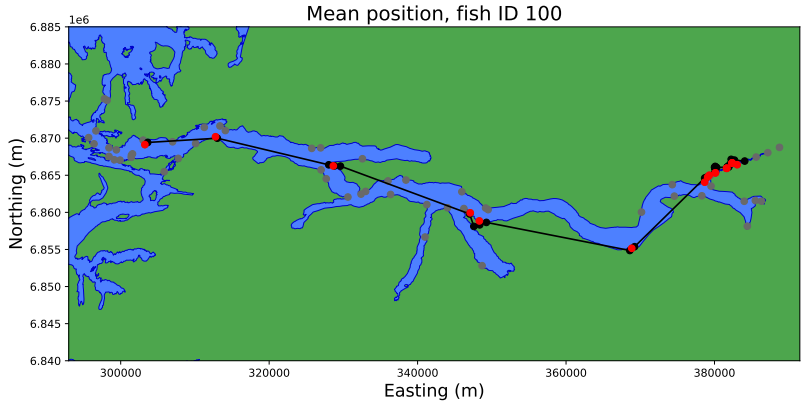


Figure 5.3: Estimated position for fish ID 100 plotted on a map of Nordfjord with $L = 0$. Black and red dots illustrate receivers without and with any detections, respectively. UTM coordinates, zone 32, are used.

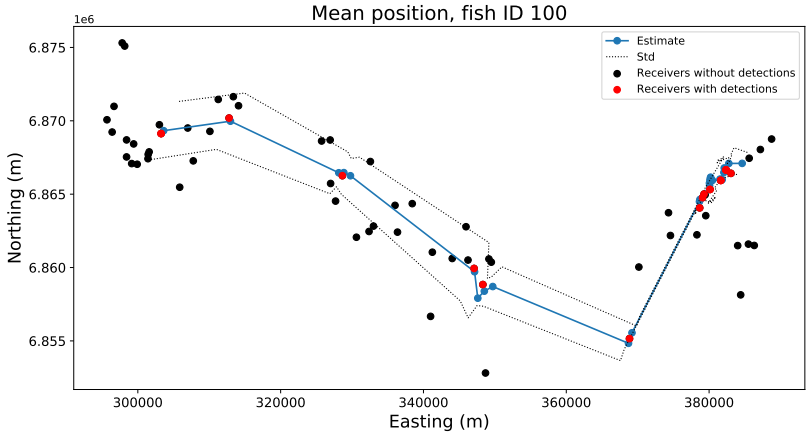


Figure 5.4: Estimated position for fish ID 100 with the standard error of the particle estimates added and subtracted to the mean with $L = 4$. Black and red dots illustrate receivers without and with any detections, respectively. UTM coordinates, zone 32, are used.

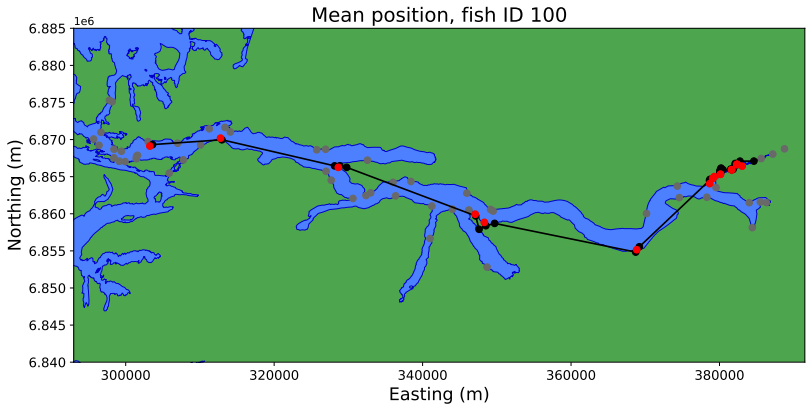


Figure 5.5: Estimated position for fish ID 100 plotted on a map of Nordfjord with $L = 4$. Black and red dots illustrate receivers without and with any detections, respectively. UTM coordinates, zone 32, are used.

Estimated velocity

The mean estimated speed in meters per second for all particles in the east and north direction with $L = 0$ are presented in **Figure 5.6** and **5.7** , respectively. The standard error of the particle estimates are added and subtracted to the mean and are also presented in the figures. The estimated mean speed in the eastward direction is negative for all time steps. This means that fish ID 100 moves westward, that is, toward the ocean, for the entire migration period. The absolute value of the speed in the eastward direction is highest in the beginning and in the end of the migration period. The estimated speed northward is negative until approximately one day remains of the detection period. For the first days, the speed is close to zero. After May 3, the absolute value of the speed increases.

For $L = 4$, the mean estimated speed for all particles in the east and north direction are presented in **Figure 5.8** and **5.9** , respectively. The estimated speed in the eastward direction is negative at all time steps except for two. The absolute value of the eastward speed is highest in the last part of the migration period. The northward speed starts out negative, which means that the salmon moves southward. From the beginning of May until May 4, the speed is positive. From May 4 to May 5, the speed is again toward south. The last day of the migration, the salmon moves northward. The absolute value of the speed in the north direction is highest on May 4.

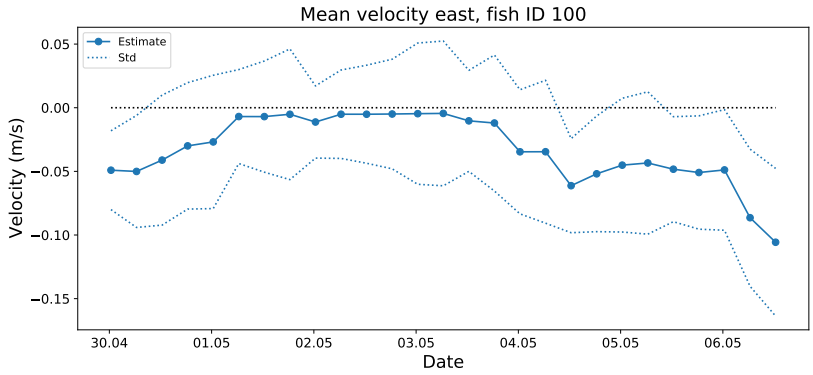


Figure 5.6: Estimated speed in east direction with the standard error of the particle estimates added and subtracted to the mean with $L = 0$. The black dotted line is a reference line at zero speed.

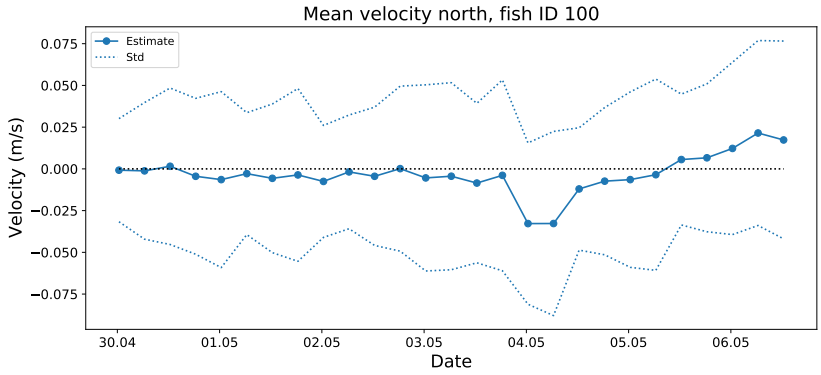


Figure 5.7: Estimated speed in north direction with the standard error of the particle estimates added and subtracted to the mean with $L = 0$. The black dotted line is a reference line at zero speed.

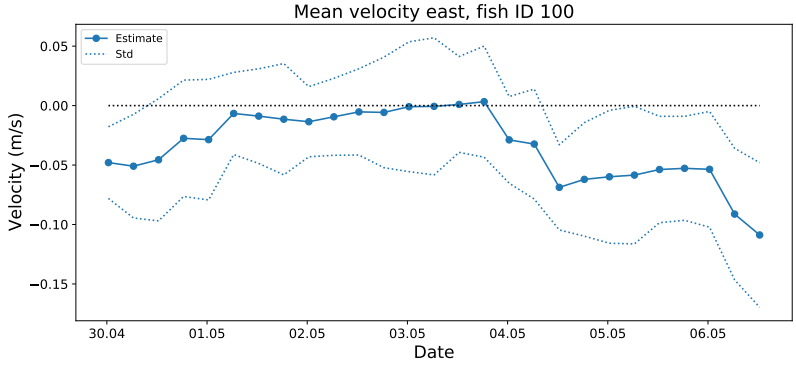


Figure 5.8: Estimated speed in east direction with the standard error of the particle estimates added and subtracted to the mean with $L = 4$. The black dotted line is a reference line at zero speed.

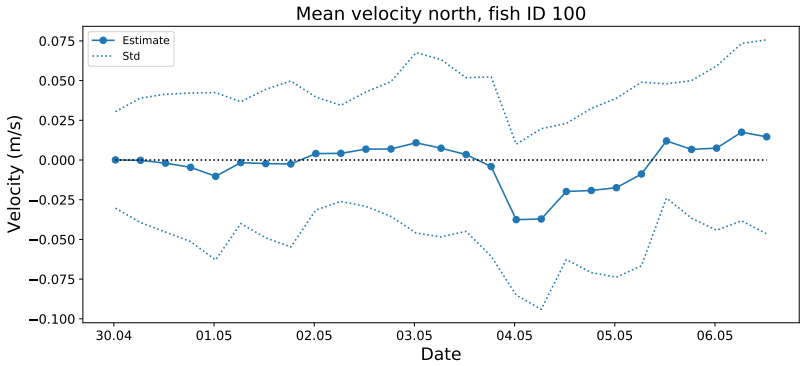


Figure 5.9: Estimated speed in north direction with the standard error of the particle estimates added and subtracted to the mean with $L = 4$. The black dotted line is a reference line at zero speed.

For $L = 0$, the absolute value of the velocity in the east and north direction is presented in **Figure 5.10** together with the direction of the speed in **Figure 5.11**. Both are plotted together with the standard error of the particle estimates added and subtracted to the mean. The speed is given in body lengths per second. The speed is somewhat increasing in the later stages of the migration period. Equivalently, the mean speed and direction for $L = 4$ are plotted in **Figure 5.12** and **Figure 5.13**, respectively. The standard error of the particle estimates are added and subtracted to the mean. The pattern of the speed and direction of the speed are similar to the results for $L = 0$.

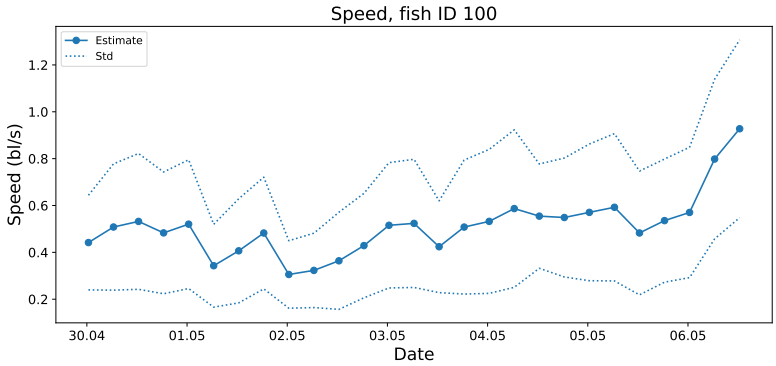


Figure 5.10: Estimated speed with the standard error of the particle estimates added and subtracted to the mean with $L = 0$.

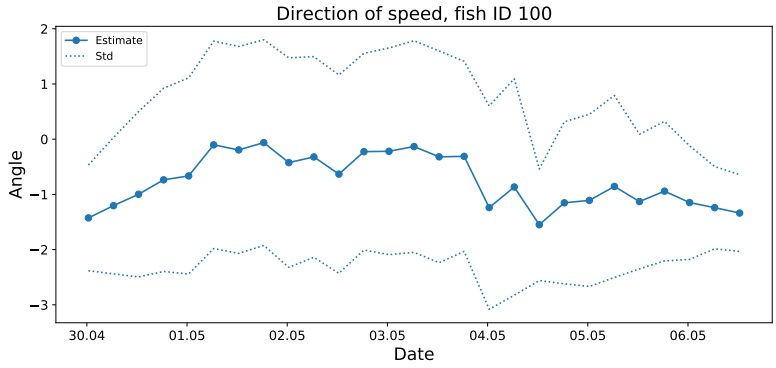


Figure 5.11: Estimated direction of the speed with the standard error of the particle estimates added and subtracted to the mean with $L = 0$.

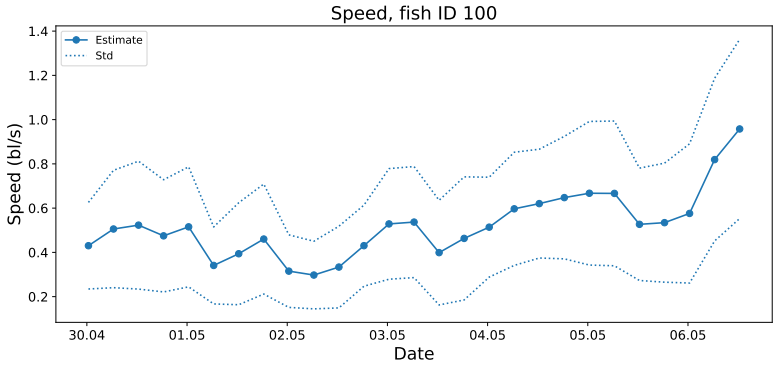


Figure 5.12: Estimated speed with the standard error of the particle estimates added and subtracted to the mean with $L = 4$.

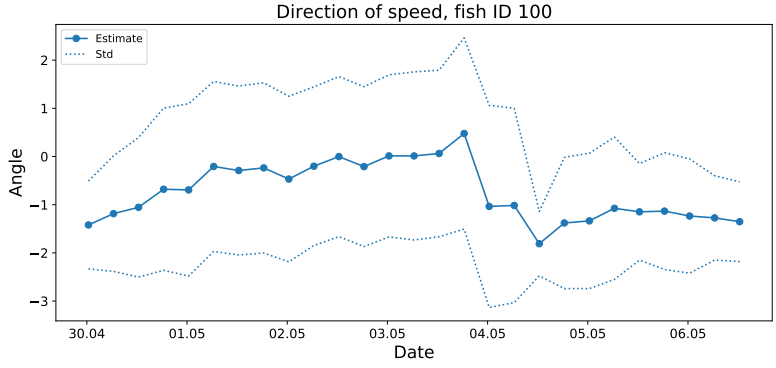
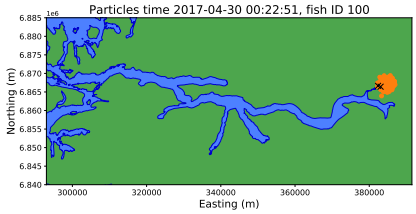


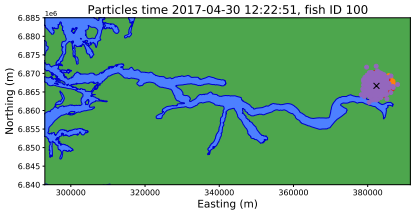
Figure 5.13: Estimated direction of the speed with the standard error of the particle estimates added and subtracted to the mean with $L = 4$.

Particles

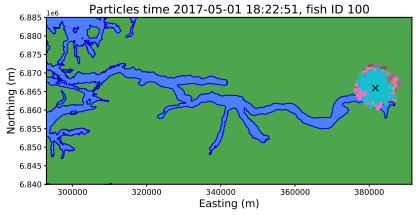
The position of the particles after six different time steps are presented in **Figure 5.14** with $L = 0$. The particles are plotted with a different color at different time steps, and each point represents one particle. A black cross represents a receiver with a detection.



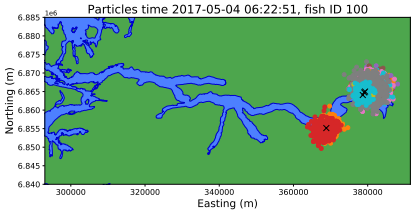
(a) Time step 1.



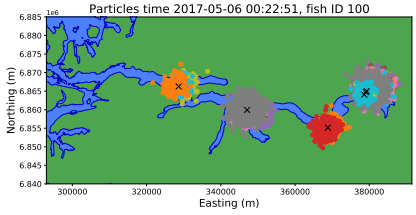
(b) Time step 3.



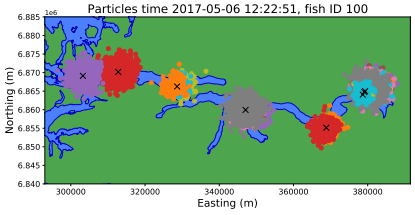
(c) Time step 8.



(d) Time step 18.



(e) Time step 25.



(f) Time step 27.

Figure 5.14: Position of particles at different time steps with $L = 0$.

The mean of the initial state corresponds to the coordinates of the receiver Run to foss, as presented in Equation (5.1). At the first time step, the particles are drawn around this point with a variance given in Equation (5.2). Fish ID 100 is detected at Sætre and Smaleveien, and in **Figure 5.14a**, we see that these receivers are further west than the initial distribution of the particles. The particles at time step 3 are presented in **Figure 5.14b**, together with the position of the receiver Sætre, which has a detection at this time step. We can see that the particles have moved westwards compared to **Figure 5.14a**. This is due to the detections and the initial velocity in the westward direction. At time step 8, fish ID 100 is detected at Holmane, as presented in **Figure 5.14c**. From time step 16 to time step 17, the estimated position moves from the estuary to Utvik. There is a detection at Blakset at time step 18, as presented in **Figure 5.14d**. In **Figure 5.14e**, we see that there

have been detections in two zones, Lote and Isane. The particles after the final detection at time step 27 are presented in **Figure 5.14f**.

The effective sample size as a function of time is presented in **Figure 5.15** and **Figure 5.16** for $L = 0$ and $L = 4$, respectively. The ESS is only calculated for the time steps with detections. The values of the ESS around 200 or lower correspond to the time steps 7, 12, 15, 20 and 24. The results are very similar for the different models.

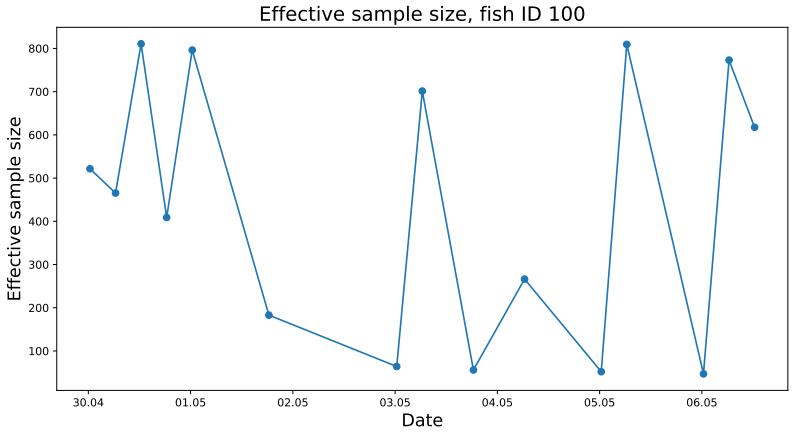


Figure 5.15: Effective sample size with $L = 0$.

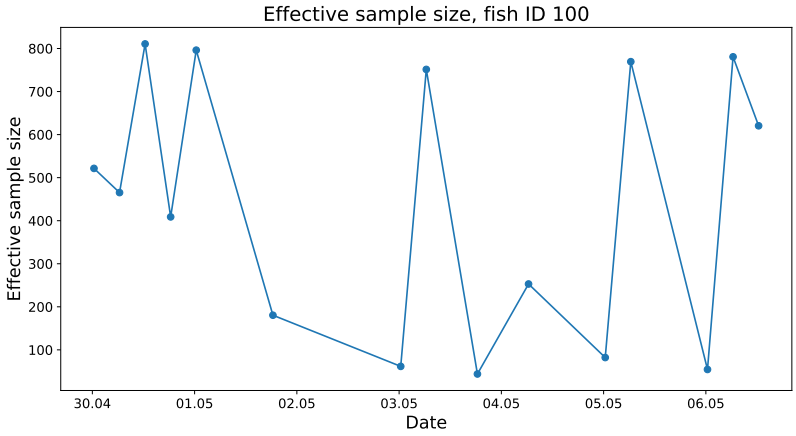


Figure 5.16: Effective sample size with $L = 4$.

Chapter 6

Discussion

This chapter covers a discussion of the results presented in Chapter 5. In Chapter 6.1, different models are compared when considering data for all fish in the study. The estimated movement pattern for fish ID 100, presented in Chapter 5.2, is discussed in Chapter 6.2. Further, the estimated velocity for all salmon are considered in Chapter 6.3. In Chapter 6.4, correlations between the estimated states and environmental data are considered.

6.1 Comparison of Models

In Chapter 5, a sequential fixed-lag smoother is applied to a particle filter for estimating the position and velocity of salmon based on acoustic telemetry data. The lag of the smoother is denoted by L. If $L = 0$, it means that there is no lag, and only the particle filter is applied to the data. In **Table 6.1**, the total approximated root mean square error for all salmon considered is presented, together with the mean effective sample size. Based on the results in Chapter 5.1, σ_z is set to 2000 and three sets of parameter values for σ_x and σ_v are tested for different values of L. To get both accurate and robust results, the RMSE should be low and the ESS high. The ESS is similar for all parameter values and values of L, varying from 320 to 336. For $\sigma_x = 1000$ and $\sigma_v = 0.01$, the sequential fixed-lag smoother applied to the particle filter with lag $L = 3$ gives the lowest RMSE and highest ESS. For $\sigma_x = 1000$ and $\sigma_v = 0.01$, the lowest value of the RMSE is obtained with $L = 2$. Overall, the RMSE is lowest for $L = 0$, with $\sigma_x = 1000$, $\sigma_v = 0.03$ and $\sigma_z = 2000$. That is, the particle filter without a sequential fixed-lag smoother adjustment. Out of the three sets of parameters, this is the combination with the highest process noise. With high process noise, the measurements have a high effect on the estimated state. This means that a small amount of the data from previous time steps are used, and hence we do not introduce a lot of bias. This could explain why the particle filter without a smoother gives the highest RMSE. It is important to keep in mind that the RMSE is approximated based on the mean position of the detections with receivers, which is not the true position of the salmon at the given time step. If we knew the true position of the salmon, it is possible that another model would be chosen as the optimal. In the following chapter, the

results presented in Chapter 5.2 for $L = 0$ will be considered.

		$L = 0$	$L = 1$	$L = 2$	$L = 3$	$L = 4$
$\sigma_x = 1000$, $\sigma_v = 0.01$	RMSE	7955	6739	6815	6414	8166
	ESS	328	334	333	334	331
$\sigma_x = 500$, $\sigma_v = 0.03$	RMSE	8219	8067	7560	7570	7597
	ESS	334	332	328	332	336
$\sigma_x = 1000$, $\sigma_v = 0.03$	RMSE	5802	6599	6903	6816	6951
	ESS	325	320	322	325	318

Table 6.1: Total approximated root mean square error for all salmon and mean effective sample size for different models.

6.2 Estimated Movement Pattern for a Specific Salmon

In Chapter 5.2, the estimated movement pattern of fish ID 100 is presented. The estimated position, as presented in **Figure 5.2**, is highly dependent on the detections. If there are detections at several receivers in the same time interval, the mean estimated position of all particles tend to be close to the mean position of the receivers. The migration is from east to west, and the estimated eastward velocity is estimated to be negative for all time steps, as presented in **Figure 5.6**. The absolute value of the speed in the east direction is highest toward the end. As we can see from **Figure 5.7**, the estimated northward velocity is somewhat negative in the beginning, and mainly positive in the end of the migration. This fits with the geography of Nordfjord, in the river and beginning of the fjord, the movement of the salmon is mainly toward south. After Lote, the fjord mainly moves toward north. Fish ID 100 reaches Finnvika, which is situated in Lote, at time step 21, after 5.25 days. This corresponds to the results in the figure, where we see that the estimated speed in the north direction turns positive during May 5. In **Figure 5.10**, the speed in body lengths per second (bl/s) is presented. The mean speed is 0.51 bl/s with a standard error of 0.13 bl/s. This is considerable lower than the mean progression speed of 1.3 bl/s found by Urke et al. (2018) and 1.27 bl/s found in a study performed in the Romsdalsfjord system (Thorstad et al., 2004). However, the results are similar to results obtained in other studies. In a study of Atlantic salmon in River Eira on the west coast of Norway, the mean speed ranged from 0.49 to 1.82 bl/s (Finstad et al., 2005). The fjord migration of Atlantic salmon was also studied by Thorstad et al. (2007), and the mean speed was found to be 0.53 bl/s.

6.3 Estimated Velocity

In the following chapter, the particle filter algorithm is applied to all salmon that are detected in the outer part of Nordfjord, see Appendix A.2 for an overview of the fish in the study. The parameters $\sigma_x = 1000$, $\sigma_v = 0.03$ and $\sigma_z = 2000$ are used, according to the results presented in Chapter 6.1. The mean speed is found to be 0.48 bl/s, with a standard error of 0.27 bl/s. This is similar to the results obtained for fish ID 100, and the predic-

tion is somewhat lower than the results obtained in previous studies (Thorstad et al., 2004; Finstad et al., 2005; Thorstad et al., 2007; Urke et al., 2018).

In **Figure 6.1** the mean speed, that is, the mean of the absolute value of the velocity in the east and north direction, of all 27 salmon at different dates is presented together with the standard error. The speed is given in body lengths per second (bl/s). The fjord can be divided into different zones, as presented in **Figure 2.2**. The mean speed and standard error in the different zones for all salmon considered are presented in **Figure 6.2**. The zones denoted by "Utvik_Lote" and "Isane_Maurstad" represent the area from Utvik to Lote and from Isane to Maurstad, respectively. For the remaining zones, the zone limits are in the middle of the two subsequent zones.

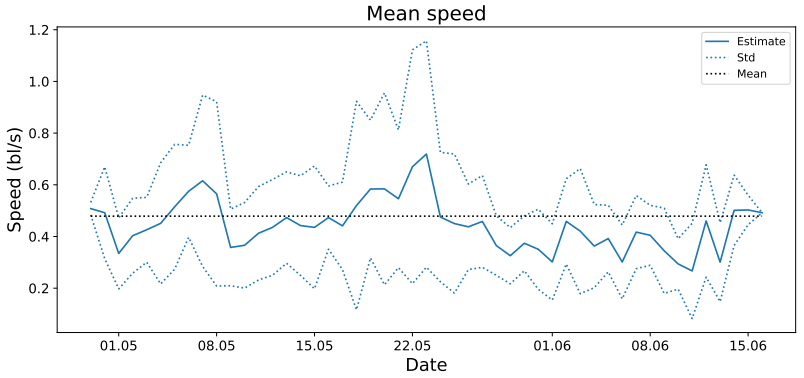


Figure 6.1: Mean speed at different dates for all fish considered with the standard error added and subtracted to the mean. The black dotted line is a reference line at the mean speed.

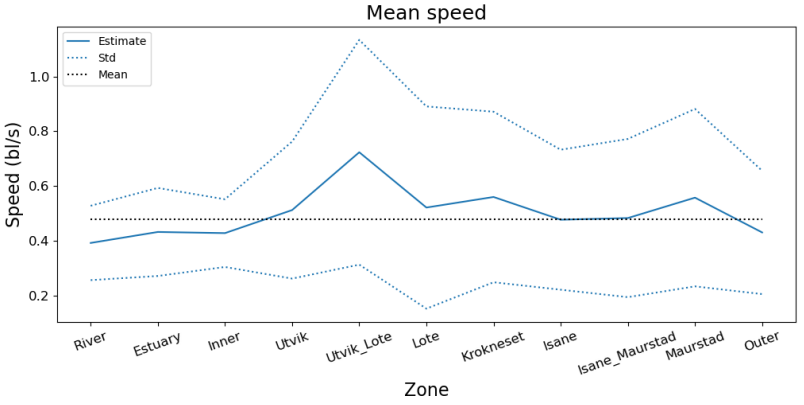


Figure 6.2: Mean speed for all fish in different zones of Nordfjord with the standard error added and subtracted to the mean. The black dotted line is a reference line at the mean speed.

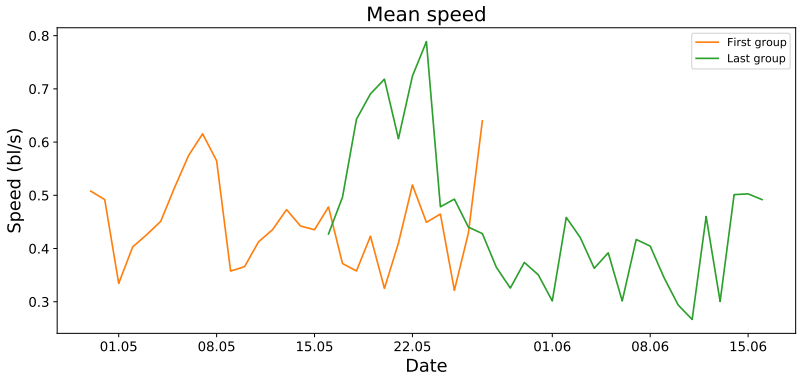
The salmon are divided into two migration groups based on the time of their first detection. The first group consists of the salmon that are detected before May 15, the last group are the salmon that are first detected on May 15 or later. In total, there are 27 fish that are detected in the outer part of Nordfjord. Of them, 17 salmon are in the first group and 10 are in the last group. In **Figure 6.3**, the mean speed at different dates for the two different groups is presented. In **Figure 6.3a**, the results for both groups are plotted. The mean speed for each group is presented separately in **Figure 6.3b** and **Figure 6.3c** for the first and last group, respectively. Here, the the standard error is added and subtracted to the mean.

In **Figure 6.4**, the mean speed in the different zones for the two groups are presented. In **Figure 6.4a**, the results for both groups are plotted. The mean speed for each group is presented separately in **Figure 6.4b** and **Figure 6.4c** for the first and last group, respectively. Here, the the standard error is added and subtracted to the mean.

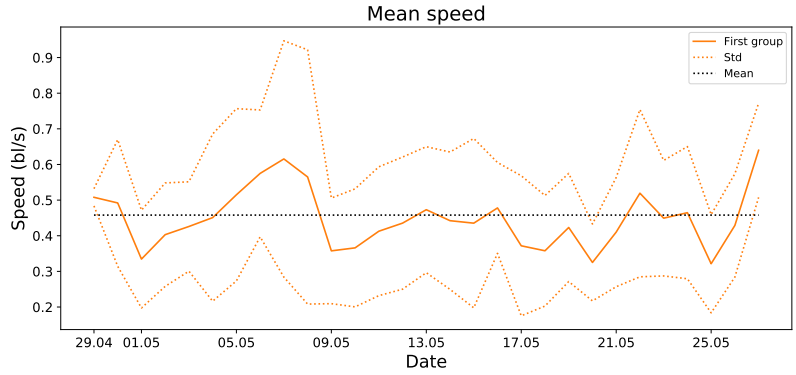
The results presented in **Figure 6.3** and **Figure 6.4** indicate that the mean speed of the last group is somewhat higher than the speed of the first group. The mean speed for the first group is 0.46 bl/sec, with a standard error of 0.23 bl/sec. The mean speed for the last group is 0.51 bl/sec, with a standard error of 0.32 bl/sec. By performing a Welch's t-test (Welch, 1947; Heumann & Shalabh, 2016) with null hypothesis of the two means being identical, the p-value is found to be 0.0029. Hence, on a 5 % level we accept the alternative hypothesis that says that the mean speed of the first and last migration group are statistically different. This means that according to the particle filter estimates, the speed of the salmon in the last migration group is found to be significantly higher than the speed of the salmon in the first migration group.

In **Figure 6.1**, we see that the estimated speed in the second half of the migration period is lower than the mean speed. There is a peak around May 22. For the salmon in the first migration group, the estimated speed is higher than the mean speed between May 5 and May 9. There are also small peaks at some of the later dates. The results can be seen in **Figure 6.3b**. The results for the last migration group are presented in **Figure 6.3c**, and we can see that the speed i highest in the first week. From May 24, the estimated speed is lower than the mean speed.

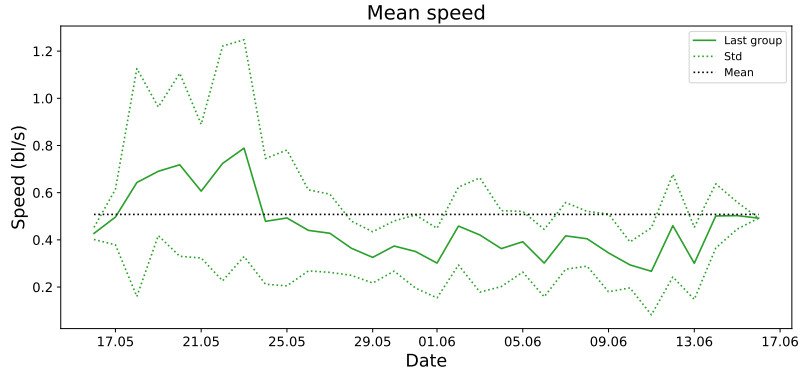
The mean estimated speed from Utvik to Krokneset is higher than the mean speed, as presented in **Figure 6.2**. There is also a small peak in Maurstad. For the salmon in the first migration group, the estimated speed is highest i Maurstad. In addition, there are peaks in the area from Utvik to Lote and in Isane, presented in **Figure 6.4b**. For the salmon in the last migration group, the estimated speed increases in the areas from Utvik to Krokneset, as can be seen from **Figure 6.4c**. The speed in the zones before and after are similar. In previous studies it has been found that the swimming speed of migrating salmon increases in the later stages of the migration, and there have been observed a pattern of increasing speed with distance from the river (Davidsen et al., 2009; Jonsson & Jonsson, 2011; Urke et al., 2013a). This does not coincide with the results found from the particle filter, as presented in **Figure 6.1** to **Figure 6.4**. Reasons for this can be differences in the local environment and the topography of the fjords where the studies have been performed.



(a) Mean speed for the first and last migration group at different dates.

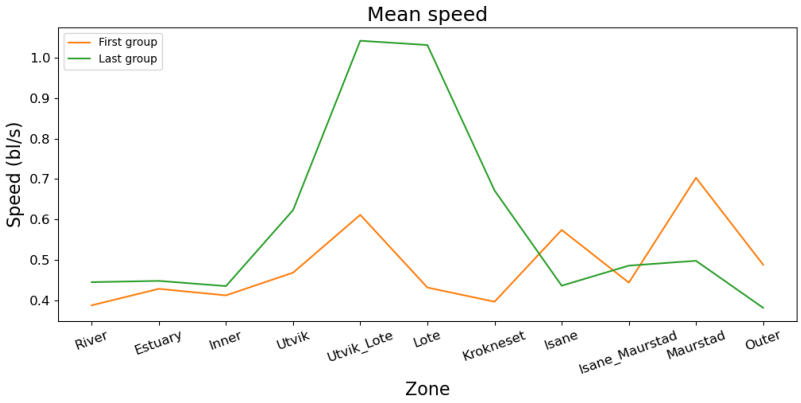


(b) Mean speed for the first the migration group with the standard error added and subtracted to the mean. The black dotted line is a reference line at the mean speed.

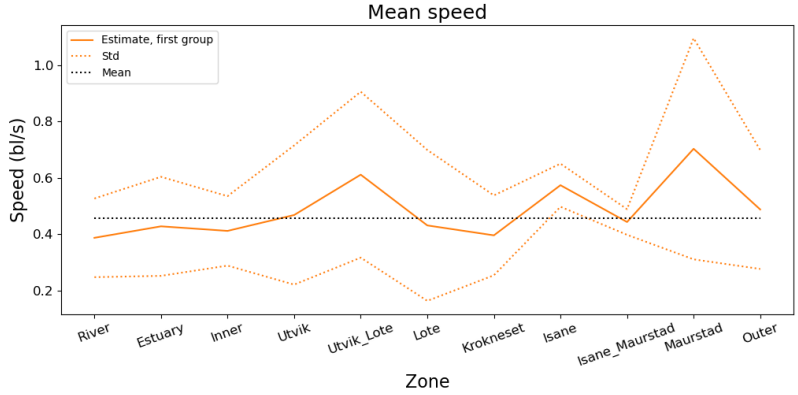


(c) Mean speed for the last migration group with the standard error added and subtracted to the mean. The black dotted line is a reference line at the mean speed.

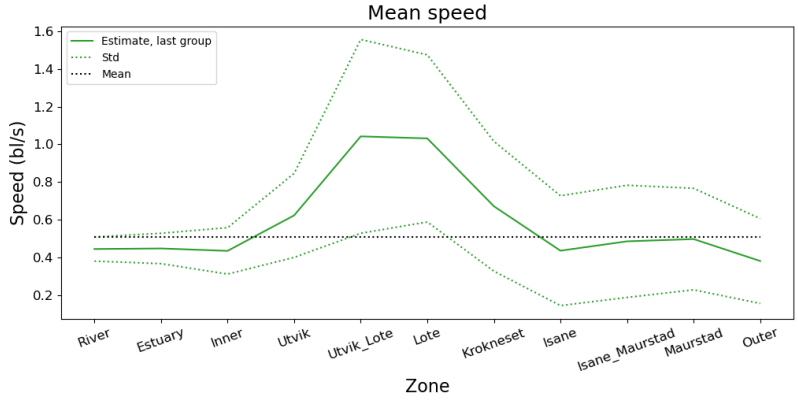
Figure 6.3: Mean speed for the first and last migration group at different dates.



(a) Mean speed for the first and last migration group in different zones of Nordfjord.



(b) Mean speed for the first migration group with the standard error added and subtracted to the mean. The black dotted line is a reference line at the mean speed.



(c) Mean speed for the last migration group with the standard error added and subtracted to the mean. The black dotted line is a reference line at the mean speed.

Figure 6.4: Mean speed for the first and last migration group in different zones.

In **Figure 6.5**, the correlations between the weight and length of the salmon and the estimated state are presented. The estimated position and speed in the east and north direction is denoted by "Easting", "Northing", "Vel_East" and "Vel_North", respectively. The absolute value of the velocity is denoted by "Speed". The variables "Weight" and "Length" are the weight and length of the salmon. The correlation is found using the Kendall rank correlation coefficient (Kendall & Gibbons, 1990). Correlations significant on a 10 % level are marked with *. In general, the correlations are weak and few variables have a significant correlation.

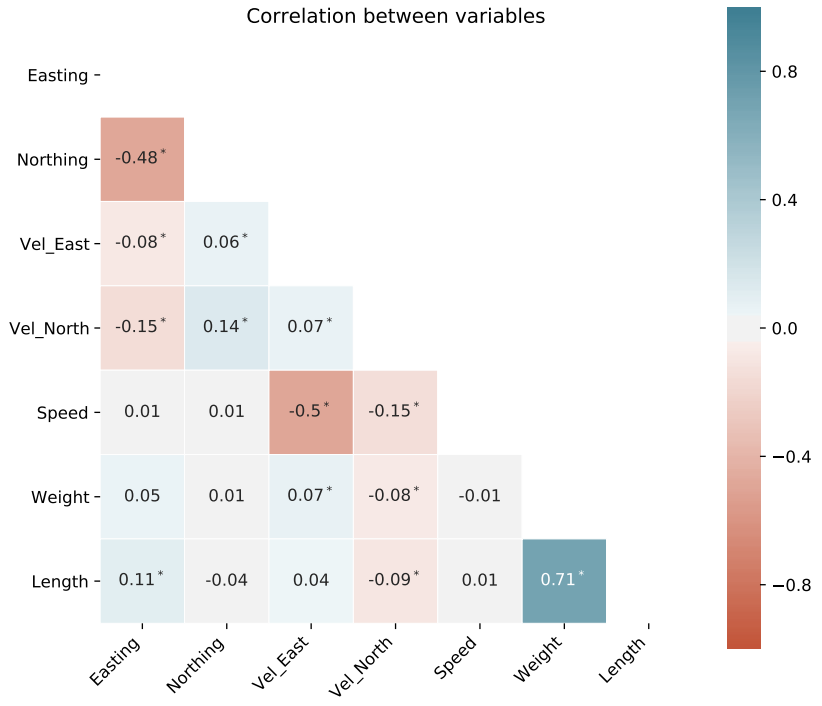


Figure 6.5: The Kendall rank correlation between the estimated position and speed in east and north direction, "Easting", "Northing", "Vel_East" and "Vel_North", the absolute value of the velocity, "Speed", and the weight and length of the salmon. Correlations significant on a 10 % level are marked with *.

The estimated speed in the east direction has a small, but significant, negative correlation with the easting position. The estimated eastward speed will mainly be negative, as we saw for fish ID 100 in **Figure 5.6**. This means that the speed in the eastward direction increases when the salmon move toward west. Since the movement is mainly toward west, it means that speed westward decreases as the salmon gets closer to the ocean. This is

the opposite of what we would expect based on previous studies (Davidsen et al., 2009; Jonsson & Jonsson, 2011; Urke et al., 2013a).

The estimated speed in the north direction is positively and significantly correlated with the northing position. This indicates that the speed northward is higher when the salmon are in the parts of the fjord further north. The northward speed has a negative and significant correlation with the easting position. This indicates that the speed in the north direction increases as the salmon move toward west. Since the movement is mainly toward west, this is equivalent to the northward speed increasing with time, which is the same pattern as we saw for fish ID 100, presented in **Figure 5.7**.

The speed, however, is not found to be significantly correlated with the easting or northing position. The speed is negatively and significantly correlated with the estimated speed in the east and north direction. This is natural as the eastward and northward speed are mainly negative, and the speed is the absolute value of these variables. The speed is found to have a very small and negative correlation with the weight, and a very small and positive correlation with the length. The correlations are not found to be significant. From an ecological view, we would expect there to be a positive correlation between these variables (Aas et al., 2011).

The length is positively and significantly correlated with the easting position. This may indicate that longer salmon tends to be detected further out in the fjord, which is supported by previous studies finding that longer smolts tend to have a higher survival rate (Saloniemi et al., 2004).

6.4 Environmental Data

The migration speed of salmon is dependent on the water temperature and salinity (Aas et al., 2011). The state-space model described in Chapter 4 does not take environmental factors like temperature, salinity and seawater velocity into account. In the following chapter, the results from the particle filter will be linked to SINMOD data, which are modelled environmental data from SINTEF, to see if there is a significant correlation between these variables and the estimated swimming speed.

The SINMOD data have a resolution of 800 meters. Data are available at different depths, starting at 3 meters. The state-space models applied to the data does not take depth into account. During the migration, salmon usually swim close to the surface, in the top 1 to 3 meters, with irregular dives (Aas et al., 2011; Urke et al., 2018). According to Plantalech Manel-La et al. (2009), the mean swimming depth of Atlantic salmon in River Eio was found to be 1.7 meters. Based on this, SINMOD data are extracted from the top layer, that is, a depth of 3 meters.

The seawater temperature is lower in the river and inner parts of Nordfjord, as presented in **Figure 6.6**. The temperature increases with time. From **Figure 6.7** it is clear that the salinity is higher closer to the ocean. The salinity decreases in the river and inner parts of the fjord with time. This is due to more melting water being discharged.

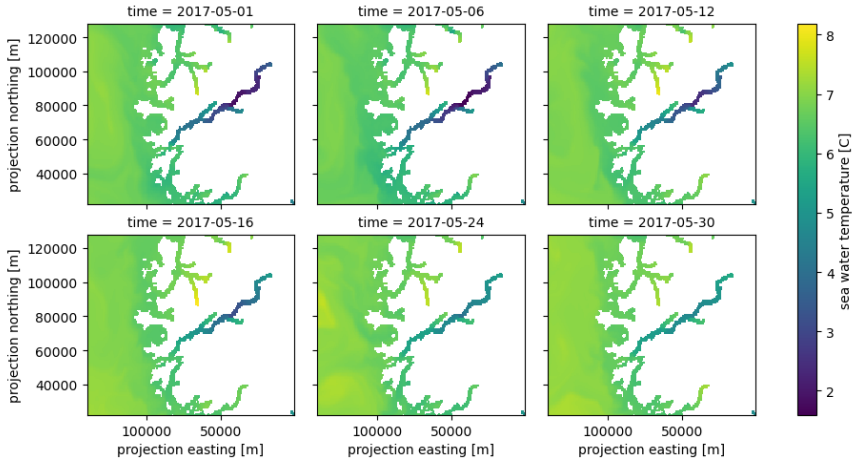


Figure 6.6: Temperature at six different times during the migration period.

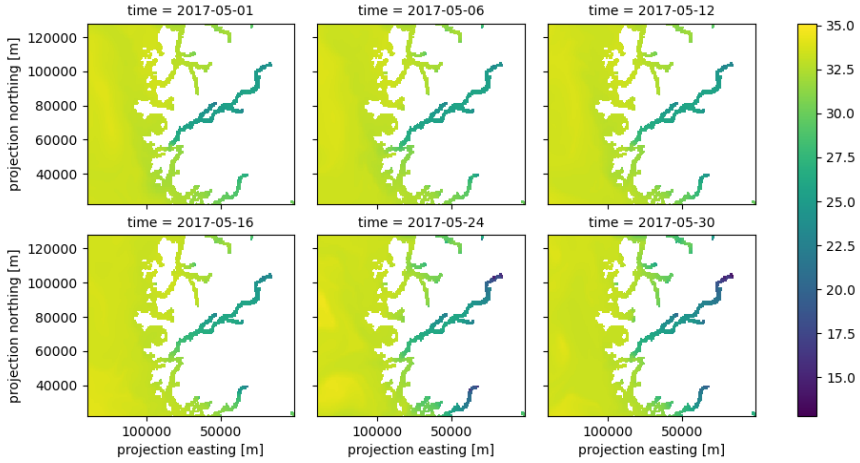


Figure 6.7: Salinity at six different times during the migration period.

The mean speed at different dates along with the mean temperature and salinity are presented in **Figure 6.8**. SINMOD data are found from the grid point closest to the estimated position. If there are no available SINMOD data at a given point, the corresponding state is not used in the mean estimation. The estimated mean speed is high in the beginning. For the first days the counts of detected salmon are low, and hence the estimates are not reliable to represent the population. There is a peak in the estimated speed around May 22, which corresponds to the results presented in **Figure 6.1**. The temperature increases with time. This is both due to the salmon being detected further out in the fjord in the later stages of the migration period, and due to the general temperature increase during the period. There is no clear pattern for the salinity.

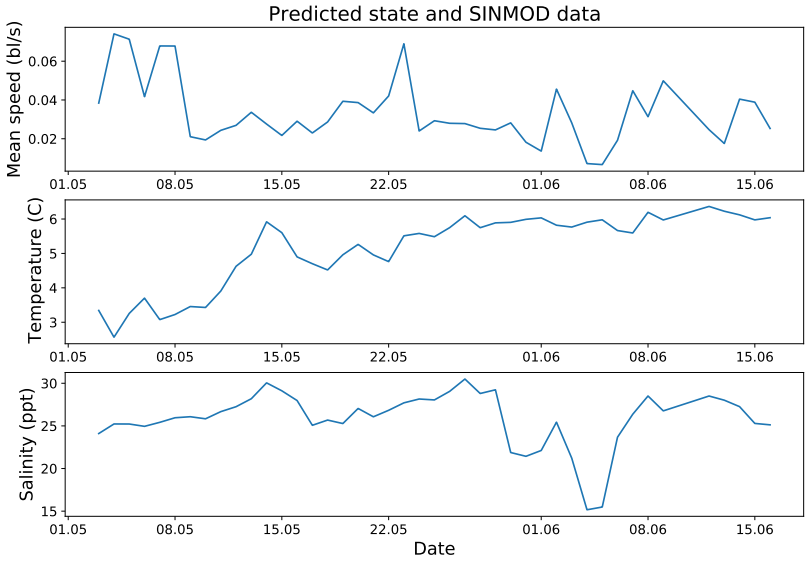


Figure 6.8: Mean estimated speed, water temperature and salinity at each day in the migration period.

The Kendall rank correlations between the estimated speed from the particle filter and the water temperature, salinity and velocity from SINMOD are presented in **Figure 6.9**. Correlations significant on a 10 % level are marked with *.

The speed, in body lengths per second, is negatively and significantly correlated with the water temperature. As presented in Chapter 6.3, the mean speed of the last migration group is significantly higher than the mean speed of the first migration group. The mean temperature at the estimated positions is found to be 4.7 °C for the first group and 5.6 °C for the last group. By again performing a Welch's t-test (Welch, 1947; Heumann & Shalabh, 2016) with null hypothesis of the two means being identical, the p-value is found to be 0.0012. Hence, we reject the null hypothesis, and we conclude that the mean temperature for the first and last migration group are statistically different. This can indicate that higher water temperature leads to increased swimming speed, which is the opposite of the conclusion drawn from the negative correlation. We would expect the swimming speed and water temperature to be positively correlated (Jonsson & Jonsson, 2011).

There is a small, negative correlation between swimming speed and salinity, but it is not significant on a 10 % level. The mean salinity for the first and last migration group are found to be 27.7 ppt and 24.2 ppt, respectively. The p-value of the Welch's t-test is 0.0013, and we accept the alternative hypothesis of the means being different. Hence, we can conclude that the first group has been exposed to more saline water than the last group. This is not consistent with the expectation that higher water salinity leads to increased swimming speed (Aas et al., 2011).

The northward water velocity is found to be negatively and significantly correlated with the estimated speed. The correlation is very small. There is a positive correlation between the eastward water velocity and the speed, but it is not significant. In a study performed in the Romsdalsfjord system, the direction of the movement of the observed salmon was not found to be dependent on the direction of the water current (Thorstad et al., 2004).

In general, the correlation between the particle filter estimates and the environmental data is low, and several of the variables does not have a significant correlation. These results are consistent with the results found by studying Atlantic salmon on the Scottish coast over a time span of 44 years. Here, the correlations between physical properties of the salmon and environmental variables were found to be weak and generally not significant (Bacon et al., 2009).

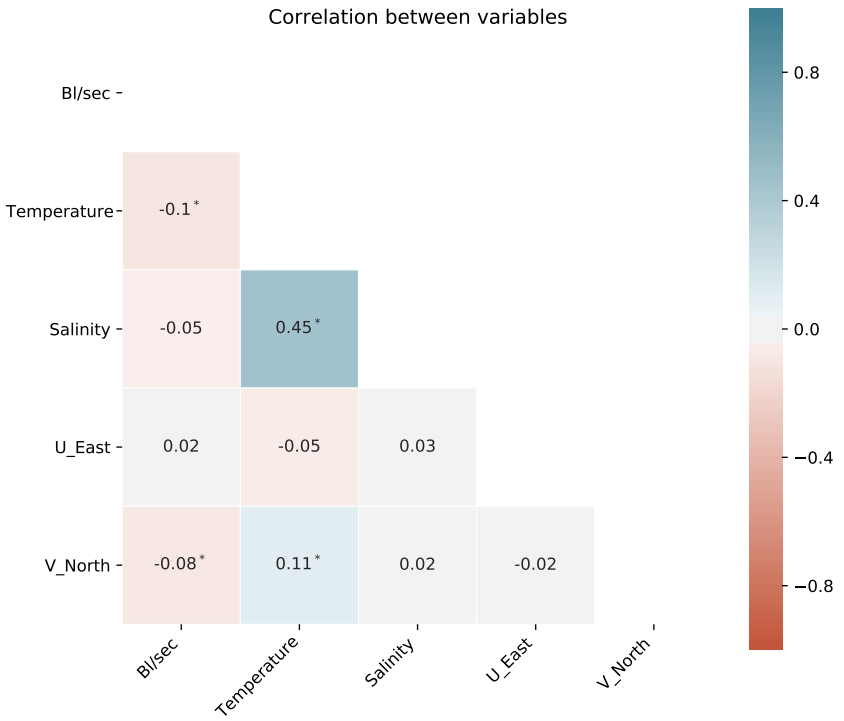


Figure 6.9: The Kendall rank correlation between the estimated velocity, temperature, salinity and seawater velocity. Correlations significant on a 10 % level are marked with *. The speed, given in body lengths per second, is denoted by "Bl/sec". The eastward and northward seawater velocity is denoted by "U_East" and "V_North", respectively.

The data from SINMOD are simulated. Actual registrations of water temperature and salinity are available from the location Haneholmen, with coordinates (347314, 6859034). We have data available from April 1 to July 31 at Haneholmen, with one registration each day at 10:00. The mean temperature from these registrations is 12.0 °C with a standard error of 3.2 °C. For each of the registrations at Haneholmen, SINMOD data from the time closest to the registration at Haneholmen is found. When averaging the data over all simulations at 3, 6 and 10 meters depth, the mean temperature of the SINMOD data from the grid point closest to the location of Haneholmen is found to be 5.8 °C with a standard error of 1.4 °C. The mean salinity from the registrations at Haneholmen is 16.3 ppt with a standard error of 7.4. By extracting the corresponding SINMOD data in a similar manner as for the temperature, the mean salinity of the SINMOD data is found to be 24.4 ppt, and the standard error is 2.0 ppt.

Temperature registrations are also available from Verpeide, with coordinates (300403, 6868886). We have registrations at 5 meters and 10 meters depth from May 26 to July 31, with one registration at each day at 08:00. The mean temperature at 5 and 10 meters depth is 12.1 °C and 11.6 °C, respectively, both with a standard error of 1.1 °C. SINMOD data is extracted from the grid point closest to the location of Verpeide. The mean temperature at both 6 and 10 meters depth is 6.9 °C, with a standard error of 0.6 °C.

The simulated temperature from SINMOD are considerable lower compared to real data at both Haneholmen and Verpeide. The measured salinity at Haneholmen is lower than the salinity from SINMOD. This could indicate that the SINMOD data are not sufficiently accurate, which could affect the correlation estimates presented above.

Chapter 7

Conclusion and Further Work

To obtain increased knowledge about the migration pattern of Atlantic salmon in Nordfjord, a state-space model for the salmon movement and the acoustic telemetry data was suggested, and a particle filter and a sequential fixed-lag smoother were applied to data collected in 2017. The goals of this project were to estimate the position and velocity of the salmon in the period when they migrate from the river to the ocean, and also to investigate if there was a correlation between these estimates and environmental data from the fjord.

118 salmon smolts were marked with acoustic transmitters and detected during the spring of 2017 in Nordfjord. 61 % of the salmon were detected at least once during the period. Only 23 % were detected in the outer fjord. The main part of the detections were made in May, and the median time between the first and the last detection was approximately 8 days.

A state-space model was considered. Each salmon in the study was considered independently of the others, and the detections of each specific salmon were the only factors influencing the estimated position and velocity. The particle filter that was applied to the data took both the presence and absence of data at the receivers into account. A fixed-lag smoother with different lag was also applied to the particle filter.

A time step of six hours was used, which meant that there could be several detections in the same interval. An approximation of the root mean square error (RMSE) was found by considering the mean of the position of all the receivers with a detection, and comparing this with the mean predicted position for all particles. Hence, we did not compare the prediction to the true position. This information was not available, and the receiver positions were used to approximate the true position of the salmon. Another measure of the model performance was the effective sample size (ESS). This is a measure of the variability of the weights. If a model has a high value of the ESS, the results of the model are more robust.

The performance of the particle filter was compared to that of a sequential fixed-lag smoother adjustment to the particle filter. A fixed lag of $L = 1, 2, 3$ and 4 were considered. The ESS of all models were relatively similar. The lowest value of the RMSE was obtained

by only applying the particle filter to the data, without a smoother adjustment. This gave a RMSE of 5802 meters and an ESS of 325. However, the model performance varied with different parameter choices. For a combination of parameters with less process noise then the combination selected for $L = 0$, applying a fixed-lag smoother adjustment to the particle filter improved the results.

By considering all the salmon that were detected in the outer part of Nordfjord during the migration period, the mean migration speed obtained from the particle filter predictions was found to be 0.48 bl/s, with a standard error of 0.27 bl/s. The salmon were divided into two groups based on the first detection time. The mean speed was found to be 0.46 bl/s and 0.51 bl/s for the salmon in the first and last group, respectively. This is somewhat lower than what we would have expected based on results from related studies.

The state-space model did not consider environmental variables like temperature and salinity explicitly. These factors are expected to have an effect on the behavior of salmon. In this case, however, most correlations between the estimated speed and the environmental variables were found to be weak and not significant. There was a small, negative correlation between the estimated speed and temperature. The correlation was significant on a 10 % level according to a Kendall rank correlation test. From an ecological view, we would expect the swimming speed to be positively correlated with the water temperature. We would also expect the speed to be positively correlated with the length and weight of the salmon. Here, the correlation was not found to be significant on a 10 % level.

There can be high local variations of the sea water temperature and salinity. In addition, the values are dependent on the depth. The state-space model did not consider depth. In addition, the resolution of the simulated environmental data (SINMOD) was low and only available at certain depths. When comparing the SINMOD data to real temperature and salinity measurements from the same period, the values differed a lot. It is possible that one, or several, of these reasons can be part of the explanation of why we did not see the expected correlations between the filter estimates and the environmental variables.

Overall, the particle filter algorithm seemed to work well on the data. The ESS was high, which should ensure robust results. The estimated movement pattern was highly dependent on the detections, and the results looked similar for the particle filter and the sequential fixed-lag smoother with different lags. A possible way to increase the model performance could be to include the environmental variables and physical properties of the salmon in the state-space model. There is ongoing work to develop techniques to enable autonomous fish tracking by placing a receiver on a moving vessel. Combined with one of the models presented in this thesis for predicting the position of salmon, this could make it possible to perform adaptive sampling. This can in turn give more detections and more accurate position measurements of the salmon.

In addition to including environmental variables in the model, the survival rate of the salmon could be taken into account to increase the model performance. All salmon are not expected to survive the migration to the outer fjord, and this should be taken into account.

The detection probability at a receiver given the predicted state of the system was approximated by an exponentially decreasing function. In practice, the detection probability was found to be high up to a certain distance. After this point, the probability decreased. It could be interesting to compare the performance of models with different expressions for the detection probability.

In this project, all detections were weighted equally. If a salmon was detected at two receivers situated in close vicinity of each other, the velocity estimate should be more reliable than it is for two receivers far away from each other. An approach to improve the particle filter could be to give the detections different weights. In addition, no boundary conditions were applied to the movement of the salmon. This meant that the particles that were drawn to approximate the movement of the salmon could be outside the fjord. Implementing boundary conditions of the fjord is another possible extension of the model.

By considering more data, the results would get more reliable. Experiments on the migration pattern of salmon smolts in Nordfjord were also performed in 2018, 2019 and 2020. Performing analysis on data from several years and comparing results could give more insight on the movement pattern of the salmon. It would also give a more holistic perspective and more knowledge about the general behavior of the salmon.

References

Aas, Ø., Einum, S., Klementsen, A., & Skurdal, J. (2011). Atlantic salmon ecology. *Wiley Blackwell*.

Anderson, B. D. O., & Moore, J. B. (1979). Optimal filtering. *Prenctice-Hall, Inc., Englewood Cliffs, N.J.*.

Arfken, G. B., & Weber, H. J. (2005). *Mathematical methods for physicists* (6th ed.). Elsevier Academic Press.

Bacon, P. J., Palmer, S. C. F., MacLean, J. C., Smith, G. W., Whyte, B. D. M., Gurney, W. S. C., & Youngson, A. F. (2009). Empirical analyses of the length, weight, and condition of adult atlantic salmon on return to the scottish coast between 1963 and 2006. *ICES Journal of Marine Science*, *66*, 844–859. (doi.org/10.1093/icesjms/fsp096)

Brownscombe, J. W., Lédée, E. J. I., Raby, G. D., Struthers, D. P., Gutowsky, L. F. G., Nguyen, V. M., ... Cooke, S. J. (2019). Conducting and interpreting fish telemetry studies: considerations for researchers and resource managers. *Rev Fish Biol Fisheries*, *29*, 369-400. (doi.org/10.1007/s11160-019-09560-4)

Clapp, T. C., & Godsill, S. J. (1999). Fixed-lag smoothing using sequential importance sampling, in Bayesian Statistics VI. *Oxford University Press*, 743-752.

Daae, K. L., Staalstrøm, A., Urke, H. A., Viljugrein, H., Jansen, P. A., & Kandal, I. (2011). Aquastrøm nordfjord kartlegging og beskrivelse av strømforhold og risiko for smittespredning. *NIVA rapport 6194*.

Davidsen, J. G., Rikardsen, A. H., Halttunen, E., Thorstad, E. B., Økland, F., Letcher, J., ... Næsje, T. F. (2009). Migratory behaviour and survival rates of wild northern atlantic salmon salmo salar post-smolts: effects of environmental factors. *Journal of Fish Biology*, *75*, 1700-1718. (doi.org/10.1111/j.1095-8649.2009.02423.x)

Doucet, A., De Freitas, N., & Gordon, N. (2001). Sequential monte carlo methods in practice. *Springer*.

Doucet, A., & Johansen, A. M. (2008). Particle filtering and smoothing: Fifteen years later.

Finstad, B., Økland, F., Thorstad, E. B., Bjørn, P. A., & McKinley, R. S. (2005). Migration of hatchery-reared atlantic salmon and wild anadromous brown trout post-smolts in a norwegian fjord system. *Journal of Fish Biology*, *66*, 86-96. (doi.org/10.1111/j.0022-1112.2005.00581.x)

Haugen, T. O., Jansen, P. A., Staalstrøm, A., Viljugrein, H., Kristensen, T., Molvær, J., ... Urke, H. A. (2014). Gyrosim sannsynlighet for spredning av gyrodactylus salaris: Kobling av 3d sirkulasjonsmodell og biologisk smittespredningsmodell. *INAQ AS Report to the Norwegian Environment Directory. 38 pages*.

Heumann, C., & Shalabh, M. S. (2016). *Introduction to statistics and data analysis*. Springer. (doi.org/10.1007/978-3-319-46162-5)

Hoar, W. S. (1988). 4 the physiology of smolting salmonids. *Fish Physiology*, *11*, 275-343.

Høgåsen, H. R. (1998). Physiologican changes associated with the diadromous migration of salmonids. *Canadian Special Publication of Fisheries and Aquatic Sciences*, *127*.

Jensen, A. J., & Johnsen, B. O. (1989). Laks og sjøaure i strynevassdraget 1982–1988. *NINA Forskningsrapport*, *4*, 1-27.

Johnson, S., D., London, M., J., Lea, M. A., & Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, *89*, 1208-1215. (doi.org/10.1890/07-1032.1)

Jonsson, B., & Jonsson, N. (2011). Ecology of atlantic salmon and brown trout: Habitat as a template for life histories. *Fish and Fisheries Series*, *33*.

Kendall, M., & Gibbons, J. D. (1990). Rank correlation methods. *Oxford University Press, New York*.

Kraus, R. T., Holbrook, C. M., Vandergoot, C. S., Stewart, T. R., Faust, M. D., Watkinson, D. A., ... Krueger, C. C. (2018). Evaluation of acoustic telemetry grids for determining aquatic animal movement and survival. *Methods Ecol. Evol*, *9*, 1489-1502. (doi: 10.1111/2041-210X.12996)

McCormick, S. D., Hansen, L. P., Quinn, T. P., & Saunders, R. L. (1998). Movement, migration and smolting of atlantic salmon (*Salmo salar)*. *Canadian Journal of Fisheries and Aquatic Sciences*, *55 (Suppl. 1)*, 77-92.

Miller, K. M., Teffer, A., Tucker, S., Li, S., Schulze, A. D., Trudel, M., ... G., H. S. (2014). Infectious desease, shifting climates, and opportunistic predators: cumulative factors potentially impacting wild salmon declines. *Evolutionary Applications*, *7*, 812-855.

Ministry of Climate and Environment. (2019). Nasjonale laksevassdrag og laksefjorder. (Retrieved 30.09.2019, from https://miljostatus.miljodirektoratet.no/tema/ferskvann/laks/nasjonale-laksevassdrag-og-laksefjorder)

Ministry of Trade, Industry and Fisheries. (2019). Regjeringen skrur på trafikklyset i havbruksnæringen. (Retrieved 04.02.2020, from https://www.regjeringen.no/no/aktuelt/regjeringen-skrur-pa-trafikklyset-i-havbruksnaringen/id2688939/)

Norgeskart. (2019). (Retrieved 05.02.2019, from https://norgeskart.no/)

Nulsen, J., Baxter, P., & Wood, T. (2015). Smoothing methods for particle filters in tracking applications. *IMA Conference on Mathematics in Defence 2015*.

Plantalech Manel-La, N., Thorstad, E. B., Davidsen, J. G., Økland, F., Sivertsgård, R., Mckinley, R. S., & Finstad, B. (2009). Vertical movements of atlantic salmon postsmolts relative to measures of salinity and water temperature during the first phase of the marine migration. *Fisheries Management and Ecology*, *16*. (doi.org/10.1111/j.1365-2400.2009.00658.x)

Saloniemi, I., Jokikokko, E., Kallio-Nyberg, I., Jutila, E., & Pasanen, P. (2004). Survival of reared and wild atlantic salmon smolts: size matters more in bad years. *ICES Journal of Marine Science*, *61*, 782–787. (doi.org/10.1016/j.icesjms.2004.03.032)

Siebert, J. R., Musyl, M. K., & Brill, R. W. (2003). Horizontal movements of bigeye tuna (thunnus obesus) near hawaii determined by kalman filter analysis of archival tagging data. *Fish. Oceanogr.*, *12:3*, 141-151. (doi.org/10.1046/j.1365-2419.2003.00228.x)

Stefansson, S. O., Björnsson, B. T., Ebbesson, L. O. E., & McCormick, S. D. (2012). Smoltification, in *Fish Larval Physiology* (Finn, R. N., & Kapoor, B. G., eds). *Enfield, NH: Science Publishers Inc*, 639-681.

Särkkä, S. (2013). *Baysian filtering and smoothing*. Cambridge University Press.

Thorstad, E. B., Whoriskey, F., Uglem, I., Moore, A., Rickardsen, A. H., & Fistand, B. (2012). A critical life stage of the atlantic salmon *Salmo salar*: behaviour and survival during the smolt and initial post-smolt migration. *Journal of Fish Biology*, *81*, 500-542.

Thorstad, E. B., Økland, F., Finstad, B., Sivertsgard, R., Bjørn, P. A., & McKinley, R. S. (2004). Migration speeds and orientation of atlantic salmon and sea trout postsmolts in a norwegian fjord system. *Environmental Biology of Fishes*, *71*, 305-311.

Thorstad, E. B., Økland, F., Finstad, B., Sivertsgård, R., Plantalech, N., Bjørn, P. A., & McKinley, R. S. (2007). Fjord migration and survival of wild and hatchery-reared atlantic salmon and wild brown trout post-smolts. *Hydrobiologia*, *582*, 99-107.

Ugedal, O., Kroglund, F., Barlaup, B., & Lamberg, A. (2014). Smolt: en kunnskapsoppsummering. *Miljødirektoratet M136–2014*, 128 pages.

Urke, H. A., Haugen, T. O., Kjærstad, G., Alfredsen, J. A., & Kristensen, T. (2018). Laks- og aurebestanden i strynevassdraget; vandringsmønsteret hjå laksesmolt og aure, ungfiskproduksjon og botndyr. *MINA fagrapport 48*, 56 pages.

Urke, H. A., Kristensen, T., Arnekleiv, J. V., Haugen, T. O., Kjerstad, G., Stefansson, S. O., ... Nilsen, T. O. (2013a). Seawater tolerance and post-smolt migration of wild atlantic salmon *Salmo salar* x brown trout s. *trutta* hybrid smolts. *J. Fish Biol.*, *82*, 206-227.

Urke, H. A., Kristensen, T., Ulvund, J. B., & Alfredsen, J. A. (2013b). Riverine and fjord migration of wild and hatchery reared atlantic salmon smolts. *Fisheries Management and Ecology 20*, 545-552. (doi: 10.1111/fme.12042)

Vemco. (2019). Vr2w - 69 khz.
(Retrieved 16.10.2019, from https://www.vemco.com/products/vr2w-69khz/?product-faqs#vr2wfaq1)

Vollset, K. W., Nilsen, F., Ellingsen, I., Finstad, B., Helgesen, K. O., Karlsen, , ... Dalvin, S. (2019). Vurdering av lakselusindusert villfiskdødelighet per produksjonsområde 2019. *Rapport fra ekspertgruppe for vurdering av lusepåvirkning*.

Vøllestad, A. (2018). Smolt. *Store norske leksikon*. (Retrieved 16.10.2019, from https://snl.no/smolt)

Welch, B. L. (1947). The generalization of "student's" problem when several different population varlances are involved. *Biometrika*, *34*, 28-35. (doi.org/10.1093/biomet/34.1-2.28)

Xydes, A., Moline, M., Lowe, C. G., Farrugia, T. J., & Clark, C. (2013). Behavioral characterization and particle filter localization to improve temporal resolution and accuracy while tracking acoustically tagged fishes. *Ocean Engineering*, *61*, 1-11. (doi.org/10.1016/j.oceaneng.2012.12.028)

Appendix

A.1 Receiver Data

Each receiver has a name and a unique ID. In **Table A.1**, the columns "Easting" and "Northing" give the UTM coordinates, zone 32, of the receiver and the column "Zone" indicates which zone the receiver is situated in. The position is found from east to west, with some manual sorting when needed, as described in Section 2.2.

Table A.1: Name, ID and position of receivers.

Name	ID	Easting	Northing	Zone	Position
Mindresunde	109708	388756	6868757	River	1
Soget	106226	387182	6868044	River	2
Run to foss	121330	385615	6867449	River	3
Saetre	121196	383056	6866420	River	4
Smaleveien	105128	382368	6866651	River	5
Holmane	121323	381637	6865940	River	6
Bill Dannat	123472	380125	6865323	Estuary	7
Tenden	105707	379445	6864956	Estuary	8
Osen midt	121199	379297	6865006	Estuary	9
Osen N	121315	379078	6864758	Estuary	10
Neset	108205	378682	6864067	Estuary	11
Stavenesvika	105718	379535	6863534	Estuary	12
Aarholen Soerside tunell	104259	378296	6862231	Inner	13
Rake	123469	384006	6861495	Inner	14
Lovika Stake	119954	385503	6861596	Inner	15
Lovika Brygge	109700	386325	6861505	Inner	16
Olden	105130	384425	6858138	Inner	17
Faleide Soer	121218	374588	6862183	Inner	18
Faleide Nord	121195	374316	6863735	Inner	19
Utvik	121197	370165	6860034	Utvik	20
Blakset	105961	368853	6855155	Utvik	21
Ysteneset	122234	349502	6860366	Lote	22
Ysteneset II	123466	349165	6860583	Lote	23
Finnvika	121211	348318	6858839	Lote	24

Continued on next page

Table A.1 Continued from previous page

Name	ID	Easting	Northing	Zone	Position
Verpeneset	121212	347089	6859937	Lote	25
Anda	121316	346237	6860509	Lote	26
Rygg II	122236	348689	6852812	Lote	27
Lote	123473	345960	6862780	Lote	28
Kviteneset	121317	344029	6860617	Lote	29
Kleppeneset	121191	340999	6856671	Lote	30
Hyeneset	109698	341244	6861045	Lote	31
Hundvikneset	122247	338429	6864354	Krokneset	32
Hammeren	123474	336363	6862413	Krokneset	33
Djupdalsneset	105959	336019	6864237	Krokneset	34
Krokneset II	103442	333003	6862824	Krokneset	35
Krokneset I	109704	332379	6862458	Krokneset	36
Hjeltneset	105695	330611	6862065	Krokneset	37
Isane anlegg	122239	327691	6864528	Isane	38
Isane fk	122244	326996	6865728	Isane	39
Aaseneset	121203	328643	6866261	Isane	40
Eidsfjorden	104258	332568	6867224	Isane	41
Staarheim Oest	121214	326936	6868693	Isane	42
Staarheim Vest	102909	325736	6868623	Isane	43
Maaurstad S	121213	314102	6871027	Maurstad	44
Maurstad V	121202	313373	6871643	Maurstad	45
Otteren	105713	312779	6870186	Maurstad	46
Totland	122229	311261	6871456	Maurstad	47
Elde	105140	310094	6869281	Maurstad	48
K. stroemmen	106229	307763	6867270	Maurstad	49
Tofteneset	122231	306990	6869508	Maurstad	50
Rugsund	121327	305864	6865473	Outer	51
Almenningsflua	102912	303258	6869130	Outer	52
Almenning	102518	303020	6869731	Outer	53
Skorpeholmen	105715	301586	6867882	Outer	54
Skorpeholmen M	109699	301449	6867712	Outer	55
Skorpeholmen S	121204	301420	6867417	Outer	56
Tangane	121208	299910	6867040	Outer	57
Biskjelneset	122230	299416	6868425	Outer	58
Risoeya	105958	299139	6867085	Outer	59
Gangsoeya	121321	298421	6867538	Outer	60
Vemmelsvika	123475	298423	6868698	Outer	61
Vaageneset	102522	296403	6869234	Outer	62
Gaasholmen	122233	295679	6870070	Outer	63
Skavoeypollen	121219	296665	6870985	Outer	64
Ulvesundet W	121216	298162	6875094	Outer	65

Continued on next page

Table A.1 Continued from previous page

Name	ID	Easting	Northing	Zone	Position
Ulvesundet E	105698	297808	6875307	Outer	66

Concluded

A.2 Fish Data

In the study, 118 pre-smolt salmon are caught and marked with transmitters. 53 salmon are equipped with transmitters with a depth sensor and have IDs between 26 and 160. The transmitter IDs of the 65 salmon with regular transmitters have IDs between 1913 and 2957. In **Table A.2**, the first and registration times of the 72 salmon with at least one detection in the processed data are presented. In addition, the weight (g) and length (mm) are given. The column "Outer" has the value "Yes" if the salmon have been detected in the outer part of the fjord, "No" if not.

Table A.2: ID, registration times and number of registrations for each fish.

ID	First Registration	Last Registration	Weight	Length	Outer
27	2017-05-04 23:13:19	2017-05-13 06:34:40	16.7	129	Yes
28	2017-05-16 21:08:15	2017-05-17 18:39:36	30.3	152	No
29	2017-05-17 03:10:17	2017-05-23 07:26:01	22.2	137	Yes
30	2017-05-04 19:37:12	2017-05-09 15:21:06	23.4	148	No
31	2017-05-10 10:39:42	2017-05-23 08:45:23	42	167	No
32	2017-05-07 02:17:07	2017-05-23 04:41:20	21.1	130	No
34	2017-05-02 23:49:27	2017-05-10 15:59:33	24.5	147	Yes
35	2017-05-20 00:15:04	2017-05-28 20:10:30	24.8	143	No
36	2017-05-07 02:51:13	2017-05-27 09:56:11	20.7	136	Yes
37	2017-05-28 00:23:38	2017-06-16 02:54:13	27.6	155	Yes
38	2017-05-05 23:06:11	2017-05-13 20:07:00	29.2	154	Yes
39	2017-05-05 21:57:46	2017-05-12 08:19:00	25.7	144	No
40	2017-05-16 20:37:38	2017-05-27 03:06:02	23.4	138	Yes
42	2017-05-07 00:57:16	2017-05-22 06:33:52	18.6	138	Yes
44	2017-05-02 21:52:53	2017-05-13 12:54:23	16.6	128	No
92	2017-05-05 22:04:37	2017-05-06 21:05:26	32	158	No
95	2017-05-04 02:18:22	2017-05-07 07:46:06	33.7	164	No
100	2017-04-30 00:22:51	2017-05-06 18:22:01	16.8	136	Yes
128	2017-05-20 00:06:01	2017-06-14 13:54:48	18.7	128	Yes
129	2017-05-17 02:56:58	2017-05-22 17:59:27	20.5	136	No
130	2017-05-16 18:54:28	2017-05-17 11:08:35	20.4	126	No
133	2017-05-20 16:13:37	2017-05-28 01:52:42	19.9	132	Yes
137	2017-05-08 19:29:09	2017-05-20 07:14:55	25	149	No
138	2017-05-17 00:06:53	2017-05-27 10:07:46	19.5	132	Yes

Continued on next page

Table A.2 Continued from previous page

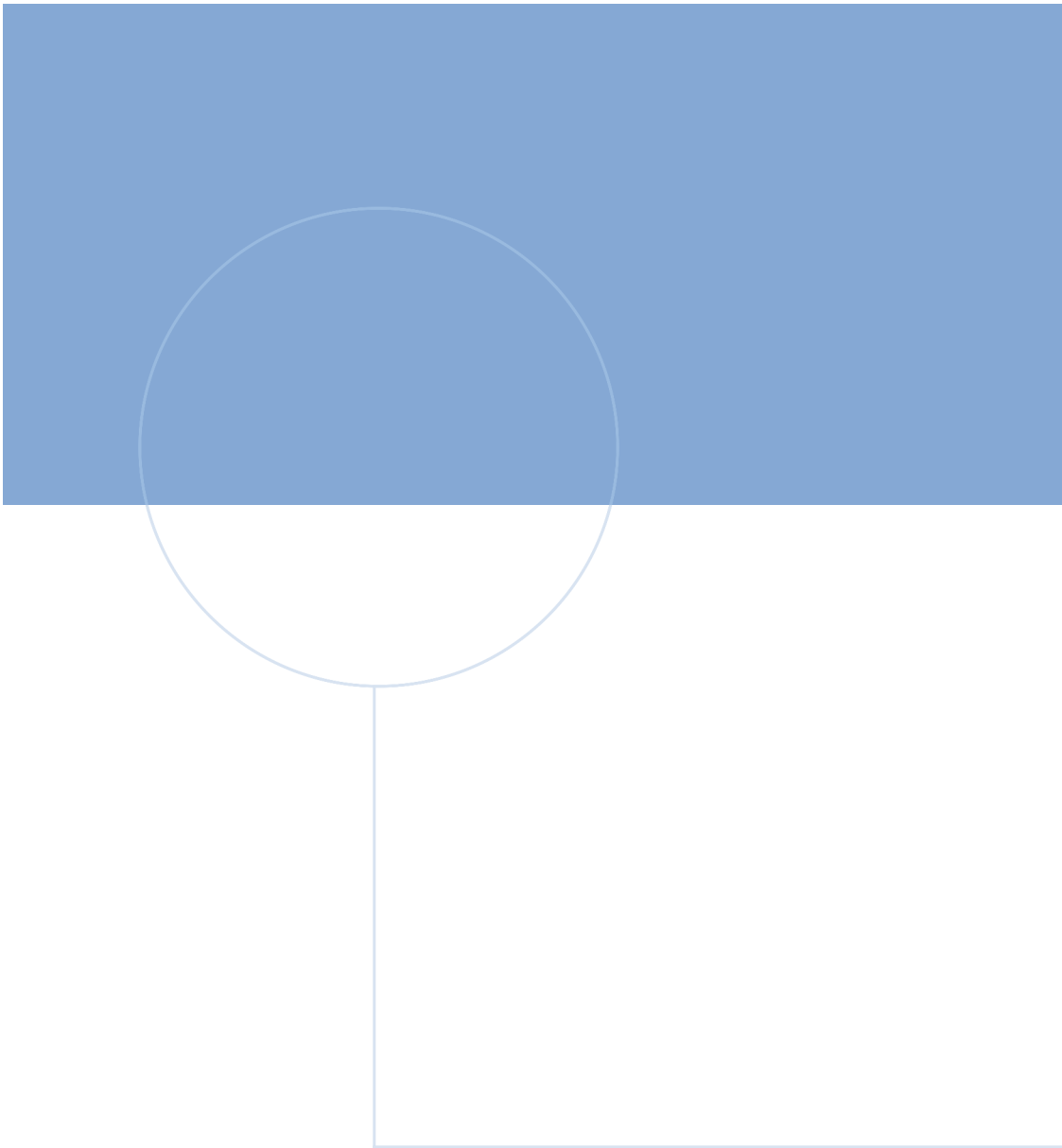
ID	First Registration	Last Registration	Weight	Length	Outer
139	2017-06-10 21:27:40	2017-06-12 21:10:59	25.9	146	No
144	2017-05-02 02:09:40	2017-05-27 15:18:29	22.3	141	No
146	2017-05-05 22:30:58	2017-05-13 06:44:52	20	140	Yes
148	2017-05-06 23:03:59	2017-05-14 12:39:17	19.3	129	Yes
151	2017-05-20 18:36:22	2017-05-27 12:45:37	29.6	152	Yes
152	2017-05-03 00:07:22	2017-05-09 10:25:38	26.6	146	No
154	2017-05-17 00:44:06	2017-05-21 12:56:00	26.6	146	No
155	2017-05-27 21:30:57	2017-06-02 18:51:29	25.1	146	No
156	2017-05-07 23:46:23	2017-05-22 06:10:30	24.1	135	No
157	2017-05-17 02:20:56	2017-06-12 22:18:57	26	136	Yes
158	2017-05-18 22:20:32	2017-05-24 19:29:40	19.5	138	Yes
159	2017-05-07 20:30:07	2017-05-16 15:52:10	22.9	143	Yes
160	2017-05-16 21:06:41	2017-05-27 15:33:57	21.2	136	No
2898	2017-05-01 23:21:16	2017-06-06 02:35:18	13.5	124	No
2899	2017-05-06 17:26:15	2017-05-12 15:54:35	18.6	134	Yes
2900	2017-04-29 22:19:53	2017-07-15 01:35:10	12.6	121	No
2903	2017-05-16 00:55:30	2017-05-16 20:35:04	15.2	126	No
2905	2017-05-04 00:11:38	2017-05-08 16:22:37	15.4	122	Yes
2906	2017-05-18 01:20:28	2017-05-23 14:25:42	16.6	126	No
2908	2017-05-06 08:46:40	2017-05-17 15:18:16	15.7	135	No
2909	2017-05-07 01:49:58	2017-05-12 23:24:08	18.5	132	No
2914	2017-05-17 00:04:19	2017-05-26 15:55:00	18.7	131	No
2915	2017-05-16 22:40:54	2017-05-23 15:28:40	18.9	134	No
2916	2017-05-03 00:02:10	2017-05-10 11:48:49	14.2	128	Yes
2917	2017-05-17 01:48:30	2017-05-27 13:30:06	16.5	126	No
2919	2017-05-17 01:13:35	2017-05-20 14:16:28	12.8	120	No
2920	2017-05-16 20:33:39	2017-05-24 09:56:10	13.1	120	No
2921	2017-05-10 20:55:36	2017-05-27 05:28:39	25	150	Yes
2922	2017-05-05 19:44:05	2017-05-14 02:03:19	27.6	152	No
2925	2017-05-14 23:51:40	2017-05-21 10:55:38	17.6	132	No
2929	2017-05-16 22:43:30	2017-05-26 07:24:34	26	142	No
2931	2017-05-13 16:37:38	2017-05-21 17:36:21	15.2	121	No
2932	2017-05-17 00:19:19	2017-05-22 07:32:04	13.4	112	No
2934	2017-05-10 02:49:30	2017-05-17 09:57:53	17.6	132	No
2935	2017-04-29 02:28:11	2017-05-14 19:28:44	13.3	128	Yes
2939	2017-05-05 23:16:51	2017-05-12 02:01:42	18	133	Yes
2940	2017-05-01 01:50:45	2017-05-11 09:43:58	16.6	129	Yes
2941	2017-05-01 23:02:56	2017-05-06 02:45:55	16.2	130	No
2943	2017-05-03 23:42:24	2017-05-23 00:30:02	12.5	120	Yes
2947	2017-05-13 18:24:35	2017-05-20 16:55:05	14.4	121	No
2949	2017-05-01 08:24:42	2017-05-20 14:26:09	16	123	No

Continued on next page

Table A.2 Continued from previous page

ID	First Registration	Last Registration	Weight	Length	Outer
2950	2017-05-16 22:47:08	2017-05-26 08:01:05	14	125	No
2952	2017-05-11 23:21:52	2017-05-19 13:38:30	13.3	124	No
2953	2017-05-25 22:23:01	2017-05-25 22:23:01	12.4	116	No
2954	2017-05-16 19:05:16	2017-05-23 09:12:49	15.7	131	Yes
2955	2017-05-16 23:06:35	2017-05-21 10:23:36	20.3	136	No
2956	2017-05-16 23:12:50	2017-05-20 11:18:58	14.6	122	No
2957	2017-05-07 23:04:13	2017-06-09 21:55:12	13	127	No

Concluded



NTNU

Norwegian University of
Science and Technology