



NTNU – Trondheim
Norwegian University of
Science and Technology

A Continuous-Time, Discrete-Space
Statistical Model for Estimating Dispersal
Rate Applied to a Population of House
Sparrows (*Passer Domnesticus*) in
Northern Norway

-

Kristin Mathilde Drahus

Master of Science in Statistics

Submission date: June 2015

Supervisor: Jarle Tufto, MATH

Co-supervisor: Henrik Jensen, IBI

Norwegian University of Science and Technology
Department of Mathematical Sciences

A Continuous-Time, Discrete-Space Statistical
Model for Estimating Dispersal Rate Applied
to a Population of House Sparrows (*Passer
Domesticus*) in Northern Norway

Master's thesis, ST3900

Norwegian University of Science and Technology (Norges Naturvitenskapelige
Universitet, NTNU)

Kristin Mathilde Drahus

June 1, 2015

Abstract

How individuals move and interact with each other is a complex matter. Even so, the knowledge of how such movements and interactions happen is important to understand in order to be able to keep genetic variation and prevent extinction of species. Also it is important to be able to limit dispersal of unwanted diseases or pests. In this thesis I have investigated different models for dispersal rates for individuals in a set of partially isolated subpopulations of house sparrows in Northern Norway. The models differ in what factors contribute to increase or decrease the dispersal rate. By using such models one can get a broader understanding of what may cause an individual to disperse from one location to another, and hence it gives a pointer in what way one should act in order to adjust the dispersal.

In my master's thesis I have used discrete localities and continuous time. The localities are defined to be the ten farms on the island where observations were made. In the study, distance as well as home locality, sex and date is tested for contributions to the dispersal rate. The expanded models seem to give a more likely result than the simple ones, suggesting that all the factors tested for make a significant improvement to the model and that the dispersal is in fact affected by many factors.

Sammendrag

Hvordan individer beveger seg og samhandler er en sammensatt sak. Allikevel er kunnskap om slik bevegelse og samhandling viktig å ha for å kunne sørge for å beholde genetisk variasjon og hindre utryddelse av arter. Det er også viktig for å kunne begrense spredning av uønskede sykdommer og skadelige arter. I denne studien har jeg undersøkt ulike modeller for spredningsrate for individer i en gruppe delvis isolerte subpopulasjoner av gråspurv i Nord-Norge. Modellene skiller seg fra hverandre i forhold til hvilke faktorer som bidrar til økning eller senkning av spredningsraten. Ved å prøve ut slike modeller kan man få bredere kunnskap om hva som kan få et individ til å reise fra en lokalitet til en annen. Dermed kan slike modeller gi en pekepinn på hvor eventuelle endringer bør gjøres for å påvirke spredningen.

I masteroppgaven benytter jeg diskrete lokaliteter og kontinuerlig tid. Lokalitetene er de ti gårdene på det geografiske området hvor dataene er samlet inn. I studien er avstand, avstand til fødested, kjønn og dato testet for bidrag til spredningsraten mellom to punkter. De mer avanserte modellene ser ut til å gi bedre resultater enn de enkle. Dette kan indikere at alle faktorene testet i denne studien gir en forbedring i modellen for spredningsrater, og ikke minst at spredningsratene er påvirket av mange faktorer.

I would like to thank my advisor Jarle Tufto for help and support throughout the process, and also my co-advisor Henrik Jensen for his help with the data set and research. I would also like to thank my parents and Per Inge for support when I needed it, and especially my mother for proof reading and offering feedback on my thesis. A special thank you to my friends and family who have followed the process of writing my thesis from the beginning.

Kristin Mathilde Drahus, Trondheim, June 2015

Contents

1	Introduction	7
2	Definitions	8
3	Data Set	10
3.1	Distances	11
3.2	Multiple Data	12
3.3	Organization of Data Set	13
4	General Model	13
4.1	Probability of Movement	14
4.1.1	Factorizing	15
4.2	Inclusion of Death Rate	17
4.3	Time Adjustment	18
4.4	Expression for Log Likelihood	19
4.4.1	Individual Likelihoods	20
4.4.2	Example	24
4.4.3	Total Likelihood	26
4.5	Expression for Rates	26
5	Models	26
5.1	Model 1: Distance Only	27
5.2	Model 2: Home-centered	27
5.2.1	Model 2a	28
5.2.2	Model 2b	28
5.3	Model 3: Differences Between Sexes?	29
5.4	Model 4: Seasonal Changes?	30
6	Simulation of Data Sets	30
7	Comparison and Testing of Models	32
7.1	Maximum Likelihood Estimators	33
7.2	The Likelihood Ratio Test	33
7.3	Standard Deviations	34
7.3.1	Standard Deviations by Fisher Information	34
7.3.2	Standard Deviations by Bootstrapping	35
7.3.3	The Standard Deviation Estimates	36
7.4	Confidence Intervals and Bias	36

8	Results	38
8.1	General Settings	39
8.2	Model 1	39
8.3	Model 2a and 2b	42
8.3.1	Model 2a	43
8.3.2	Model 2b	46
8.3.3	Model 2a and Model 2b	49
8.4	Model 1 and model 2b	49
8.5	Model 3	49
8.6	Model 2b and model 3	52
8.7	Model 4	53
8.8	Model 3 and Model 4	57
8.9	General	57
9	Discussion	58
10	Conclusion	61
A	The Log Likelihood Function	68
A.1	Implementation of the Log Likelihood Function	68
A.2	Issues in and Adjustments to the Log Likelihood Function . .	69
A.2.1	Rounding of Probabilities	69
A.2.2	Imaginary Values	69
A.2.3	Scaling of Parameters	70
A.2.4	Choice of Method	71
A.2.5	Limitations on Parameters	72
A.3	Notes on the Different Models and Simulations	73
A.3.1	Model 1	73
A.3.2	Model 2a	73
A.3.3	Model 2b	74
A.3.4	Model 3	74
A.3.5	Model 4	74
A.4	R-code for Log Likelihood Function	75
B	Calculations	78
B.1	The Haversine Formula	78
C	A Note on R	78

1 Introduction

For several years large amounts of data have been collected on house sparrows (*Passer domesticus*) on the coast of Northern Norway. The house sparrow is a small bird often found close to human settlements where food resources like farming waste is accessible. For practical reasons, the geographical area in this study is limited to the island Hestmannøy, which is on Helgelandskysten outside the city of Mo i Rana.

The island is scarcely populated and has ten farms where data is collected. Thus it is likely that the sparrows live close to the farms as they can supply food and shelter. In this study I have assumed that as long as an individual is alive, it will be on a farm. The bird can be at different farms during its lifetime, but at any one point in time, each individual is on a farm if it is alive. Ringsby et al. (2006) study the extinction time of a house sparrow population after the last farm on another island in the area is shut down. In the study they find that after the farm is closed, the adult death rate increased and new recruitment decreased, which are important reasons for the extinction. Therefore, the assumption of discrete locations is justified in that it is likely that any sparrow at any one time will be at one of the ten farms.

In addition to resources, genetic variation is essential for a population to survive. An important source for large genetic variation is migration, in that individuals move from one population to another and breed there, thus possibly extending the genetic pool in the population where they arrive. In populations completely cut off from others of the same species, inbreeding may become a serious problem if the population is not of sufficient size to maintain a large enough genetic pool. This could cause inbreeding depression. This means that a large number of negative recessive conditions being expressed in an individual as explained by Halliburton (2004, p. 286-293) . This is the result of closely related parents breeding and thus increasing the risk of their offspring receiving a double dose of negative recessive genes. As discussed in Halliburton (2004, p. 290-292), this could potentially contribute to extinction in small populations. The genetic pool on an island is typically smaller than on the mainland (Jensen et al., 2013), increasing the risk of a small genetic pool in this case. An endangered population is not likely to be very large, and the distance to another population of the same species might be too large for individuals to travel. This illustrates the importance of having a thorough knowledge of how genes can be passed between populations, as this knowledge could help prevent extinctions. This could for instance be done by aiding movement between populations so a certain genetic exchange is present. A large enough genetic pool is also important to prevent extinc-

tion as it increases the probability that some individuals will be able to adapt to possible changes, either in climate, or in their environment.

Because of the assumption that the birds are at one of the ten farms at all times, Hestmannøy creates a tiny model of several subpopulations of sparrows that are partially isolated. That is, the subpopulation on one farm has limited migration to and from other farms. Hence the setup can be used to model how populations with limited contact spread their genes and interact with other populations. Thus the model is useful for situations where subpopulations have limited contact, as may be the case for endangered species.

On the other hand, knowledge of how individuals move does not only help aid in the migration of individuals, it can also help prevent the spread of unwanted elements. For instance, if a population has individuals carrying a contagious lethal disease, it could be important for the management of other populations. If the infected population has individuals who can disperse to other localities, such knowledge is important to prevent spread from the contaminated population or to ensure that proper precautions are taken. Even if the disease is not lethal, knowledge of the dispersal rate and distance can help prepare the governing body in surrounding areas of the oncoming disease so that they can take the necessary precautions.

The aim of my thesis is to attempt to find a good model for the dispersal rate of house sparrows. Ovaskainen (2004) conducted a similar study on butterflies in continuous space. In his study the entire area is divided into different habitat types according to how individuals move, with adjustments for border areas and habitat areas. In my study I have used discrete localities, the ten farms on Hestmannøy, and continuous time. In the study I gradually include new factors to the function of the rate, to see what factors give an improved result.

2 Definitions

In this thesis the use of the word observation will normally refer to a case where an individual is either caught or seen. The total number of observations made at the different locations is shown in figure 1 for the reduced data set for model 1.

The term sampling point is used to indicate a pair of a locality and a date at which observations were taken. In other words that someone was present to make observations of the birds at that location at that date. Each location, that is farm, is given a location number, see table 2.0.1. Similarly, each date is denoted by a t_s , which is the date for sampling point s . Note

Locality number	Locname
1	Isachsen
2	TorMathisen
3	Randecker
4	Heen
5	BirgerMathisen
6	Klausen
7	Hjørdis
8	Asbjørn
9	Danielsen
10	Sletthågen

Table 2.0.1: Locality name and number

that the definition of a sampling point means that two sampling points may have the same date so each date does not have one unique s , but they will then not have the same locality. This will be the case if observations were made on the same day at more than one locality. For instance, observations were made on on the 6th of May 2011, at the Danielsen farm. Hence this time and place pair is a sampling point, denoted by $(l_2, t_2)=(9,20110506)$, as the Danielsen farm is farm number 9 and time-location pair is the second sampling point at the chronological list of points.

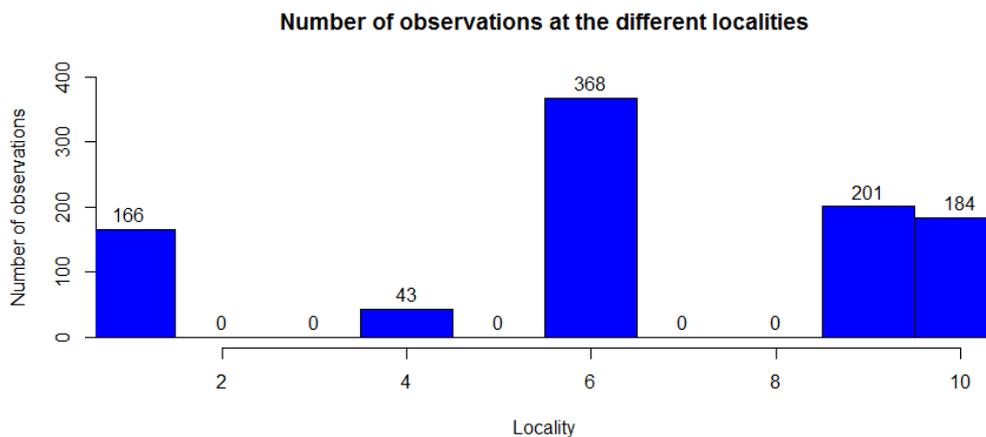


Figure 1: Histogram of observations at each location

An individual's catch-mark-resight history is denoted by Y_i . The history consists of a series of variables, $X_{i,s}$, one for each sampling point after the

individual enters the study. The value of $X_{i,s}$ is 0 if individual i is not observed at sampling point s , and 1 if it is observed.

3 Data Set

The data set was provided by Henrik Jensen at the Department of Biology at NTNU. For practical reasons, the data used for this study has been limited to the data collected on one of the islands, Hestmannøy, located outside of Mo i Rana in Northern Norway. The data set is collected through the catch-mark-resight method, where each individual is caught, marked, released and later might be caught or observed again. The marking makes it possible to identify the different individuals, hence tracing each individual's history.

The observations of the birds are made at different times throughout the year. That is, the birds are not continuously followed with a computer chip or something of the sort. This means that the only times an observation may take place is when data was actually collected. Hence, individuals may have moved more than what is indicated by the data, or have gone through other localities between two observation times. An individual is considered to enter the study when the first observation of it is made.

The data set includes information on many individuals. Each bird is identified by a ring number, and data is collected and sorted according to this unique identifier. This means that for each sampling point, an observation is either made of an individual or not, simply depending on whether the bird is observed at the sampling point. The fact that the bird is not seen or caught (observed), does not necessarily mean that it was not there, something taken further into account later.

To further limit the data set, only juvenile individuals were taken into account, and only data collected in 2011 was used as part of the data set, resulting in only individuals born in 2011 being used in the project. This narrowed the data set down to data on $I = 192$ individuals, and 962 observations (not including multiple data). Note that all sample times and locations from 2011 were included, also if juveniles were not observed at that particular time and place, as this also gives useful information to the models. This leaves a total of $S = 142$ sampling points to be considered. Note also that the number of individuals can differ depending on what model is used, because some individuals lack data that is needed for certain models.

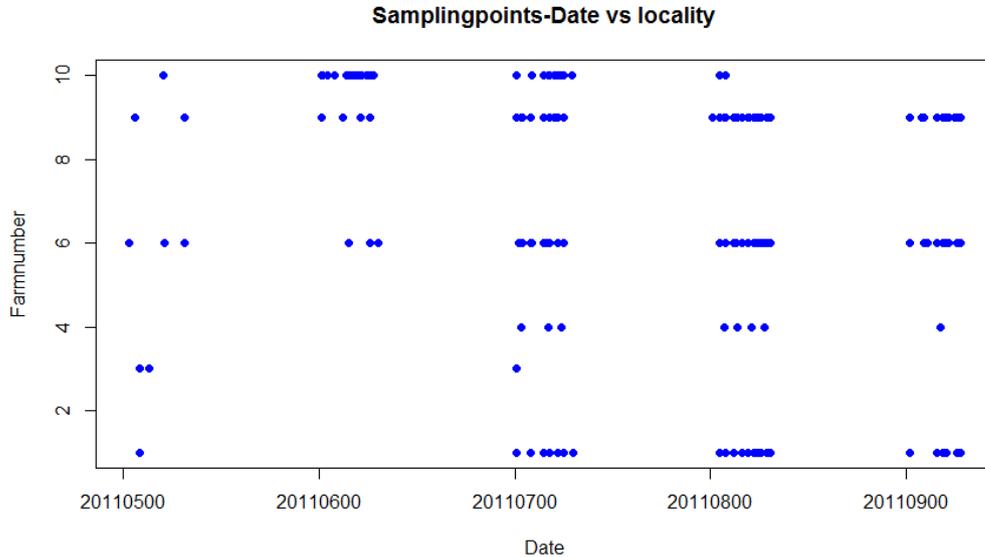


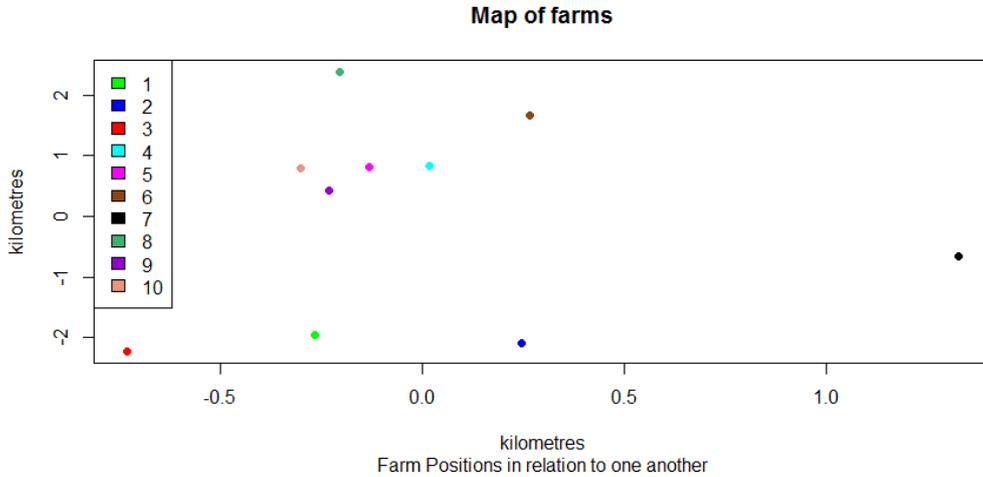
Figure 2: The distribution of the sampling points by location and date(yyymmdd).

3.1 Distances

When finding the distance between two locations, I used the values of the longitude and latitudes of the data collected at each farm. Henceforward, the term collected data refers to the data that is real, that is, collected in the field. This is meant to be the opposite of the "simulated data" which is computer generated, and not the result of observations. Each farm only had one latitude and one longitude value, so it is assumed that all individuals observed on a farm is observed at the same coordinates. Note that these coordinates are based on the complete data set, including double or multiple observations, as not all locations were observed at in the limited data set. This should not affect the results as the distance should be constant.

The function I used to calculate the distance used the mean of the longitudes and latitudes collected as each farms longitude and latitude value. The use of the mean allows for small differences in the longitude/latitude measurements, for example within the area of a large farm. However as mentioned, in the data set used for this project, all data from the same farms have the same longitude and latitude, so use of the mean values would not have been necessary in this case, and thus has no effect as the mean is the same as all the measurements for each farm.

From the calculated mean position for each farm, I calculated a distance



matrix giving the distances between the farms, based on the Haversine formula (Movable Type Ltd. (2014) , see appendix B.) which calculates the distance between two points (on a sphere, not elliptical) given by the radius of the sphere (the earth) and the longitudes and latitudes of the two points. The distance matrix is thus a symmetric $n_l \times n_l$ matrix, where n_l is the number of locations, here $n_l = 10$, in which element i, j denotes the approximate distance (in meters) between location i and location j .

3.2 Multiple Data

While working with the data set, it became clear that certain individuals were observed more than once at certain sampling times, either at different localities, or in the same locality. Since the time units are days, it is reasonable that it is possible to observe an individual twice within the course of a day, if different localities are sampled from in the same day. However, this causes an issue. From the model, an assumption that the different dates do not overlap follows, as each date is seen merely as one point in time, not as an interval. Therefore, it does logically not make sense that an individual is at several farms during one date, and thus such data has to be omitted from the set.

Multiple observations are therefore removed by reducing them to one observation, that is, only one observation is permitted per individual per day. This is simply done in the case of multiple observations at the same sampling point, and for the case where an individual is observed at the same date and at different localities, I have adjusted the data so that only one time-location observation, randomly chosen between the observations (with

equal probabilities), is present in the capture history data used. A total of 17 observations were removed for one of these two reasons, leaving the data set to include 962 observations. This means that each observation had to be checked to see if the individual was already observed on that date, and if that was the case, the two (or more) observations had to be randomly chosen from.

3.3 Organization of Data Set

To organize the data set in a practical way, the variable $X_{i,s}$ is introduced. The variable indicates whether or not individual i is observed at sampling point s . If the individual is observed at sampling point s , $X_{i,s}$ takes the value 1. Otherwise, the value is 0.

From this, an observation matrix could be made. In this matrix, each row represents one individual, and each column represents one sampling point. Note that one time (date) can have several columns in cases where several localities were sampled from in the same day, recall that a sampling point is a date and location pair.

In the observation matrix, each element is one of the X-s mentioned in section 2. All the $X_{i,s}$ -s in one row belong to the same individual, and together they represent that individual's capture-mark-resight history, that is that individual's Y . In other words, element i, s in the observation matrix is 1 if individual i was observed at sampling point s , and 0 if not. This organization of the data hence gives a quite simple overview of each individual's catch -release history.

4 General Model

The goal of the study is to find a statistical model that fits the data well, so multiple models should be tested to find out what works best. As a basis for the different models in this study, a general model is used, which is adjusted for different model.

To find out what models work well, it is essential that one is able to calculate how well the data set fits the model. The value of the log likelihood can be used for this purpose. In this study, $p_i(t)$ indicates the probability of an individual being in locality i at time t . There are several individuals, so the total probability of observing the observed data given a correct model depends on all these individuals' probabilities.

4.1 Probability of Movement

The probability of an individual being in a location at a given date is assumed to follow a certain model. If one thinks of the probability of an individual being at a locality at a certain time, it would be reasonable to assume that this probability depends on where the individual was at the previous time. If this previous location is also unknown, this could also be represented by probabilities and so on. At the time of sampling point s , one can then assume that the probability of being at locality i is the sum of the probabilities of moving from the previous locality to locality i , and subtracting the probability of moving away from i if i was the previous locality. Note that the dispersal rate $\lambda_{i,j}$ is the dispersal rate from location j to location i , that is the rate at which an individual in locality j moves to locality i (see section 4.5). This gives the following differential equation (note that $p_i = p_i(t)$):

$$\begin{aligned} \frac{d}{dt}p_i(t) = & \lambda_{i,1}p_1 + \lambda_{i,2}p_2 + \dots + \lambda_{i,i-1}p_{i-1} + \\ & \lambda_{i,i+1}p_{i+1} + \dots + \lambda_{i,10}p_{10} - \sum_{j \neq i} \lambda_{j,i}p_j \end{aligned} \quad (1)$$

where the $p_i(t)$ is the probability of an individual being in location i at time t , given the history until time t . This scenario forms the basis for all the models in my thesis, the differences between the models being in the expressions for the rates. From the expression for the derivative above one can see that the positive additions in the start of the expression represents contributions that come from individuals moving to location i from the other locations, while the last sum is the contributions from location i to the other locations.

Rewriting this to matrix form, one gets the simple expression

$$\frac{d}{dt}p(t) = Ap(t). \quad (2)$$

This expression thus has to be solved. By the definition of an exponential matrix e^{At} where A is a matrix,

$$\begin{aligned} e^{At} &= I + At + \frac{1}{2}(At)^2 + \frac{1}{6}(At)^3 + \frac{1}{24}(At)^4 + \dots \\ &= I + At + \frac{1}{2}A^2t^2 + \frac{1}{6}A^3t^3 + \frac{1}{24}A^4t^4 + \dots \end{aligned} \quad (3)$$

The derivative of expression (3) with respect to t can now be found

$$\begin{aligned} \frac{d}{dt}e^{At} &= 0 + A + \frac{1}{2}2A^2t + \frac{1}{6}3A^3t^2 + \frac{1}{24}4A^4t^3 + \dots \\ &= A(I + At + \frac{1}{2}A^2t^2 + \frac{1}{6}A^3t^3 + \dots) \\ &= Ae^{At} \end{aligned} \quad (4)$$

Hence, because of equation (4) it is clear that

$$p(t) = e^{At}p(0) \quad (5)$$

is a possible solution of the differential equation because

$$\begin{aligned} \frac{d}{dt}p(t) &= \frac{d}{dt}(e^{At})p(0) \\ &= \left(\frac{\partial A t}{\partial t} e^{At}\right)p(0) \\ &= A e^{At}p(0) \\ &= A(e^{At}p(0)) \\ &= Ap(t). \end{aligned}$$

So (2) has the solution $p(t) = e^{At}p(0)$ where $p(t)$ is a 10x1 column vector of the p_i (first element p_1 , second p_2 , etc), and A is a 10x10 matrix with the coefficients from above. More specifically, the matrix A takes this form,

$$\begin{bmatrix} -\sum_{j \neq 1} \lambda_{j,1} & \lambda_{1,2} & \lambda_{1,3} & \lambda_{1,4} & \dots & \lambda_{1,9} & \lambda_{1,10} \\ \lambda_{2,1} & -\sum_{j \neq 2} \lambda_{j,2} & \lambda_{2,3} & \lambda_{2,4} & \dots & \lambda_{2,9} & \lambda_{2,10} \\ \lambda_{3,1} & \lambda_{3,2} & -\sum_{j \neq 3} \lambda_{j,3} & \lambda_{3,4} & \dots & \lambda_{3,9} & \lambda_{3,10} \\ \vdots & & & \ddots & & \vdots & \vdots \\ \vdots & & & & & \ddots & \lambda_{9,10} \\ \lambda_{10,1} & \lambda_{10,2} & \dots & & & \lambda_{10,9} & -\sum_{j \neq 10} \lambda_{j,10} \end{bmatrix}$$

where the exact expression for A is determined by the chosen model for the $\lambda_{i,j}$ s.

4.1.1 Factorizing

The basic difference between the log likelihood function for the different models is how the A-matrix is calculated. This is done outside the log likelihood function itself by calling a different function which calculates the correct λ s. What function is called upon depends on what model is used. As is clear from section 4.1, the diagonal elements can then be found by simply subtracting the sum of all the other elements in the appropriate column.

The A-matrix is different for each individual for models more complicated than model 1. This indicates that a new calculation has to be done for each of the individuals when it comes to calculating the exponential of the

matrix. The definition of an exponential matrix is an infinite sum of matrix multiplications as noted in section 4.1: $e^A = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 + \dots$. Logically, this requires some computation since it has to be repeated for each individual and time step (from equation 4 that it is e^{At} that has to be calculated) and during an optimization for several parameters. Hence any short cut to limit the computation may be of great help.

A possible way to reduce computations and limit the use of exponential matrices is to factorize the matrix At by using the eigen values and eigen vectors. At may be factored into a product of three matrices, UDU^{-1} , where U is a matrix where the columns are eigen vectors, and D is a diagonal matrix of the eigen values. By using this for the A matrix, the matrix exponential becomes

$$\begin{aligned}
e^{At} &= e^{UDU^{-1}t} \\
&= I + UDU^{-1}t + \frac{1}{2}(UDU^{-1}t)^2 + \frac{1}{6}(UDU^{-1}t)^3 + \dots \\
&= I + UDU^{-1}t + \frac{1}{2}(UDU^{-1}t)(UDU^{-1}t) \\
&\quad + \frac{1}{6}(UDU^{-1}t)(UDU^{-1}t)(UDU^{-1}t) + \dots \\
&= UU^{-1} + UDU^{-1}t + \frac{1}{2}UD^2t^2U^{-1} + \frac{1}{6}UD^3t^3U^{-1} + \dots \\
&= U(I + Dt + \frac{1}{2}D^2t^2 + \frac{1}{6}D^3t^3 + \dots)U^{-1} \\
&= Ue^{Dt}U^{-1}. \tag{6}
\end{aligned}$$

The use of the factorized A-matrix hence only requires e^{Dt} to be calculated for each time step. Since D is a diagonal matrix of the eigen values, the elements in D can easily be calculated without using matrix multiplication by simply multiplying the eigen value vector by t , the time difference, and then setting up a diagonal matrix of the result. However, a diagonal matrix multiplied by itself again makes a diagonal matrix, where each element is simply the square of itself. Similarly, for a product of three diagonal matrices, the product is a diagonal matrix where each of the elements is the original element to the power of three. From this it follows that e^{Dt} is simply, by definition of an exponential for a single value ($e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots$), a diagonal matrix. The elements on the diagonal are the exponential values of the time difference multiplied by the different eigen values, $e^{Dt} = \text{diagonal}(e^{ev \cdot tdf})$ (ev and tdf are the eigen values and the time difference respectively). By utilizing this, the need for an exponential matrix calculation vanishes, replaced by a simple matrix multiplication. Hence in the log likelihood function, this

factorization of the A-matrix is used. Furthermore, noting that if $t = 0$, $e^{At} = I$, the identity matrix, the need of the calculation is eliminated. This occurs whenever there are several sampling points in the same day, as the time difference between at least two sampling points is then 0.

4.2 Inclusion of Death Rate

It is likely that not all individuals will survive to move to a different location. This implies that some sort of death toll adjustment should be done. Expression (5) assumes that an individual is alive, and hence $p(t)$ contains the probabilities for an individual's possible positions given that the individual is alive. Assuming that death rate is independent of location and movement, the probability of movement and death should be independent probabilities. The assumption of a constant death rate may be a controversial assumption as some farms may offer better living conditions for the birds than another. This is discussed further in section 9. For an example of locality dependent death rates, see for example the article by Ovaskainen (2004) on dispersal rates in continuous time and space.

In order to obtain an expression for the probabilities which includes an adjustment for possible deaths, some assumptions has to be made about the lifetime of a house sparrow. The exponential distribution can be used to model the lifetime of an individual, as it reflects a decreasing probability of being alive after a certain age. Assuming that the lifetime of an individual bird is exponentially distributed, one can find the probability of survival in a given time interval for the bird. Let the variable T denote the lifetime of a random individual. Then

$$P(T = t) = \alpha e^{-\alpha t}, t \geq 0, \quad (7)$$

where α is the death rate. The cumulative distribution is given by

$$\begin{aligned} P(T < t) &= \int_0^t P(T = x) dx \\ &= [-e^{-\alpha x}]_0^t \\ &= -e^{-\alpha t} + e^0 \\ &= 1 - e^{-\alpha t}. \end{aligned} \quad (8)$$

The probability needed in the model is the probability of an individual remaining alive between two sampling points, if we know it is alive at the first sampling point. This can be written as $P(T > t_{s+1} | T > t_s)$. This is easily

calculated by Bayes' theorem,

$$\begin{aligned}
 P(T > t_{s+1} | T > t_s) &= \frac{P(T > t_{s+1})}{P(T > t_s)} \\
 &= \frac{e^{-\alpha t_{s+1}}}{e^{-\alpha t_s}} \\
 &= e^{-\alpha(t_{s+1}-t_s)}.
 \end{aligned} \tag{9}$$

So, the probability of an individual being alive at the end of a certain time interval, given that it was alive at the beginning, is simply dependent on the time difference between the two sampling points. This is due to the memoryless property of the exponential distribution, which states that if you know that an individual is still alive at the beginning of a time interval, the probability of the individual dying in the following time interval is the same, no matter how old the individual was at the beginning of the interval.

Having this probability, the calculation of the adjusted probability for the movement is simply calculated by multiplication of the probabilities of being alive and of being in a certain locality at a certain time. The probability of a bird being in any position at time t given that it is alive, is given in $p(t)$. This is independent of the probability of being alive, which depends on when the bird was last observed alive. From now on letting $p(t)$ mean the probability of the individual being alive and at the different localities, the expression becomes $p(t) = e^{At}p(0)e^{-\alpha t}$ starting at $t = 0$. For a time interval t_{s-1} to t_s , the t is simply replaced by the length of the interval, and $p(0)$ is replaced by the probability at the beginning of the interval (a new zero-point is set at $t = t_{s-1}$), so the expression becomes

$$p(t_s) = e^{A(t_s-t_{s-1})}p(t_{s-1})e^{-\alpha(t_s-t_{s-1})}. \tag{10}$$

4.3 Time Adjustment

Ideally, the birds' whereabouts would be known at all times. This is however not the case, and a stepwise approach is thus used. The way the data set is organized makes it natural to use the sample points as steps in the calculation of the probabilities as indicated above. This means that the $p(t)$ probability must be adjusted to account for the fact that we do not know exactly where the individual birds are at all times. Furthermore, the later sample points, the more information we have on the birds, and hence the probability should be updated each time new information is obtained, that is, at each sampling point.

In order to calculate the probability under the approximation that the sampling points are single points in time, probabilities are calculated right

before a sampling point, then directly after. Note that this assumes that an individual stays in the same locality between two sampling points. That is, if an individual is at locality i at time t_s , it is assumed to stay there until (but not necessarily at) t_{s+1} , even if this is several days after. To calculate the time difference, the actual dates are used, not the day before or after. For instance, for a hypothetical time interval were the two following dates are May 6th and May 16th, the time difference would be 10 days. The probabilities for a given sampling date is calculated by using the value given right before the sampling point. This way, one can take into account that time actually has (or might have) passed between the two sampling points, and adjust the probability of an individual being at a certain probability accordingly.

Throughout the calculations, expression 10 is used to find the next probabilities by using the available information from the previous sampling point. Defining t_{s+1-} to be the time immediately before sampling point $s + 1$, and t_s+ to be the time immediately after t_s , one can express the probabilities of the different localities for a sampling point. For example, when sampling the new position of an individual for the simulated data sets, the probabilities sampled from are found by using

$$p(t_{s+1-}) = e^{A(t_{s+1}-t_s)}p(t_s+)e^{-\alpha(t_{s+1}-t_s)}. \quad (11)$$

Note that since the probability is based on the time between the two sampling points, the time is simply the time difference between the previous and the current sampling point.

4.4 Expression for Log Likelihood

The likelihood of a certain result or observation is by definition the probability of getting the result you got (or making the observation you did), given that the model you are testing out is the "correct" model (no model will be perfect, as they are all models). This means that to calculate the likelihood for each of the models, I need an expression for the likelihood that can be applied. In this thesis, I assume that the individuals are independent. Hence, the total likelihood is the product of all the individual likelihoods. If one lets there be I individuals, and the complete set of observations for individual i is denoted by y_i , the the expression for the total likelihood is

$$P(Y_T = y_T) = P(Y_1 = y_1)P(Y_2 = y_2) \cdots P(Y_I = y_I). \quad (12)$$

Hence, the individual likelihoods, $P(Y_i = y_i)$, have to be found so one can multiply them or sum their logarithms to get the total likelihood or log likelihood respectively.

4.4.1 Individual Likelihoods

The goal is to be able to find an estimation of the probability $P(Y_T = y_t)$, which can be done by considering one individual at a time. Let y_i be the history of the i 'th individual. When the individual enters the study, all we know is where it was at the time of the first observation. After this first observation, all observations or non-observations will contribute to the log likelihood of the individual.

As mentioned in section 2, y_i consists of several values, referred to as the $X_{i,s}$ -s. Each of the sampling points after an individual enters the study has an X. If individual i is observed at the sampling point s , then $X_{i,s} = 1$. Otherwise, $X_{i,s} = 0$. This means that the individual's likelihood can be expressed in terms of its x's,

$$P(Y_i = y_i) = P(X_{i,o_i} = x_{i,o_i}, X_{i,o_i+1} = x_{i,o_i+1}, \dots, X_{i,S} = x_{i,S}), \quad (13)$$

where S is the total number of sampling points in the study, and o_i is the sampling point number where individual i is first observed.

By the chain rule of probability, this can be rewritten as

$$\begin{aligned} P(Y_i = y_i) = & P(X_{i,o_i} = x_{i,o_i})P(X_{i,o_i+1} = x_{i,o_i+1} | X_{i,o_i} = x_{i,o_i}) \\ & P(X_{i,o_i+2} = x_{i,o_i+2} | X_{i,o_i+1} = x_{i,o_i+1}, X_{i,o_i} = x_{i,o_i}) \cdots \\ & P(X_{i,S} = x_{i,S} | X_{i,o_i} = x_{i,o_i}, X_{i,o_i+1} = x_{i,o_i+1}, \dots, \\ & X_{i,S-1} = x_{i,S-1}). \end{aligned} \quad (14)$$

By obtaining an expression for the probabilities above, one can thus find the individual likelihoods. Note that by calculating the probabilities above stepwise through all the earlier probabilities, the later probabilities are conditioned on the earlier probabilities, creating the conditioned probabilities above. Since the probabilities are dependent on the past probabilities through the present probability only, they constitute a Markov Chain. A Markov chain is defined by that the probability for the next state depends on the history only through the previous state (see for instance Ross (2010, p. 191-192) for more detailed definition).

Calculating the probabilities of the different $X_{i,s}$ can be done by using the law of total probability. First however, some probabilities need to be stated. Let $L_{i,s}$ denote the true position of individual i at sampling point s (time t_s and location l_s). Then,

$$P(X_{i,s} = 0 | L_{i,s} \neq l_s) = 1, \quad (15)$$

the probability of not observing the individual, given that it is not at l_s at time t_s , is naturally one, and an observation could only be registered if a

misidentification or other mistake is made. Consequently

$$P(X_{i,s} = 1 | L_{i,s} \neq l_s) = 0, \quad (16)$$

the probability of observing the individual given that it is not present at l_s at time t_s is zero as an individual cannot be observed if it is not present. Further, introducing the recapture/resighting probability β ,

$$P(X_{i,s} = 1 | L_{i,s} = l_s) = \beta, \quad (17)$$

the probability of observing an individual who is present at location l_s at t_s , defining β to be the probability of observing an individual that is present. From this it follows that

$$P(X_{i,s} = 0 | L_{i,s} = l_s) = 1 - \beta, \quad (18)$$

is the probability of not observing an individual who is present. If the individual is present, it must either be observed or not, so together the two probabilities sum to one. If $L_{i,s} = 0$, the individual is taken to be dead, and hence $P(X_{i,s} = 0 | L_{i,s} = 0) = 1$. Note also that the probability of the individual dying is

$$P(L_{i,t} = 0) = 1 - \sum_{g=1}^{g=n_l} p_g(t), \quad (19)$$

since this is one minus the probability of the individual being alive, under the assumption that the individual must be at one of the n_l localities at all times while alive. This also implies that if an individual is observed, then it is still part of the study, and counted as not dead. Thus $P(L_{i,s} = 0 | X_{i,s} = 1) = 0$.

After observations are taken into consideration for a sampling point, one conditions the following probability on this result. For instance, if the individual is observed, the probabilities for the next sampling points are conditioned on this observation. Hence, if the individual is observed at sampling point s the probability is 1 for its presence at l_s , or $p_{l_s}(t_s+) = 1$. This means that $p_{l_k \neq l_s}(t_s+) = 0$. As before the $+$ in t_s+ indicates immediately after the sampling at sampling point s , and p_j is $P(L = j)$.

If a bird is observed at a sampling point, the probability is given for the following step (a vector of a 1 and 0, where the 1 is the probability of being at the observed location). In mathematical terms, this means that

$$P(L_{i,s} = j | X_{i,s} = 1)(t_s+) \quad (20)$$

is known for all localities j , as they are either 0 or 1. However, to be able to calculate the contribution to the log likelihood, one needs to be able to

say something about how likely an observation of an individual is at a given place and time, and not only if the individual was observed at the previous sampling point.

The probability of observing a bird which is known to be present is β (see equation (17)). The probability of a bird being present at locality l_s at time t_s is calculated by $p_{l_s}(t_s-)$. Hence, the probability of an observation of a bird at sampling point s is

$$\begin{aligned} P(X_{i,s} = 1) &= P(X_{i,s} = 1|L_{i,s} = l_s)P(L_{i,s} = l_s) + P(X_{i,s} = 1|L_{i,s} \neq l_s)P(L_{i,s} \neq l_s) \\ &= P(X_{i,s} = 1|L_{i,s} = l_s)P(L_{i,s} = l_s) \\ &= \beta p_{l_s}(t_s-). \end{aligned} \tag{21}$$

Thus the non-observation of an individual is

$$1 - \beta p_{l_s}(t_s-), \tag{22}$$

due to that the two possibilities are complements of each other (one or the other has to happen). The expressions 21 and 22 hence constitute the contribution to the likelihood from an observation or non-observation of an individual respectively. The missing component in these expressions are thus $p(t_s-)$. These probabilities are found from expression (10). The remaining is thus to find the probability $p_s(t_{s-1}+)$, so expression (10) can be calculated. This expression is known for cases where the individual is observed at $s - 1$ (see previous paragraph), but not for cases where the individual is not observed at $s - 1$. The probability that needs to be found for individual i is thus

$$P(L_{i,s} = j|X_{i,s} = 0)(t_s+) \tag{23}$$

for all localities. Together, with each used in the appropriate situation, expressions (20) and (23) offer all probabilities needed to continue the chain of probabilities for all sampling points.

The probability of not observing an individual is known from expression (22). The probability in this expression is $P(X_{i,s} = 0)$. By then using Bayes' theorem one can find the probability $P(L_{i,s} = j|X_{i,s} = 0)$, which is the probability used to advance the probabilities after an observation or non-observation is made.

By Bayes' theorem, for locations $l_k \neq l_s$ (locations where sampling is not

done at sampling point t_s),

$$\begin{aligned}
P(L_{i,s} = l_k | X_{i,s} = 0) &= \frac{P(X_{i,s} = 0 | L_{i,s} = l_k)P(L_{i,s} = l_k)}{P(X_{i,s} = 0)} \\
&= \frac{1p_{l_k}(t_s-)}{1 - \beta p_{l_s}(t_s)} \\
&= \frac{p_{l_k}(t_s-)}{1 - \beta p_{l_s}(t_s)}, \tag{24}
\end{aligned}$$

where t_s- indicates the time right before sampling point s . For the location where the sampling is being done, l_s ,

$$\begin{aligned}
P(L_{i,s} = l_s | X_{i,s} = 0) &= \frac{P(X_{i,s} = 0 | L_{i,s} = l_s)P(L_{i,s} = l_s)}{P(X_{i,s} = 0)} \\
&= \frac{(1 - \beta)p_{l_s}(t_s-)}{1 - \beta p_{l_s}(t_s)} \tag{25}
\end{aligned}$$

From the expressions one can see that the probability of the individual being in location l_s if it is not observed, is decreased by a factor of $1 - \beta$ compared to the other probabilities. The rest are increased by the division of something between 0 and 1, and not reduced. Logically, this makes sense as the individual is not present at locality l_s with probability $1 - \beta$, while no such restriction is present on the other localities. Hence, by not observing an individual, the probability of it being present is smaller than for a locality where no attempt at observations have been made. This means that as the recapture/resight probability, β , increases, the probability of an individual being present, but not observed, decreases, which is what one would expect.

Each of the probabilities depend on the probabilities right before the current sampling point through $p_j(t_s-)$. Furthermore, this probability is calculated by use of $p_j(t_{s-1}+)$, the probability immediately after the previous sampling point. This depends on $p_j(t_{s-1}-)$, which depends on $p_j(t_{s-2}+)$ and so on until the very first observation of the individual. By this, it is clear that the conditioning on the previous observations in expression 14 is in fact done. The process thus goes as follows: a sampling is done at sampling point s , the probability is updated to $p(t_s+)$. This probability is updated to account for the time between two sampling points, giving $p(t_{s+1}-)$. The probability is again updated by using the new sampling data, giving $p(t_{s+1}+)$ and so on. The chain is started by using that the probability immediately after the first observation is set to 1 for the location where the individual was observed ($p_{l_0}(t_0+) = 1$), and 0 for the rest.

4.4.2 Example

By adding together the logarithm of the appropriate cases above, one gets each individual's contribution to the log likelihood of the entire sample, that is, one gets $\log(P(Y_i = y_i))$. For example, consider an hypothetical individual who is observed at sampling point 3, 4 and 6, that is at localities $l_3 = 1, l_4 = 3$ and $l_6 = 10$ in the first six sampling points. Recall that $p(t_s-)$ and $p(t_s+)$ are column vectors which denote the probability of an individual's position immediately before and after sampling point t_s respectively, and that element k in these vectors is the probability for locality k position.

The calculation goes stepwise after a sampling, first $p(t_s+)$ is calculated conditioning on the observation (or non-observation). Then $p(t_{s+1}-)$ is calculated, adjusting for the time passed between the two adjoining sampling points. This entails that the probability of an observation at sampling point t_s is $p_{l_s}(t_s-)\beta$, and hence for a non-observation $1 - p_{l_s}(t_s-)\beta$, and so the log likelihood contribution is simply the logarithm of the appropriate of these. The calculation is then done as seen in table 4.4.1 for the first three sampling points after the first observation, and so on until sampling point S .

As soon as the computations for individual i is completed for all sampling points, the likelihood can be obtained. The log likelihood is the sum of all the contributions, which can be calculated parallel to the computations, and the likelihood can thus be calculated by $P(Y_i = y_i) = e^{\log(\text{likelihood})}$. I have chosen to use the log likelihood in place of the likelihood for comparisons, optimizations and calculations. One can do this because of the monotony of the functions (if a is greater than b , then $\log(a)$ is greater than $\log(b)$). Note that the first sampling point for an individual is taken to be when the individual is first observed, hence the number of sampling points that contribute to an individual's log likelihood can be different for different individuals depending on the first observation of them.

	Example		
Sampling-point ($X_{i,s} =$)	Probability	Contribution to log-likelihood	Details
s=3 (1)	$p(t_3+) = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$	$\log(1) = 0$	First observation.
s=4 (1)	$p(t_4-) = e^{A(t_4-t_3)}p(t_3+)e^{-\alpha(t_4-t_3)}$ $p(t_4+) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$	$\log(\beta p_3(t_4-))$	adjustment for time difference between s=3 and s=4. $l_4 = 3$
s=5 (0)	$p(t_5-) = e^{A(t_5-t_4)}p(t_4+)e^{-\alpha(t_5-t_4)}$ $p(t_5+) = \frac{1}{(1-\beta p_3(t_5-))} \begin{bmatrix} p_1(t_5-) \\ p_2(t_5-) \\ (1-\beta)p_3(t_5-) \\ p_4(t_5-) \\ p_5(t_5-) \\ \vdots \\ p_{10}(t_5-) \end{bmatrix}$	$\log(1 - \beta p_3(t_5-))$	time adjustment. $l_5 = 3$, so element 3 is used in the fraction and the third element is different (see (25)).
s=6 (1)	$p(t_6-) = e^{A(t_6-t_5)}p(t_5+)e^{-\alpha(t_6-t_5)}$ $p(t_6+) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$	$\log(\beta p_3(t_6-))$	time adjustment. $l_6 = 10$

Table 4.4.1: Example of calculation of log likelihood.

4.4.3 Total Likelihood

As soon as the log likelihood is calculated for an individual, it is added to the total sum. The sum is 0 at the beginning, and when the log likelihoods of each of the I individuals are added, the total log likelihood for the entire sample is obtained. This log likelihood can then be used to compare different models, optimize the value of the log likelihood, and by repeatedly calculating the log likelihood value for different parameters, numerically obtain the maximum likelihood estimates of the different parameters. The optimization of the log likelihood function is done by use of the function `optim` in R.

4.5 Expression for Rates

The remaining expression is now that of the rate. $\lambda_{i,j}$ denotes the dispersal rate from location j to location i . The dispersal rate is, as can be seen in equation 1, a value that reflects how much the probability of being in a location changes. The dispersal rate from i to j gives the weight of how much the probability of being in location i affects the change in probability of being in j with time. The aim of my thesis is to find a good expression for these rates, an expression that shows how the rates change with different factors.

In the thesis I have used the general equation

$$\lambda_{i,j} = \theta e^{\eta_{i,j}} \quad (26)$$

where θ is a parameter calculated for each model and $\eta_{i,j}$ is a linear function of distances and other specifics of the individual and locality the rate is calculated for. For example, $\eta_{i,j}$ could be $\eta_{i,j} = -\theta_1 r_{i,j} + \theta_3 (r_{j,h} - r_{i,h})$, where $r_{i,j}$ is the distance from locality i to locality j , and h is the home locality of the individual in question. The values of the parameters θ_1 and θ_3 are thus the weight given to these covariates. The example mentioned is model 2b, which is described further in section 5.2.2.

5 Models

As mentioned in section 4, several models are tested out to find out which one gives the best picture of how the dispersal rates change. In this section I present the different models I have attempted to fit to the data. The first model is pretty simple, depending only on distance between two sampling points. For the later models, more variables are added to try to create a better fit. A more complicated model often will cause more computations.

Therefore, for each new model, the new model is compared to the simpler one to check if the improvement is significantly large to justify the use of a more complicated model.

5.1 Model 1: Distance Only

The first model is the simplest one, adjusting only for the distance between two localities. It is based on the assumption that the dispersal rate from locality j to locality i , $\lambda_{i,j}$, is given by:

$$\lambda_{i,j} = \theta_0 \exp^{-\theta_1 r_{i,j}}, \quad (27)$$

where θ_0 and θ_1 are constant parameters. This means that the rate depends on the distance between the two locations in question. If the distance is large, the dispersal rate is small (assuming that $\theta_1 > 0$). Accordingly, if the distance is small, the rate is larger. So the rate decreases with increasing distance, which seems plausible assuming that birds are more likely to move shorter distances than larger ones. Also, as long as the death rate is constant, the dispersal rate $\lambda_{i,j}$ is constant given the distance $r_{i,j}$ between locality i and locality j , and between individuals. This means that it is only necessary to calculate the matrix once, as all individuals can use the same matrix.

5.2 Model 2: Home-centered

The second model is a little more complicated in that it includes an extra parameter. In models 2a and 2b the birds' hatching locality plays a role. The term home or home locality will henceforward mean the location (farm) at which an individual was hatched. The models are made under the assumption that the sparrows have a stronger connection to their home locality, than to the other farms, so that they are more likely to disperse towards this location, and less likely to leave it for other locations. Thus it is a type of focal point attraction model as discussed by Börger et al. (2008), where the home locality is the focal point. A focal point attraction model is a model in which individuals are drawn to a specific point (see Börger et al. (2008) for further details on such models and home range behaviour). I have suggested two different models, both based on this assumption.

A consequence of using this type of model in this case is that the data set is reduced. Only individuals where the hatch locality is known can be used. This reduces the data set from 192 individuals and 962 observations to 105 individuals and 538 observations. In addition to this, the home locality will be different for each individual, the rates will differ between individuals.

This entails that a new A-matrix has to be calculated for each individual, increasing the computation required.

5.2.1 Model 2a

In the simplest model, the expression for the dispersal rate is adjusted depending on whether or not i or j is the individual's home locality. If j is their home locality, the dispersal rate to i is reduced. If i is their home locality, the dispersal rate from j to i is increased, and if neither of i or j is the individual's home locality, the dispersal rate is in between the other two options. This model does not take into account whether the location i (where the individual might be moving to) is closer to the home location than location j , which seems like a reasonable assumption. This possibility is however taken into account in model 2b.

The expression I have used for the dispersal rate to reflect this scenario is

$$\lambda_{i,j} = (\theta_2)^{\delta_{i,j}} \theta_0 \exp^{-\theta_1 r_{i,j}}$$

where $\delta_{i,j}$ is 1,0 or -1 if i , neither i nor j , or j is the home locality respectively. For this model to reflect an increased probability of returning home, it follows that $\theta_2 > 1$.

5.2.2 Model 2b

Model 2b takes into account the distance from each possible locality to the home locality, h . By this I mean that the rate for a bird to move from j to i is greater if the distance from i to h is smaller than the distance from j to h , and lower if the distance from i to h is greater than the distance from j to h . In other words, that the birds move towards their home at a higher rate than they move away from it.

To reflect this, I have introduced a new parameter θ_3 which is to be the weight given to the difference in the distances. Hence, the new expression for the dispersal rates is

$$\lambda_{i,j} = \theta_0 e^{-\theta_1 r_{i,j} + \theta_3 (r_{j,h} - r_{i,h})},$$

where h is the home locality.

Here, the difference between the distance from i to h and from j to h causes the rate to increase or decrease depending on whether the distance from i to h is smaller than the distance from j to h . If $r_{i,h}$ is greater than

$r_{j,h}$ the rate will decrease, since the distance to the home locality is smaller from the position where the bird initially is (j). If it is the other way around, the opposite is the case and the rate will increase. This of course assumes that $\theta_3 > 0$, an assumption which is therefore part of model 2b. At the same time, the rate is reduced by an increasing distance as in model 1 and 2a.

5.3 Model 3: Differences Between Sexes?

Tufto et al. (2005) discuss the effect of sex on dispersal range. Although they do not find any difference between dispersal range between the two sexes for house sparrows, they do find a difference in tits (great tits and blue tits), where females dispersed longer distances than males. Based on the idea of a possible difference between the sexes, I chose to further expand the model by adding a variable depending on the sex of the individual, even though no significant difference between the dispersal ranges of male and female house sparrows was found by Tufto et al. (2005).

Tufto et al. (2005) considered the dispersal distance. To see if there is a difference in dispersal rate, that is if there is a reason to believe that females disperse more often than males, or the other way around, parameter θ_4 is added. This parameter is to reflect the potential difference caused by sex. For female individuals, the parameter will be present, for male individuals not. It is incorporated in the model by adding $\theta_4 s x_p$ to the expression in the exponential, where $s x_p = 0$ if individual p is a male, and $s x_p = 1$ if individual p is female. It follows from this that if the value of θ_4 is negative, females disperse less often than males, and if θ_4 is positive, females disperse more often than males. Accordingly, if θ_4 is 0, there is no difference between males and females.

By comparing the first models (model 1, 2a and 2b), I chose to work further with model 2b, as this seemed to give the best results (see sections 8.3.3 and 8.4). Therefore, model 3 is simply made as an expansion of this model. The rates are thus calculated by the following expression, where $s x_p$ is as defined above,

$$\lambda_{i,j} = \theta_0 e^{-\theta_1 r_{i,j} + \theta_3 (r_{j,h} - r_{i,h}) + \theta_4 s x_p}. \quad (28)$$

The inclusion of a sex-parameter requires that data on each individual's sex is available, which could mean a further reduction of the data set. However, in this case all individuals where nestling locality is known, also have data on their sex, and hence no further reduction is needed compared to models 2a and 2b.

5.4 Model 4: Seasonal Changes?

The house sparrow is known to have seasonal changes in its dispersal pattern (Anderson, 2006). These changes are present in the dispersal distance in that the birds fly longer distances in late summer and autumn to find food (Anderson, 2006, p. 336). To check if there is a seasonal change in the dispersal rate, I have chosen to include a model in which the date of the sampling has a part in adjusting the dispersal rate. To do this, a new parameter is introduced to model 3, θ_5 , which is multiplied by an indicator variable, d_s , which is 0 if the date is before a certain date limit, θ_6 , and 1 if the current date is after the limit. This date limit, θ_6 , can either be a variable itself, or a set date. This makes the dispersal rate function

$$\lambda_{i,j} = \theta_0 e^{-\theta_1 r_{i,j} + \theta_3 (r_{j,h} - r_{i,h}) + \theta_4 s x_p + \theta_5 d_s}, \quad (29)$$

where the value of d_s indicates whether the current sampling point is before or after the date limit. The inclusion of this date parameter increases the number of computations that need to be done, as the A matrix has to be calculated more frequently. For one set value of the date limit, the number of A matrix calculations is in the best case doubled, because there has to be one matrix for each individual before and one matrix after the date limit for each individual. If the value of θ_6 is not a set date, further computations can be necessary. I have chosen to use a set date for θ_6 as this limits the number of additional computations.

6 Simulation of Data Sets

To be able to compare the data set to a data set actually based on the models, I have also simulated data sets by use of bootstrapping. How they are simulated depends on what model they are being based on, but the general idea is the same.

Bootstrapping is a statistical method used to obtain information about a distribution or a variable when the distribution is partially unknown. A version of bootstrapping is parametric bootstrapping. The idea of parametric bootstrapping is to repeatedly sample from a distribution with parameters estimated from the original data, calculate equivalent estimates for the parameters based on the sample, and from the collection of results make inferences about the distribution itself (the parameters). Details of such bootstrapping is explained by Devore and Berk (2007, p.339). More specifically, in this case the maximum likelihood estimators (MLEs) resulting from the calculations on the collected data are set as the true parameter values in the simulation

of n data sets based on the model in question. Then a new observation matrix is made, and a new likelihood optimization through the `optim`-function in R is done, and new parameter estimates reached. This is done for each of the n data sets, so one has n estimates for each parameter. From these estimates one can then calculate properties of the parameter estimators, for instance the expected value of the estimator of the parameter, which is the mean of these n MLEs. The set of results from the bootstrapping procedure can also be used to create confidence intervals for the expected values of the parameters and standard deviations for the estimates.

For example, if one does $n = 100$ simulations, one simulates 100 data sets with the true values set to be the result from the collected data. For each of the data sets one calculates the MLE of the parameters in question. Note that the simulated data sets are in fact based on the model in question, so one would perhaps assume that the simulated data sets give a very well fit. However, given the randomness involved, this will not necessarily be the case.

After the 100 simulations are done, one has 100 estimates of each parameter. Then, one can calculate the mean of these estimates to obtain an estimate of the expected value of the maximum likelihood (ML) estimator. Also, since the 100 estimates will differ because of the randomness involved, they can provide insight into how much variation there is expected to be in the parameter estimates. Therefore one can calculate the standard deviation of these 100 MLEs to obtain an alternative estimate of the parameter estimate standard deviation.

The simulated data set consists of the same number of individuals as the collected data set. However, in the simulated data set each individual's position at each time is simulated based on the model in question, with exception of the first observation for each individual which is the same as for the original data set. Naturally, the distances and sampling points are also the same as for the original data set, as well as the home location and sex. The starting values are set to be the true values of that simulation, and so is the scaling (see appendix A). The results, that is, parameter estimates, log likelihood and convergence of the optimization, are then stored in a matrix. The process is then repeated n times, with a new simulated data set for each time.

Each individual's new position is sampled, with probability $p(t)$, which depends on the dispersal rates (through matrix A, see section 4.1), and on what the previous position was, through a vector of the probabilities of the location of the individual immediately after the previous observation was made. The probabilities right after a sampling point at time t is denoted by $p(t+)$, and correspondingly the probabilities of position right before a

sampling point at time t is denoted by $p(t-)$. The previous probability is a vector of length 10 with all zeros except for element j , which is 1, when element j is the position at the previous sampling point.

Multiple observations are avoided automatically by the way the data are sampled. If the time difference to the next sampling point is 0, then the probability that the bird is in the same locality will be 1, following expression (10) (and this will be 1 for the locality of the bird at sampling point s). Hence, if the time difference is 0, the position of the bird at the next sampling point is the same as the last, and multiple observations are avoided.

After each individual has had its movement history simulated, the observation matrix is constructed. If the position of an individual at a certain time is the same as the position where sampling was done at that time, then the simulated individual was observed with probability β . Hence a 1 appears in the observation matrix with a probability β for that sampling point. All the other elements are 0, including the points where the individual was present, but not observed (happens with probability $1 - \beta$). Since the simulated observation matrix has the same meaning as the observation matrix based on actual data, one can simply use it the same way as the data-based observation matrix.

Each data set is simulated using the model in question. However, when and where the observations are done is independent of the simulated data. Also, because the probability of observing an individual even though it is present is less than one, the data may not give as much information to the optimization as expected. Hence, the simulated data may not appear to be as well fitted to the model in question as one would expect, since not all of the movement history is taken into account. This is also the case for the collected data, but one would perhaps expect the simulated data to fit better, as it is actually based on the model.

7 Comparison and Testing of Models

In order to expand the models, each model needs to be evaluated in some way in order to compare it with other models. This is necessary to know if an extended model is worth the extra computation needed for a more complicated model. In this section I discuss the methods used to evaluate the different models.

7.1 Maximum Likelihood Estimators

The estimates of the parameters are obtained by optimizing the log likelihood function through the **optim** function in R. Thus, this makes the results approximate maximum likelihood (ML) estimates, and hence properties of these can be used. For instance, as the number of samples goes to infinity, the distribution of maximum likelihood estimators (MLEs) approaches a normal distribution (see for instance Devore and Berk (2007, p.369)). Thus each of the values that result from the optimization is approximately normally distributed.

The results from the bootstrapped/simulated data are also estimates of the MLE for their respective data sets. Thus, the optimized parameters for the simulated data set is also approximately normal. Since the simulated data results are the result of different simulated data sets based on the same true parameters (found from the collected data), the results of the simulated data can be used to show how much a MLE from a data set with the same initial or true parameters can differ from the true parameter values behind the data set.

7.2 The Likelihood Ratio Test

Each of the models, except model 1, are expansions of a simpler model with only an extra parameter added. This makes it possible to use a likelihood ratio test, as described by for instance Casella and Berger (2002, p.490). The likelihood ratio test is a statistical test used to compare two models. The two models must be such that one model (the larger one) is equal to the other one when some of its parameters are within a restricted parameter space compared to the parameter space for the large model. The likelihood ratio test tests if it is reasonable to assume that the two models are different. Hence the null hypothesis is that the extra parameters in the larger models are within the area they have to be for the two models to be equal. The alternative hypothesis is that the restricted parameters are not within the restricted parameter space. For instance, one could test a model where $\eta = 0$ (the smaller model) against a model where $\eta \neq 0$ (the larger model, the parameter η is free and not limited to the restriction $\eta = 0$).

The likelihood ratio test uses that the value $-2\log(LR(Y))$ is χ_{df}^2 , when the number of observations goes to ∞ . Here, LR is the ratio of the likelihood under the null hypothesis (or more generally, the highest possible value of the log likelihood under the null hypothesis) to the maximum value of the likelihood under all possible parameters in the parameter space of the parameter of interest. Note that this may be simplified to $-2[\log(\text{likelihood}_0) -$

$\log(\textit{likelihood}_1)]$ where the 0 indicates the null hypothesis, and the 1 the alternative hypothesis. The degrees of freedom, df , is the number of parameters restricted by the null hypothesis or in other words, the difference between the number of free parameters over all the parameter space and under the null hypothesis.

The above means that having obtained the values of the likelihood of the different models, the models can be compared to one another by simply choosing a level for the test and using the LR-test variable. For example, in comparing model 1 and 2b, the null hypothesis would be $H_0 : \theta_3 = 0$ which is model 1, and the alternative hypothesis is $H_1 : \theta_3 \neq 0$, or in this case $\theta_3 > 0$, which is model 2b. Then one can calculate the test statistic, $-2[\log(\textit{likelihood}_0) - \log(\textit{likelihood}_1)]$, and reject the null hypothesis if the obtained value is greater than or equal to the relevant χ^2 -value. Alternatively, the p -value can be computed. For this test the p -value is the area under the χ^2 curve to the right of the computed test statistic value. This is the approach I have chosen to use.

The p -value is by definition the probability of receiving the result you got or a more extreme result given that the model you assumed is correct. Hence, a large p -value would indicate that the null hypothesis model is a good fit as the probability of getting a more extreme result under the null hypothesis is large, and one is within a reasonable range of the null hypothesis model. On the other hand, a small p -value suggests that the probability of getting a more extreme value is rather small under the null hypothesis. This suggests that the value of the test statistic is on the outskirts of the model, and hence the null hypothesis model is perhaps less likely. Another way to define the p -value is as the lowest level of test for which the null hypothesis is rejected (Devore and Berk, 2007, p.450). For instance, if the p -value is 0.01, all tests of a level higher than 0.01 will reject the null hypothesis, while all tests of a level lower than 0.01 will reject the alternative hypothesis.

7.3 Standard Deviations

The standard deviations of the estimates can be used to see how accurate the estimate is. A smaller standard deviation indicates that the estimate is quite accurate, while a larger standard deviation indicates greater variability of the estimate, and a less accurate estimate.

7.3.1 Standard Deviations by Fisher Information

One method to find the standard deviation for the different parameters from the collected data set is to use the Fisher information (FI) and the Hessian

matrix, a matrix that contains the double derivatives of the log likelihood with respect to the different parameters. The Fisher information for a parameter κ can be defined by $-E[\frac{\partial^2}{\partial \kappa^2} \log(f(X; \kappa))]$ (Devore and Berk, 2007, p.365), and is called so because it says something about how much information is contained in a data set. The expression $\log(f(X; \kappa))$ is simply the log likelihood when viewed as the function of the parameter κ . The Fisher information can be approximated by use of the observed Fisher information, which is defined by $-\frac{\partial^2}{\partial \kappa^2} \log(f(X; \kappa))$, and consequently the observed FI is found in the Hessian matrix. In this thesis I have used the observed Fisher information, so this is an approximation to the Fisher information itself. Note however that Efron and Hinkley (1978), argue that the estimate based on the observed FI is a better estimate of the standard deviation than the estimate based on the expected FI. If this is indeed the case, the use of the observed FI should be justified.

The estimate based on the FI is approximately the value of the standard deviation for the limiting distribution of the parameters. For large sample sizes the limiting distribution of the MLEs is approximately normal with variance equal to the inverse of (-1) times the Fisher information (Devore and Berk, 2007, p. 369), or in the case of more parameters, (-1) times the diagonal elements on the inverse of the Fisher information matrix.

I have found the Hessian matrix by use of the **numDeriv**-package in R, with the values from the result and the function as parameters. Originally I used the built in Hessian function in **optim**, but the **numDeriv** function allows more control and is more reliable, since the parameters are scaled.

By use of the Hessian matrix one can also find further support for that a point is actually a maximum point. If the point is a maximum, the Hessian matrix should be negative definite (alternatively, negative semidefinite), ("non-positive definite", de Jong and Heller (2008, p. 69)), which is the case if all the eigen values are negative (Lay, 2012, p. 405-406) (for semidefinite, some may be 0). This works as an extra control, but cannot alone prove that the point is a maximum, so further controls have to be used.

7.3.2 Standard Deviations by Bootstrapping

In addition to saying something about the variability of the estimate, one can compare the standard deviation of the estimates obtained by use of the observed Fisher information on the collected data, to the standard deviation obtained by simulation and bootstrapping.

As described in section 6 one can reach an estimate of the standard deviation of the estimators by using the simulated data. This is done by calculating the standard deviation between the optimized results from the simulated

data sets. The simulated data sets are all based on the model in question, and hence the results are all a MLE of the corresponding parameter in the true model. The estimated MLEs thus represent a set of results one could expect to get if the model is correct. Therefore the standard deviation of the estimator can be estimated by using the standard deviation of the n simulation results.

7.3.3 The Standard Deviation Estimates

The standard deviation estimates based on the bootstrapped data sets are not based on the Hessian matrix, and do not use that the limiting distribution of an MLE is normal with variance $\frac{1}{I_F}$ (Devore and Berk, 2007, p. 369). According to the Cramér-Rao inequality, any unbiased estimator of a parameter will have a variance (and hence a standard deviation) which is equal to or greater than $\frac{1}{I_F}$, where I_F is the Fisher information for the sample (Devore and Berk, 2007, p. 367). Note that the sample space cannot depend on the parameter in question. This means that in theory, the standard deviation based on the Fisher information should have the smallest possible variance for an unbiased estimator. Note however that I have used the observed Fisher information in this study (see 7.3.1), and that the Cramér-Rao inequality uses the expected Fisher information.

7.4 Confidence Intervals and Bias

To determine the bias of the MLEs and compare the standard deviations obtained from bootstrapping to the ones obtained from the observed Fisher information, confidence intervals for the expected value of the parameter estimators and the standard deviations are calculated. A confidence interval is an interval which illustrates a possible area for an estimated value. When many confidence intervals of level α are made, approximately $(1 - \alpha)\%$ of them will cover the true parameter value. Confidence intervals for the expected values of the parameters can be obtained by using the simulated or bootstrapped data results. The bootstrap estimates are approximate maximum likelihood estimators, as they are obtained by numerically optimizing the log likelihood value. As mentioned, a maximum likelihood estimator for a large sample size will approach a normal distribution. The results calculated from the simulated data hence constitutes an approximately normal sample, and hence properties of a normal sample can be used. The confidence interval is based on a sample, and the width of the interval is dependent on the number of samples, as can be seen in the calculation below. The larger the number of samples, the narrower is the interval and hence less uncertainty

in the result.

A normally distributed sample X_1, \dots, X_n has the property that $\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ is t distributed with $df=n-1$ (Devore and Berk, 2007, p. 315), where $\frac{S}{\sqrt{n}}$ represents the estimated standard deviation of the sample. Using this as a starting point, it is simple to construct a confidence interval of size α for the true expected value, μ ,

$$\begin{aligned}
t_{n-1, 1-\frac{1}{2}\alpha} &< \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{n-1, \frac{1}{2}\alpha} \\
\frac{S}{\sqrt{n}} t_{n-1, 1-\frac{1}{2}\alpha} &< \bar{X} - \mu < \frac{S}{\sqrt{n}} t_{n-1, \frac{1}{2}\alpha} \\
-\bar{X} + \frac{S}{\sqrt{n}} t_{n-1, 1-\frac{1}{2}\alpha} &< -\mu < -\bar{X} + \frac{S}{\sqrt{n}} t_{n-1, \frac{1}{2}\alpha} \\
\bar{X} - \frac{S}{\sqrt{n}} t_{n-1, 1-\frac{1}{2}\alpha} &> \mu > \bar{X} - \frac{S}{\sqrt{n}} t_{n-1, \frac{1}{2}\alpha} \\
\bar{X} - \frac{S}{\sqrt{n}} t_{n-1, \frac{1}{2}\alpha} &< \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{n-1, \frac{1}{2}\alpha}
\end{aligned} \tag{30}$$

Note that the t-distribution is centered around 0 and symmetric, from which it follows that $t_{n-1, 1-\frac{1}{2}\alpha} = -t_{n-1, \frac{1}{2}\alpha}$.

By the above it is simple to find confidence intervals for the parameters based on the simulations. By comparing the mean of the estimates from the simulated data (the expected value of the simulated sample) to the true value of the model behind the simulated data sets (the value reached from the collected data optimization), one can determine whether it is a reasonable estimate, or whether it is biased in either direction. The bias is the difference between the expected value of an estimator and the true value. That is

$$E(\hat{\theta}) - \theta. \tag{31}$$

In the study, the bias is sometimes given in percent. This is calculated by

$$\frac{E(\hat{\theta}) - \theta}{\theta}. \tag{32}$$

If the estimate based on the collected data is unbiased, it should be found in the corresponding confidence interval. Similarly, one can find confidence intervals for the standard deviations. However, the standard deviations here represent an alternative estimate to the Fisher information estimate, and not the expected value. The estimated variance for a normal sample, S^2 is related to the χ^2 distribution by $\frac{(n-1)S^2}{\sigma^2}$, which is χ_{n-1}^2 (see Devore and Berk (2007,

p. 314)). This makes it simple to construct a confidence interval of level α for the standard deviation, based on $P(\chi_{n-1,1-\frac{1}{2}\alpha}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{n-1,\frac{1}{2}\alpha}^2) = 1 - \alpha$.

$$\begin{aligned}
\chi_{n-1,1-\frac{1}{2}\alpha}^2 &< \frac{(n-1)S^2}{\sigma^2} < \chi_{n-1,\frac{1}{2}\alpha}^2 \\
\frac{1}{\chi_{n-1,\frac{1}{2}\alpha}^2} &< \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi_{n-1,1-\frac{1}{2}\alpha}^2} \\
\frac{(n-1)S^2}{\chi_{n-1,\frac{1}{2}\alpha}^2} &< \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1,1-\frac{1}{2}\alpha}^2} \\
\sqrt{\frac{(n-1)}{\chi_{n-1,\frac{1}{2}\alpha}^2}} S &< \sigma < \sqrt{\frac{(n-1)}{\chi_{n-1,1-\frac{1}{2}\alpha}^2}} S
\end{aligned} \tag{33}$$

See alternatively Devore and Berk (2007, p. 402-403). By constructing a confidence interval for the standard deviation of data actually based on the model, one can compare it to the size of the standard deviation obtained by the collected data, which is based on the limiting distribution of the MLE.

8 Results

The results reached when optimizing the different models are presented in this section. The models are all built up as an extension of the first model. The main idea is that by setting the null hypothesis to be that a parameter is equal to a specific value, one gets the new extended model equal to the old model. The likelihood ratio test can be used in just these situations, since the null hypothesis is then that the extra parameter, for example θ_3 in model 2b, is set ($\theta_3 = 0$ in model 2b), since that makes the first (1) and second (2b) models the same.

Running the models yield results on the model itself, as well as on what model is better suited when compared to others. Below, the individual results for a model are presented first, before two models are compared. Note that to be able to use the **optim**-function in R, slight adjustments had to be made between the different models as to how many parameters are to be optimized, and that the log likelihood function was modified to work for the more complex model between the different models. These changes should not matter to the results other than they allow results for more models.

8.1 General Settings

The settings used for running the different models can be seen below in the copy from an optimization of model 2b. Similar settings are used for the other runs as well. The first element in **optim** is the vector of the starting points for the parameters to be optimized. The length of this vector thus varies with the model used, as the number of parameters differ. `logl` is the name of the log likelihood function, and the three following elements are settings that indicate model, the observation matrix, and a vector of the nesting localities. The method used by the **optim** function is set to be the BFGS method (note that the Nelder-Mead method was also considered, see appendix section A.2.4). McKinnon (1998) shows that in certain cases, the Nelder-Mead method may converge to a nonstationary point, and this supports the decision to use the BFGS method. The final element, `control`, is a list containing the scaling of the different parameters (see appendix section A.2.3), the maximum number of iterations permitted and finally `fnscale=-1` indicates that the function is to be maximized, not minimized which is the default for the **optim** function.

```
om2b1BF=optim(stpar2b1,fn=logl,m=3,obsmat=l2$mat,hm=nest2,
              method="BFGS",control=list(parscale=stpar2b1,maxit=max,fnscale=-1))
om2b1BFh=hessian(logl,om2b1BF$par, method="Richardson",
                 method.args=list(eps=om2b1BF$par*10^(-4)),
                 m=3,obsmat=l2$mat,hm=nest2)
```

The arguments for the **hessian** function mostly explain themselves, after the explanation for the **optim** function. Otherwise the method is set to the default (Richardson), and `logl` is the function for which the Hessian is required for the points `om2b1BF$par` which are the resulting parameters for the optimization done above.

When attempting to run the optimization of the methods, several issues occurred. Some could be fixed by simple adjustments, while others were more complicated. This was also the case for optimization of the simulated data. The issues were mostly dealt with and fixed before the optimizations presented in this section. For further details of issues that occurred at some point during the process, see appendix A.

8.2 Model 1

The first model has four parameters to be optimized, namely θ_0 , θ_1 , α and β . To control that the optimization worked properly, I tried several starting

Parameter	Starting value	Estimate	Standard deviation	Log-likelihood
θ_0	0.01	0.75852	0.23566	-2290.065
θ_1	0.0001	0.00232	0.00018	
α	0.1	0.01358	0.00166	
β	0.1	0.74873	0.03936	
θ_0	0.1	0.78289	0.24555	-2290.059
θ_1	0.005	0.00234	0.00018	
α	0.01	0.01362	0.00167	
β	0.06	0.75169	0.03965	
θ_0	0.1	0.78399	0.24773	-2290.062
θ_1	0.006	0.00234	0.00018	
α	0.0001	0.01348	0.00165	
β	0.9	0.75153	0.03980	

Table 8.2.1: Model 1 results for collected data, BFGS method.

values for the parameters. This was to ensure that the results applied no matter where the procedure was started.

Results from the optimization of model 1 for three sets of starting values can be seen in table 8.2.1. The horizontal lines between β and θ_0 indicates different runs, in other words, that the results from a new run start below the line with the estimate for β . The scaling of the parameters for the runs in the table are done by the starting values. It is however not necessarily the case that this scaling works. For instance, when running the optimization for starting parameters (0.09, 0.005, 0.033, 0.02) and scaling with these there was an error. When scaling with $sc=c(0.5, 0.0001, 0.01, 0.5)$ however, the result is similar to the others (see appendix A.2.3).

From the results in table 8.2.1 there seems to be an agreement about a maximum point. That the same point (or approximately the same point) is reached from different starting values suggests that this is a maximum. To back this up, the corresponding Hessians all have negative eigen values.

However, not all the results were the same. This was for instance the case for the Nelder-Mead method for the first run and third runs above. The resulting values from these runs could however not be maximums, as the log likelihood values were smaller than for the results listed above. They did not agree either, suggesting that they ended up at points which were not a very global maximum at the best, or that the optimization failed somehow. The Hessian matrix of these results did have negative eigen values, suggesting that perhaps the points are along a ridge, or a very local maximum. Another

Parameter	Mean	Standard deviation
θ_0	0.79614	0.128
θ_1	0.00235	0.00013
α	0.01364	0.00164
β	0.75078	0.02898

Table 8.2.2: Model 1 results of simulation, $n = 1000$ simulations.

Parameter	95% confidence interval for $E(\hat{\theta})$, based on bootstrapping	True Parameter value/ ML estimate
θ_0	[0.78819, 0.80408]	0.78289
θ_1^*	[0.0023403, 0.0023562]	0.0023434
α	[0.01354, 0.01375]	0.01362
β	[0.74898, 0.75257]	0.75169

Table 8.2.3: Model 1, confidence intervals based on bootstrap results for expected values of the estimators. *Two extra digits included to be able to determine difference.

possible explanation is that the Nelder-Mead method did not manage to find sufficient change in the log likelihood method to continue searching for a maximum, or that the convergence went to the wrong point.

The standard deviations of the parameters are quite good, except for the one for θ_0 , which appears to be very large for all the runs above. The standard deviation is here about one third of the size of the parameter, suggesting that the values are not very exact. Otherwise, the size of the standard deviations seem to be very small, suggesting precise estimates.

The simulated data is built on the results from the collected data, through the method described in section 6. The mean values of the parameters reached by simulated data sets can be seen in table 8.2.2. As can be seen, the values are not far away from the true parameters (estimates from collected data), so this suggests that the optimizations actually find the values they are supposed to find.

As mentioned, the simulated results may also be used to make confidence intervals (see section 7.4). The confidence interval for the expected values of the estimators can be seen in table 8.2.3. If the estimates (ML estimates) of the expected value are unbiased, they should lie within the confidence interval of the expected value. As can be seen in the table, this is the case

Standard deviation	95%CI based on bootstrapping	Estimate based on FI
σ_{θ_0}	[0.12263, 0.13387]	0.24555
σ_{θ_1}	[0.00012, 0.00013]	0.00018
σ_{α}	[0.00157, 0.00171]	0.00167
σ_{β}	[0.02776, 0.03031]	0.03965

Table 8.2.4: Model 1 confidence intervals for standard deviation, based on simulation results.

for all the parameters with exception of the first parameter, θ_0 . The estimate of θ_0 appears to be slightly biased to the left side. This bias is however small, approximately 0.01325 (1.692%), so for a slightly larger level of the confidence interval, it might have been within the interval itself as a higher level indicates a broader interval.

In table 8.2.4 the confidence intervals and the FI based estimates represent two possible estimates of the standard deviations of the parameters. The confidence intervals are based on the standard deviation in the parameter estimates from the bootstrapped sample, as described in section 6, while the FI based estimate is based on the assumption of normality for the MLE of large sample sizes. As can be seen, the estimate based on the FI tends to be greater than the estimate based on the simulated data. This is the case for all the parameter standard deviations except α . The value of $\hat{\sigma}_{\alpha}$ is closer to the value from the simulated data as it is within the limits of the confidence interval. The difference between the two estimates has an absolute value of approximately 0.00003, so it is quite small.

8.3 Model 2a and 2b

The second models are an expansion of the first, with one extra parameter. I have chosen to use the estimates of θ_0 , θ_1, α and β from the first model results for further calculations. Since the second models are an expansion, it is reasonable to use the estimates from the first model as the starting point for the next models. The new parameter is simply started with a few different values as starting points. The starting values, rounded to five digits, can be seen in table 8.3.1.

θ_0	θ_1	α	β
0.78289	0.00234	0.01362	0.75169

Table 8.3.1: Starting values for models 2a and 2b, rounded to five decimals.

Parameter	starting value, (θ_2)	Estimate	Standard deviation	Log-likelihood
θ_0	1	0.26998	0.08326	-1248.801
θ_1		0.00161	0.00020	
θ_2		1.57018	0.10223	
α		0.01372	0.00222	
β		0.60977	0.03239	
θ_0	1.05	0.26991	0.08351	-1248.802
θ_1		0.00161	0.00020	
θ_2		1.57003	0.10220	
α		0.01370	0.00222	
β		0.60963	0.03241	
θ_0	1.5	0.27009	0.08325	-1248.801
θ_1		0.00161	0.00020	
θ_2		1.56967	0.10218	
α		0.01372	0.00222	
β		0.60986	0.03239	

Table 8.3.2: Model 2a, results from collected data.

8.3.1 Model 2a

Model 2a is a simple model that changes the rate only dependent on whether the localities involved are the individual's home locality or not. For further details on the model, see section 5.2.1.

The results for a few runs with different starting values can be seen in table 8.3.2. Note that only the starting values for θ_2 are indicated in table 8.3.2, as the starting values for the remaining parameters are in table 8.3.1, as stated .

The results in table 8.3.2 seem to converge towards the same point. In particular, the value of θ_2 is approximately 1.57. This suggests that the change in rate due to whether or not the bird is flying towards its nesting locality is an increase of approximately 57 % for flying to the home locality, and a decrease of about 36 percent in going away from their home locality compared to other locations. Also, the value of the log likelihood is very

Parameter	Mean of Estimates	Standard Deviation in estimates
θ_0	0.27672	0.05636
θ_1	0.00162	0.00015
θ_2	1.57329	0.10593
α	0.01381	0.00221
β	0.60821	0.03265

Table 8.3.3: Model 2a, results from simulated data. n=1000 simulations.

reduced compared to the first model, suggesting that a model which adjusts for home localities may be a good idea.

The runs in the table represent starting values at three points positioned differently compared to the value they converge to for θ_2 . That all three of them converge to the same value suggests that the value is in fact a maximum. When optimizing the function for a higher starting value of θ_2 ($\theta_2 = 2$) the results are similar to the ones in the table above. However, probably because of the limited accuracy of a computer, some of the probabilities contained imaginary parts that were not 0 (see appendix A.2.2). As explained in the appendix, this probably does not have an effect on the result, even though it causes a warning to be issued. For the runs in table 8.3.2 all the Hessian matrices have negative eigen values, supporting the theory that this is a maximum point.

The simulation of the results for model 2a is similar to that of model 1, but with the difference that the data is now based on model 2a, and the optimization is done with respect to the model 2a log likelihood. Some summary statistics can be seen in table 8.3.3, while confidence intervals can be seen in table 8.3.4 and 8.3.5.

The mean of the simulated data results, as seen in table 8.3.3, are quite similar to the results reached by the collected data calculation in table 8.3.2. θ_0 is however a bit larger. In general, similar results suggest a small bias of the estimates based on the collected data if any. The means in table 8.3.3 represent the expected value of the ML estimator, and is thus a way to check for biases as the true parameter values are known for the simulated data sets.

From the bootstrap results in table 8.3.4, the only parameters which are biased are θ_0 and θ_1 . The bias of θ_0 is small, approximately 0.00663, which means a percentage of approximately 2.455. The bias is small, but nevertheless it might show a tendency for the ML estimator to give a biased estimate of the true value. This could therefore mean that the estimate from the collected data is also biased.

Parameter	95% CI based on bootstrapping	True Parameter value/ ML estimate
θ_0	[0.27322, 0.28022]	0.27009
θ_1^*	[0.0016099, 0.0016284]	0.0016065
θ_2	[1.56671, 1.57986]	1.56967
α	[0.01367, 0.01395]	0.01372
β	[0.60619, 0.61024]	0.60986

Table 8.3.4: Model 2a, confidence intervals for expected values of parameter estimators.

Standard deviation	95% CI based on bootstrapping	FI estimate
σ_{θ_0}	[0.05400, 0.05895]	0.08325
σ_{θ_1}	[0.00014, 0.00016]	0.00020
σ_{θ_2}	[0.10149, 0.11079]	0.10218
σ_{α}	[0.00212, 0.00232]	0.00222
σ_{β}	[0.03128, 0.03414]	0.03239

Table 8.3.5: Model 2a, Standard deviation Confidence intervals based on simulated data.

The bias for θ_1 is approximately 0.00001 (0.621%), and is thus very small. Although this could also be caused by a tendency of the MLE to be biased, this is perhaps less likely as a smaller difference than for θ_0 would have made it unbiased. It could also be the case that the θ_1 bias is caused by computational limits, but either way, a slightly broader confidence interval would probably have included the θ_1 estimate. Consideration of the accuracy of computers must be taken into account, so even though the estimates are on the outside of the CI, one should bear in mind the proximity to the interval. The other estimates are within their respective confidence intervals, and are thus unbiased.

As can be seen in table 8.3.5, the estimates based on bootstrapping and the estimates based on FI do not quite agree for all the parameters. For three of the parameters the FI estimate is not within the confidence intervals at all. The FI information estimates are based on the normal assumption of the MLE for large sample sizes, and in addition to this, the FI used is the observed FI, and not the Fisher information to which the normal behaviour of the MLEs apply (note however that Efron and Hinkley (1978) argue for

the use of the observed Fisher information). The normal assumption and the observed Fisher information could be reasons that the two estimates are so different.

8.3.2 Model 2b

In model 2b, the distance to the home locality is taken into account. This is done by increasing the rate for travelling from i to j if j is closer to the home locality than i , and decreasing it if the distance is smaller from i to the home locality than from j to the home locality. The idea of this model is thus that a bird moves towards its home at a higher rate than away from it.

What value one starts with for a parameter is particularly important for models such as this. The reason for this is that as the rather large distances are measured in meters and there is no subtraction from the meters, the parameter can quickly blow up the rate to a very large or shrink it to a very small value. If the distance had been in for instance kilometres this could be avoided to a certain degree. However, the parameter value would have had to have been adjusted, and perhaps have been very small. This is also the case in model 1, but in model 1 there is only one parameter which has the ability to do this in the exponential, while for model 2b there are two. However, this could also be an advantage as the two contributions might cancel one another to a certain degree as one is negative and one is positive. However, one has to consider that the value of θ_3 is in fact the effect in the exponential of the rate per meter. A choice of parameter of for instance 1 could lead to a change of e^{4000} in the rate if the distance is 4000 meters. This change is probably too large. Hence it is important that the starting values are not too large, as this might cause the rate to become uncontrollably high (see appendix section A.3.3).

All the parameter values in table 8.3.6 seem to converge to the same point, suggesting that this is a maximum. The new parameter, θ_3 , is estimated to be 0.00044, which indicates a change in the rate of 0.44 per kilometre of distance. The fact that the value of θ_3 is assumed to be positive seems to give a reasonable result, which perhaps would not be the case if movement tended to be away from the home locality.

Comparing table 8.3.7 and table 8.3.6 it is clear that the mean of the estimates from the simulated data is close to the value reached for the collected data. This suggests that the bias is small for the estimator. To further consider the bias, the confidence intervals reached for the expected value of the estimator are shown in table 8.3.8. Table 8.3.9 shows the estimates of the standard deviation based on the simulated data and the Fisher information

Parameter	Starting value	Estimate	Standard deviation	Log-likelihood
θ_0	0.0001	0.09872	0.03651	-1248.538
θ_1		0.00094	0.00027	
θ_3		0.00044	0.00006	
α		0.01328	0.00220	
β		0.59212	0.03181	
θ_0	0.0005	0.09868	0.03649	-1248.538
θ_1		0.00094	0.00027	
θ_3		0.00044	0.00006	
α		0.01328	0.00220	
β		0.59211	0.03181	
θ_0	0.00001	0.09872	0.03652	-1248.538
θ_1		0.00094	0.00027	
θ_3		0.00044	0.00006	
α		0.01328	0.00220	
β		0.59212	0.03181	

Table 8.3.6: Model 2b, results from collected data.

Parameter	Mean of estimates	Standard deviation in estimates
θ_0	0.10002	0.02156
θ_1	0.00093	0.00018
θ_3	0.00044	0.00006
α	0.01335	0.00231
β	0.59262	0.03009

Table 8.3.7: Model 2b, summary results from bootstrapped/simulated data, $n = 1000$ simulations.

Parameter	95% CI based on bootstrapping	True parameter value/estimate
θ_0^*	[0.0986810, 0.1013573]	0.0986833
θ_1^*	[0.0009226, 0.0009447]	0.0009401
θ_3^*	[0.0004335, 0.0004405]	0.0004351
α	[0.01320, 0.01349]	0.01328
β	[0.59076, 0.59449]	0.59211

Table 8.3.8: Model 2b, Confidence intervals for expected values of parameters based on simulated data.

Standard deviation	95% CI based on bootstrapping	FI estimate
σ_{θ_0}	[0.02066, 0.02255]	0.03649
σ_{θ_1}	[0.00017, 0.00019]	0.00027
$\sigma_{\theta_3^*}$	[0.0000536, 0.0000586]	0.0000616
σ_{α}	[0.00221, 0.00242]	0.00220
σ_{β}	[0.02883, 0.03147]	0.03181

Table 8.3.9: Model 2b, standard deviation confidence intervals based on simulated data, $n = 1000$ simulations. *7 digits used to show difference.

respectively.

By the confidence intervals of the expected values of the estimators, all the estimates based on the collected data are within the confidence intervals. Thus, the comparison indicates that all the estimators are unbiased when measured with a 95% CI. Hence the estimates are not assumed to be consequently too low or too high compared to their expected value.

From the confidence intervals for the standard deviation in table 8.3.9 it is once again clear that in general, the FI estimate tends to estimate a higher standard deviation than the bootstrap estimate. The opposite is the case for σ_{α} , where the value is slightly lower for the FI estimate. However, this estimate is only just below the confidence interval, so the difference is not very big.

8.3.3 Model 2a and Model 2b

The two models based on the principle that birds move home at a higher rate seem to both have improved the first model a lot when considering the log likelihood values. When performing a likelihood ratio test on the two selected runs for model 2a and 2b respectively versus the selected run for model 1, the p-value came out as 0 for both of the tests.

Thus a test of any level would reject the null hypothesis which is $H_0 : \theta_2 = 1$ and $H_0 : \theta_3 = 0$ for model 2a and 2b respectively. This suggests that including an increased rate for movement towards home is a significant improvement on the model. This is reasonable when you consider the very large difference in the log likelihood value. The log likelihood values reached for model 2a and 2b are much higher than for model 1, but they are quite similar to each other. This makes the choice of model 2a or 2b more difficult.

As can be seen, the difference between the log likelihood values for model 2b and 1 is larger than that between model 2a and 1. This results in a slightly larger value for the chi-squared test variable of the likelihood ratio test, placing the value even further away from the chi squared distribution which applies if the null hypothesis is true.

I have chosen to continue with model 2b as the log likelihood value is slightly larger, and the specification of direction is more accurate in that it increases the rates for all movements toward home and not only for the movements specifically to the home locality.

8.4 Model 1 and model 2b

Comparing model 1 to model 2b gives a p-value of 0, so model 2b is a significant improvement at any level for the data. Therefore I have chosen to use model 2b as a basis for the next models. Note that model 2b also incorporates the details from model 1.

8.5 Model 3

Model 3 is an extension of model 2b. It allows for a difference between rates for male and female individuals through the introduction of the parameter θ_4 (see section 5.3). As in model 2b, model 3 allows for an adjustment for increased rate to home locality and an adjustment based on distance from model 1. Hence model 3 incorporates all the factors from model 1 and model 2b in addition to the new parameter.

As for model 2a and model 2b, the starting parameters for the "old" parameters are set to be the result of the run for the previous model. In this

Parameter	Starting value	Estimate	Standard deviation	Log likelihood
θ_0	0.5	0.18911	0.07876	-1243.541
θ_1		0.00110	0.00026	
θ_3		0.00042	0.00006	
θ_4		-0.78009	0.25237	
α		0.01332	0.00219	
β		0.59407	0.03214	
θ_0		3	0.18932	
θ_1	0.00110		0.00026	
θ_3	0.00042		0.00006	
θ_4	-0.78029		0.25255	
α	0.01332		0.00219	
β	0.59413		0.03216	
θ_0	-3		0.18974	0.07919
θ_1		0.00111	0.00026	
θ_3		0.00042	0.00006	
θ_4		-0.78004	0.25257	
α		0.01332	0.00219	
β		0.59436	0.03218	

Table 8.5.1: Model 3, results from collected data.

case these are the results from model 2b (table 8.3.6). The starting values of the new parameter for a few runs positioned differently compared to the resulting estimate can be seen in table 8.5.1. From the results in table 8.5.1 it seems like the value of θ_4 is approximately -0.78. Since the value is negative, it further decreases the rate for individuals where it is present. Since males do not have θ_4 ($sx = 0$, so $\theta_4 sx = 0$), this suggests that females have an overall lower dispersal rate than males. This can easily be seen by that for a female the rate is reduced by a further multiplication of e^{θ_4} , which is smaller than 1, and hence reduces the overall rate compared to the rate for male individuals. In fact, the value of $e^{-0.78}$ is approximately 0.46, which means almost halving the dispersal rate of a female compared to a male in the same position and with the same home locality.

As mentioned, Tufto et al. (2005) writes that female blue and great tits disperse longer distances than males according to their results. In the book *Biology of the Ubiquitous House Sparrow* by Anderson (2006), it is stated that male house sparrows have greater foraging distances than females in average (Anderson, 2006, p. 337). The results above (table 8.5.1) indicate that female house sparrows disperse less often than males, but does not say

anything about distances when they do. However, if males and females are thought to be travelling the same amount of time, a lower dispersal rate for females would indicate that the females fly greater distances than the males, which is not the case according to Anderson. Another possibility is that either females stay at their home farm more, or that they are more difficult to observe. Of the 105 individuals with data recorded on them for all the models, there are 56 females and 49 males, so the distribution of individuals on the two sexes is quite equal. However, when it comes to the number of observations, 308 of the observations used were of females, and only 230 of males. This suggests a difference in the mean number of observations of each individual to be nearly one higher for females than for males, indicating that perhaps females are easier to observe. Note that this is not a formal test, and that further research would have to be done to establish this.

Again, the different starting points seem to converge to the same point. This suggests that this is a maximum over a larger area, and not just a local maximum for one of the starting points. In addition to this, the runs have negative eigen values on their Hessians, which supports the claim that this is indeed a maximum point.

Parameter	Mean of estimates	Standard deviation in sample
θ_0	0.19620	0.04940
θ_1	0.00111	0.00019
θ_3	0.00042	0.00005
θ_4	-0.79338	0.21595
α	0.01337	0.00227
β	0.59425	0.03090

Table 8.5.2: Model 3, summary results from simulated data, $n = 1000$ simulations. Note warning, see appendix, section A.3.4.

The results seem to be quite similar for the simulated data. The means in table 8.5.2 are similar to the results reached for the collected data (which are the true values for the simulated data sets). When running the simulations, a warning appeared. Even though this warning appeared, the results appear to be quite as expected, so the results were used anyway. For more details on the warning, see appendix section A.3.4.

From the confidence intervals based on the simulated data in table 8.5.3 it is clear that all the parameters are unbiased except for θ_0 . The bias of the θ_0 parameter is of size 3.749%, so it is not very large, albeit larger than it was for model 1 and 2a for the same parameter (here they were 1.692% and 2.455% respectively). The bias here is about 4% so it seems that the ML

Parameter	95% CI for expected value of estimate	True value
θ_0	[0.19313,0.19926]	0.18911
θ_1^*	[0.0010993,0.0011228]	0.0011046
θ_3^*	[0.0004202,0.0004270]	0.0004211
θ_4	[-0.80678,-0.77998]	-0.78009
α	[0.01323,0.01351]	0.01332
β	[0.59234,0.59617]	0.59407

Table 8.5.3: Model 3, confidence intervals for expected value of parameter estimator, n=1000 simulations. Note Warning, see appendix section A.3.4. *7 digits used to determine difference.

Standard deviation	95% CI based on bootstrapping	FI based estimate
σ_{θ_0}	[0.04733,0.05167]	0.07876
σ_{θ_1}	[0.00018,0.00020]	0.00026
$\sigma_{\theta_3^*}$	[0.0000526,0.0000575]	0.0000611
σ_{θ_4}	[0.20689,0.22586]	0.25237
σ_{α}	[0.00217,0.00237]	0.00219
σ_{β}	[0.02961,0.03232]	0.03214

Table 8.5.4: Model 3, estimates of standard deviations. *7 digits used to determine difference.

estimator tends to overestimate the true value by almost 4%. The parameter estimate of θ_0 seems to be out of the confidence interval quite often, as this is the case in model 1 and 2a as well.

Table 8.5.4 of the standard deviation estimates shows that the FI based estimate again seems to have the highest estimate for the standard deviation for the parameters. The estimates for σ_{α} and σ_{β} are however within the confidence interval of the bootstrapping results, so at this level the two results are not in disagreement. Overall, the standard deviations appear to be quite good, but perhaps a bit large for the parameter θ_4 .

8.6 Model 2b and model 3

From the log likelihood values of the runs of models 2b and 3 it is clear that the improvement is nowhere near the improvement the second models made on the first model. The log likelihood values are however a little smaller for

model 3, and to test if they are significant enough the likelihood ratio test is used. Note that the test in this case takes the form $H_0 : \theta_4 = 0$, which is model 2b, versus $H_1 : \theta_4 \neq 0$, which is model 3.

The likelihood ratio test gives a p-value of approximately 0.00157. The p-value is quite low, below any value which is normally used for a test (that is, 0.05 or 0.01), so tests on these levels, and any other level higher than the p-value, would accept the alternative hypothesis. The alternative hypothesis is thus accepted, and the adjustment in the model based on an individual's sex is a significant improvement of model 2b.

8.7 Model 4

From the results for the comparison between models 2b and 3 it was clear that model 3 offered a significant improvement to the dispersal rate model. Therefore model 3 is used as the basis for model 4, and the starting values used for the parameters present in both the models as the result from model 3 (table 8.5.1).

To run model 4 I had to choose what value to give θ_6 , the date limit (see section 5.4), in addition to the starting values for θ_5 . To coincide with the increase in the dispersal distance noted by Anderson (2006), I chose θ_6 to represent a date in late summer. As the limit I used 16th of August 2011. This corresponds to the index 91 in the list of ordered sampling points. This way, all sampling points which have a sampling point with index lower than 91, gets $d_s = 0$, and equal or greater than 91 gets $d_s = 1$. This splits the sampling points into two uneven groups when it comes to size (90 sampling points and 52 sampling points), but the number of sampling points is of reasonable size on both sides of the limit (note that this is the number of sampling points, not the number of observations).

The results for model 4 for three runs where the starting value of θ_5 is positioned differently to the estimate reached for θ_5 , can be seen in table 8.7.1. From the results, it is clear that all the runs converge to the same point. They all agree that θ_5 takes a value of approximately -1.99. This value indicates a decrease in the dispersal rate in late summer because the dates which have θ_5 present, the ones in late summer, are multiplied by e^{θ_5} compared to those who do not have it present. This means that the result suggests a decrease in dispersal rate for individuals in late summer and autumn. The result could fit with longer distance dispersal in the late summer and autumn (Anderson, 2006), as longer distances could mean fewer dispersals as the distances flown are more time-consuming.

The standard deviation estimates based on the observed FI are varying. The standard deviation is mostly good, that is quite small, but for certain

Parameter	Starting value	Estimate	FI estimate of s.d.	Log likelihood
θ_0	2	0.18994	0.09504	-1218.056
θ_1		0.00017	0.00033	
θ_3		0.00048	0.00005	
θ_4		-0.75467	0.30675	
θ_5		-1.99458	0.33075	
α		0.01141	0.00214	
β		0.56164	0.02712	
θ_0	-0.5	0.19019	0.09525	-1218.056
θ_1		0.00017	0.00033	
θ_3		0.00048	0.00005	
θ_4		-0.75491	0.30695	
θ_5		-1.99591	0.33106	
α		0.01141	0.00214	
β		0.56165	0.02712	
θ_0	-4	0.19134	0.09622	-1218.056
θ_1		0.00017	0.00033	
θ_3		0.00048	0.00005	
θ_4		-0.75775	0.30732	
θ_5		-1.99630	0.33133	
α		0.01141	0.00214	
β		0.56164	0.02712	

Table 8.7.1: Model 4, results from collected data.

parameters it is very bad. To take the worst case first, the standard deviation of θ_1 is larger in absolute value than the parameter itself. This means that the estimate is very uncertain, and drawing conclusions from this estimate can be risky.

Further, the standard deviations of θ_1 and θ_4 are quite large as well. The values of the standard deviation estimates are approximately half and one third of the estimated parameter value. Hence one has to use caution if drawing critical assumptions from this.

The log likelihood values for the three runs in table 8.7.1 are all equal, and the estimates all tend to converge to the same point. The eigen values of the Hessians are all negative, which supports the maximum point assumption. The log likelihood value is larger than the one for model 3, suggesting that there is a difference in time.

The simulation and optimization of the model 4 data sets require more computations than the equivalent process for the other models. This is be-

Parameter	Mean of estimates	Standard deviation
θ_0	0.22370	0.10107
θ_1	0.00022	0.00021
θ_3	0.00049	0.00004
θ_4	-0.77064	0.25165
θ_5	-2.04532	0.27216
α	0.01150	0.00219
β	0.56367	0.02884

Table 8.7.2: Model 4, summary statistics for simulated data. $n = 425$ simulated data sets.

cause of the recalculation of the A matrix mentioned in section 5.4. Due to this, the number of simulations done for model 4 is lower than the 1000 simulations done for the other models. This causes the confidence intervals to be broader, and the estimates less accurate, as the simulation is less representative (not as many possibilities are taken into account). As opposed to the other models, some of the runs did not converge for model 4. All of the runs are based on the same conditions, so either the function does not work for certain simulated data sets, or the optim function must have failed. After removing the results from runs which did not converge (results from 35 runs were removed), the sample size for model 4 is $n = 425$.

The results from the simulated data sets seem to differ a bit from the results from the collected data. The values are all greater than their true value (the results from the collected data), which suggests a tendency for the estimator to be biased. However, some of the optimizations for model 4 failed with the BFGS method. The Nelder-Mead method is then used instead as explained in appendix section A.2.4. This is the case for 84 of the used simulations, while the remaining 341 are calculated with the BFGS method.

The standard deviations among the simulated results are rather large for θ_0 , θ_1 and θ_4 as it was also suggested in the calculations for the collected data. This again can suggest that these parameters, and particularly θ_1 as the standard deviation is nearly as large in absolute value as the estimate itself, are very uncertain.

From table 8.7.3 it is clear that some of the estimates are within a reasonable range of the true value, and are hence unbiased. This is the case for parameters θ_3 , θ_4 , α and β . The remaining parameters are biased when measured at the 95% level. The biases are 17.8%, 29.4% and 2.5% for θ_0 , θ_1 and θ_5 respectively. The very large biases for the two first parameters

Parameter	95% CI from simulations	MLE from collected data
θ_0	[0.21406,0.23333]	0.18994
θ_1	[0.00020,0.00024]	0.00017
θ_3^*	[0.0004824,0.0004908]	0.0004842
θ_4	[-0.79463,-0.74664]	-0.75467
θ_5	[-2.07127,-2.01937]	-1.99458
α	[0.01129,0.01171]	0.01141
β	[0.56092,0.56642]	0.56164

Table 8.7.3: Model 4, CI for expected value of estimator, and estimate based on collected data (true value for simulation optimization). *7 digits used to show difference.

Parameter	95% CI from simulations	FI based estimate of sd
σ_{θ_0}	[0.09470,0.10836]	0.09504
σ_{θ_1}	[0.00019,0.00022]	0.00033
$\sigma_{\theta_3}^*$	[0.0000411,0.0000471]	0.0000492
σ_{θ_4}	[0.23580,0.26981]	0.30675
σ_{θ_5}	[0.25501,0.29180]	0.33075
σ_{α}	[0.00205,0.00235]	0.00214
σ_{β}	[0.02702,0.03092]	0.02712

Table 8.7.4: Model 4, estimates of standard deviation. CI for estimate based on simulation, and estimate based on FI.*7 digits used to show difference.

can be seen in the light of the very large standard deviation present in table 8.7.1. A very large standard deviation suggests a very uncertain estimate. This could be the case for the simulated data as well, suggesting that the different parameter estimates for these two parameters are vastly different, as is also reflected in table 8.7.2 where the size of the standard deviations are very large compared to the estimates.

The estimates of the standard deviations in table 8.7.4 again seems to show that the estimates based on the observed Fisher information are higher than those based on the simulated data. For θ_0 , α and β however, the two estimates seem to be in reasonable agreement with the value of the observed Fisher information estimate inside the 95% CI of the simulated data estimates. For the remaining four parameter estimates, the observed Fisher information estimate seems to be larger than the upper limit of the

confidence intervals based on the simulated data, and hence the observed Fisher information estimator seems to estimate a higher standard deviation than the standard deviation present in the results from the simulated data.

8.8 Model 3 and Model 4

As already mentioned, the log likelihood value of model 4 is a bit larger than the log likelihood value of model 3. This suggests that model 4 offers a better fit to the data. However, some of the parameter estimates in model 4 had very large standard deviations. This suggests that the model is perhaps not the best model when it comes to stability, as the uncertainty of the results are quite large. Nevertheless, a likelihood ratio test gives a p-value of approximately $9.3778 \cdot 10^{-13}$, which suggests that model 4 offers a significant improvement of model 3. If doing a test of level 5% it is now clear that the alternative hypothesis ($\theta_5 \neq 0$) would have been accepted.

The large standard deviations in model 4, which were not present in the earlier models to such a great extent, suggests that perhaps the use of model 3 is a safer choice. One will have to weigh up the two models and compare them in each case to find out which is better suited. In cases where accuracy is essential, model 3 might be a better choice, while in models where it is not crucial that the parameters have small standard deviations, perhaps model 4 would be better.

8.9 General

Throughout the estimates from the different models it seems that the estimates of α and β are quite similar. This should make sense because the death rate and the resighting probability should be independent of the dispersal rate, and they should be the same for the same data set. Note that the two parameters are not present in the expression for the dispersal rates (equation 26 in section 4.5), and only enter the calculations through the probabilities calculated and the log likelihood expressions respectively.

In figure 3 one can see how the different model estimates are compared to each other. From the plots it is clear that there are differences between the different models, but that the values of the parameter estimates are not that different. There is however some variation, as is to be expected. One would expect such a change when there are more parameters and covariates which can explain changes in the rate.

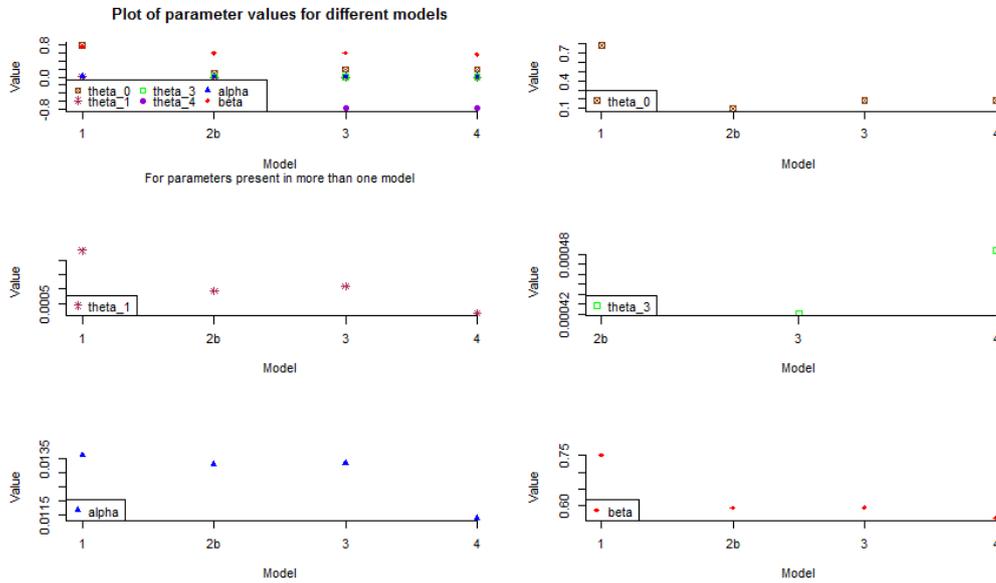


Figure 3: Estimates for different models.

9 Discussion

The assumption of the death rate to be equal at all localities and all times is perhaps controversial. As von Post mentions in her PhD thesis, modernization of farm work and less leftovers from farming may make birds' life more difficult at farms (von Post, 2013, p.16). Changes in agriculture or farming is believed to be a major reason for an observed decline in house sparrow populations Anderson (2006, p.343). Hence a possible addition to the model is to take into account possible differences in the living conditions at the different farms. However, the largest distance between two farms in this study is approximately 4840m (4841.12m between farms 3 and 8). Anderson (2006, p. 336) states that house sparrows are quite sedentary, but that typical flight distances from a bird's breeding site can be up to a few kilometres at certain times of the year (late summer/autumn). Anderson mentions several studies, and from the distances mentioned there, typical foraging distances varies from a few hundred meters from the breeding site to 5 km, although 3.2 km is also mentioned as a maximum distance (Anderson, 2006, p. 337).

Hence, the conclusion on the distances between the farms is that although the direct distance between some farms are perhaps out of normal range for an individual, it should certainly be possible for an individual to move between all the farms in search of food. Also, it must be taken into account that

although an individual is assumed to always be on a farm, this is probably not the case in real life. It is possible that an individual will start on a journey longer than what one would perhaps assume normal, and eat on the way. This suggests that the birds will be able to travel between any two farms, although they might not do so too frequently for farms separated by large distances.

One should also keep in mind that as the data are collected at certain times and not continuously, the birds may have used a completely different route than what appears from the data. In addition to that the bird may have flown differently than what is randomly chosen for the multiple data, this could also be the case for other data. For example, a hypothetical individual could have travelled between farms 1 and 4 in many ways, for instance 1-2-6-3-2-4. If the bird is only observed at farms 1 and 4, it will appear from the data that the individual flew directly from farm 1 to farm 4, although that was not the case. This is particularly important to keep in mind since the farms are so close to each other that an individual should easily be able to reach any farm from its position. This makes such cases not only possible, but also rather likely. For larger study areas, for example when considering subpopulations on different islands as Jensen et al. (2013) does, this is perhaps less likely to happen.

The birds used in this study are all juveniles, meaning that they might disperse away from their nestling site as well to find a new place to live (*natal dispersal*, Anderson (2006, p. 338)). This could therefore increase the number of dispersals in this study compared to a population where all individuals are part of the study. On the other hand, the natal dispersal in house sparrows is only about 10% (Pärn et al., 2012)), so the effect of this if present, might not be very large. If the value of the natal dispersal is higher for such short distances as used in this study, it might however have an effect.

Another potential inaccuracy is the constant death rate when it comes to time. Very young individuals may have a higher death rate than older individuals, as they are new to the environment around them and more helpless. On the other hand, younger individuals may have other advantages over older ones. In addition to this, seasonal changes may affect the death rate, as it may be more difficult for birds to find food during certain periods, for example in the winter. As mentioned, only data on juvenile birds are used in this project, again justifying the constant death rate with respect to age as all the individuals are of similar age. Also, all the data used is collected in the period May through September. Hence any great seasonal changes might not be reflected in the data. This can make it hard to give a good picture of movement for all of the year, but then it also justifies the use of a constant death rate. I have used the assumption of a constant death

rate, but this is something that should be further researched. For further research on the death rate it might be reasonable to also use data from the winter half of the year in order to detect potential differences. One way to do this is to split the year into units of similar weather conditions and then have a death rate for each unit. Alternatively, the death rate could be a function of the weather (or other factors, for instance age) if one finds a way to sufficiently tie together weather conditions and death rate. By inclusion of biologists with a thorough knowledge of how house sparrows are affected by the weather, this could be an interesting possible field of study.

Further research should also be done into the differences of the dispersal rate in time. Model 4 takes a very simple version of this into account, but only for one date limit. The value of when such a limit should be set is perhaps something that should be discussed, and also whether there should be an increase in the number of such limits. The one limit set in model 4 is set with the goal of capturing a potential difference between spring/early summer and late summer/autumn. There might be similar differences close to the breeding season as well, if one imagines that the birds are busy feeding their offspring. Although this might not affect the dispersal distance, it could affect the dispersal rate, which is what is being studied here.

An alternative to the set date time difference used here is to instead let the time affect the rate through a function. It is rather unlikely that suddenly in the course of one day, the rate should change drastically. A more plausible scenario is that the gradual changes in weather and the environment that naturally comes as the seasons change, have an effect on the rate. A way to do this is to let the date affect the rate through a continuous function that has an appropriate value at different times of the year so the change between following days is gradual, and not abrupt as in model 4. To make such a model, an appropriate function for the seasonal changes must be found. This is a challenge in itself, but by no means impossible, and with the help of a meteorologist and a biologist it should be possible.

The dispersal rate can also be dependent on population size indirectly (Pärn et al., 2012)). Pärn et al. argues that a larger population size may restrict the resources available for each individual. This is logical, as the number of individuals competing for the food and other resources is greater, and hence more individuals may not get the needed resources. At Hestmannøy, the distances are so small that individuals will in theory be able to travel between all farms to find food, so if there is not enough, they might fly to the neighbouring farm just to get food, and then return. This could also increase the number of dispersals.

Another possible adjustment is to add an adjustment based on memory effects, which is discussed by Börger et al. (2008). Börger et al. discuss

examples where a home range is defined by the help of memory. Such adjustments are for instance adjustments on whether an animal will or will not revisit previously visited sites more or less often than other sites. Perhaps an incorporation of an idea of this type could also be useful in a dispersal rate model.

When simulating the data sets, some values have to be chosen as the true values. The simulations are based on the the model in question, but this also means that the value of the parameters have to be chosen before the simulation is done. The values I chose to use as true values in this thesis are the values that result from the optimization of the collected data. These values are as mentioned estimates of the maximum likelihood estimates, and hence a very reasonable value to base the simulations on. Under the hypothesis that the collected data is based on the model in question, these MLE values would be likely values for the true parameters of the model behind the collected data. One would similarly expect the MLE estimates based on the simulated data set to be similar to their true value if the log likelihood function and optimization work as they should. The difference is that for the simulated data, the true parameters and the model behind the data sets are known.

Because of the limitations in my data set on year and age, several of the localities do not have any data on individuals travelling to or from them. This could either be because observations were not made on that particular farm in 2011 (the case for farm number 1, 5, 7 and 8), or because no juvenile individuals were observed at these farms at any of the sampling points (the case for farm 3). This means that all the rates to and from these farms are based on the rates between other farms. Therefore, the rates may not be very representative for rates to and from the "missing" farms. To amend this, a possible solution is to use a larger data set, perhaps the data set for juveniles for two years, or however many years are needed for all farms to be represented by data. Note that to avoid any age difference disturbance one should only use juveniles for each year (or another age group) or adjust the rates to account for a possible difference. If the resulting data set proves to be too large, it might be sufficient to follow a few individuals from each year, or perhaps a random sample of individuals from each year.

10 Conclusion

In this study I have attempted to fit five different models to a data set consisting of house sparrow movement history. The goal has been to find a well-suited model for the dispersal rates of the house sparrows, and thus reach

a model that can easily be adapted to other partially isolated populations.

It was clear from the study that the models improved the more detail was included in the expression for the rate. An exceptional improvement was seen when each individual's home locality was adjusted for, through a very great increase in the log likelihood value. The home locality was adjusted for in a manner that increased the dispersal rate for travelling towards the home locality. This suggests that distance between two points as a basis for the dispersal rate between them does not alone offer the best possible basis for the rates.

When expanding the model, more data on the individuals is used. The first expansion, the inclusion of a home locality, suggested that the rate was indeed higher for movement toward the home locality than away from it. This does not necessarily suggest that the birds only fly to where they were hatched, but might indicate that a bird prefers not to disperse too far from their hatching locality. The second expansion, which included a parameter depending on the sex of an individual, pointed in the direction that there might be a dispersal rate difference between males and females. The females seemed to disperse less often than males. This result might be controversial, so this should be studied further.

The last element included in the final expansion is difference in time. The log likelihood ratio test indicated that the inclusion of difference in time is significant. However, the model used in this study is very simplified, with the difference set for dates after a specific date limit. In addition to this, the standard deviation estimates were very large for some of the parameters. It is possible that other dates would be more suited, or that more dates should be included. A perhaps better possibility is to let the time be a function to allow gradual changes in the rate throughout the year. The dispersal rates' dependence on time is a field of study with much potential, as it is reasonable that the birds move differently at different seasons as the hunt for food or shelter would be different.

I conclude that to create a model for the dispersal rate, several factors should be taken into account. The model seemed to improve for every new parameter included, although it was clear that not all new expansions to the model should be accepted unquestionably. If utilizing the models tested in this study without further improvement, I would probably recommend model 3 or 4. Although the fourth model has an increased log likelihood to the third model, the fourth model has a few unreasonably large standard deviations. More research on the model could probably improve it further. What parameters are taken into account in a possible future study is up to the researches, but I would recommend to further use distance to home locality, or for older individuals perhaps change this to the site of their adult

nests. Another possibility is to adjust for memory, as mentioned by Börger et al. (2008). In particular, I think the use of a function to represent the contribution based on the different seasons could be an interesting field of study for further research.

References

- Anderson, T. R., 2006. *Biology of the Ubiquitous House Sparrow*. Oxford University Press. Chapter 8.
- Börger, L., B. D. Dalziel, and J. M. Fryxell, 2008. Are there general mechanisms of animal home range behaviour? A review and prospects for further research. *Ecology Letters* 11:637–650. DOI:10.1111/j.1461-0248.2008.01182.x.
- Casella, G. and R. L. Berger, 2002. *Statistical Inference*. International Student Edition, 2. ed. Duxbury.
- Devore, J. L. and K. N. Berk, 2007. *Modern Mathematical Statistics with Applications*. International Student Edition ed. Duxbury, Thomson Brooks/Cole.
- Efron, B. and D. V. Hinkley, 1978. Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information. *Biometrika* 65:457–482. DOI:10.2307/2335893.
- Halliburton, R., 2004. *Introduction to Population Genetics*. Pearson, Prentice Hall.
- Jensen, H., R. Moe, I. J. Hagen, A. M. Holand, J. Kekkonen, J. Tufto, and B.-E. Sæther, 2013. Genetic variation and structure of house sparrow populations: is there an island effect? *Molecular Ecology* 22:1792–1805. DOI:10.1111/mec.12226.
- de Jong, P. and G. Z. Heller, 2008. *Generalized Linear Models for Insurance Data*. Cambridge University Press, New York.
- Lay, D. C., 2012. *Linear Algebra and Its Applications*. International Edition, 4. ed. Pearson, Addison-Wesley.
- McKinnon, K. I. M., 1998. Convergence of the nelder–mead simplex method to a nonstationary point. *Siam Journal on Optimization* 9:148–158. DOI: 10.1137/S1052623496303482 (published online: 2006).
- Movable Type Ltd., 2014. Calculate distance, bearing and more between Latitude/Longitude points. <http://www.movable-type.co.uk/scripts/latlong.html>. Accessed first 8. of September 2014.

- Ovaskainen, O., 2004. Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model. *Ecology* 85:242–257. DOI: <http://dx.doi.org/10.1890/02-0706>.
- Pärn, H., T. H. Ringsby, H. Jensen, and B.-E. Sæther, 2012. Spatial heterogeneity in the effects of climate and density-dependence on dispersal in a house sparrow metapopulation. *Proceedings of the Royal Society B* 279:144–152. DOI: 10.1098/rspb.2011.0673 (published online: 2011).
- von Post, M., 2013. Effects of farmland heterogeneity at multiple spatial and temporal scales on house sparrow (*Passer Domesticus*) population ecology. Ph.D. thesis, Lund University.
- Ringsby, T. H., B.-E. Sæther, H. Jensen, and S. Engen, 2006. Demographic Characteristics of Extinction in a Small, Insular Population of House Sparrows in Northern Norway. *Conservation Biology* 20:1761–1767. DOI:10.1111/j.1523-1739.2006.00568.x.
- Ross, S. M., 2010. *Introduction to Probability Models*. 10. ed. Elsevier.
- Tufto, J., T.-H. Ringsby, A. A. Dhondt, F. Adriaensen, and E. Matthysen, 2005. A Parametric Model for Estimation of Dispersal Patterns Applied to Five Passerine Spatially Structured Populations. *The American Naturalist* 165. E-Article. DOI:10.1086/426698.

List of Tables

2.0.1 Locality name and number	9
4.4.1 Example of calculation of log likelihood.	25
8.2.1 Model 1 results for collected data, BFGS method.	40
8.2.2 Model 1 results of simulation, $n = 1000$ simulations.	41
8.2.3 Model 1, confidence intervals based on bootstrap results for expected values of the estimators. *Two extra digits included to be able to determine difference.	41
8.2.4 Model 1 confidence intervals for standard deviation, based on simulation results.	42
8.3.1 Starting values for models 2a and 2b, rounded to five decimals. . .	43
8.3.2 Model 2a, results from collected data.	43
8.3.3 Model 2a, results from simulated data. $n=1000$ simulations. . . .	44
8.3.4 Model 2a, confidence intervals for expected values of parameter estimators.	45
8.3.5 Model 2a, Standard deviation Confidence intervals based on simulated data.	45
8.3.6 Model 2b, results from collected data.	47
8.3.7 Model 2b, summary results from bootstrapped/simulated data, $n = 1000$ simulations.	47
8.3.8 Model 2b, Confidence intervals for expected values of parameters based on simulated data.	48
8.3.9 Model 2b, standard deviation confidence intervals based on simulated data, $n = 1000$ simulations. *7 digits used to show difference.	48
8.5.1 Model 3, results from collected data.	50
8.5.2 Model 3, summary results from simulated data, $n = 1000$ simulations. Note warning, see appendix, section A.3.4.	51
8.5.3 Model 3, confidence intervals for expected value of parameter estimator, $n=1000$ simulations. Note Warning, see appendix section A.3.4. *7 digits used to determine difference.	52
8.5.4 Model 3, estimates of standard deviations. *7 digits used to determine difference.	52
8.7.1 Model 4, results from collected data.	54
8.7.2 Model 4, summary statistics for simulated data. $n = 425$ simulated data sets.	55
8.7.3 Model 4, CI for expected value of estimator, and estimate based on collected data (true value for simulation optimization). *7 digits used to show difference.	56

8.7.4 Model 4, estimates of standard deviation. CI for estimate based on simulation, and estimate based on FI.*7 digits used to show difference.	56
--	----

List of Figures

1	Histogram of observations at each location	9
2	The distribution of the sampling points by location and date(yyyymmdd).	11
3	Estimates for different models.	58

Appendix A The Log Likelihood Function

A.1 Implementation of the Log Likelihood Function

The implementation is principally done in such a way that each function has one task. The main function is the function that calculates the log likelihood value based on the observation matrix.

The log likelihood function calculates the value of the log likelihood based on the parameter values and the observation matrix. The input given to the log likelihood function varies with the model used, for instance model 1 requires only four parameters, while model 2b requires five parameters, as well as a list over the home localities. Because of this, what model one wishes to use is also given as input with default set to model 1. In addition to this, the likelihood function is updated to accommodate later models, but in general, the results from earlier models would not have been affected by the changes made.

Copy of section of log likelihood with tests

```
if(tdf==0){
    pm=pn #To avoid using A2
} else {

    A2=diag(exp(D*tdf)) #

    pm =U%%A2%%Uinv%%pn*exp((-1)*alpha*tdf)

    if (sum(abs(Im(pm))>(1*10-15))>0){
        print('Imaginary value error')
        print(sum(abs(Im(pm))>(1*10-15)))
        print(t(round(Im(pm),digits=16)))
    } else {
        pm=Re(pm)
    }

    pm=round(pm,digits=15)
}
```

During the computation of the results, I have tried to change the log likelihood function as little as possible. However, certain runs have caused issues, and to further unveil these, some changes have been made. In general I have attempted to make sure that all the results presented in the thesis

for one model is run with the same code. There is however one case where there is a difference for the simulated and collected data, but this should not have affected the results. Note however, that these are not changes that would likely have affected the calculations of earlier models, either because they are not relevant for the earlier model, or because no warning or error was obtained for earlier runs to indicate that something was amiss. This is for instance the case for the imaginary values issue discussed below, which occurred for model 2a. Furthermore, the log likelihood function has been changed to work for new models, but in a way that should not affect older models.

A.2 Issues in and Adjustments to the Log Likelihood Function

While running the log likelihood function, several issues turned up. In this section, some of the issues are discussed, and how they were resolved are presented.

A.2.1 Rounding of Probabilities

It became clear that some of the runs resulted in probabilities which were out of the legal range for probabilities ($[0,1]$), as they sometimes occurred as negative. This resulted in an attempt to calculate the logarithm of a negative value, which of course cannot be done. Some of the probabilities were of size 10^{-17} and thereabouts when this occurred. It is likely that probabilities this small were in fact supposed to be 0, but were not because of computational errors caused by the limitations in the number of digits used in the calculations on the computer. Therefore, when discovering that the probabilities were wrong and of what size they were, I chose to round the probabilities to 15 decimal places. This causes extremely small probabilities to be rounded to 0, and hence avoiding the problem occurring with negative probabilities. With the rounding of the probabilities to 15 decimal places, the problem did not occur any more, suggesting that these probable computational effects were indeed the problem.

A.2.2 Imaginary Values

For some of the runs of model 2a, an issue with imaginary values occurred. This was the case when the probability of movement contained imaginary values in addition to real values. From the values I observed, it appeared to be the case that in very many of these instances, the imaginary value actually

had value 0, and hence the number was in fact real, but it was not registered as such. In order to attempt to correct this, I chose to check the vector of probabilities if any imaginary values were different from 0, and later if the imaginary value's absolute value was greater than $1 \cdot 10^{-15}$, to avoid further complications. If this was not the case, then only the real part was used, otherwise, a message was printed out together with some information. At first this was the probabilities, however, the probabilities still came out with an imaginary value of 0, even though the test had claimed that was not the case.

I later tried a few runs where the function printed out the number of imaginary parts different from 0 as well, and eventually, only the imaginary part in hope of getting more decimals. The result of this was that it appeared that the values giving out an imaginary value were very small, on level 10^{-15} , so they were just large enough to be greater than the set limit and therefore very likely also a result of computational limitations. This raises the question of where such a limit should be set.

Note that after the message stating that the imaginary value was different from 0 was printed, the function continued as usual, and the imaginary parts were discarded. As a result of this warning messages are issued from R, stating that the imaginary values were discarded. However, assuming that these imaginary values are indeed very small, and in any case should have been 0, this should not matter for the results reached.

A.2.3 Scaling of Parameters

By first running the data, it became clear that the **logl** function, which calculates the log likelihood, did not work as planned. This is because it caused very large numbers for the rates, which in turn caused issues when attempting to do calculations with the A matrix. At first, I attempted to scale down the model for model 2b, as the problem did not appear very clearly in model 1. However, it turned out that this did not work. After reading the warnings received when running the first model (that NaN were produced), I chose to run the whole process again, from the first model.

The **optim**-function in R did not take into account what values the different parameters were permitted to take, and this caused another problem. I had not taken this into account while writing the function. This was therefore adjusted, by returning **-Inf** as the log likelihood if any of the parameters were out of bounds (see A.2.5 for further details). Particularly for α the **optim** function did not work with larger starting values (for example 0.3), and it appeared that one of the parameters did not converge, even though the **optim** function claimed that it did. In the end, it turned out that this

last problem could be solved by scaling the parameters so appropriate step lengths were taken for each parameter. The first problem is probably caused by too large values not making sense or causing too large values in the function. This is avoided by making sure the starting values are not too far away from a reasonable value.

Scaling the parameters entails that when the function is optimized, the step sizes for the different parameters are not the same for each parameter, but a step of a reasonable size depending on what scale one expects the result to be. For instance, if one parameter is of size 1000 and another of size 0.001, changing the parameters by the same value of for instance 0.1 when trying to find the maximum, might not give an effect on the log likelihood. This is because it might not change enough for the large parameter, and too much for the small parameter to find a difference in the log likelihood. For scaled parameters it is up to the scaling whether each parameter is changed little or much.

A.2.4 Choice of Method

In order to optimize a method with the **optim** function in R, a method for the optimization had to be chosen. As several parameters were to be optimized, a method which could handle this was necessary. Two of the methods implemented in R which can handle this is the BFGS method and the Nelder-Mead method. Eventually, mostly the BFGS method was used, but also the slower Nelder-Mead method was used in some cases.

Initially I used the Nelder-Mead method in the **optim** function, however, there turned out to be a problem. The Nelder-Mead method did not give the same results for all scalings at least in one case, although mostly the same values were reached. According to McKinnon (1998), the Nelder-Mead method will in some cases not converge to the stationary point one might expect, but instead to a nonstationary point. This was part of the reason why I chose to use the BFGS method. Also the BFGS method sometimes gave results that was in accordance with other runs when the Nelder-Mead method did not.

I chose to run all the runs (different starting values) for model 1 with both the BFGS and the Nelder-Mead method, to ensure that they were in agreement. If the Nelder-Mead method was not in agreement, I chose to see which one agreed with the other runs, and to double check that it was in fact a maximum. For the later models I chose to run the BFGS method as a first choice, and to run the Nelder-Mead method if an issue occurred. The issue could be that a result was not in agreement with the other runs or that the run failed.

For optimizing the simulating data, I chose to use the BFGS method, but to try the Nelder-Mead method if the BFGS method failed. This was to prevent that the entire simulation loop failed if the BFGS method failed. If both the methods failed for a simulated data set, the simulated data set would have been replaced with another, so that the correct number of simulations were calculated. This is potentially an error in the method, if specific data sets lead to failing. However, in the end, for this study, no data sets failed for both the Nelder-Mead and the BFGS method.

When optimizing the simulated data sets for model 4 it also became particularly clear that the BFGS method is a lot faster than the Nelder-Mead method. This was because for model 4 the calculations took some time in the first place, and for cases where the BFGS method failed and the Nelder-Mead method used, it became very clear that the BFGS method was a lot faster.

A.2.5 Limitations on Parameters

The log likelihood function rejects illegal parameter values by returning **-Inf** before any calculations are done, and ending the calculation. The return of minus infinity is reasonable because it is the worst possible log likelihood value a set of parameters can give, and so the parameter set will not be accepted as a maximum as any other returned log likelihood value would be better (or as bad).

As mentioned, the results I used are the BFGS result, as far as possible. However, when attempting to simulate data, it became clear that the BFGS method failed quite often. Because of the fails when running the optimization with the BFGS method, I also chose to only allow values of θ_1 that were greater than or equal to 0, as the method seemed to get warnings when θ_1 was negative. This is reasonable as the assumption is for the rate to decrease with increasing distance. These warnings were issued by the probabilities being **NaN**, which might have been caused by errors because the rates in the A-matrix were of level infinity. In other words, for model 1, if θ_0 or $\theta_1 < 0$, $\alpha \leq 0$ or if β is not in the interval $[0,1]$, the log likelihood came out as **-Inf** if an illegal parameter value is attempted used by the **optim**-function.

Copy from code illustrating settings for running the `optim` function in R for first parameter set for model 1.

In the code section, the function `hessian` is from the `numDeriv`-package in R, and the value `-1` for `fnscale` makes `optim` find the maximum of the log likelihood function `logl`.

```
#Model 1
om11NM = optim(spar1,fn=logl,obsmat=l1$mat,
              method="Nelder-Mead",
              control=list(parscale=spar1,maxit=max, fnscale=-1))
om11NMh = hessian(logl,om11NM$par,
                 method.args=list(eps=om11NM$par*10^(-4)))
om11BF = optim(spar1,fn=logl,obsmat=l1$mat,method="BFGS",
              control=list(parscale=spar1,maxit=max,fnscale=-1))
om11BFh = hessian(logl,om11BF$par,
                 method.args=list(eps=om11BF$par*10^(-4)))
```

A.3 Notes on the Different Models and Simulations

A.3.1 Model 1

The simulated data is simulated using the results from the original data set, second run in table 8.2.1, as the true values. I chose to use the second run in the table above as the basis, because it had the highest log likelihood value. The values of its standard deviation are not the best for all parameters, but they are not the worst for all parameters either. However, the three results displayed in table 8.2.1 are at so similar that which one of them is chosen would probably not have much effect on the final results.

A.3.2 Model 2a

For the simulations, the results for the third run in table 8.3.2 were used as the true values and starting values. Note however again the similarity of the runs-either one would probably give the same result.

Note that during the simulation, there were several instances of imaginary value occurrences, see section A.2.2. A total of 7 warnings were issued because of this. When controlling this, the imaginary parts were very close to zero, that is of order 10^{-17} and smaller. I therefore assumed that they were caused by computational limitations. However, the result did not appear to be affected by it.

A.3.3 Model 2b

The second run in table 8.3.6 was used for the simulations of the data and further calculations. For the collected data, attempting to run the function for larger values of θ_3 , for instance $\theta_3 = 0.01$, $\theta_3 = 1$ or $\theta_3 = 0.01$, did not work. This is probably because too large values are created, and therefore the calculations cannot be carried out (for instance if a value is infinite). As discussed in section 8.3.2, if the value of θ_3 is too large it will when multiplied by possibly thousands of meters cause the rate to get extreme. For lower values the errors did not occur. Too large values of for instance θ_3 can cause very large changes in the rate, and possibly an error when attempting to calculate $e^{-\theta_1 r_{i,j} + \theta_3 (r_{j,h} - r_{i,h})}$ if $\theta_3 (r_{j,h} - r_{i,h})$ is much greater than $-\theta_1 r_{i,j}$.

A.3.4 Model 3

For the simulated data, the first run results in table 8.5.1 are used as the true values, although there probably would not have been much of a difference if any of the other results had been used. At the end of the simulation, one warning occurred. It stated that for $\log(\text{beta} * pm[ctli])$ NaNs were produced, where pm is the vector with the probabilities right before the sampling point, and $ctli$ is the location in question. When investigating this by running a few additional simulations with some printouts from the function, it appeared that the same warning was caused by negative values in the probabilities. However, the values that were printed out were negative values of 0, which does not make sense. My theory is therefore that the values are very small and negative, but for some reason either did not get rounded so it is still registered as negative, or that the prefix for some reason did not change when the value became 0. The small negative numbers are probably a result of numerical limitations in the calculations. I chose to use the simulated data set even though this warning occurred.

A.3.5 Model 4

For model 4, the results in the first run in table 8.7.1 are used for the simulations and further calculations. Several of the runs for model 4 did not converge according to the convergence indicator in the `optim` function. These data sets were hence removed from the set of simulated results before any further calculations were done. The number of data sets optimized with the BFGS method is 341, while for the remaining 84 data sets, the BFGS method failed and the Nelder-Mead method was used. As the computations for model 4 in general took some time as more computations were required, the relatively frequent use of the Nelder-Mead method (see ?? and its use of time (see section A.2.4), further contributed to limit the number of simulations.

A.4 R-code for Log Likelihood Function

This section contains a copy of the code for the log likelihood function. Some comments are added for the benefit of the reader.

```
logl = function(par,obsmat=l1$mat,m=1,hm=0, th2=0,th3=0,
               th4=0,th5=0, sl=slist,datloc=c1,tim=tm){
  if (par[4]>1 || par[4]<=0 || par[3]<0 || par[1]<0 || par[2]<0){
    return(-Inf)
  } #control of parameter values, all models
  th0 = par[1]
  th1 = par[2]
  alpha = par[3]
  beta = par[4]
  if (m==2) {      ##model 2a, control of parameter values, model specific.
    if (par[5]<=0){return(-Inf)}
    th2=par[5]
  } else if (m==3){ #model 2b
    if (par[5]<0) {return(-Inf)}
    th3=par[5]
  } else if (m==4){ #model 3
    if (par[5]<0){return(-Inf)}
    th3=par[5]
    th4=par[6]
  } else if (m==5){ #model 4
    if(par[5]<0){return(-Inf)}
    th3=par[5]
    th4=par[6]
    th5=par[7]
  }

  nlt = ncol(obsmat) #number of sampling points

  prdct= log(1)
  if(m==1){ #for model 1, the same matrix for all individuals.
    A = matrixA1(th0,th1)
    Ae = eigen(A)
    U = Ae$vectors
    Uinv=solve(U)
    D = Ae$values
  }
}
```

```

for (ind in 1:nrow(obsmat)){ # For each individual
  s=0
  pn = rep(0,10) #probabilities, plus(after a sampling point)
  pm = rep(0,10) #probabilities, minus(before a sampling point)
  cll = log(1) #contribution to log likelihood for individual i
  st=match(1,obsmat[ind,]) #finding first observation for individual

  if(m>1 && m<=5 && s==0){#other models than model 1,
    #only once for each individual
    A=matrixA1(th0=th0,th1=th1,th2=th2,th3=th3,th4=th4,
              m=m,hm=hm[ind],sl=sl[ind])
    #more complicated models require
    Ae=eigen(A) #a separate A matrix for
    U = Ae$vectors #each individual
    Uinv = solve(U)
    D= Ae$values
    s=1
  }
  pn[datloc[st,2]]=1 #first observation of individual
  for (i in (st+1):nlt){#For each time/location pair,
    #after individual enters study.
    if (i>nlt) {#otherwise the previous observation is in
      #the last sampling time/place,
      break() #to prevent possible errors if st=last observation
    }
    if(m==5 && i>=91 && s<2) { #for model 4, only calculated
      #once after datechange
      A=matrixA1(th0=th0,th1=th1,th2=th2,th3=th3,
                th4=th4,th5=th5,m=m,hm=hm[ind],
                sl=sl[ind],spl=i)
      Ae=eigen(A)
      U=Ae$vectors
      Uinv=solve(U)
      D=Ae$values
      s=2
    }
    tdf=tim[i-1,i] #Time difference between two sampling times

    if(tdf==0){
      pm=pn #To avoid using A2
    } else {

```

```

A2=diag(exp(D*tdf)) #

pm =U%*%A2%*%Uinv%*%pn*exp((-1)*alpha*tdf)

if (sum(abs(Im(pm))>(1*10^(-15)))>0){ #Checking for
                                     #larger imaginary values
  print('Imaginary value error')
  print(sum(abs(Im(pm))>(1*10^(-15))))
  print(t(round(Im(pm),digits=16)))
} else {
pm=Re(pm)
}

pm=round(pm,digits=15)
if(sum(pm<0)>=1){print(pm)} #Check if any elements are negative
}
ctli = datloc[i,2] #current location
if (obsmat[ind,i]==1) {#if individual is observed
  pn=rep(0,10)
  pn[ctli]=1

  cll=cll+log(beta*pm[ctli])

} else { #individual not observed
  pn=pm

  pn[ctli]=(1-beta)*pm[ctli] #probability that individual is not observed
  pn=pn/(1-beta*pm[ctli])
  cll = cll+log(1-beta*pm[ctli]) #Contribution to log likelihood
}

} #end i-loop
prdct=prdct+(cll)#adds in each individual's contribution
} #end individual calculation

return(prdct)
}

```

Appendix B Calculations

B.1 The Haversine Formula

The coordinates used in the formula are measured in radians, and can easily be obtained from the longitude and latitude by $c_r = 2 * \pi(\frac{c_d}{360})$, where c_r is the coordinate in radians, and c_d in degrees (longitude or latitude).

In the Haversine formula below, the abbreviation la means latitude and lo means longitude. The index 1 indicates the first point, and 2 indicates the second point, so the output is the distance in meters between points 1 and 2. The radius is that of the earth and in this study I have used $R = 6341000$ (meters), as stated on Movable Type Ltd. (2014).

$$\begin{aligned} s_1 &= \left(\sin\left(\frac{c_{2,la} - c_{1,la}}{2}\right)\right)^2 \\ &\quad + \cos(c_{1,la}) \cos(c_{2,la}) \left(\sin\left(\frac{c_{2,lo} - c_{1,lo}}{2}\right)\right)^2 \\ s_2 &= 2 \arctan(\sqrt{s_1}, \sqrt{1 - s_1}) \\ \text{distance}_{1,2} &= s_2 \cdot \text{radius} \end{aligned}$$

Appendix C A Note on R

For the calculations in this thesis, I have used the programming language R. I have mainly used this through the program R-Studio, which allows you to program in R. The version of RStudio I have used is 0.98.501, and the version of R has varied through the process. I would also particularly like to mention that I have used the **numDeriv** package in R, and to thank contributors to R for making many of the functions I have used available.