MSc thesis: Analyzing GPS data of inland breeding Eurasian Oystercatchers using an Interactive Approach to Clustering

Marije Faas

First version, January 10, 2023 – Current version: August 16, 2023

Abstract

The Eurasian Oystercatcher is one of the most well-studied bird species and was the first species to be equipped with the GPS trackers from the University of Amsterdam Bird Tracking System. The available GPS data has been used to intensively study the oystercatchers that live and breed in the Wadden Sea in The Netherlands, but not as much research has been conducted on the individuals that move inland during the breeding season and nest there. In this thesis, a new interactive clustering approach was designed and implemented to study these individuals using various analysis methods that can be applied to the created clusters. Fourteen adults from a large collection of GPS data from oystercatchers caught in Vlieland were selected. The analysis was aimed at gaining a better understanding of the different locations the birds spend time at during their time inland as well as their nesting period. When oystercatchers move inland, they first aggregate at traditional locations next to lakes or other bodies of water referred to as 'clubs'. The research aimed to identify and further investigate possible clubs and how they were used throughout the season, as well as throughout a 24-hour period. Some interesting behaviour was identified, including birds who continued to spend time at the water away from their territory even in the middle of the breeding system when they were nesting. However, the majority only did so during the start and end of the season which is more in line with the expectations. The observed group of birds showed rather variable behaviour regarding the time of day during which they spend time in clubs. Nesting site analysis based on density suggests that eleven out of fourteen individuals were able to hatch chicks, nesting in fields or on rooftops, and was able to retrieve almost the exact nesting period of around 30 days for each bird. Overall, the completed work consists of a user-guided analysis that provides both detailed and easy to read results for GPS data that can be used for future research into oystercatchers as well as other GPS data in which various distinct locations are visited. In addition, the analysis that was performed in this thesis was able to identify interesting behaviour of specific individuals as well as give an overview of all studied birds in their use of clubs and nesting sites. As a result, the developed tool proved fully functional while providing a solid base for future additions and improvements that could provide even more information as well as smooth out various processes in the tool.

Contents

Chapter 1

Introduction

The Eurasian oystercatcher is a well-studied shorebird that can be found throughout Europe and northern Africa. They generally live in coastal areas, though they can also be found locally inland. Their numbers have declined strongly in the last few decades, leading to concerns about the populations in The Netherlands [16]. Since 2008, research into oystercatchers in The Netherlands intensified and in this year the first trials using the GPS trackers from the University of Amsterdam Bird Tracking System (UvA-BiTS)[3] were completed. Together with gulls, oystercatchers were the first species with which this was tried [14]. In the years after, several more studies using these GPS trackers were performed on oystercatchers caught on various islands in the Dutch Wadden Sea. The data from all these studies has been published as open data under the Creative Commons Zero license [8]. Since then, many studies have been completed investigating oystercatchers that live and breed in the Wadden Sea (for example, [9], [2]), where behaviour such as locations and timing of roosting, breeding and foraging are well described and most research is focused on the impact of disturbance on the birds. However, many oystercatchers move inland during the spring to breed there and these have not been studied as much as those that breed on the islands. Moving inland means their daily rhythm is no longer dependent on the changing of the tide, which may influence where they spend time throughout the day. In addition, they now have access to different environments for nesting, such as rooftops. As not much is known about the breeding period inland yet, a good first step would be to evaluate where they spend their time throughout the season as well as throughout the day. Specifically, oystercatchers are known to aggregate and form groups at the edge of water, called clubs, which would be interesting to investigate. Next, finding the location of their nest can give valuable information on nesting behaviour and allow for comparing birds that nest in fields to birds that nest on rooftops. This information can also be beneficial for conservation efforts to ensure oystercatchers have places to nest, reducing further decline. The large majority of the Dutch oystercatcher, an estimated 75 percent of the population, breeds inland [6].

The goal of this study is to find out how these inland-breeding birds make use of their home range inland, from shortly after their migration until the end of their stay. This includes the extend in which they still have a connection to the inter-tidal mud flats, the use of clubs, their territory and their nesting behaviour. To do so, different methods have been developed. Through use of an advanced clustering setup, using multiple algorithms, user interaction and domain-based rules, data at the inland home range will be clustered and analyzed to investigate where the birds spend their time throughout the day and throughout the season.

1.1 Research questions

In order to learn more about the inland-breeding birds, several research questions have been defined which will be answered throughout this thesis. There are two main topics that these questions are divided under, namely, 1) the time oystercatchers spend close to water and 2) what we can learn about nesting.

1. When and where do the birds spend time at the edge of water bodies?
    (a) During which part(s) of the season do the birds spend time at the water?
    (b) When are these locations visited throughout a twenty-four hour time period?
2. What can we learn about the birds' nesting behaviour?
    (a) Where is their nesting site?
    (b) When do they nest and for how long?
    (c) How many birds appear to have successfully hatched chicks?

1.2 Eurasian oystercatchers

Nowadays, oystercatchers can be found breeding in most environments: along the coast on saltmarshes, dunes and dikes and inland in cities, on agricultural land and protected meadow areas, but they do not occur in forests. During their stay in the Wadden Sea, they mostly forage on shellfish, but also take worms and crabs. When foraging in grasslands, they will go for earthworms or insects, mainly tipulids. While many oystercatchers winter in the Wadden Sea and the Delta, some birds migrate further south to South-West Europe or Africa. During prolonged freezing weather, there can be casualties under the birds that winter on the Wadden Islands [13].

During the breeding season, only a minority breeds near the intertidal areas in the Wadden Sea, with the majority migrating inland where they breed in fields, meadows or urban areas [15]. Oystercatchers start breeding relatively late, at 3-8 years of age and breed from half of April to the end of June, laying 3-4 eggs. The incubation time for the eggs is close to thirty days, and another 28 days after hatching the chicks are able to fly. They are faithful to their breeding site, meaning most will return to the same location every year with the same partner [12]. The pair jointly defends the territory and both parents take part in the parental duties such as incubating, chasing off predators and feeding the chicks [15]. Inland breeding birds arrive in their breeding territory in February or March [13], which they will leave soon after the breeding season in June to August [12]. Based on visual inspection of the Vlieland dataset used in this project, all birds have moved back to the islands by the end of July and some earlier. During the breeding season, oystercatchers are very territorial and will defend their territory from other oystercatchers.

Interestingly, oystercatchers that move inland during the spring do not always move straight to their breeding territory. Early in the breeding season, around March, they are seen forming groups with other oystercatchers, usually on the edge of water, called clubs. Here, they relax, sleep and sometimes display aggressive behaviours. After this period, they will move to their breeding territory to lay and protect their eggs. It is possible that parents with their young may briefly visit the location again before they migrate back to the coast, which would be somewhere in July for most individuals [15]. While the Dutch population of oystercatchers initially increased, since the late 20^{th} century their numbers have been declining [8]. Many studies have looked into potential causes for this decline, which involve both the breeding season and the wintering season. During the winter, birds have to deal with a lack of food due to over-fishing in the Wadden Sea. During the spring, intense use of agricultural land means the birds are not raising enough chicks succesfully [13]. Various studies have also looked into other potential causes, such as disturbances caused by tourism [11] and aircrafts [10].

1.3 Data

The dataset used for this research is the GPS tracking data of Eurasian oystercatchers (Haematopus ostralegus) from the Netherlands and Belgium collected by van der Kolk et al [8]. This is a collection of six datasets, of which the O_Vlieland dataset is the largest with almost 5 million GPS records. This dataset was chosen for this research, due to the large number of individuals that were tracked with a number of them moving inland during spring. The O_Vlieland dataset includes biometrics of the tracked birds such as bill length, shape and width, and further contains behaviour tags that were applied using a random forest model on the accelerometer data. The five classified behaviours are as follows: flying, walking, foraging, preening and inactive. Typical breeding behaviours are not distinguished because the training data was acquired during the non-breeding season; therefore, incubating behaviour is classified as inactive behaviour [8]. For the research, each bird will be studied from the moment they arrive in their home range inland from the moment they migrate back to the islands. After visual investigation of the tracks, the period from February to July will be used as a start, as this period covers the migrations from and back to the islands for all birds. However, only time spent at a predefined (based on GPS tracks) home range inland will be considered for the individual birds.

1.3.1 Data collection

Data collected by the trackers was stored in internal memory, and transmitted remotely to a base station, sometimes via in-between relay stations. These stations were set up around nesting sites and high tide roosts. For birds that were outside of the station network, mobile base stations were sometimes used to download the data. If birds left areas with stations and did not return, the data was not downloaded except through mobile base stations or when trackers were retrieved from dead birds.

The trackers were solar powered, meaning that during winter, they were often drained which paused data collection leading to many birds having no (or very little) data from November to January. Frequency of GPS fixes could be changed whenever the tracker connected to a base station. Generally, more data was collected when birds resided within the area covered by receiving stations and the battery was fully charged.

1.3.2 Data exploration

Initial exploration of the Vlieland data has located fourteen individuals that move inland at some point during their tracking period. These individuals are all adults - juveniles and subadults have not been included, as they do not breed at this age. There is one exception, as there was one bird caught as a subadult who was known to be nesting in the springs after being caught. Eight of the birds appear to move to agricultural environments while four move to an urban area, and two other birds spend time in both. Note that the birds that have their territory in agricultural land are mostly found next to a town and may also spend time there. Nesting site analysis will have to determine where exactly these birds were nesting. There are five males and nine females in the dataset.

The selected individuals have been checked to ensure their data is not lacking large periods of time during spring, by looking at the number of days that have at least one data point. By far, most months had data for each day, indicating that there appears to be plenty of data available. This makes sense, as the solar-powered batteries receive plenty of sunlight in this period.

Chapter 2

Literature Review

2.1 Oystercatchers

When clubs are discussed in current literature, this is often regarding younger, non-breeding oystercatchers. During the first years when the birds do not breed yet, they are known to spend time in clubs together with other oystercatchers during the Spring [6]. Generally, they are located at traditional spots on the edge of lakes or canals. In addition, oystercatchers that do breed inland are also seen sleeping in these clubs at the beginning and end of the breeding season. There have been quite a few efforts to count the number of oystercatchers at clubs and sleeping places. Note that the distinction between the two is not always clear [6], as the location used for sleeping at the start and end of the season by breeding oystercatchers can be the same as where the non-breeding oystercatchers remain throughout the season. Either way, both show a clear patterns throughout the breeding season regarding the number of birds counted: the arrival of the first birds around the beginning of March, followed by a peak at the end of March, low numbers throughout the summer (May - July) and a slight increase in July, likely from breeding oystercatchers returning to the club. Further research, based on the reading of ringed oystercatchers, suggests that the sleeping areas are largely used by local breeding birds. They appear to be faithful to a specific sleeping location that's within 4.5km and 7.0km from their breeding territory [6].

At the start of the breeding season, in February or March [13], breeding male and female oystercatchers return to their territory, which they will jointly defend against intruders. The timing of breeding can vary between pairs, possibly differing by over 6 weeks [16]. The quality of the habitat can vary greatly, referred to as high quality (HQ) or low quality (LQ) territories [7]. The key difference is that in HQ territories, chicks can follow the parents from the nesting area to the feeding territory which is directly adjacent to it, making it relatively easy to feed the chicks. In LQ territories, the parents have to transport food individually to the chicks due to the nesting and feeding area being separated from each other. This continues until the chicks are able to fly and thus join the parents. As a result, parents in LQ territories raise fewer chicks than those in HQ territories due to them being unable to feed the chicks sufficiently [7].

Oystercatchers are seen breeding in a variety of open habitats like saltmarshes, agricultural fields and grasslands, but they can also nest in residential areas on flat rooftops [16]. When nesting on rooftops, the rooftops are usually covered in gravel and other small stones or pebbles. The nest then consists of a small spot from which the gravel was removed. While there appear to be benefits to nesting on rooftops, such as the lack of ground predators, it comes with risks as well. Chicks can fall off rooftops, leading to injury and vulnerability to ground predators, or death. On the roof, they can also be vulnerable to predators such as crows [5]. Overall, though, if there was a convenient food supply nearby, studies show that the performance of roof nesters was quite good compared to birds nesting on the ground [5].

Interestingly, various research has shown that the oystercatcher populations during the breeding season often contain a high number of adult birds that are not breeding, despite being old enough to do so. It can take many years before an adult oystercatcher breeds for the first time. In the meantime, they spend their time during the breeding season by intruding on territories, feeding in undefended areas and attending clubs. They may also choose to group up with other nonbreeders and feed in defended territories [7]. It is assumed that nonbreeders have a trade-off to make, as they can either settle for a LQ territory relatively quickly, or wait a possibly very long time to obtain a HQ territory, but do not seem to be able to dedicate themselves to more than one territory. Studies suggest that nonbreeding oystercatchers can tell the difference between HQ and LQ territories, as intrusion rates seen in HQ territories are considerably higher than those of LQ territories. Either way, some studies suggest that they need to commit to one site, in order to become familiar with the current owners and possible neighbors, which provides them with an advantage over other nonbreeders that are less familiar with the site [7]. Sometimes, nonbreeders successfully take over an existing territory resulting in the original breeders having to settle for a different territory, which on average takes them 2.4 years. The new territory is seen to be located close to the original one [7], suggesting that they will return to the same general area until they successfully resettle.

As mentioned before, oystercatchers are known to share the care for the eggs by alternating nesting 'duties', such that at any given time, one parent is responsible for the nest and the other parent is responsible for defending the territory [7]. Research has been done into what has been called 'incubation bouts', the time that one parent is responsible for the nest, and found that the incubation bout length can vary greatly both between and within species [4]. This goes even for pairs of the same species that breed in the same area. Between the observed species, incubation bout length varied greatly from one bout lasting half a day to bouts lasting only a couple hours. Besides length, the period of the incubation rhythms was determined which also varied between species from a strict 24 hour period to ultradian (shorter than a day) and infradian (longer than a day) rhythms. One possible ecological factor that could explain the observed variation in bout length proposed by this research is the relation to anti-predation strategies. This suggests that species where parents rely on avoiding detection (parental crypsis) would benefit from having less activity near the nest, as this can give away the nest location to predators. As a result, these species may prefer longer incubation bouts which leads to less activity at the nest as the parents change duties. These species tend to remain on the nest when approached by a possible predator until it is nearly on top of them. On the other hand, species that rely on actively defending against predators through keeping watch, attacking or distracting, leave the nest long before the predator is near. For this behaviour, there is no real benefit to minimizing activity at the nesting site, meaning their bout lengths can be shorter without creating a disadvantage. This theory was tested by observing the distance at which the parent left the nest when approached by a human, called the escape distance. The experiment showed that the escape distance correlated negatively with the length of incubation bouts, meaning species with a short escape distance were observed to have long incubation bouts and vice versa [4]. For oystercatchers, known to stand watch and actively defend their territory [7], this would suggest they may afford shorter incubation bouts. The previously mentioned research included several oystercatchers and observed a median bout length of around 2 hours, which is indeed short compared to the other observed species, placing them as the 8th shortest bout length out of 32 species observed [4].

2.2 Interactive Clustering

Clustering is an unsupervised machine learning task commonly used to find natural structures in unlabeled data sets by automatically grouping the data into clusters. As an unsupervised technique, there are no labels to tell the algorithm whether or not the clustering is 'correct', which can complicate their use compared to classification tasks. Standard clustering algorithms cannot utilize domain-specific or user-specific inputs which can greatly limit their use in cases where these kinds of inputs are essential. Several reasons for this include the fact that clustering is an intrinsically subjective task because it depends on the specific needs and goals that the user has in mind [1]. As standard clustering algorithms do not have access to these specifics, the results could be far off from what the user is looking for. To solve these issues, several interactive clustering methods have been explored, where the user and system interact with each other to complete the task in a way that suits the users needs.

A survey has been conducted on interactive clustering approaches [1], grouping existing approaches based on three criteria: 1) at which stage the interaction is happening, 2) which interactive operations are involved for both the user and machine and 3) how the user's feedback is incorporated to improve the clustering model.

The two most common approaches to user interaction differ in what kind of input is given by the user. The most advanced approach requires the algorithm to do most of the work, meaning the user identifies mistakes and imperfections in the given results which provides the algorithm with hints about the user's preferences. This feedback is then used to figure out a better optimized clustering that takes the information given by the user into account. The cycle of a provided solution, followed by feedback from the user, resulting in a new solution, can be repeated many times until the user is satisfied. The second approach differs in the fact that the user provides more direct information on which parameters must be changed in order to get the results that are desired, whereas in the first, the algorithm has to figure out how the given information would translate to different parameters. Here, once the initial clustering is performed, the user can re-run the clustering with different parameters, such as the number of clusters or the similarity threshold parameters. In many existing interactive clustering methods, it is possible to interact with both the model (through changing parameters) and the results, for example by deleting, merging, moving or reclustering the results. Note that, while the survey describes how most papers in the interactive clustering field work in this manner - a user-guided process - the opposite can also be true, where the algorithm takes initiative, for example by presenting the user with suggestions and possible changes with the user accepting or declining the proposed change [1].

Throughout existing interactive clustering tools there have been many possible interactive operations. This includes a common method to allow the user to split and merge clusters, often directly through an interface such as dragging and dropping operations to edit the clusters. In some cases, tools start out with a large amount of clusters and only allow the user to remove or merge clusters. This is often done through a visualization, most commonly a 2D scatter plot, though other options such as adjacency matrices and treemaps also exist. Another interaction that many existing applications use is the ability to add clusters, through changing a 'number of clusters' parameter, splitting or re-grouping clusters. Some existing applications allow the user to select which data to cluster, allowing for local changes to generated clusters [1].

Lastly, the user's feedback has to be incorporated in some way. The survey describes several strategies that are used in current literature, such as directly modifying the cluster's structure through split and merge commands or by changing the number of clusters. A different method is by utilizing the feedback to adjust distance matrices, similarity functions or other parameters of the algorithm. Lastly, the feedback can be used to constrain the clustering algorithm, usually in terms of 'must-link' or 'cannot-link' constraints between pairs of data points [1].

According to the survey, evaluating interactive clustering methods remains an open challenge but many different quality measures have been utilized by existing research. A distinction is made between objective and subjective measures, with objective measures being further divided between unsupervised and supervised methods. Examples of unsupervised objective measures are cluster cohesion, cluster stability and cluster separation, while supervised measures could be accuracy, recall or F-score. Subjective measures revolve around user satisfaction and joint performance between the user and machine [1].

Chapter 3

Methodology

3.1 Data pre-processing

Pre-processing the data starts with defining an inland home range for each bird and then using it to exclude data outside of this home range. If birds have data for multiple years, these are separated as well, and each year will be treated as a separate dataset from here on. While removing unnecessary data, it is noted if, when and for how long the bird leaves this defined area. This information could be interesting if the bird regularly leaves the area, moving far away from its territory and possibly even back to the islands. In addition, datapoints that are marked as outlier either manually or automatically are excluded from the data.

After defining the home range for each bird, which is done manually based on Google Earth tracks of the bird's data, data outside of it is excluded and separated per year, if applicable. This lead to a total of 24 separate breeding seasons or datasets. As the Vlieland dataset contains accelerometer data which was used to create behaviour tags, this data is also added. About half the datasets are complete with behaviour tags, the others are missing all or a number of tags. When tags are available, sometimes one entry in the GPS data has multiple behaviour tags attached. This can lead to an increase in data in the dataset due to points being duplicated for each associated behaviour. To mitigate this, during analysis, double entries for the same timestamp are excluded. Lastly, one additional attribute is created for each dataset, namely the daytime attribute. This attribute indicates whether the corresponding timestamp sits between sunrise and sunset, meaning it is day (daytime = True) or not, in which case it is night (daytime = False).

For some individuals, the area that should be defined as the inland home range is a bit ambiguous. This is the case when the bird spends a significant amount of time (based on the number of tracks) in multiple locations inland, which are far apart from each other. In these cases, first an area is defined that included all these locations and the corresponding dataset was created. Then, based on the scatterplot of this data as well as a basic analysis to determine where the bird spends its time during the summer, a new area is defined which zooms in on the location of the territory. This then creates a second dataset which is more suitable for analyzing the territory and home range for the purpose of this research.

3.2 Clustering

3.2.1 Goal

The goal of the clustering method is to geographically separate the GPS data of a bird's home range inland during the breeding season and label it accordingly. When each data point is assigned a cluster label, it becomes quick and easy to automatically analyze when the bird spends its time in which clusters.

3.2.2 Clustering algorithms

The developed interactive clustering approach uses two different clustering algorithms, namely, agglomerative clustering and DBSCAN. The main prerequisite is that the method must work without specifying the number of clusters beforehand, as each bird will have a different number of locations that it visits and we don't wish to figure out exactly how many clusters we want before running any algorithms. Instead, the method should present a decent first 'guess' and then allow changes to be made from there. To avoid overcomplicating the process and using many different clustering algorithms that all require different parameters to be tested and optimized, two algorithms were chosen which will be described in more detail below. To ensure robustness to different kinds of GPS data that will vary in shape, density and number of desired clusters, a range of parameter combinations are tried whenever a set of data is being clustered and the best clustering is chosen out of the different results according to a general score and a set of 'best clustering' update rules.

Agglomerative clustering algorithm

The agglomerative clustering algorithm works in a bottom-up manner, by starting out with each data point being its own cluster. At each iteration of the algorithm, the two most similar clusters are merged together. This process can continue until there is only one cluster left, or use a distance threshold to determine when to stop merging two clusters. When this leads to no more clusters being allowed to merge, the algorithm is also finished. There are various way to compute the similarity between clusters, of which using the Euclidean distance is a commonly used method and often the default option. It is also a simple and intuitive method, which is why it was chosen during this project. Next, the linkage criterion takes these distances and uses them to determine which two clusters should be merged. In the used setup, the linkage criterion uses a distance threshold, meaning that if the distance is at or above the threshold, two clusters will not be merged. Five evenly spaced distance values are dynamically determined based on the data that is being clustered. The step size between the values is calculated using $(step\_size = max\_dist - median\_dist)/7$, which is then used to calculate the distances in a loop using $dist = dist + step$ which is run $x = 5$ times, with $dist = min_{dist}$ at $x = 1$. These values cover a range of different clusterings from many small clusters to a few larger clusters, while avoiding values too close to the maximum distance, as this would lead to too many clusterings with only a single cluster.

There are four different linkage criteria, those being complete, average, single and ward, which determine what distance to use when comparing two clusters. Complete uses the maximum distance between all the observations in the two clusters, while single uses the minimum distance and average uses the average distance of each observation between both clusters. Intuitively, this means that using the complete linkage criterion will result in the least amount of clusters, as the linkage distance between two clusters will be highest with this criterion. Single results in the largest number of clusters, and average will sit somewhere in between. On the other hand, ward aims to minimize the variance of the two clusters.

For the approach in this project, the complete, average and single linkage criteria are used, each paired with five possible distance threshold values. Ward is left out, as the use of variance rather than raw distance values would require it to be paired with different distance thresholds. The complete, average and single linkage criteria, combined with five distance thresholds, also ensure an adequate variety in the resulting clusters to account for different types of data without needing to spend more computation time.

DBSCAN algorithm

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clusters, as the name suggests, based on density. It starts by locating core points with a high density and expands clusters from there. Two variables play an important role: epsilon and min_points. Epsilon is a distance value which specifies the radius around a point within which other points are considered to be in the vicinity, which is used to calculate the density of the point. Or, in other words, if the distance between two points is equal to or lower than epsilon, they are considered neighbours. Min_points specifies the number of points that should be in this vicinity in order for a point to be considered a core point. A point p is *directly reachable* from a core point q if it is within the epsilon distance from q. A point p is *reachable* from a core point q if there is a path between the two points, where each point along the path is directly reachable from the previous point in the path. As a point p can only be directly reachable from another point q if q is a core point, this means that all points along the path must be core points, except for possibly the final point. A cluster is formed from any core point p along with all points that are reachable from p. Non-core points, which don't reach the minimum number of neighbours, form the edge of the cluster.

In this project, min_points is set to 1, meaning that every point starts out as a core point and there are no non-core points. While this is unusual for this algorithm, it is done to avoid points being labeled as noise, which is the case when a point has no neighbours at all. Points labeled as noise are given a label of -1, so that they can be identified and removed. However, in the used setup, which does not aim to remove these points as they may actually be relevant points, it would consider all 'outliers' to be in the same cluster, causing weird results. This does change the working of the DBSCAN algorithm slightly, but in the context of this project, this did not hamper its results.

Epsilon is the most important parameter of DBSCAN, with a higher value creating larger clusters whereas a lower value creates smaller clusters. Normally, a lower value would also create more outliers, but as the min_points parameter is set to 1, it would instead create many clusters of size 1, if there are areas with very low density data. Like agglomerative clustering, five evenly spaced distance values are dynamically calculated from the data. The stepsize is calculated with $step = (max_{dist} - min_{dist})/15$, where max_dist and min_dist are calculated from the data to be clustered. Then, the five distance values to be used are calculated in a loop using $dist = dist + step$ which is run $x = 5$ times, with $dist = min_{dist}$ at $x = 1$. This setup evenly covers the lower third of the total range in distance within the data, which allows for variety while sticking to lower epsilon values, avoiding values that would create too few clusters.

3.2.3 Method

Interactivity

Due to the geographical nature of the data, standard clustering approaches are not sufficient because the optimal clusters cannot be determined solely based on the standard ways to score clusterings such as density of the clusters or distance between clusters. The information we want to learn from the data will influence what are considered 'good clusters', meaning that the 'best' clustering in the context of the analysis can vary even within the exact same data set, depending on the questions that are asked. This makes it practically impossible to achieve useful clusterings for the analysis based on only on results provided by standard clustering algorithms.

To solve this problem, an interactive clustering approach was designed and implemented. An initial clustering is done on the data using agglomerative clustering, after which the user is able to make changes.

It is possible to recluster one or more clusters or merge multiple clusters together. By default, reclustering applies the DBSCAN clustering algorithm to the data of the selected clusters. DBSCAN is faster than agglomerative clustering with the setup that was chosen, making it more suitable for reclustering, which may have to be done a few times. In addition, DBSCAN bases its clusters on differences in density, making its results more predictable and intuitive compared to agglomerative clustering, which can produce more variable results due to the different linkage types that it can use. However, it is possible to recluster using agglomerative clustering instead, if DBSCAN does not produce adequate results, such as creating too few clusters. Regardless of which algorithm is used, it is not possible to choose which clusters will be created due to the nature of clustering techniques. However, various update rules are implemented which together decide whether a new clustering will update the current best and consequently, which clustering is presented to the user after each parameter combination has been tried. These rules were designed to make the interactive clustering process as flexible and smooth as possible.

Selecting the best clustering

In order to start comparing different clusterings, a simple score needs to be assigned to each, which gives a general indication of how good the clustering is. For this, the silhouette score is used. This score ranges from -1 to 1, where 1 would be the optimal score, indicating that the means of all clusters are properly distanced from each other. Normally, this score would give a good indication of the quality of the clustering. However, as mentioned earlier, the nature of this GPS data means that there are more relevant factors which determine the quality of the clustering. Therefore, instead of choosing the clustering with simply the highest score, various rules are applied to determine if the current best will be updated or not. These rules were designed based on four aspects: domain knowledge, convenience during interactive clustering, experimentation and the need to not over-complicate the process. Below is the list of rules that is applied when evaluating a clustering, which will be further explained in the rest of the section.

1. Clusterings created by the agglomerative clustering algorithm that only have two clusters receive a 15% penalty on their silhouette score before evaluating.
2. A new clustering that has a lower number of clusters and a significantly higher score than the current best, will overwrite the current best.
3. A new clustering which has a higher number of clusters and a higher score than the current best, will overwrite the current best.
4. If the current best has more than ten clusters and a new clustering has a higher score than the current best, the new clustering will overwrite the current best.

1. Clusterings created by the agglomerative clustering algorithm that only have two clusters receive a 15% penalty on their silhouette score before evaluating. During the initial clustering with agglomerative clustering, the data of the full home range is being clustered. Based on what is known of oystercatchers, it is highly unlikely that a bird will only visit two defined locations during the whole season, as there will usually be at least a territory, one or more different clubs and one or more trips outside of the territory. As such, an initial clustering with only two clusters will almost always be incorrect, leading to a heavy penalty being applied when this is the case, automatically favouring all clusterings that have three or more clusters.

2. A new clustering that has a lower number of clusters and a significantly higher score than the current best, will overwrite the current best. In the interactive clustering process, merging clusters gives a perfectly predictable result, as it is a simple matter of assigning all selected clusters to the same cluster. On the other hand, reclustering will always have some variability, some aspect of uncertainty, regarding which clusters will be created. As a result, merging clusters is preferred over reclustering, which leads to a preference for clusterings with more clusters. In order to accept a clustering that reduces the number of clusters we have, the silhouette score assigned to it should not only be better than the current best, but significantly better. This ensures that while a number of clusters will be lost, there is more certainty that it will lead to a preferred clustering.

To determine how significant the improvement in silhouette score should be, a total of 250 experimental clusterings was run on various datasets selected for this project, in order to get some statistics on the distribution of silhouette scores. Note that each parameter combination counts as a a separate experimental clustering, so running agglomerative clustering on a dataset once accounts for fifteen different clusterings. In addition, a mix of initial clusterings using agglomerative clustering and reclusterings on one or more clusters using DBSCAN were performed. This setup ensured that the results are robust for many differently shaped datasets and for each part of the interactive clustering process. For each of the 250 clusterings, the number of clusters (cluster count) and silhouette score were saved and used to calculate the minimum, maximum and average scores that were obtained for each cluster count. Next, a standard deviation of 0.127 over all 250 silhouette scores was obtained from this data. This number held even when considering only the most common cluster counts, as these together have a standard deviation of 0.121, only slightly lower. With the minimum silhouette score sitting around just over 0.50 and the maximum around 0.95 for the most common cluster counts, 0.127 is a reasonable threshold to use. This means that if a new clustering has less clusters than the current best, its silhouette score should be at least 0.127 points higher than the current best, in order to overwrite it. As a result, even if it has less clusters, a good clustering will still be able to overwrite a current best. On the other hand, a decent clustering with more clusters that could work very well for this data, will not be overwritten as easily as before.

3. A new clustering which has a higher number of clusters and a higher score than the current best, will overwrite the current best. Since a higher cluster count is easier to work with during interactive clustering, there is no reason to not simply accept a clustering with more clusters and a higher score than the current best.

4. If the current best has more than ten clusters and a new clustering has a higher score than the current best, the new clustering will overwrite the current best. This rule counters an issue where it could be too difficult to overwrite a current best that has more clusters than ideal. While there is a preference towards many clusters over few clusters, having to merge excessive numbers of clusters hinders the interactive clustering process, slowing it down. For this reason, if the current best has more than ten clusters, the threshold explained in rule 3 falls away. A clustering with lower cluster count no longer has to have a significantly better score, as it is actually preferred to lower the cluster count to below 10.

3.3 Cluster Analysis

This section will describe the various types of analysis performed on the clustered datasets and the plots created during the process. Each analysis is designed with a specific goal in mind regarding what kind of questions it can answer about the data. In order to give a complete explanation of the qualities of each analysis, a short description of this goal is provided, before going into the methods used.

3.3.1 Proportion analysis

Goal

The goal when performing proportion analysis is to provide a broad overview of where the bird spends their time throughout a longer period, in this instance the entire breeding season from the moment the bird arrives at their home range inland until they leave. This can be used to, at a glance, determine during which parts of the season the bird spends time at locations such as clubs. This information could already give new insights on its own, but can also be used to provide a reference for more specific analysis, such as the interval analysis, by providing an overview of interesting periods to examine further.

Methods

The proportion analysis is computed based on the time the bird spends in each cluster on a given day. This is done by calculating how much time passes each time it visits a cluster, when moving through the data in chronological order. The first data point that is assigned to cluster A serves as the start time, after which we move through the data until a cluster B is encountered. The last point to be assigned to cluster A serves as the end time, and the time spend in cluster A is calculated as the difference between the start and end times. When the stay in a cluster consists of only one datapoint, the start and end time for this stay are the same. In this case, a duration of 10 minutes is assigned (the minimum time between two points in this dataset). Without this change the stay at the cluster would be ignored entirely. At the end of a day, for each cluster, all durations of stays in the cluster are summed up which form the total time spend in this cluster ($duration_{cluster}$). In addition, all calculated durations across every cluster are summed up to calculate the total duration of the day ($duration_{day}$. Lastly the proportion is calculated for each cluster as $\frac{duration_{cluster}}{duration_{day}} * 100\%$ and plotted for each day. This method is the most robust to missing or infrequent data, as it does not make any assumptions about where the bird is when moving between clusters, something that could skew the results. There are limitations, for example, even with high frequency data points and no missing data, $duration_{day}$ will never be 24 hours due to the time between one cluster's end time and the next clusters start time not being considered as we do not know where the bird was during this time. In addition, wen the stay in a cluster consists of only one datapoint, the start and end time for this stay are the same. In this case, a duration of 10 minutes is assigned, but this is simply an estimate as is not certain how long the bird was actually there.

3.3.2 Interval analysis

Goal

The main purpose of interval analysis is to more closely observe how the bird spends its time throughout the day at 1 hour intervals. Where proportion analysis can give a broad overview of where the bird spends time throughout the season, interval analysis shows where the bird spends time throughout the day. This can answer questions such as, during which part(s) of the day is the bird in a club?

Methods

Interval analysis identifies at each whole hour in which cluster the bird currently is. Because the GPS points don't necessarily happen at the whole hour, a range of fifteen minutes before and fifteen minutes after the whole hour is tolerated. If there are multiple points within this range, the one that is closest to the whole hour is used.

3.3.3 Nesting site analysis

Goal

The nesting site analysis is targeted at a specific location within the assumed territory of the bird, the nesting site. After finding this location, it can be used to estimate how long and during which weeks the bird spends time on its presumed nest. This information can give an idea of how likely it is that the bird was successful in hatching the eggs, or whether the nesting site may have been disturbed.

Methods

The nesting site analysis relies on density methods to determine the peak where the bird has spend most of the time within the territory. The assumption is that when a bird is nesting, it will generate a huge number of points at this rather specific location, creating a considerably higher density here compared to anywhere else. The Gaussian Kernel Density is computed on the territory cluster to get the estimated density at each point, which is then used to calculate the percentile that each point is in. To provide a general overview of the density, a grid the size of the territory is created and the estimated density is computed for each point, based on the density from the data. This is then plotted to show one or multiple peaks to the user before they continue the analysis. Next, the user has to determine which data to use as the nesting site. This is done by selecting a threshold for the percentile and evaluating the data at this percentile. The analysis is assumed to be most accurate when the percentile threshold is as low as possible while keeping the data condensed to one spot. With the final data, a proportion plot and interval plot are made. The proportion plot shows for each day, what percentage of the time available in that day was spent at the chosen nesting site. The interval plot shows, at each hour, whether or not the bird was was at the nesting site. The methods for both plots are similar to those described in section 3.3.1 and 3.3.2. Like the proportion and interval analysis, the start and end times can be chosen by the user.

3.4 Interactive Clustering and Analysis Methods

This section describes in more detail how each individual dataset will be analyzed for the research questions in section 1 using the interactive clustering approach. As mentioned before, the kind of clustering that is the 'best' for a dataset greatly depends on the questions that are asked. Since a plots readability goes down the more different colors are involved, it is best to keep the number of clusters fairly low. This is done by focusing on specific questions and creating clusters accordingly. For example, if the goal is to learn about the use of clubs and territory throughout the season and day, then it is not necessary to cluster each individual location where the bird is not at the water or at its territory as an individual cluster. Grouping different locations together does reduce how much can be learned about those locations, but it improves readability of plots and allows focusing on the areas that are relevant to the research question. If there are multiple distinct research questions that target entirely different topics, it may be better to separate those and create separate clusterings for each.

There are two situations that require extra attention or analysis steps. If the bird spends a considerable amount of time in two or more different locations after leaving the island which are a considerable distance apart, cluster each location and perform proportion analysis to zoom in on the correct area with the bird's home range. This is the location where the bird remains (almost) the entire time during the breeding period (March-June) after it arrives there, meaning this location is used for further analysis steps. If there is enough data to analyze multiple breeding seasons of the same bird, aim to keep clusters consistent, where possible, between the different datasets to allow for easy comparisons.

Keeping the above in mind, the following steps will be taken.

1. **Research topic 1: clubs**
   (a) When selecting the data for clustering in the case that behaviour tags are available, exclude the 'fly' tags but include all other behaviours. This makes the GPS data easier to cluster.
   (b) Create a clustering that separates the territory and each location that is at a body of water as much as possible, such as lakes, rivers or the sea.
   (c) Perform proportion analysis on this data to determine during which part(s) of the season the bird spend time at bodies of water, hypothesized to be clubs. Also observe how this changes over time.
   (d) Perform interval analysis on the time period(s) that the bird spend time in clubs to determine during which part(s) of the day this is the case.
2. **Research topic 2: nesting site**
   (a) Use the cluster which is assumed to be the bird's territory for nesting site analysis.
   (b) When choosing the percentile threshold, aim for a value that is as low as possible while keeping the area condensed in one spot.
   (c) Take note of the exact location of the nesting site, and determine in what kind of location it is, such as in a field or on a flat roof.
   (d) Use the proportion plot to provide a first estimation on whether or not the bird successfully incubated its eggs.
   (e) The interval plot can be used to support the previous hypothesis by observing the bird's present at the nesting site at each hour.

Chapter 4

Results

The map below indicates the home range location of the birds after moving inland. The majority of birds (8/14) are located in the province of North-Holland. Four birds move to various other places in The Netherlands, and two individuals move to Germany. For each bird, during preprocessing, data was gathered to check if the individual moved back to Vlieland during the season. This was the case for only two individuals, one of which (5519058) traveled back and forth many times. The other, 5519081, spend a few days inland in 2018, then moved back to the coast of North-Holland for about two weeks, before going back to the home range and remaining there for the rest of the season.

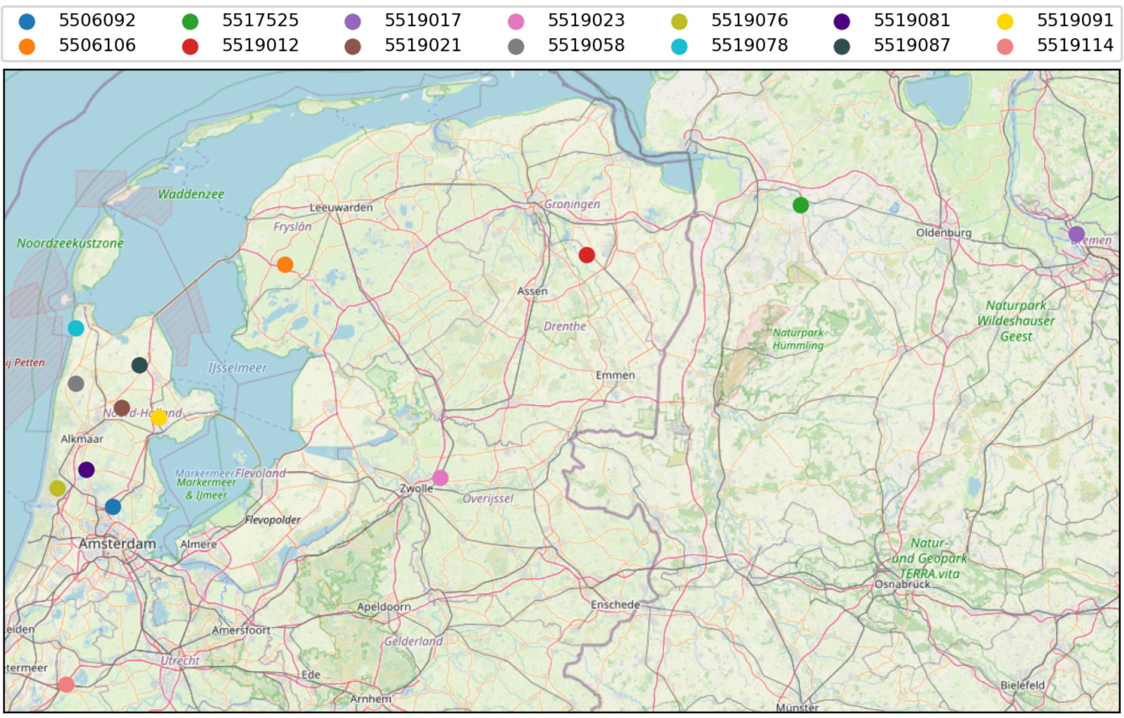
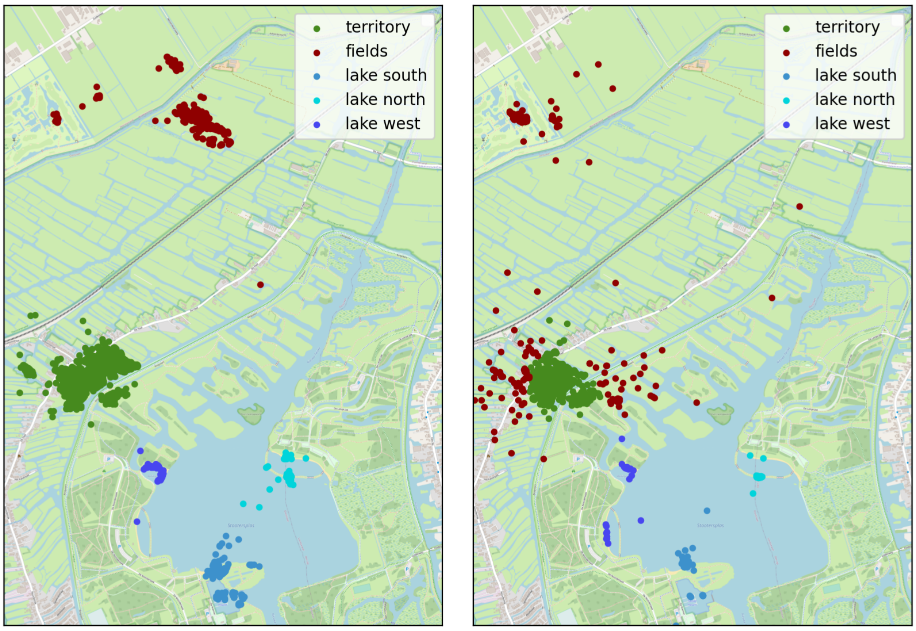


Figure 4.1: Overview of the inland-home range location of all birds included in the research

4.1 Interactive Clustering

All thirteen birds were clustered at least once, aiming to separate different bodies of water from each other, as well as separate a possible territory. Data was gathered on the total number of adjustments needed to finish a clustering as well as how this total was comprised from reclusterings, merges and undoes. An average of 6.3 adjustments were necessary to complete a clustering, with slightly more merges needed than reclusterings: average 3.5 merges and 2.7 reclusterings. This is in line with the intentions, as merging is preferable over reclustering. In addition, most areas that were not the territory or clubs were merged together, unless they were very distinct locations. This also results in more merges.



(a) Individual 5506092, year 2019. Fly data is excluded **(b)** Individual 5506092, year 2020. All data included

Figure 4.2: Comparison of data with and without fly data for the same individual in different years.

Nine datasets were finished with less than 5 adjustments, which can be considered quite fast. Of these, five did not require any reclusterings, showing the benefits of having a bias towards more clusters at the initial clustering. Five datasets required more than 10 adjustments with a maximum of 15 adjustments. Three of those datasets belonged to the same bird and contained several high density areas next to each other, which is hard to separate using clustering methods and therefore required multiple repeated reclusterings. Overview data (datasets where the bird spend a considerable amount of time inland in a location that is not near the territory) were the easiest to cluster, requiring only 2 or 3 adjustments, since the desired clusters were far apart and not many clusters were needed. Mostly, some clusters had to be merged together. Fly data required an average of 8 adjustments whereas data without fly points required an average of 6.1 adjustments, showing that removing the fly data makes the process easier, as it creates less areas with low density around the higher density areas, resulting from the bird flying around. See Figure 4.2 to see the difference.

4.2 Clubs

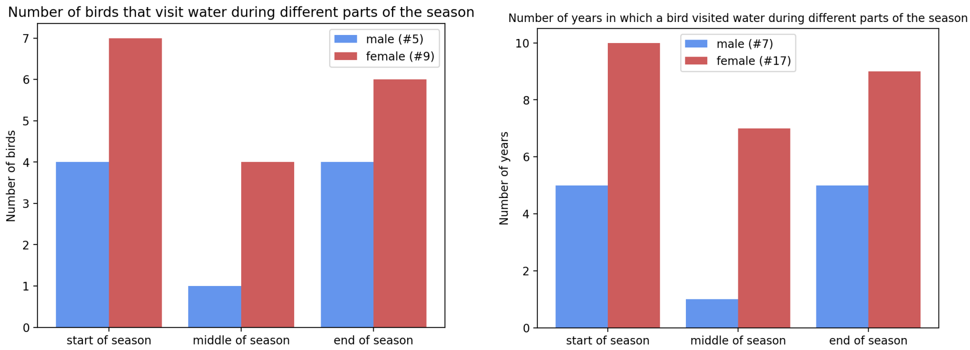
Which birds spend considerable time at one or more bodies of water, and at what kind?

Thirteen out of fourteen birds spend time at a body of water at some point during the season. One individual, 5519091 does this only the last one or two days before moving away again, at a man-made body of water. As only 2 hours were spend here at most in total, is was not deemed to be useful to include in the analysis. Individual 5519114 does not visit any bodies of water in the available data. This means twelve individuals spend some amount of time at a water body accounting for 18 different data sets, when different years of the same individual are considered separate data sets. The most unique locations at a body of water that a bird spend time at is 5, which happened only once, and the least unique locations was 1, with an average of 2.2 and median of 2. Note that in some cases, a bird spends time at distinctly different parts of one large body of water such as a lake, in which case they are clustered and counted separately.

The types of water bodies that were identified were lakes (5/12 individuals), the coast (4/12 individuals), rivers or canals (3/12 individuals), harbors (2/12 individuals) and wet/marshy fields (2/12 individuals). Time spend at the coast or in the sea where interval analysis shows the the bird was (almost) exclusively here during low tide, was not included as this would indicate that the birds were there to forage, and the location won't be a club.

During which part(s) of the breeding season do these birds spend time at bodies of water?

For each bird, the breeding season is split into three parts, which can be seen in Figure 4.4. The different parts were defined based on the bird's activity in different clusters retrieved from the proportion plot and, where needed, the interval plot. The start of the season is the time from arriving at the inland home range until a clear shift happens, generally with the bird no longer leaving the territory or doing so considerably less. The middle of the season follows, and always contains the full nesting period, as well as any time before or after this period that does not fit in the start or end of the season. The end of the season starts when another shift happens, usually with the bird leaving its territory more often again, and ends when the bird leaves the home range and does not return. Note that the exact division of the season is partially subjective, as not all birds show a clear transition with different interpretations as a result. However, this is not expected to influence the results too much. In addition, in some cases there was truly no change in behaviour visible from the plots in which case Figure 4.4 may only show two periods, as defining a third would have been an entirely random choice.



(a) Counting individual birds

(b) Counting separate years

Figure 4.3: Overview of the number of birds that visited bodies of water during different parts of the season. Of the 14 birds included, 5 are male and 9 are female. When considering all years separately, 7 are male and 17 are female.

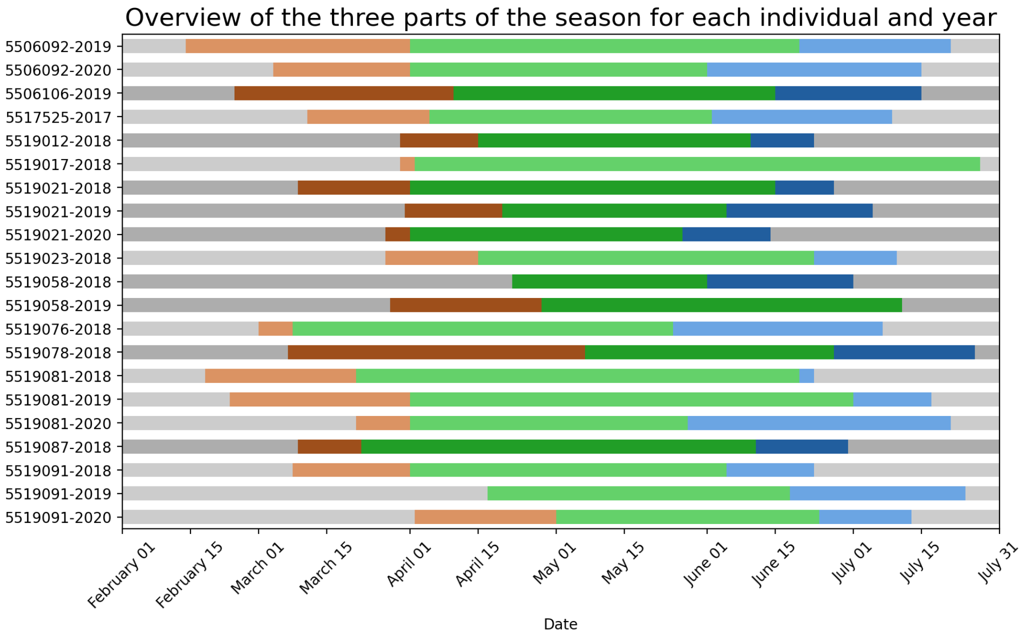


Figure 4.4: A general overview of how each season was split into three parts for each bird and each year. Orange/brown = first part of the season, green = middle of the season, blue = end of the season. Different shades separate different birds for ease of viewing.

Overall, birds spend time at water bodies throughout every part of the season, see Figure 4.3. For the sake of clarity when counting, very few and short trips to water were not included, as those don't really fit with the idea of clubs and sleeping places and are more likely unrelated short visits. In any particular part of the season, the bird should spend at least 3 hours (3 consecutive colored squares on the interval plot) for it to be included in analysis. 11/12 do so at the start of the season, 5/11 do so during the middle of the season and 10/12 do so at the end of the season. The decrease during the middle of the season can easily be explained by the birds breeding during this time, thus being more faithful to their nesting duties and social visits to clubs being less relevant. Similarly, it can be questioned whether the birds still visiting bodies of water during the middle of the season are actually visiting clubs, especially if they are also nesting.

In most cases, a bird does spend time at water during the start and end of the season and skips the middle part, but this is not true for all, see figure 4.5. Individual 5519017, which nests on a rooftop in Germany close to a harbor and river, only spend a few days at those locations at the very start of the season and did not return there at all. The same goes for individual 5506092 in 2020 and 5519076 in 2018, who both show only a couple visits of one hour throughout the rest of the season, which were excluded in analysis due to how infrequent and short they are. Also note that individual 5519058 only arrives at the inland home range in April, and was therefore considered to have missed the start of the season entirely - thus not visiting bodies of water during that time.

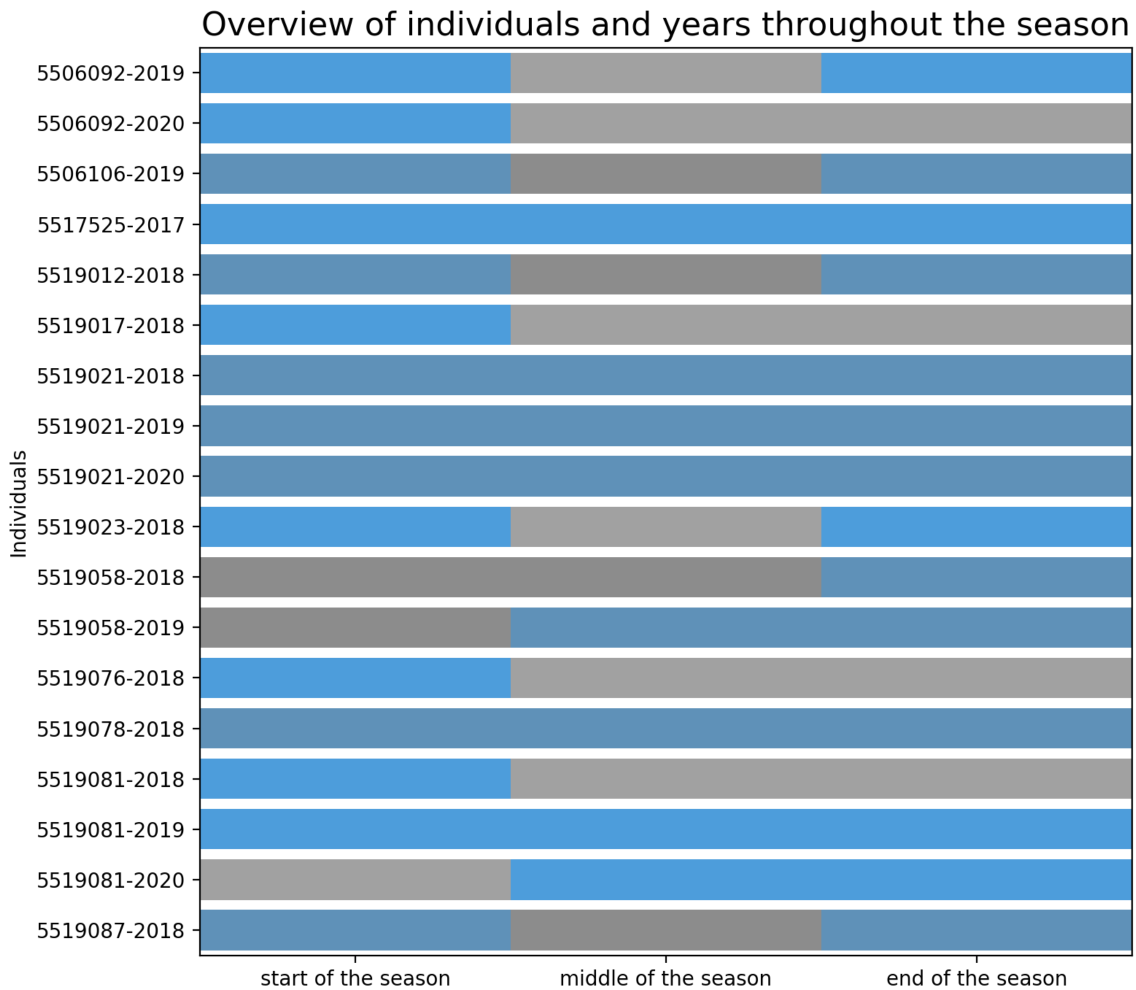
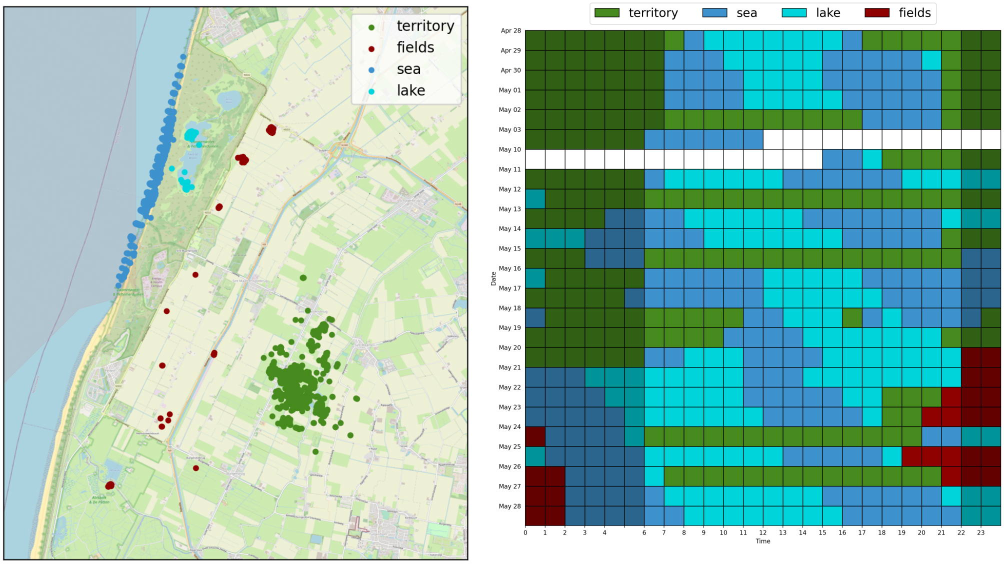


Figure 4.5: Overview each bird and each year, indicating whether or not they spend time at water during each part of the season. Blue indicates yes, gray indicates no. Different color shades differentiate different birds

There are also a few birds with multiple years of data, and they are not necessarily consistent throughout the different years. This is the case for individual 5519058, located near the coast in North-Holland, who visited a nearby lake during the middle of the season in 2019 but not in 2018. In 2019, this bird alternates between the coast and the lake (see Figure 4.6), showing a pattern following the tides. Comparing the data to data on water levels found on Rijkswaterstaat, the bird was at the sea cluster during low tide, and at the lake cluster during high tide. Because of this, the coast is assumed to be a foraging location, but the lake could be a club where the bird spends time while waiting for low tide to return, as the location would be appropriate. Based on available field observations, it cannot be confirmed that the lake is indeed a club. This behaviour is absent in the previous year.

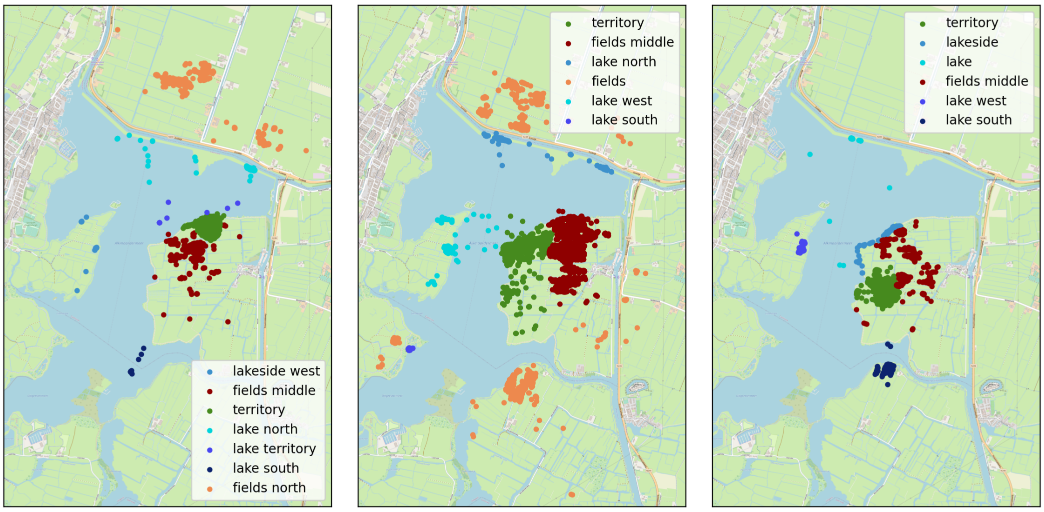
The data for individual 5519081 is a bit ambiguous, as its territory is located at the edge of a large lake, as shown in Figures 4.7. This bird appears to have moved to different territories in each of the three years with data. In 2018, there were two large clusters of data, one of which the territory (green), the other the fields next to it (dark red), which touch right up to the edge of the water. As a result, it is not possible to distinguish between the bird being at the water or in the fields for those clusters and the bird may have spend considerable time at the edge of the lake that went unseen. This means the analysis for this year may be lacking in this area, though it is reasonable to assume that there is no club in or right next to the bird's territory.



(a) The clustered data of individual 5519058 in 2019.

(b) Interval plot from April 28 to May 28 of individual 5519058 in 2019.

Figure 4.6: Clustered data and interval analysis of 5519058 in 2019



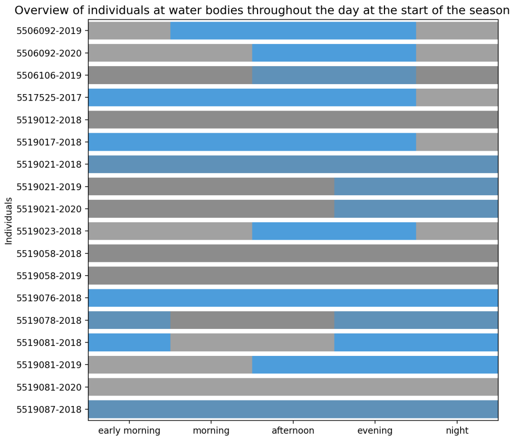
(a) The clustered data of individual 5519081 in 2018.
(b) The clustered data of individual 5519081 in 2019.
(c) The clustered data of individual 5519081 in 2020.

Figure 4.7: All three years of 5519081, clustered data

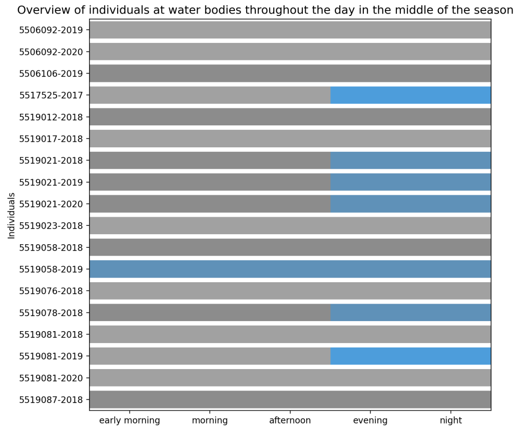
How do birds spend time at bodies of water during a twenty-four hour time period?

Overall, birds spend time at the water at any time in a 24hr period, but there are some differences between different parts of the season and between different birds, though there are also similarities. In addition, the duration of the visits also varies wildly, from sporadic visits of around one hour to whole nights or whole days. During analysis, a distinction is made between short durations (1-2 hours, shown as 1-2 consecutive colored squares on the interval plot), medium durations (several hours, part of a day or night) and long durations ((almost) the whole day or whole night). Like before, the 'short' duration visits are not included, but to get a better idea of during what time of the day these birds visit possible clubs, it's also important that they do this more than just once. For that reason, the following analysis considers repeated visits of at least 3 hours.

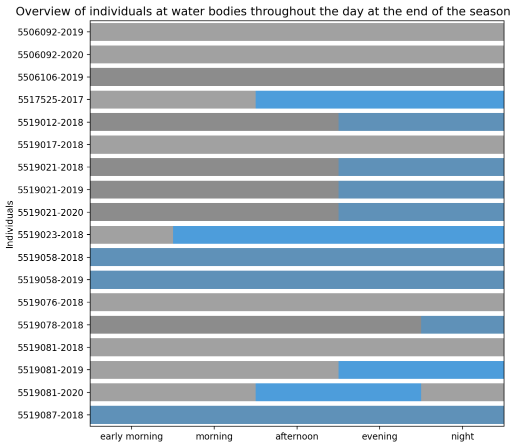
Based on the gathered data and common knowledge, the 24-hour day was divided into 5 parts: early morning (4am to 6am), morning (7am to 12pm), afternoon (1pm to 5pm), evening (6pm to 9pm) and night (10pm to 3am). The visits mentioned earlier are assigned to one or more (depending on the duration) time slots. As birds don't follow an exact schedule, they are assigned to the timeslot that fits best: e.g. a visit from 5pm to 9pm is considered to be evening, even though the first hour technically falls under afternoon, saying it was there during the afternoon would be misleading. This way, a clear overview is created that works around the 'noise' from sporadic visits or visits that sit in between timeslots, while still giving a good idea of how the birds spend time at bodies of water throughout the day in different parts of the season.



(a) Start of the season



(b) Middle of the season



(c) End of the season

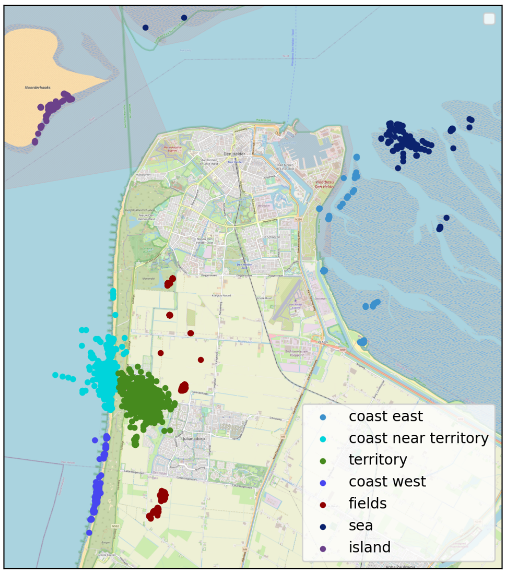
Figure 4.8: Overview indicating for each individual year, whether the bird spend considerable time at water during different portions of the day, separated by part of the season. Different color shades differentiate different birds

The plots above show the same as previous analysis, that the amount of visits to bodies of water goes down considerably in the middle of the season, and increases again at the end of the season. During the first part of the season, visits happen at any time of the day overall, though the evening seems to be most common, as every bird that does go to the water does so during the evening, though not exclusively.

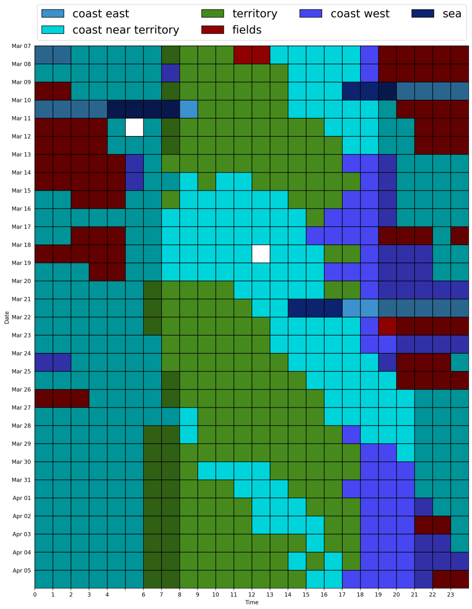
During the middle of the season, the few birds that keep visiting the water only do so during the evening and night, with one exception. Note that the former do not stay until the morning: a distinction was made between night and early morning specifically because these birds move back to their territory around 2-4am, several hours before sunrise. This also happens during the first and last part of the season, though it is most notable during the middle of the season because it makes for very consistent visits overall. One bird that stands out is individual 5519058 in 2019, who spend most of its time at the lake and nearby coast, alternating with the tides. As a result, it is at the lake during every part of the day overall, but not for full days.

During the last part of the season, roughly the time that the birds are done nesting, more birds return to bodies of water for different parts of the day, though the majority only does so at the end of the day.

In two instances, for individuals 5519058 and 5519087, birds shift their focus to spend the majority of the 24 hours at the water.



(a) Clustered data of individual 5519078

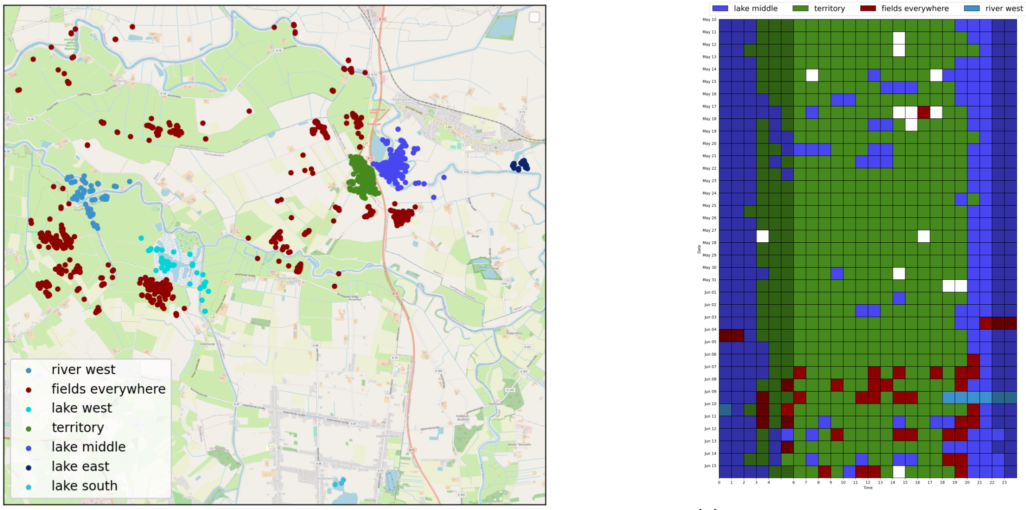


(b) Interval plot of the first few weeks of the season

Figure 4.9: Map overview (left) and interval plot (right) of individual 5519078, located in North-Holland

As seen in Figure 4.8, four birds (all female) appear to use these locations to spend the night throughout all or some of the season. One of these individuals (5517078), has a territory (green) in an urban area in North-Holland close to the coast (blue) seen in Figure 4.9a, and starts out spending the whole night at the coastline during until the second half of April. In addition, it appears to forage there during the day, based on the time coinciding with low tide, seen in the interval plot in Figure 4.9b. The frequency and duration of these visits reduce during the middle of the season and increases again at the end of the season, though not for entire nights. Currently available field observations were not able to confirm if this is a sleeping location or club.

Individual 5517525, which migrates to Germany (Figure 4.10a), has its territory (green) close to a lake and accompanying river (lake middle: dark blue/purple). At the start of the season until early April, the whole day is spent at the lake/river cluster while nights are spent in the territory and various fields around it. Note that the lake/river cluster also contains some data points of fields around the river. Starting in the middle of April, full nights are spend at the river area (few hours before dusk until 2-3am) with shorter visits on a few days. Interestingly, this behaviour continues into the period that the bird is nesting shown in Figure 4.10b (elaborated on in Section 4.3). Towards the end of June, the bird appears to move slightly west, away from its territory and lake area, spending time at a lake (lake west: light blue) and a pond next to a river (river west: blue) before returning to The Netherlands.



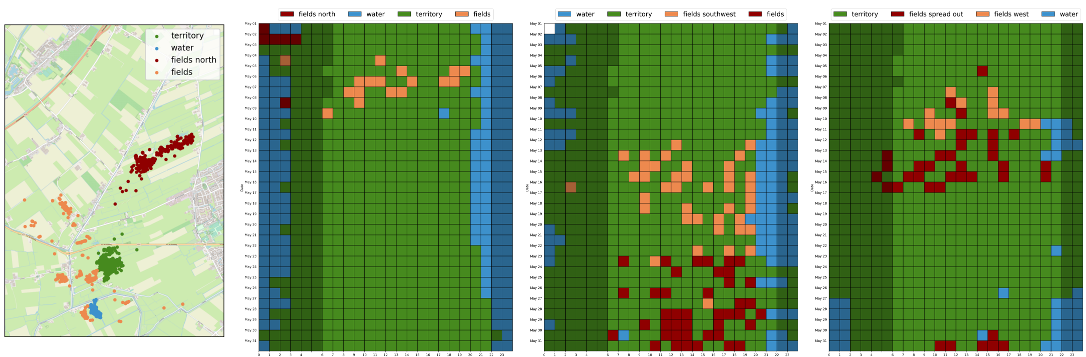
(a) Clustered data of individual 5517525

(b) Interval plot showing some weeks in the middle of the season

Figure 4.10: Map overview (left) and interval plot (right) of individual 5517525, located in Germany

One individual which has three years of data (5519081), spends the night at a small pond in a field on the other side of the lake (see Figure 4.7c 'lake west'), but only in one of the three years, in 2019. Starting in May and lasting for the rest of the season, it is there from around an hour before dusk until 2am, something that is not seen in any of the other two years. Based on available field observations, this location is a known sleeping area for oystercatchers, which explains why the bird spend the night here - though it is still interesting that it does so even as a breeding adult.

Lastly, individual 5519021 (Figure 4.11, also with three years of data, spends the night at a wet field area next to a ditch south of the territory throughout the entire season and across all three years. However, the frequency and duration goes down with each year, which can be seen in the interval plots below. This location is a known sleeping area for both oystercatchers and black-tailed godwits based on observations in the field. As this bird was breeding during each year, it is interesting to note that it made use of the sleeping area despite having a territory, and it is also unclear why the use of this sleeping area reduced in later years.



(a) Clustered data of individual 5519021 in 2018

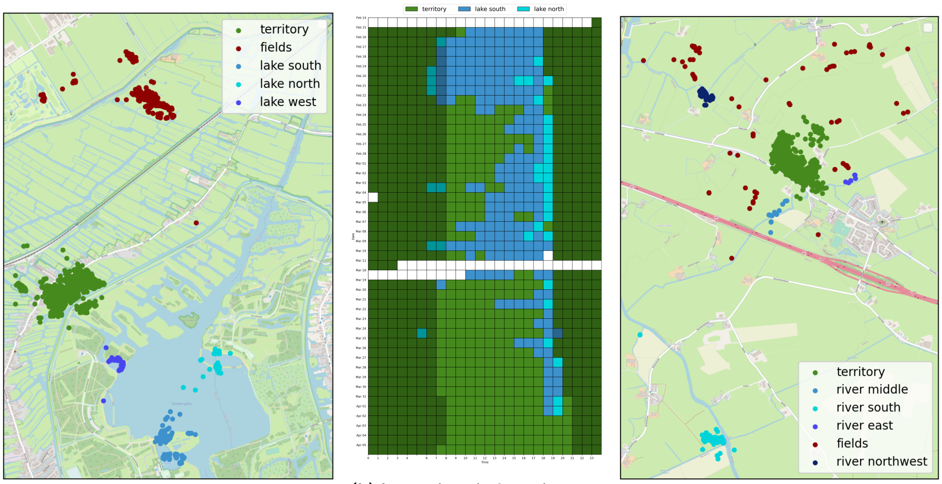
(b) Interval plot of May 2018

(c) Interval plot of May 2019

(d) Interval plot of May 2020

Figure 4.11: Map overview (left) and interval plots (right) of individual 5519021, across three different years.

Based on the performed analysis and knowledge of oystercatchers, various locations can be identified that may be clubs based on the bird's behaviour and the location at the edge of water. Some of those were confirmed based on sightings of oystercatchers in various years, as was already mentioned throughout this section. Here are the other locations identified as possible clubs that were confirmed to be either clubs, sleeping areas or both based on existing field observations.



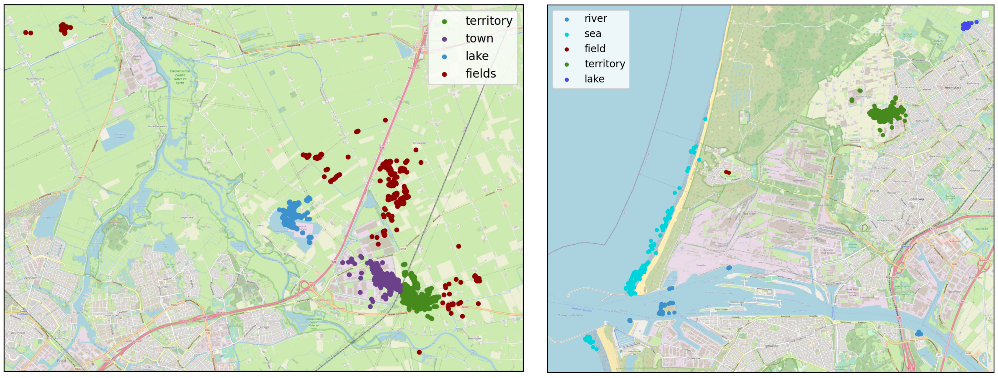
(a) Clustered data of individual 5506092 in 2019.

(b) Interval analysis at the start of the season for 5506092 in 2019.

(c) Clustered data of individual 5506106 in 2019.

Figure 4.12: Clustered data and interval plot of 5506092 (left, middle) and clustered data of 5506106 (right)

In Figure 4.12a, the bird's territory is near a large lake which shows three different locations visited at the edge of the lake. The upper area ('lake north') was identified as a sleeping area while the lower area ('lake south') was identified as a club, with no sleeping oystercatchers spotted at this location. Based on the interval analysis, the bird spend time at the southern club for various amounts of time (ranging from 2 hours to a full day, shown in Figure 4.12b) after arriving inland, in February and March, during daytime. The northern sleeping area was visited only very briefly (no more than 2 hours) just before sunset. Individual 5506106, data shown in Figure 4.12c and additional plots in Appendix 1, included multiple locations close to water but only one was considered to be significant enough for further analysis based on the time spent. This location south near a river ('river south') was confirmed to be a club based on field observations. During a few days in March, the bird spend several hours at this location during the afternoon and evening, before sunset. Next, individual 5519023 visits one large lake during the start and end of the season, location shown in Figure 4.13a and additional plots in Appendix 2. At the end of March and early April, it is here on several days for 2-7 hours largely during the afternoon. In the first few days of July, it spends a large portion of the day around this lake, only leaving it to visit fields or the nearby town for several hours. Field observations confirm that the location at this lake functions as both a club and sleeping area for oystercatchers. Lastly, individual 5519076 spends four full days in early March at two different locations near water, both seen in Figure 4.13b with additional analysis plots in Appendix 3. One is an island in the canal known as the Forteiland, which is a known club and sleeping area for oystercatchers. According to field observations, the birds can also spend time on the beach when it's quiet, though this appears to be mostly for foraging.



(a) Clustered data of individual 5519023 in 2018.

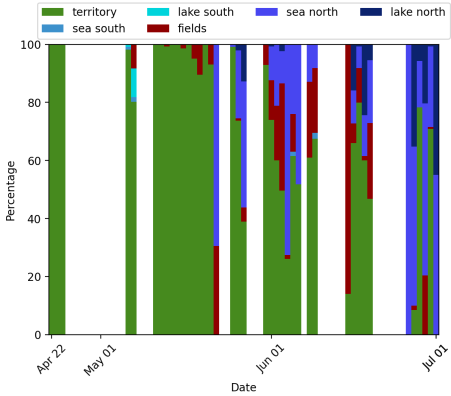
(b) Clustered data of individual 5519076 in 2018.

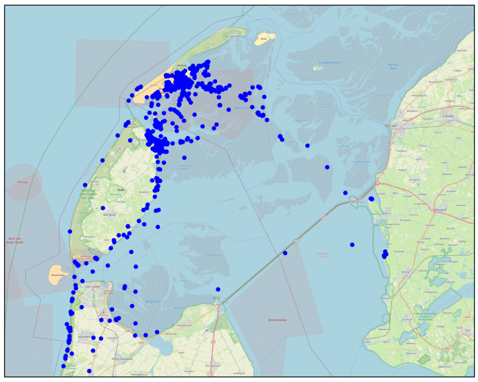
Figure 4.13: Clustered data of various birds

4.3 Nestingsite

To determine whether a particular bird has been nesting, a few factors are taken into account after locating the highest density spot: 1) the duration of the stay in this spot and 2) the relative amount of time spend there per day, retrieved from the proportion plot, 3) the distribution of visits based on the interval plot and 4) the location of the highest density point as a reference for the exact nesting site. Here, the first and second points are most telling: when a bird spends a considerable portion of its day ($> 30\%$) for close to 30 days in a small specific location, it is very likely the bird was nesting as there is not much reason for this behaviour otherwise. The location of the possible nest can then support this hypothesis, if the location corresponds to a reasonable place for the bird to nest, and the interval plot can give final confirmation if it shows regular stays of one to several hours throughout a 24-hour time period, as Oystercatchers are known to alternate with their partner for nesting duties. Based on known information on Oystercatcher nesting behaviour and the observed results across all birds, a nest is considered to be successful when:

- The proportion plot shows at least 28 consecutive days of non-negligible activity at the nesting site in the months April, May or June
- High proportion of time spend at the nesting site, around $> 30\%$ average
- The interval plot confirms frequent visits of one or several hours at this site
- The nesting location, determined as the data point of the nesting site with the highest density, is appropriate based on current knowledge of Oystercatchers.



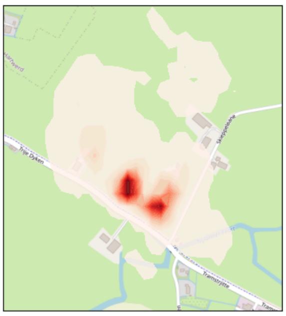


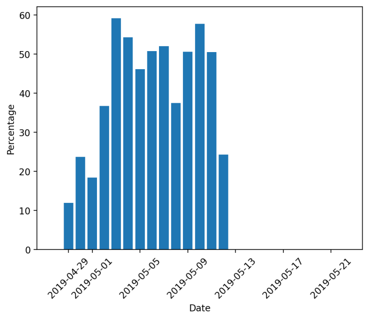
(a) The proportion plot for individual 5519058 in 2018, with gaps where the bird left the area.

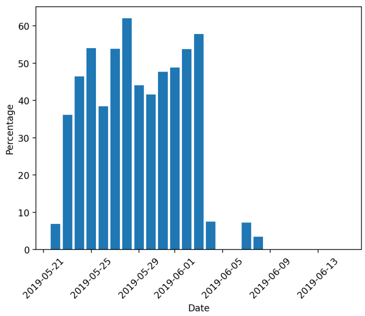
(b) The data points outside the inland home range from its visits during from April to July in 2018.

Figure 4.14: Proportion plot of 5519058 (left), nesting analysis of 5519081 in 2019 (right)

Based on the defined criteria, out of the fourteen birds analyzed for nesting behaviour, two have almost certainly not hatched chicks. Of these, one bird (5519058) clearly made no attempt, as it regularly flew back to the Wadden Sea throughout the season and stayed away for several days or longer each time, which can be seen from the gaps in Figure 4.14a and the data in 4.14b. The other bird (5506106) has data that suggests two attempted nests (estimated density peaks in Figure 4.15a), both failed. The first is regularly visited in line with expectations for a nesting bird, from April 29 to May 12, only fourteen days, shown in Figure 4.15b. A little over a week later, a second attempt occurred in a different part of the field, from May 22 to June 4 (Figure 4.15c), again only fourteen days. As this attempt occurs a little over a week after the first, it is plausible that this is indeed a second attempt at nesting, which was not successful either. Lastly, one bird (5519081) had two successful nests in 2018 and 2020, but the nest in 2019 was abandoned after 24 days, as shown in Figure 4.16, which is just short of the duration expected for a successful hatching and about six days less compared to the other birds. The bird moved to another high density spot after the failed nest, but data does not confidently suggest that another nesting was attempted.





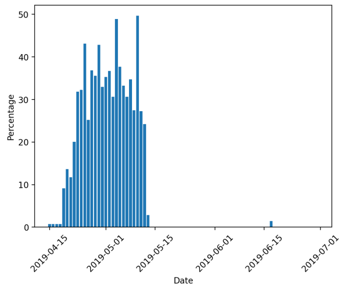


(a) The density estimation for individual 5506106 throughout the whole period, focused on the assumed territory, showing two peaks.

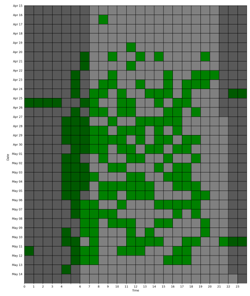
(b) The proportion of time spend on the nesting site for 5506106 in 2019, between April 28 to May 22

(c) The proportion of time spend on the nesting site for 5506106 in 2019, between May 22 and June 15

Figure 4.15: Nesting analysis of 5506106



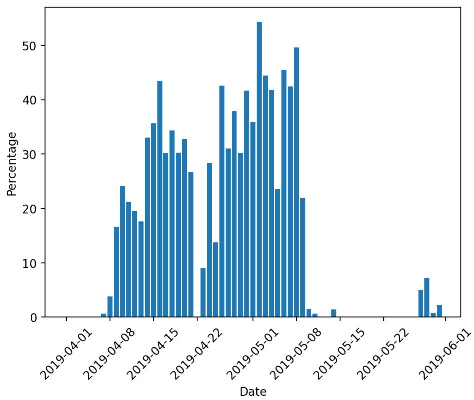
(a) The proportion of time spend on the nesting site for 5519081 in 2019



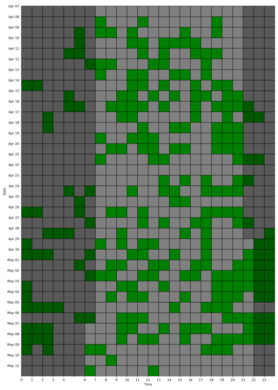
(b) The interval analysis of 5519081 in 2019.

Figure 4.16: Nesting analysis of a barely failed nest for 5519081 in 2019

For a few birds, the data is a bit remarkable, not quite showing the expected circumstances for nesting but also not showing obvious failure. One of these instances occurs in 2019 for individual 5506092, which shows a near-perfect proportion plot with 30 days of time spent at the nesting site as seen in Figure 4.17a. The percentage is a little low, which is also shown in the interval plot (Figure 4.17b) which has some bigger gaps than usual, including a full day the bird did not visit the nesting site at all. This could be a case of the bird relying more on its partner to stay at the nest, as the duration looks good and the amount of time spend at the nest is still far too high to be entirely unrelated to nesting.



(a) The proportion of time spend on the nesting site for 5506092 in 2019

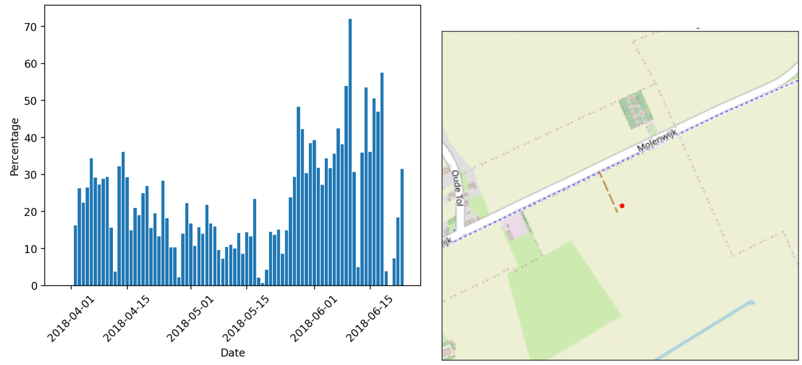


(b) The interval plot for 5506092 during the 30-day nesting period, indicating at each hour whether the bird is at the nest (green) or not (grey)

(c) Location of the presumed nesting site for 5506092

Figure 4.17: Nesting analysis of 5506092 in 2019

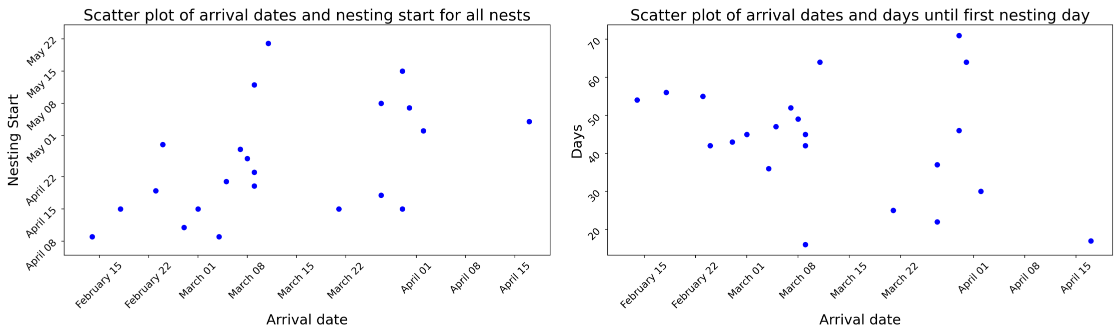
Another odd situation is that of individual 5519012 in 2018, of which the highest density area (location shown in Figure 4.18b) is frequently visited throughout the entire breeding season from April 2 until June 23 (Figure 4.18a), a few days before the bird leaves for the islands. Based on the interval plot, there appears to be a switch around May 28. Before this date, the percentage is lower and the bird only stays here during the night. From May 28, the site is also fairly regularly visited throughout the day and the pattern looks more like a nesting bird, though with large gaps. As the bird leaves for the island less than 30 days after this point, it could not have had chicks - they would not have been able to fly already if we assume the nesting period to be starting around May 28. It is highly unlikely it was nesting before this date, as its partner would have had to account for the nest for over 18 hours a day. As such, based on the analysis available it is uncertain whether there was an attempt at nesting during this period, but it is highly unlikely that chicks hatched, even if there was a nest. For this reason, this bird is not included in further analysis and is considered to have 'failed' at hatching the chicks.



(a) Proportion of time spend at the possible nesting site throughout the season for 5519012 **(b)** The location of the possible nesting site of 5519012

Figure 4.18: Nesting analysis of 5519012

This leaves birds where data analysis showed a successful nesting period of around 30 days or more with regular visits. This is the case for the remaining 11/14 birds, though one bird has one unsuccessful year, resulting in 19/24 separate breeding periods having successfully hatched chicks, based on the available analysis. Figure 4.20 shows the successful nestings, indicating the time period that the bird visits the nest at high frequencies. In some cases, this period is longer than 30 days, which means that it was not possible to differentiate the exact time frame where the bird was nesting from the proportion plot. For the analyzed birds with a successful nest, the minimum time on the nest was 28 days and the maximum 58 days. The mean is 35.95 days with a median of 32 days and mode of 29 days, indicating that the distribution is right-skewed. This makes sense, as birds that nested less than 28 days were not included at this step of the analysis, knowing they did not manage to hatch the eggs. On the other hand, it is possible, though less common, for the parent to keep visiting the nest even after the chicks hatched, causing a skew to the right.



(a) Scatter plot comparing the arrival date to the first day of nesting for each attempted nest. **(b)** Scatter plot comparing the arrival date to the number of days until the first day of nesting for each attempted nest.

Figure 4.19: Scatter plots showing possible correlation between arrival and nesting timing

To investigate the time between arriving and nesting, figure 4.19a shows the date that each bird arrived on the x-axis compared to the date it started nesting on the y-axis. This plot includes the instances where a clear nesting attempt was present, but the eggs did not seem to hatch based on the analysis. From the plot, there appears to be a possible positive correlation, indicating that arriving earlier correlates to nesting earlier in the season. The Pearson correlation coefficient for this data is 0.50 which has a p-value of 0.02, confirming a positive correlation in this data. Do note that the sample size is still relatively small and that we cannot confirm causation based on the correlation seen here. However, it does suggest that early arrivals may be more likely to start nesting earlier in the season as well. Secondly, to approach this from a different angle, Figure 4.19b shows the arrival date and the number of days between this date and the first day of nesting. Based on the plot, there may be a very slight negative correlation. To check this, the Pearson coefficient is -0.32 which would indeed indicate a negative correlation, meaning late arrivals would spend less time waiting until they start nesting, perhaps to avoid nesting too late in the season. However, the correlation coefficient of -0.32 only has a p-value of 0.15, meaning we do not have significant proof that this correlation indeed exists. On average, there are 43 days between arrival and the first nesting day with a median of 45, indicating a slightly left-skewed distribution.

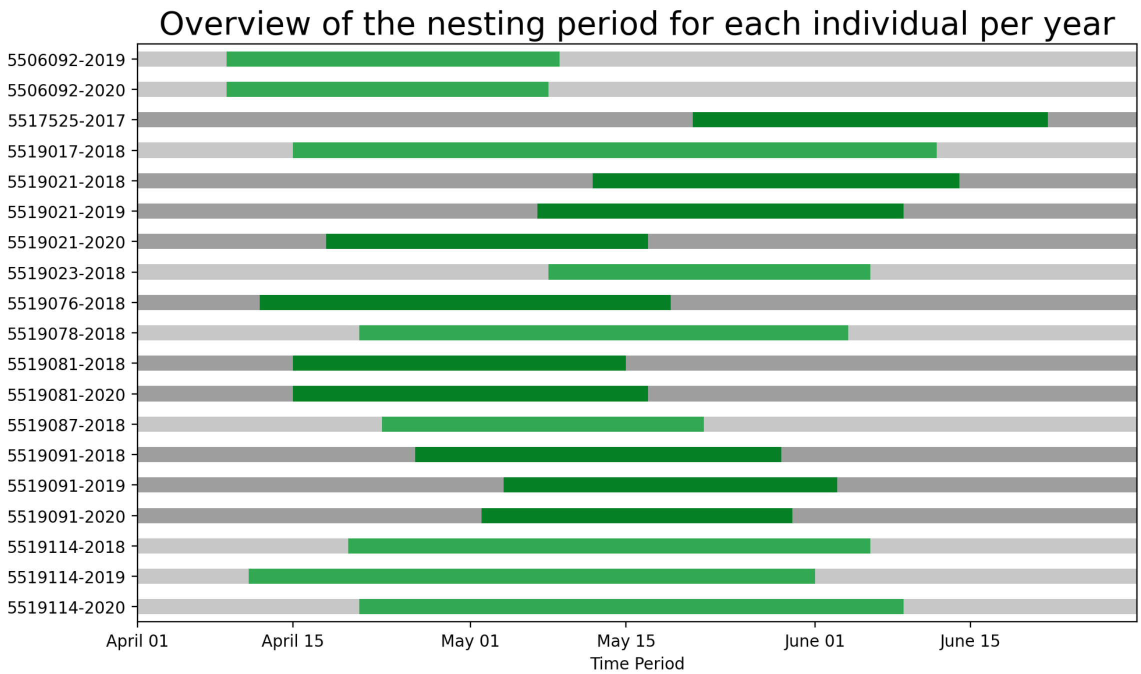
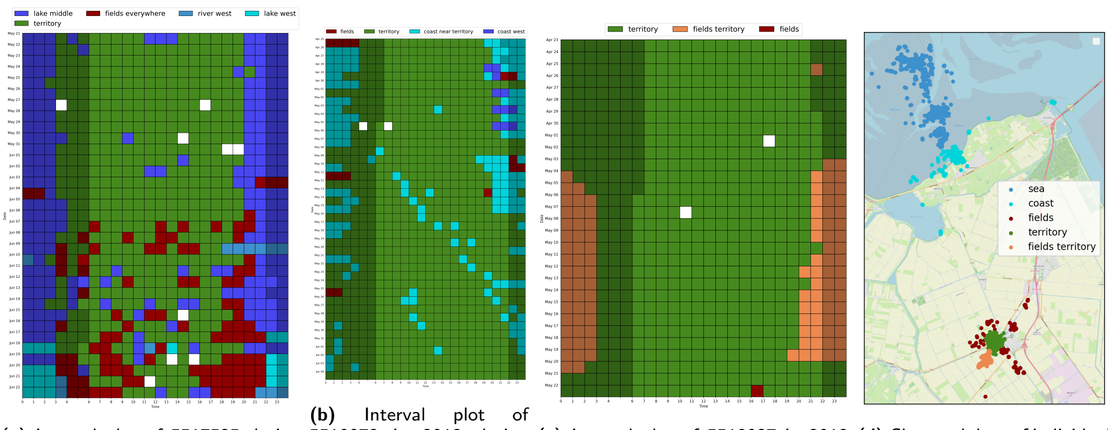


Figure 4.20: Overview of the time period in which each bird is frequently visiting their nest, separated by year. Different color shades differentiate different birds

Birds that breed for multiple years stay faithful to their nesting location, at least to the extend that the type of location is the same. Four birds nested on flat rooftops, one nested at an industrial area and the other six nested in grassy fields, as verified for each bird on Google Earth or Google Maps. Birds nesting on rooftops may be at a disadvantage, as the chicks may not be able to leave the roof to follow their parents for food. This can also be seen in Figure 4.20. Rooftop nesting birds in this dataset are 5519017, 5519076, 5519078 and 5519114. These birds visit the nesting site at high frequency for 58, 37, 44 and 49 (average of 3 years) respectively, all longer than the usual 30 days, for 47.7 days on average. Comparing this to the other seven, who are not limited to rooftops, the average for field-nesting birds is lower: 30.5 days. The bird nesting in an industrial-like area has been included in this count, as their location is similar to the fields in that nothing is preventing the chicks from following their parents right away. Performing a Welch t-test, as the variances are not equal (41.2 for rooftop and 3.0 for field) to check if the average time at the nest for rooftop nesting birds is indeed higher, gives a p-value of 0.0008, thus rejecting the null hypothesis of equal means and supporting the alternative hypothesis that rooftop-nesting birds in this data set spend considerably longer visiting the nesting site.

A few more things stand out when looking at Figure 4.20. When comparing different years of the same bird, where available (recognize-able by the same shade of colors), shows that most are very consistent in their nesting behaviour when it comes to timing and duration. For example, 5506092 and 5519081 start nesting at the exact same day both years and differ a couple days at most in the duration. Taking into account that the analysis doesn't know exactly when a bird is nesting (but rather looks at the time spend at the presumed nesting location) and that the data included as the nesting site also depends on the available GPS data, the results are remarkably consistent. Though individual 5519114 starts nesting noticeably earlier in 2019 compared to the previous and next year, the durations across all three years remain consistent still. It is unclear why the timing varies, as this bird is female and as such is less likely to lay eggs at a different time if their partner changes. One notable individual is 519021, who also nests for a consistent duration across all three years, but starts nesting earlier each year, starting around May 15 in 2018 and April 15 in 2020, almost a full month earlier. The reason can't be deduced from the available data, but it's possible that the bird experienced some sort of benefit from nesting a few days earlier in 2019, and decided to start even earlier in 2020 as a result. Alternatively, perhaps the bird acquired a new mate who influenced the start of the nesting period for the pair, though it should be noted that the territory remained the same each year. However, as this bird is also a female, this seems less likely.



(a) Interval plot of 5517525 during their nesting period

(b) Interval plot of 5519078 in 2018 during their nesting period

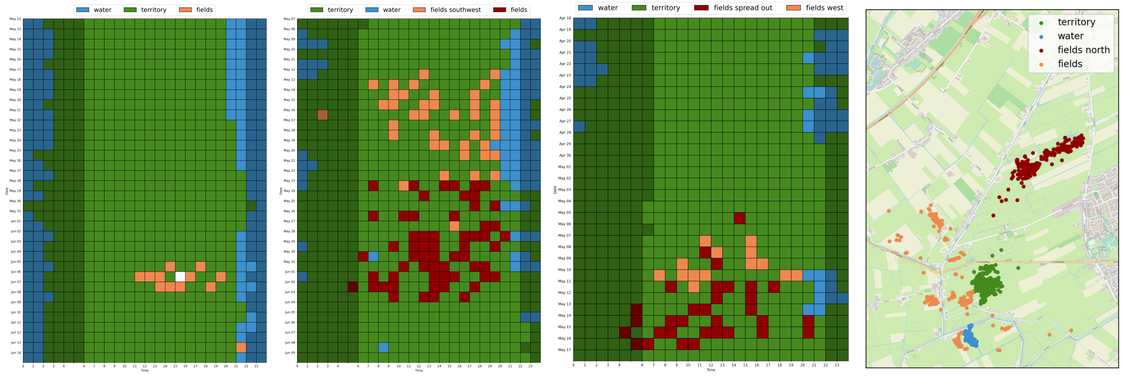
(c) Interval plot of 5519087 in 2018 during their nesting period

(d) Clustered data of individual 5519087

Figure 4.21: Interval plots of various birds (left) and clustered data (right)

In the previous section, four birds were discussed that would spend the night at a body of water, some of which continued to do so during the breeding season from April to June. This raises the question whether these birds were also nesting and if so, if their leave during the night had any consequences for nesting. Previous analysis already showed that each bird appears to have a had a successful nesting period, but an interval plot of the nesting period will show if a bird spent considerable time away during this period as well. Based on the Figures in Figure 4.21 and 4.22, some birds definitely do this, to more or lesser extent. Individual 5519021 (Figure 4.22), is an interesting case as each next year shows less time spent at the club at night. As discussed previously, the nest analysis shows a confident window of 30 days that the nest is frequently visited, meaning the visits do not seem to affect this. Interestingly, this bird also starts nesting earlier each year, which could be related to the decreasing amount of time spend at the club during the night, though it is not possible to determine if there is a connection and if so, why. In 2019 and 2020, the bird also has brief visits to other fields during the day, possibly for foraging.

5517525 (interval plot of nesting period seen in Figure 4.21a) is another individual who spends the entire night at the river area throughout the whole nesting period, and at the end even went to a lake much further west. In addition, there's some sporadic visits to various fields, likely the ones nearby the territory, in the second half, possibly to forage. Individual 5519078, interval plot in Figure 4.21b also spends time at the water during the night, though the frequency and duration reduces throughout the presumed nesting period. Lastly, individual 5519087 does not visit water bodies during the nesting period, but does spend time in nearby fields south-west of the territory during most of the night for roughly the second half of the season as can be seen in Figure 4.21c. This is likely to forage for food, something that could be determined in future analysis through the behaviour tags.



(a) Interval plot of 5519021 in 2018 during their nesting period

(b) Interval plot of 5519021 in 2019 during their nesting period

(c) Interval plot of 5519021 in 2020 during their nesting period

(d) Clustered data of individual 5519021 in 2018

Figure 4.22: Interval plots across different years and clustered data of individual 5519021 in 2018

Chapter 5

Discussion

5.1 Interactive clustering

Overall, the performance of the interactive clustering tool that was designed and implemented was good. It was able to adequately cluster the desired data sets with minimal 'losses' (accepted clusterings with unwanted/less than ideal clusters) as it was possible to separate territory and water locations quite well which allowed for fairly detailed analysis of different possible club locations. The interaction works smoothly due to an emphasis on merging smaller clusters over reclustering larger clusters, and in most cases the process does not take too long after the initial clustering, which takes a few minutes. However, there are some limitations and possible improvements to the current setup.

First, the user's ability to give feedback is rather limited. When reclustering, the only possible input is 'cluster these clusters again' but no further indication can be given as to the desired outcome of this reclustering, such as the number of clusters. One way to influence this is through selecting the clustering algorithm, which is either DBSCAN or Agglomerative. While the original intention was that DBSCAN would be the main algorithm for reclustering and Agglomerative is more of a backup, both turned out to have their own strengths. In general, DBSCAN gives predictable results when there are well defined clusters of similar density with clear gaps between them. Agglomerative is less intuitive, but tends to create more clusters in situations where DBSCAN cannot find enough clearly defined clusters, thus ending up creating a larger cluster with several very small ones. Knowing this, the user can play into these differences, but it does require some experience with the tool to get a feel for which algorithm may work best. It could be beneficial to extend the reclustering feature to allow for additional input, such as the desired number of clusters. The tool would then need to ensure that the presented clustering has at least this many reasonably sized clusters; meaning, one large cluster and a cluster consisting of five points is not adequate. More clusters may be tolerated if the desired amount isn't found, as they can then be merged to shape the clusters the user wanted.

Another aspect that could be improved upon is a learning method for the tool. In the current implementation, the tool does not learn from the user's answers, it simply carries out the required reclusterings or merges. Implementing a method for the tool to learn from clustering sessions may over time eliminate repeated mistakes. However, allowing the tool to learn should be done with caution, as desired results can vary greatly from one session to another. What is considered a mistake during one session may be the desired outcome for another depending on the geographical features and the user's goals, so allowing the tool to learn should be done with care to prevent it causing issues later on. On the other hand, if it was possible to identify consistently undesirable clusterings and avoid them, it could speed up the process.

5.2 Analysis

The implemented analysis methods were successful in creating an easy to read overview of the places the bird spend time in throughout a long period of time, in this case the whole breeding season from February to July, though it is expected to work for both shorter and longer time periods as well. The overview provided by the proportion plots can then be used to identify specific periods to investigate more closely, using an interval plot showing at each hour in what cluster the bird was located. This strikes a balance between a more detailed insight into the bird's daily schedules while the plot remains easy to read with as little missing data as possible. However, as these data sets have data for every 10 minutes in some cases, future methods may focus on using this to gain more information on the exact duration of a stay in a cluster, and on trips shorter than an hour that are not represented as well in an interval plot. In addition, the removal of fly-labeled data could potentially influence the analysis and is worth looking into in future improvements for the tool. While the fly-data was excluded to make clustering easier by removing low density 'clouds' of points around clusters, it also removed fly data within the clusters, ultimately leading to gaps of unknown sizes in the dataset that is analyzed. There may be ways this can be handled better, for example by using a type of k-nearest neighbours method to assign the correct cluster labels to fly data within a cluster, while preventing the low density 'noisy' fly data from being labeled by keeping k sufficiently high. Another option is to implement a method to specifically target the low-density data, such as DBSCAN's noise detection, which would label points that do not have a high enough density as noise, though it then still needs to be decided how to handle this data.

One major aspect that can provide even more information for this particular data set is to make use of the behaviour classification that is included with the Vlieland Oystercatcher data. Similar methods of analysis can be used, but rather than visualizing the different clusters, one cluster can be chosen to visualize the different behaviours the bird shows in this cluster. Lastly, an additional type of analysis focused on duration may be useful, as the interval plot can give a rough idea of the duration of a stay, but it is not precise and requires manual work. An algorithm that calculates the time between the bird entering and leaving the cluster could save this data allowing it to be analyzed later, calculating total duration in a cluster, averages and others.

The nesting analysis methods, while indirect, do seem to provide a good idea of the bird's nesting location and period. Due to the density-based approach and provided options, there are a few factors that influence the results: the chosen cluster, time period and percentile threshold are all important factors. Especially the latter can greatly influence the results as it determines which GPS points are considered to be at the nesting site. Selecting an area that's too small would result in points that should be included, being excluded, which in turn affects the calculations for the duration that the bird is on the nest, possibly giving a wrong idea of the nesting behaviour. For this reason, it would be good to look into a more solid way to determine the points to be included, either via improving the selection of the percentile threshold or in some other way. Taking into account the distance that the GPS can be off from the true location can help with this, for example through selecting a range in meters and including all points in this range from the highest density point. If a reliable range can be determined, this might give more consistent results compared to manually deciding on a threshold through trial and error and inspecting scatter plots of the included data points, though it would require distance calculations based on coordinates. However, despite this aspect of the analysis being relatively variable due to the user having to make a separate choice for each instance, the results of the nesting site analysis are remarkably consistent across different birds, especially in terms of duration. The percentage of time spent on the nest each day is more variable, but the nesting period stands out very strongly on these plots. A clear difference is seen between bird nesting on rooftops, where chicks will stay at or around the nest for longer and continue to receive food from parents in this location compared to field-nesting birds who leave the nest behind immediately at the end of the nesting period. In addition, the interval plots can give some extra insight into when the bird visited the site throughout the day.

Besides improving consistency, a feature of the nesting analysis can be improved is to allow for an easier way to investigate multiple high density points. In some cases, there was more than one density peak, and these cases can be even more interesting to investigate as it may be the case that the bird attempted to nest more than once. Right now, this would require manually 'separating' the peaks by changing the time period such that only one peak is included, but this is not easy as it requires trial and error to find which peak happened in which time period. In addition, if the bird is visiting both peaks simultaneously, the method does not work at all and cannot separately analyze either one, crippling the analysis greatly. One way would be to include data from both peaks and separating them using DBSCAN clustering methods, as the result should be two (or more) very well defined clusters with fairly even density throughout. Afterwards, both clusters can be analyzed separately using the methods implemented here.

Another angle that can be explored regarding nesting analysis is to more closely investigate the distribution of nesting tasks between the tracked bird and its unknown partner. If consistency of the tool can be improved, this can be done by closely tracking the stays of the bird in its nesting location as well as the time spend away from the nest. As oystercatchers don't usually leave the nest undefended, it can be assumed that when the tracked bird is away, their partner will be on the nest. From here, it is possible to calculate the duration of the incubation bouts (discussed in section 2) of both male and female. These can then be compared between males and females and/or to other pairs to investigate within-species variability further. Especially for birds where the distribution of nesting duties seems skewed, it can be interesting to investigate this in more detail as well as finding possible explanations, such as location, availability of food or other factors. The average time spent on the nest could also be used to compare successful and unsuccessful nesting attempts.

Finally for the nesting analysis, it may be possible to further explore the success or failure of a particular nest. Currently, only the time and frequency spend on the nest is considered, but the behaviour of the parents after the assumed nesting can also say something about the likelihood that they're raising chicks during this time. For example, it would take over 30 days for the chicks to be able to fly, meaning departure to the islands within 30 days after the final nesting day is almost certain to indicate that the chicks, if there were indeed any, did not survive long enough.

Lastly, when taking into account the available behaviour tags, it would be especially interesting to investigate the bird's foraging behaviour. This can be done by analyzing where and when they forage, but also by taking into account the vegetation or ground cover that was present during their period inland. This can show a preference for different types of vegetation of this species, but it may also vary within-species. Rooftop-nesting birds may find themselves in a location with completely different types of vegetation compared to those that nest in fields, and as a result may show different foraging behaviour. Regardless, both groups will have to adapt to their new location away from the sea and tides, meaning a complete change in their daily rhythm as they no longer need to await low tide in order to forage. It can be interesting to compare each individual's foraging behaviour on the mudflats to that inland, and to what extend this change is immediate or perhaps more gradual after arriving.

Chapter 6

Conclusion

The research performed was able to provide a user-guided analysis to investigate the behaviour of Eurasian Oystercatchers. An interactive clustering method was designed and implemented which gives the user the freedom to define clusters based on the desired research goal, supported by various 'best clustering' update rules based on existing knowledge on Oystercatchers and more generally, easy of use. The finished clusterings were analyzed using various methods including a proportion analysis to observe the bird's time spend in clusters over various months and an interval analysis to observe where the bird spend time throughout the day at each hour. Lastly, a density-based nesting analysis is able to locate the bird's nesting site in their territory and identify the nesting period based on the amount of time spend in close proximity of this site.

Several research questions were defined regarding the use of clubs and nesting behaviour. Out of the fourteen birds included in analysis, twelve spend notable time at one or more bodies of water, some of which were confirmed to be clubs based on real life observations. Spending time at water was mostly done during the start and end of the season, though several individuals continued to spend time there even in the middle of the season, though only during the evening and first part of the night. Overall, the studied birds were at the water during any part of the 24-hour day, with a slight preference for the evening and night. The relatively small sample size, combined with a high variability between different birds, did not reveal clear patterns besides nightly stays during the middle of the season.

Nesting site analysis showed that thirteen out of fourteen birds were breeding birds, having attempted to nest at least once, with eleven birds having succeeded in hatching chicks at least once, based on the available analysis methods. Four nested on rooftops, one in an industrial area on the ground and the other six nested in fields. The ground-nesting birds spend an average of 30.5 days at their nesting site while the rooftop-nesting birds spend an average of 47.7 days visiting their nesting site, indicating that the ground-nesting birds abandoned their nest after hatching their chicks whereas the rooftop-nesting birds continued to utilize it to more or lesser extent. Most birds with multiple years of data were very consistent in their duration and timing, though some interesting exceptions were also seen.

Future work can improve and continue the methods that were implemented here by adding additional ways to provide feedback and interact with the iterative clustering process to make the process faster and smoother. Analysis methods and future research can be extended to include behavioural tags available in the Vlieland dataset, which would allow a deeper insight into how the bird is spending time in each cluster. This would also allow research into the foraging behaviour throughout the season, such as the timing and type of vegetation at the foraging locations. Lastly, nesting analysis could be improved to provide a more consistent method and future research could investigate nesting behaviour more closely.

Bibliography

[1] Juhee Bae et al. "Interactive Clustering: A Comprehensive Review". In: *ACM Computing Surveys* 53 (2020). DOI: https://doi.org/10.1145/3340960.

[2] Wiene Bakker et al. "Connecting foraging and roosting areas reveals how food stocks explain shorebird numbers". In: *Estuarine, Coastal and Shelf Science* 259 (2021), p. 107458. ISSN: 0272-7714. DOI: https://doi.org/10.1016/j.ecss.2021.107458. URL: https://www.sciencedirect.com/science/article/pii/S0272771421003115.

[3] W. Bouten et al. "A flexible GPS tracking system for studying bird behaviour at multiple scales". In: *Journal of Ornithology* 154 (2013), pp. 571–580. DOI: https://doi.org/10.1007/s10336-012-0908-1.

[4] Martin Bulla et al. "Unexpected diversity in socially synchronized rhythms of shorebirds". In: *Nature* 540 (Dec. 2016), pp. 109–113. DOI: 10.1038/nature20563.

[5] A. Duncan et al. "Roof and ground nesting Eurasian Oystercatchers in Aberdeen". In: *Scottisch Birds* 22 (June 2001).

[6] B.J. Ens et al. "Scholeksters in de knel: onderzoek naar de oorzaken van de dramatische achteruitgang van de Scholekster in Nederland." In: *SOVON-onderzoeksrapport 2011/13* (2011), pp. 1–120.

[7] Bruno Ens, Martijn van de Pol, and John Goss-Custard. "Chapter Eight – The Study of Career Decisions: Oystercatchers as Social Prisoners". In: *Advances in the Study of Behavior* 46 (Dec. 2014), p. 343. DOI: 10.1016/B978-0-12-800286-5.00008-0.

[8] Henk-Jan van der Kolk et al. "GPS tracking data of Eurasian oystercatchers (Haematopus ostralegus) from the Netherlands and Belgium". In: *ZooKeys* 1123 (2022), pp. 31–45. ISSN: 1313-2989. DOI: 10.3897/zookeys.1123.90623. eprint: https://doi.org/10.3897/zookeys.1123.90623. URL: https://doi.org/10.3897/zookeys.1123.90623.

[9] Henk-Jan van der Kolk et al. "Shorebird feeding specialists differ in how environmental conditions alter their foraging time". In: *Behavioral Ecology* 31.2 (Nov. 2019), pp. 371–382. ISSN: 1045-2249. DOI: https://doi.org/10.1093/beheco/arz189. eprint: https://academic.oup.com/beheco/article-pdf/31/2/371/32957323/arz189.pdf.

[10] Henk-Jan van der Kolk et al. "Spatiotemporal variation in disturbance impacts derived from simultaneous tracking of aircraft and shorebirds". In: *Journal of Applied Ecology* 57 (Aug. 2020). DOI: 10.1111/1365-2664.13742.

[11] Henk-Jan van der Kolk et al. "The hidden cost of disturbance: Eurasian Oystercatchers (Haematopus ostralegus) avoid a disturbed roost site during the tourist season". In: *Ibis* 164 (Nov. 2021). DOI: 10.1111/ibi.13035.

[12] Vogelbescherming Nederland. *Factsheet Scholekster*. Accessed April 16 2023. URL: `https://www.vogelbescherming.nl/docs/58d018d0-e537-4518-8cb5-69b467eb58dd.pdf`.

[13] Vogelbescherming Nederland. *Scholekster*. Accessed April 16 2023. URL: `https://www.vogelbescherming.nl/ontdek-vogels/kennis-over-vogels/vogelgids/vogel/scholekster`.

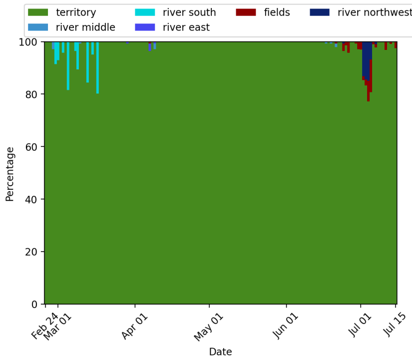
[14] J. Shamoun-Baranes et al. "From sensor data to animal behaviour: an oystercatcher example." In: *PLoS ONE* 7, e37997 (2012), pp. 1–120. DOI: `https://doi.org/10.1371/journal.pone.0037997`.

[15] UvA Bird Tracking System. *Oystercatcher*. Accessed April 16 2023. URL: `https://www.uva-bits.nl/species/oystercatcher/`.

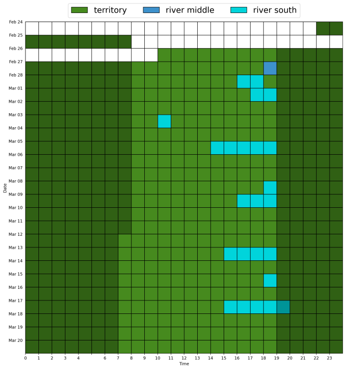
[16] M. Van de Pol et al. "A global assessment of the conservation status of the nominate subspecies of Eurasian oystercatcher (Haematopus ostralegus ostralegus)". English. In: *International Wader Studies* 20 (2014). Reporting year: 2014 Metis note: 5544; AnE Data archiving: no data; review, pp. 47–61. ISSN: 1354-9944.

Appendix

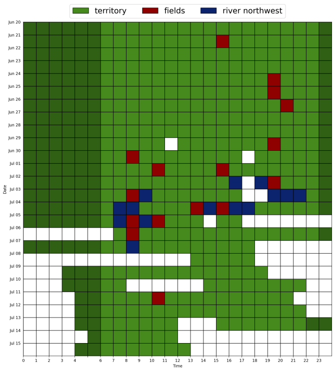
Appendix 1



(a) Proportion analysis of individual 5506106



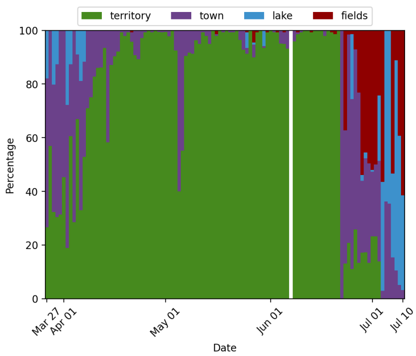
(b) Interval analysis at the start of the season for 5506106



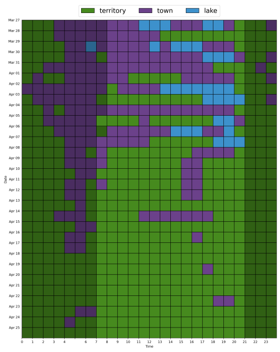
(c) Interval analysis at the end of the season for 5506106

Figure 1: Proportion and interval plots of individual 5506106

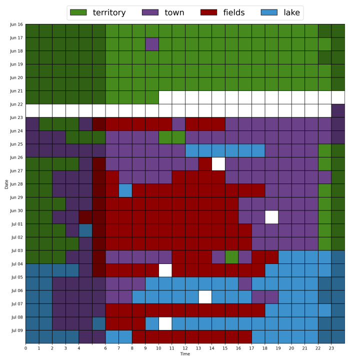
Appendix 2



(a) Proportion analysis of individual 5519023



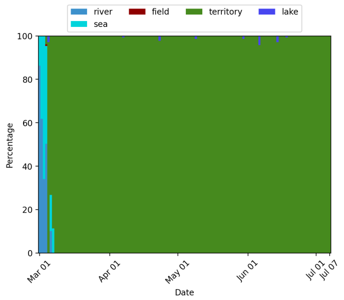
(b) Interval analysis at the start of the season for 5519023



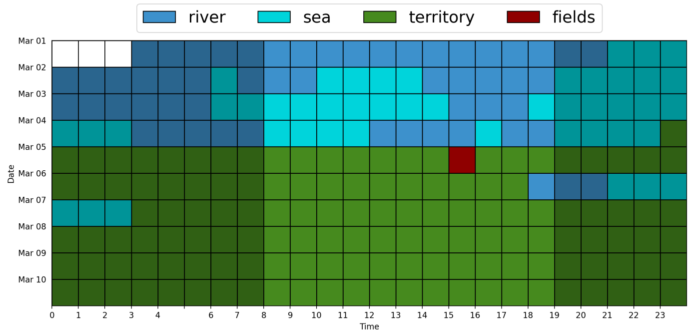
(c) Interval analysis at the end of the season for 5519023

Figure 2: Proportion and interval plots of individual 5519023

Appendix 3



(a) Proportion analysis of individual 5519076



(b) Interval analysis at the start of the season for 5519076

Figure 3: Proportion and interval plots of individual 5519076

Appendix 4

animal id	ring id
5506092	BLB-OTB2
5506106	BLB-OTBP
5517525	RB-CLLP
5519012	BLR-O8BA
5519017	BLR-O8BL
5519021	BLR-O8BT
5519023	BLR-O8BX
5519058	BRR-REYL
5519076	BLP-RSY4
5519078	BLP-RSY6
5519081	BLP-RSY9
5519087	BLP-RSYL
5519091	BLP-RSYT
5519114	BLG-WCB7

Table 1: Overview of animal id and ring ids

Appendix 5

animal id	gender
5506092	male
5506106	male
5517525	female
5519012	female
5519017	male
5519021	female
5519023	male
5519058	male
5519076	female
5519078	female
5519081	female
5519087	female
5519091	female
5519114	female

Table 2: Genders for each bird in the used data

Appendix 6

During this project, a proportion plot, one or multiple interval plots and various nesting plots were created for each bird and each year. All these plots can be found at the Google Drive link below and may be viewed and downloaded for research purposes or personal use. If posting the plots somewhere public, I would appreciate it if you give me credit for the work on creating them. Keep in mind that I do not guarantee that this content is 100% accurate and am not responsible for any issues the use of this content may cause.

`https://drive.google.com/drive/folders/12z1hW7s4yJnOY8-7cbyA_VVaYL_P0QW8?usp=sharing`

In addition, the Interactive Clustering Tool that was developed for this project, along with the corresponding preprocessing method and analysis methods, is available on Github. The link can be found in the above Google Drive folder as well. The tool may be used for future analysis and may be improved and/or added to for research purposes, assuming that I will be credited for the original work on the tool. The tool may not be perfect and may still contain some mistakes. I am not responsible for any issues resulting from the use of the tool.