University of Zurich[UZH]

Identifying Anthropogenic Feeding Sites from GPS Tracking Data: A Case Study for Red Kites (Milvus milvus) in Western Switzerland

GEO 511 Master's Thesis

Author
Nathalie Heiniger
15-707-458

Supervised by
Dr. Martin Grüebler (martin.gruebler@vogelwarte.ch)
Patrick Scherler (patrick.scherler@vogelwarte.ch)

Faculty representative
Prof. Dr. Robert Weibel

30.09.2020
Department of Geography, University of Zurich

University of Zurich[UZH]

vogelwarte.ch

MASTER'S THESIS

GEO 511

Identifying Anthropogenic Feeding Sites from GPS Tracking Data: A Case Study for Red Kites (*Milvus milvus*) in Western Switzerland

Author
Nathalie HEINIGER
15-707-458

Supervisor and faculty member
Prof. Dr. Robert WEIBEL
University of Zurich

External supervisors
Dr. Martin GRÜEBLER
Swiss Ornithological Institute
Patrick SCHERLER
Swiss Ornithological Institute

Geographic Information Systems (GIS)
Department of Geography
University of Zurich

30 September 2020

Contact

Prof. Dr. Robert Weibel
University of Zurich
Department of Geography
Winterthurerstrasse 190
CH - 8057 Zurich

Dr. Martin Grüebler
Swiss Ornithological Institute
Seerose 1
CH - 6204 Sempach

Patrick Scherler
Swiss Ornithological Institute
Seerose 1
CH - 6204 Sempach

Nathalie Heiniger
nathalie.heiniger@uzh.ch

Abstract

Today, animal movement data can be used in a remote sensing approach to obtain environmental information and detect human activities. Using the example of red kites *(Milvus milvus)* in western Switzerland the aim of this thesis was to develop a location-based, data-driven method to identify locations of anthropogenic feeding sites. Anthropogenic feeding of red kites is a rather common habit within the rural and urban population of Switzerland. Survey-based research has located and quantified anthropogenic feeding within a study area in western Switzerland. However, this is a very time-consuming approach, especially for larger geographic areas. The developed individual-based approach combined context variables, kernel density estimation and revisitation analysis in a modelling framework. A key finding is that the methodology works best with data collected on breeding birds. For breeding birds 63% of potentially used anthropogenic feeding sites could be detected, with 60% and 75% detection rates for small and large sites, respectively. Non-breeding bird data, while performing well for large sized anthropogenic feeding sites with a detection rate of 80%, only detected 31% of potentially used anthropogenic feeding sites overall. In conclusion, while it is possible to detect small and large size anthropogenic feeding sites with breeding bird data, data of non-breeding birds only delivers reliable results for large size feeding sites. These findings highlight the behavioural differences of the age classes and indicate that the behaviour of breeding birds is more predictable by repeated visits to a location. In general, the results show how GPS tracking data of individual animals offer a new way to remotely sense environmental information.

Acknowledgements

This master's thesis would not have been possible without the thematic, technical, and personal support of various people who accompanied me through the process.

First of all, I would like to thank Prof. Dr. Robert Weibel from the Department of Geography at the University of Zurich, Dr. Martin Grüebler and Patrick Scherler from the Ecological research group at the Swiss Ornithological Institute in Sempach, for the many helpful meetings, their engagement, their expert advice, and their patience. A special thanks goes to Patrick Scherler for being so dedicated and responsive, for all the additional meetings and the provision of valuable *R* scripts.

I would also like to thank the Swiss Ornithological Institute for allowing me to use their large GPS movement data set of red kites.

Furthermore, my gratitude goes to Dr. Peter Ranacher from the University of Zurich for his excellent *R* programming support.

Many thanks go to my family and friends for all the support, understanding and encouragement during the process of working on this thesis. And finally, Kelvin Salmon, for his patience and moral support during this long process.

Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

1.1 Motivation

Today, animal movement data are used in a remote sensing approach to obtain environmental information and detect human activities (Navarro *et al.*, 2016). The red kite (*Milvus milvus*) is a widespread species in Switzerland that is known to exploit anthropogenic food sources (Mougeot, Garcia and Viñuela, 2011; Orros and Fellowes, 2015). Therefore, the GPS tracking data of red kites serve as a good example to develop a methodological approach in the context of anthropogenic bird feeding and its detection.

The Swiss breeding population gained international importance because red kite populations show declining or stagnating trends in their core distribution areas Germany, Spain and France (Aebischer, 2009; Swiss Ornithological Institute, 2019). However, in Switzerland red kites showed a continuous increase in the past decades and today they breed not only in the Swiss Plateau but also at altitudes above 800 metres (Aebischer, 2009; Knaus *et al.*, 2018; Swiss Ornithological Institute, 2020). The reasons for the increase in Switzerland, particularly in light of the contrasting development of surrounding countries, are widely unknown. To gain an understanding of the factors driving this recent population development of red kites, the Swiss Ornithological Institute has initiated a research project in 2015.

This master's thesis was conducted in collaboration with the Swiss Ornithological Institute and it focuses on anthropogenic feeding in Swiss private households or farms (Cereghetti *et al.*, 2019). Anthropogenic feeding affects the food availability of red kites, which in turn might affect the reproduction, survival and, through the spatial distribution of food also their likelihood to disperse, and ultimately the population development in Switzerland. The development of a location-based, data-driven approach to identify anthropogenic feeding sites in the study area is the key objective of this thesis, which aims to complement survey-based studies that identify locations of anthropogenic feeding sites. The use of ongoing technological advances offered by global positioning system (GPS) sensors for the collection of animal movement data and thus for gaining insight into environmental information, specifically human activities, is a key aspect of this thesis (Cagnacci *et al.*, 2010; Kays *et al.*, 2015; Navarro *et al.*, 2016).

1.2 Research Objectives

Changes in food availability have major impacts on the dynamics on a population and on an individual level, such as individual behaviour and movement decisions. As a result, anthropogenic feed-

ing sites become locations of ecological interest (Plummer *et al.*, 2015). Studies on bird feeding concluded that garden feeding leads to reduced winter migration (Plummer *et al.*, 2015). Therefore, specific behavioural changes of red kites might be decreased migration due to anthropogenic feeding or a trend towards communal roosting sites close to anthropogenic feeding sites in winter. Furthermore, individual movement patterns could change due to the attractiveness of anthropogenic feeding sites in search of food. However, large-scale and long-term influences on the ecology of bird communities are still unknown (Plummer *et al.*, 2019). Moreover, the extent of anthropogenic feeding is often unclear (Cereghetti *et al.*, 2019). Therefore, the aim of this master's thesis is to develop a method to identify and thus, quantify anthropogenic feeding sites based on GPS tracking data of red kites. Results serve as a basis for further research on impacts of anthropogenic feeding and allow the identification of anthropogenic feeding sites in different regions in Switzerland.

1.3 Research Gaps

Although observational research in Switzerland has found evidence that feeding of red kites by private households occurs regularly, especially during winter time, little information is available about the extent of these anthropogenic feedings and their spatio-temporal occurrence (Cereghetti *et al.*, 2019). Therefore, it is crucial to localise and quantify them to then further investigate their importance for the Swiss red kite population (Cereghetti *et al.*, 2019). Survey-based research to identify anthropogenic feeding sites has its limitations and is confined to a small spatial extent where time-consuming door-to-door surveys can be conducted. A remote sensing approach using GPS tracking data is less limited by the spatial extent and therefore represents an interesting opportunity to quantify anthropogenic feeding for larger areas. As will be detailed in Chapter 2, animal movements have been used to gain information about the environment in which the movement occurs and to detect human activities (Navarro *et al.*, 2016). But no location-based, data-driven approach has been developed to identify anthropogenic feeding at private households or specifically for red kite movement data. The methodology is developed for, and tested in, a study area in western Switzerland, with the intention of applying it to other regions of Switzerland.

1.4 Research Questions and Hypotheses

This master's thesis addresses the above-mentioned research gaps guided by the following research questions and hypotheses.

RQ A *What methods are required to identify daily anthropogenic feeding sites and those with lower feeding frequencies using GPS tracking data?*

The hypothesis is that anthropogenic feeding sites can be detected using red kite GPS data by applying a combination of density-based measures, particularly kernel density estimation, and recursive movement characteristics integrated with a revisitation approach. Higher GPS point densities are expected at anthropogenic feeding sites since most currently known anthropogenic feeding sites have a regular feeding pattern and most private households have been feeding the red kites for

several years (Cereghetti *et al.*, 2019). Furthermore, a revisitation-based approach is considered suitable, as Welti, Scherler, and Grüebler (2019) concluded that red kites respond more strongly to repeated placement of carcasses than other facultative scavengers. Hence, they exploit sites with high predictability of food occurrence and are therefore expected to regularly return to the same anthropogenic feeding site.

RQ A.1 *Do the different frequencies of anthropogenic feeding sites influence their identifiability?*
There is a great diversity within the feeding frequency and the amount of feeding of the different anthropogenic feeding sites. The hypothesis is that this variety of frequencies in food supply affects the density of GPS points and the revisitation rate at a location, thus influencing the identifiability of a feeding site. Therefore, an appropriate method might require different parametrisations or even different approaches. A solution could be to address this problem by creating frequency classes for the feeding sites (e.g., daily, three and more times a week, once a week, monthly) and the use of two different approaches. First, by using an individual-based approach, which attempts to identify anthropogenic feeding sites based on an individual's movement. This might be a good solution for identifying anthropogenic feeding sites that provide food on a very regular (i.e. daily) basis. Second, an approach aggregating over several individuals could be used to identify anthropogenic feeding sites that cannot be detected on an individual basis.

RQ A.2 *Does the identifiability vary amongst age classes of red kites?*
Red kites undergo behavioural changes over the course of their lives; therefore, individuals of different age classes may show different movement behaviours. This heterogeneity in behaviour potentially influences the frequency with which an individual makes use of an anthropogenic feeding site and the number of feeding sites used. Therefore, the hypothesis is that the identifiability of anthropogenic feeding sites is dependent on age classes (e.g., first year (fledgling), non-breeding, breeding individuals) of the red kite providing the GPS data.

RQ B *What environmental context variables must be considered to identify anthropogenic feeding sites?*

The hypothesis is that environmental context variables based on ecological knowledge about red kites and characteristics of anthropogenic feeding sites are crucial. It is expected that this additional information allows for a differentiation between natural feeding sites, nests, roosting sites, and anthropogenic feeding sites which are all sites with potentially high density of GPS localisations. Frequently visited forest areas are used by red kites for sleeping and resting (Aebischer, 2009) and are therefore not of interest for this thesis. Hence, including information on forest areas could reduce the number of locations indicating nests or roosting sites and focus on feeding grounds. Anthropogenic and natural resources display different temporal characteristics. While natural resources are usually not restricted to a certain time period during the day, anthropogenic resources tend to exhibit different levels of ephemerality. Anthropogenic food is usually only available for a short time during daytime due to human involvement (Yoda *et al.*, 2012). Since feeding takes place during the

day and red kites are diurnal animals only daytime GPS points are considered relevant for the task. Spatially, anthropogenic feeding takes place in or near residential areas due to human participation. Therefore, the inclusion of information on human-made structures at locations with high GPS point density seems appropriate and is considered essential for the successful identification of anthropogenic feeding sites. It is expected that this reduces the number of frequently visited areas by focussing on the anthropogenic aspect of the feeding. Since the data have a low, 1 hour temporal resolution, GPS locations indicating a visit to an anthropogenic feeding site might not have been recorded at the exact location of the site, and therefore the polygons of the kernel density estimation could be slightly displaced. Hence, the use of a buffer around settlement areas might be a reasonable option to only include highly used areas in proximity to buildings into further analysis.

1.5 Outline

- Chapter 2 provides background knowledge on the ecology of red kites and the impact of anthropogenic feeding on bird populations. Furthermore, necessary background information about movement ecology and the state of research is given.
- Chapter 3 introduces the study area and the data used in this master's thesis.
- Chapter 4 explains the methodological workflow developed to identify locations of anthropogenic feeding sites.
- Chapter 5 presents the results.
- Chapter 6 discusses the results and the applied methodology. It also answers the research questions.
- Chapter 7 provides conclusions and an outlook on future research.

Chapter 2

2 Theoretical Background

2.1 Red Kite Ecology

2.1.1 Physiology, Habitat and Reproduction

The red kite *(Milvus milvus)* is a diurnal raptor and the third largest native bird of prey in Switzerland (Swiss Ornithological Institute, 2019). Red kites (Figure 2.1) are around 56 to 73 cm long and wingspans in adult birds can reach up to 170 cm. Nevertheless, red kites weigh only 860 to 1400 grams, with females being significantly heavier than males (Aebischer, 2009). Red kites have narrow, long and angled wings. Their plumage is mainly reddish-brown, especially the body, the upper tail, and the wing coverts. Other typical features are the large white spots on the underside of the primary feathers and the greyish head of adult birds as well as the long, deeply forked tail. Their short legs are yellow with black claws and their beak is horn-coloured, only the base is coloured yellow (Aebischer, 2009).

Figure 2.1. First calendar year red kite (*Milvus milvus*) in flight. Image credit: P. Scherler.

While red kites need trees for breeding and resting, open landscapes such as fields, meadows, pastures, lakes, rivers, and forest edges are vital to access food resources. Therefore, preferred habitats are diversified and fragmented landscapes with pastures, heath lands, farmland mixed with trees or forests (Carter, 2001; Aebischer, 2009; Mougeot, Garcia and Viñuela, 2011; Heuck *et al.*, 2013). Red kites occur less commonly in regions with large forest areas or without trees. In Switzerland, red kite nests are found at heights of 15 to 30 metres, mostly on forest edges (Aebischer, 2009). In the

wild, birds can reach an age of up to 29 years (Aebischer, 2009). Reproduction starts between 2 – 4 years of age, with most birds breeding for the first time at the age of 3. Generally older birds, being more experienced breeders, have higher breeding success than young red kites (Aebischer, 2009; Mougeot, Garcia and Viñuela, 2011).

The annual cycle of the red kite population can be classified into different phases. The cycle starts with spring migration back to Switzerland initiating in January, with continuing migration and the occupation of breeding territories in February. After nest building, eggs are laid in late March, beginning of April, and incubated until chicks hatch after 30 days (Carter and Powell, 2019). In May the chicks are reared, until they fledge with 47 – 78 days of age in June and beginning of July (Bustamante, 1993). The young red kites become independent in August. July and August are part of the post-breeding season. In September and October, the autumn migration begins. November and December are spent in the non-breeding range before spring migration is initiated again (Carter and Powell, 2019).

2.1.2 Population and Distribution

Changes in animal population densities and distributions are strongly influenced by human activities causing changes in habitat structure and resource availability. Such impacts have also been observed in the development of red kite populations (Blanco and Montoya, 2004).

Red kites generally benefit from anthropogenically modified landscapes because the species is well adapted to open land and fragmented landscapes (Blanco and Montoya, 2004). However, other anthropogenic actions such as persecution, poisoning and nest plundering decimated the red kites populations throughout Europe in the 19th century and the first half of the 20th century (Aebischer, 2009). Between 1970 and 1990 a new decrease in several European countries occurred due to a variety of reasons, such as strong industrial agriculture in eastern Germany, accidental poisoning by poison meant for water voles in north-eastern France and illegal hunting in Spain (Aebischer, 2009). In contrast to the negative population trends in most European countries, Switzerland showed a continuous increase in the red kite population since the 1950s (Aebischer, 2009).

The distribution area of red kites is limited to central, western, and southwestern Europe. Scotland and southern Sweden mark the northern border of the red kite breeding area, whereas Spain and Italy set the southern boundary (Aebischer, 2009). Germany, Spain, and France are the three most important countries for the red kite and account for about 85% of the world's population (Mougeot, Garcia and Viñuela, 2011). Recently, the world's red kite population has been estimated to a size of around 25,000 – 33,000 pairs (Swiss Ornithological Institute, 2019).

In Switzerland, a continuing recolonisation of breeding areas and an expansion into altitudes above 800 metres and alpine valleys can still be observed today (Aebischer, 2009; Knaus *et al.*, 2019). For the period 2013 – 2016, the population size of red kites in Switzerland was estimated at around 2,800 to 3,500 breeding pairs, compared to only around 150 in the 1970s. The Swiss red kite population amounts to approximately 10% of the world's red kite population (Aebischer, 2009; Knaus *et al.*, 2018, 2019).

The reasons for the unparalleled increase in Switzerland, particularly in light of the contrasting development of surrounding countries, are widely unknown and the reason for increased research interest (Swiss Ornithological Institute, 2019).

Red kites are residents or short-distance migrants (Swiss Ornithological Institute, 2019). While breeding pairs in Great Britain and southern Italy are mostly sedentary, red kites breeding in Central Europe migrate to southern European countries due to adverse environmental conditions in Central Europe (Aebischer, 2009). However, since the 1960s the number of red kites and the number of communal roosting sites in winter in Central Europe has continuously increased (Aebischer, 2009). Communal roosting sites are located on single trees or tree rows, where red kites gather to spend the nights in winter (Figure 2.2). In Switzerland sometimes a dozen to up to 300 birds spend the nights on a few trees (Aebischer, 2009). The Swiss wintering population has grown to more than 3,000 individuals (Knaus *et al.*, 2018). Interestingly, the red kite's population has mainly grown in countries where large numbers of individuals spend the winter. Whether there is a causal relation between the growing breeding population and the wintering areas is unknown (Aebischer, 2009).

Figure 2.2. Red kites at a communal winter roosting site. Image credit: P. Scherler.

2.1.3 Foraging Behaviour

Red kites travel long distances when searching for food resources. Meadows, pastures, arable land, ploughed grain fields, and urban areas are used for foraging (Aebischer, 2009; Orros and Fellowes, 2015). To cover long journeys red kites use thermals to gain height and then glide for long distances. In their search for food, they usually fly at heights below fifty metres above ground scanning the land below for food. However, large suitable feeding locations can also be spotted from flight altitudes of several hundred metres (Aebischer, 2009). Red kites have excellent eyesight and it is their most important sense when searching for food (Carter and Powell, 2019). When foraging they usually approach a site carefully before grabbing a piece of food with their feet and fly to a safe place to

eat it while small prey can be eaten in flight (Aebischer, 2009). During the breeding season red kites show an increased need for food to feed their young (Hötker, Krone and Nehls, 2017).

Red kites are opportunistic scavengers and therefore have a very broad diet. Specifically, red kites feed on insects, small mammals, birds, reptiles, amphibians and fish species as well as carrion and are known to exploit anthropogenic food sources (Mougeot, Garcia and Viñuela, 2011; Orros and Fellowes, 2015). In the literature, this wide range of food consumption has been connected to the geographically dependent availability of different food sources, individual specialisation, and seasonal changes. The variety of food used by red kites is seen as a major advantage when food is scarce (Mougeot, Garcia and Viñuela, 2011; Carter and Powell, 2019).

2.2 Anthropogenic Feeding

2.2.1 Types of Anthropogenic Feeding and Impacts

The feeding of wild birds in gardens or backyards is a common human-wildlife interaction (Reynolds *et al.*, 2017). This leads to birds being influenced by the availability of anthropogenic food (Cereghetti *et al.*, 2019). Often the intention of private persons to feed birds is to improve survival during winter months (Reynolds *et al.*, 2017). But other motivations such as seeing birds from up close, feeling relaxed and connected to nature, are also important (Cox and Gaston, 2016; Reynolds *et al.*, 2017). While the public might associate anthropogenic food with feeding birdseed in gardens specifically for small passerines, also avian facultative scavengers such as corvids, gulls and raptors exploit anthropogenic food. However, these species feed on waste, carrion or livestock afterbirths (Reynolds *et al.*, 2017; Cereghetti *et al.*, 2019).

Changes in food availability have major impacts on the ecology and dynamics of bird populations which makes research on anthropogenic feeding important (Plummer *et al.*, 2015). By altering the food availability anthropogenic feeding sites can influence the size, distribution and behaviour of the population (Plummer *et al.*, 2019). Studies on other bird species concluded that supplementary feeding as well as milder winter conditions in the Northern Hemisphere are potential factors changing the migratory behaviour (Plummer *et al.*, 2015).

In Spain large feeding stations exist where dead livestock is dumped. Those places called 'muladares' attract primarily vultures but also red kites (Mougeot, Garcia and Viñuela, 2011). In research, scavenger feeding stations have been widely used in conservation programmes although their behavioural effects are still under discussion (Monsarrat *et al.*, 2013). Monsarrat *et al.* (2013) analysed the foraging behaviour of Griffon vultures (*Gyps fulvus*) in the Grands Causses (France). They classified possible feeding sites into three groups: (1) 'light feeding stations' characterised by carcasses placement by farmers at their farms, (2) 'heavy feeding stations' where carcasses from different farms are accumulated and (3) open grasslands, where carcasses appearance is random. Those sites show different levels of predictability. While (1) 'light feeding stations' and (2) 'heavy feeding stations' deliver spatially predictable food resources, (3) open grasslands are unpredictable. Monsarrat *et al.* (2013) concluded that predictable feeding stations were used preferably by the vultures compared to other, less predictable foraging areas within their habitats. Furthermore, there seemed

to be a tendency towards light feeding stations due to less intraspecific competition compared to heavy feeding sites. Bartumeus *et al.* (2010) found that seabirds or in general omnivorous animals alternate between natural and human-related food sources based on their availability. Yoda *et al.* (2012) stated that black-tailed gulls showed more ballistic movement when using anthropogenic food resources than when searching for natural ones, indicating their movement strategies are adapted to the food source. This served as an indication that gulls show knowledge of the spatial and temporal distribution of anthropogenic food resources and adapted their foraging strategies and movement patterns to the higher spatial predictability of anthropogenic sources (Yoda *et al.*, 2012).

The conclusions about the preference of predictable food resources, the effect of intraspecific competition at feeding sites, and the differing movement patterns, although made for Griffon vultures and gulls, are interesting to consider in the case of red kites in Switzerland (Yoda *et al.*, 2012; Monsarrat *et al.*, 2013). This suggests that red kites potentially show different movement patterns when using anthropogenic or natural feeding sites, respectively.

However, anthropogenic feeding in Switzerland takes place on a smaller scale than in 'muladares' or 'heavy feeding stations' and should be more associated with the 'light feeding stations' classified by Monsarrat *et al.* (2013). For the red kite study area in the Sensebezirk FR and Schwarzenburgerland BE it was found that intentional anthropogenic feeding does not necessarily lead to big aggregations of scavengers at the sites. Only at feeding sites that provide large amounts of food or do so during the winter season, such aggregations might occur (Cereghetti *et al.*, 2019). Cereghetti *et al.* (2019) concluded that a maximum daily average of almost 0.9 kg of food per km^2 was distributed within the study area. In this region 12.7% of rural and 4.6% of urban households that were surveyed frequently provided food for red kites. Furthermore, only 20% of the feeding events took place intentionally while 80% were unintentional. In Switzerland, the feeding rarely follows an exact time pattern because it happens in private households or on farms. Hence the food availability is dependent on factors such as the disposal of kitchen leftovers or afterbirths, which in turn are related to livestock birth. This indicates that whereas the time is unpredictable the location of private feeding sites is predictable (Cereghetti *et al.*, 2019). Therefore, one could conclude that red kites also slightly favour predictable, anthropogenic feeding sites over unpredictable resources. The preference of predictable resources would lead to a frequent use visible in GPS tracking data. Therefore, the detection of anthropogenic feeding sites should be possible using red kites tracking data.

2.2.2 Anthropogenic Feeding Sites

To be able to develop a methodology that can detect anthropogenic feeding sites in a Swiss context, it is necessary to first define the characteristics of such a site on a conceptual level. There are two different dimensions to be considered. First, the landscape characteristics specific to an anthropogenic feeding sites and second the behavioural aspect of red kites.

The focus of this thesis lies on spatially predictable anthropogenic feeding that takes place in an anthropogenic environment, particularly at or close to farms or private houses. Depending on the

amount of food and the frequency of feeding occurring at an anthropogenic feeding site it is classified to a different level (small or medium, large, unknown). The feeding can be intentional (e.g., placement of food in the backyard) or unintentional (e.g., compost heap), but in this thesis no distinction is made, as spatial predictability is the important aspect. Since red kites alternate between anthropogenic and natural food sources (Orros and Fellowes, 2015), it is important that a method can distinguish between locations of those two types. Hence, the connection of red kite locations to human-made constructions is crucial.

Temporally, there are regular and irregular anthropogenic feeding sites. The methodology focuses on the detection of anthropogenic feeding sites with regular feeding in sufficient quantity, as such locations are likely to attract red kites. Consequently, this leads to a high density of GPS localisations of an individual and possibly higher revisitation rates at such sites. Hence, locations expected to be feeding sites meet the following characteristics: a high density of GPS points, proximity to human-made constructions, and recurring visits.

The above-mentioned characteristics lead to the following aspects to be considered in an approach to detect anthropogenic feeding sites:

A *Density of GPS localisations*

B *Anthropogenic environment*

B.1 *negative buffer around forests (resting and nesting)*

B.2 *positive buffer around settlement areas*

C *Recurring, predictable feeding spots*

C.1 *Revisitation of an individual to the same area*

2.3 Movement Ecology

2.3.1 Movement Definition

Moving objects can be defined as 'entities whose positions or geometric attributes change over time' (Dodge, Weibel and Lautenschütz, 2008, p. 240). Animals or humans are dynamic objects that move in geographic space and are geographically referenced. Sequences of locational points, when ordered in time, result in trajectories (Dodge, Weibel and Lautenschütz, 2008; Buchin, Dodge and Speckmann, 2012). Hence, trajectories can be defined as 'discrete time series of measured locations' (Demšar *et al.*, 2015, p. 2). Or in other words, trajectories represent 'paths through space and time' executed by a moving object (Dodge, Weibel and Lautenschütz, 2008, p. 240). While the movement takes place continuously, location data is collected at discrete time intervals and according to specific temporal schedules. As a result, the location between two updates remains unclear, causing uncertainty about the course of the trajectory (Demšar *et al.*, 2015; Zheng, 2015).

2.3.2 Recording Animal Movement

In the past, data collection from wild animals was too limited to describe the real-world phenomena of animal movement behaviours. As a result, the field of animal movement fell toward the margins of ecological research (Kays *et al.*, 2015).

Patterns in animal movements provide valuable information on behaviour between individuals, foraging strategies, learning excursions, seasonal migration and critical resources such as food (Nathan *et al.*, 2008; Kays *et al.*, 2015; Slingsby and van Loon, 2016). With movement being the spatio-temporal expression of an animal's behaviour, it contains behavioural information and hence affects the animal's life history (Kays *et al.*, 2015; Gudmundsson, Laube and Wolle, 2017). Therefore, it is crucial to gain a better understanding of movement patterns, mechanisms as well as their causes and consequences to obtain profound knowledge about populations, but also individuals (Nathan *et al.*, 2008).

Not only movement in general, but especially recursive movement, which is defined as returns to already visited places, plays an important role in movement ecology (Berger-Tal and Bar-David, 2015). Empirical studies indicate that a large variety of animal species on different trophic levels show recursive movement patterns (e.g., bees (Williams and Thomson, 1998), birds (Garrison and Gass, 1999), large felines (Laundré, 2010) and ungulates (Bar-David *et al.*, 2009)) (Berger-Tal and Bar-David, 2015). Furthermore, the resource recovering rate might be the influencing factor controlling the temporal scale of the recursive movement pattern (Berger-Tal and Bar-David, 2015). In case of anthropogenic activity, the resource recovering rate could be regarded as the regularity with which a person displays food at a spatially predictable site. Recursive movement behaviour is only possible if an animal meets certain prerequisites (Berger-Tal and Bar-David, 2015). First, the animal must have basic cognitive abilities to perceive and identify locations with resource availability. This enables a targeted movement towards such places (Nathan *et al.*, 2008; Berger-Tal and Bar-David, 2015). In addition, the ability to remember the locations and their quality of food is crucial. Therefore, a good spatio-temporal memory to predict and be able to estimate the rate of resource recovery is important for animals with recursive movement behaviour (Berger-Tal and Bar-David, 2015). A second prerequisite for recursive movement behaviour is the degree of environmental predictability because memory does not prove useful if the environment is unpredictable (Berger-Tal and Bar-David, 2015). Anthropogenic feeding sites show those characteristics of increased predictability of food availability spatially and to a certain extent temporally (Monsarrat *et al.*, 2013).

For a long time, the analysis of animal movements relied on radio telemetry, especially VHF (very high-frequency) technology (Cagnacci *et al.*, 2010). In ornithological research radio telemetry was used to quantify mortality rates, habitat use or home range size and habitat selection (Grüebler *et al.*, 2008; Menz, Mosimann-Kampe and Arlettaz, 2009; Pfeiffer and Meyburg, 2015). However, a main disadvantage of this method is that receivers must be relatively close to an animals' transmitting device in order to determine its location.

Modern generations of tracking devices deliver new opportunities in the field of animal tracking research. The emergence of satellite telemetry and devices incorporating a global positioning system (GPS) receiver allow researchers to remotely track animal locations and movements (Cagnacci *et al.*,

2010; Pfeiffer and Meyburg, 2015). Telemetry using GPS led to (Cagnacci *et al.*, 2010; Kays *et al.*, 2015; Pfeiffer and Meyburg, 2015):

1. *The coverage of large spatial extents without time-consuming efforts*
2. *High precision and accuracy of positional data, increasing the spatial resolution*
3. *The possibility to record movements of animals with an increased temporal resolution and therefore, insights into short-term behaviour*

This growing amount of collected high-resolution spatio-temporal movement data led to research into animal tracking becoming increasingly important for the study of animal behaviour and the analysis of movement patterns (Nathan *et al.*, 2008; Kays *et al.*, 2015; Gudmundsson, Laube and Wolle, 2017). GPS and other sensors (e.g., accelerometer) have been integrated into miniaturised tracking tags to minimise the effects on an animal's behaviour and survival. A sensor should not weigh more than 5% of an animal's body weight (Kays *et al.*, 2015). However, there are trade-offs between the size of a sensor and its performance, such as spatial accuracy, temporal resolution and battery life (Nathan *et al.*, 2008). With the miniaturisation of sensors, it is now possible to collect trajectory data for a variety of animals, including most medium- or large-sized vertebrates, which has led to a better description and study of animal movements (Damiani, Issa and Cagnacci, 2014; Kays *et al.*, 2015). Furthermore, the decreasing costs of sensors capable of tracking movements resulted in higher numbers of tagged individuals (Gudmundsson, Laube and Wolle, 2017). Today, GPS tracking data can be remotely downloaded using satellite or cell phone communication networks, which makes real-time data about the tagged animals available (Kays *et al.*, 2015).

The technological advances led from data scarcity to an abundance of movement data with high levels of spatial and temporal detail (Demšar *et al.*, 2015; Gudmundsson, Laube and Wolle, 2017). However, with the increasing amount of movement data, new challenges regarding data management, processing, analysis and visualisation of these data emerged (Millspaugh and Marzluff, 2001; Nathan *et al.*, 2008; Demšar *et al.*, 2015; Gudmundsson, Laube and Wolle, 2017). Developing new, appropriate analysis and visualisation methods attracted the interest of various research fields (Demšar *et al.*, 2015). An interdisciplinary collaboration between movement ecology and information science shows a good opportunity to improve methodological developments (Demšar *et al.*, 2015). Furthermore, online repositories such as Movebank (www.movebank.org), hosted by the Max Planck Institute of Animal Behavior, facilitate access to large ecological and geographic databases and therefore improve the management, sharing, protection, analysis and archiving of movement data (Nathan *et al.*, 2008; Wikelski, Davidson and Kays, 2020).

Today, the movement data are increasingly applied not only to study the behaviour and movement patterns of animals, but also to gain insights into the environment in which the movement takes place. Hence, animals can be used as an instrument to deliver remotely sensed environmental information. This approach relocates the research's focus from the animal's movement behaviour towards the collection of environmental knowledge. A practical example using animals' spatial movement as remote sensing information is Navarro *et al.* (2016)'s approach to detect illegal dumping

activities in southern Spain. Based on foraging activity recorded by real-time GPS data from yellow-legged gulls the authors pinpointed an illegal waste dump. They uncovered that the site was repeatedly used by approximately 25% of the tagged birds (Navarro *et al.*, 2016). Another possible application is the use of albatrosses equipped with devices that detect radar emissions and record positions of vessels and birds to monitor illegal fisheries. In the future, this could allow the transmission of real-time information on the whereabouts of illegal fishing vessels to authorities (Weimerskirch *et al.*, 2018). This novel approach directs the research interest towards the detection of anthropogenic activities such as environmental crime and away from the study of the species itself. The detection of anthropogenic feeding sites used by red kites in the study area can also be assigned to this emerging field of research. However, in this master's thesis conclusions about the behaviour of red kites, the influence of anthropogenic feeding on red kites are still of interest and essential for the development of a valid method.

2.3.3 Utilisation Distribution

Ecologists are particularly interested in the analysis of GPS tracking data to analyse spatio-temporal dynamics in movement patterns, estimate utilisation distributions, classify or identify behaviour and link movement data with an environmental context (Demšar *et al.*, 2015).

Utilisation distributions (UDs) are a frequently applied method in ecological research. The utilisation distribution is 'the distribution of an animal's position in the plane' (Worton, 1989, p. 164). Another definition used by Van Winkle (1975) is 'the two dimensional relative frequency distribution for the points of location of an animal over a period of time' (Van Winkle, 1975, p. 118). According to these two definitions, UDs can be used to examine animal movement behaviour in space. The estimated utilisation distribution is widely used in home range research (Worton, 1989). The estimation of home ranges is conducted with the aim to extrapolate where an animal will go if it continues to move with similar movements (Fleming *et al.*, 2016). In this way an animal's space use pattern can be quantified. Home range areas are defined as a certain percent coverage region. Commonly 95% of the probability distribution of all localisations is used (Fleming *et al.*, 2015). Core areas of activity are often described by the 50% home range area (Harris *et al.*, 2012; Hötker, Krone and Nehls, 2017). Kernel density estimations (KDEs) of trajectory points are frequently used to estimate a probability density function, by placing a probability function on each observed location. These probabilities are then summed up into a surface which displays the utilisation distribution (Lyons, Turner and Getz, 2013; Demšar *et al.*, 2015). While traditional utilisation distribution approaches provide a static representation of space use there have recently been developments to add a more dynamic characteristic (Benhamou and Riotte-Lambert, 2012). Benhamou and Riotte-Lambert (2012) point out that additional information such as residence time or visit frequencies to areas of special interest can help to identify areas within a home range which are intensively used or repeatedly visited. This approach provides new insight into the movement behaviour of animals and the environmental characteristics influencing it.

2.4 Methods

2.4.1 Kernel Density Estimation

Kernel density estimation (KDE) is a widely used method of point pattern analysis in GIScience and for home range calculation in ecology (Downs, 2010; Benhamou and Riotte-Lambert, 2012; Lyons, Turner and Getz, 2013). The spatial variation in the density of points is turned into a smooth and continuous surface by a KDE. By using a kernel density estimator, the spatial intensity can be assigned to each position in a study area. Thereby, locations of unusually high spatial intensity of points, called 'hot spots', can be identified (Downs, 2010). KDE is a non-parametric method to estimate utilisation distributions (Worton, 1989). The bivariate kernel density estimate f at any point x can be mathematically expressed as (Downs, 2010, p. 17):

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{|x - X_i|}{h}\right) \tag{2.1}$$

In Equation (2.1) sample size n contains points $X_1, X_2, \ldots, X_n$, h is the bandwidth and $K(y)$ is the kernel function (Downs, 2010). In applications of the utilisation distribution $\hat{f}(x)$ is the estimated utilisation distribution at a defined position x. The kernel function is based on the weighted sum of the target points X_i within the selected bandwidth h. The kernel's bandwidth h influences the degree of smoothing and thus the distance within which a point influences the density estimate. A point close to a relatively large number of other points receives a high intensity, whereas a point with greater distance is given less weight (Downs, 2010). While the choice of the kernel function (e.g., Gaussian, Epanechnikov) which influences the kernel's shape is considered secondary in the literature as most functions deliver similar results, choosing the bandwidth h is crucial when dealing with kernel density estimation (Silverman, 1998; Downs, 2010; Fleming *et al.*, 2015). Choosing h too small results in spurious fine structures becoming visible, hence an under-smoothing of the estimate, whereas a too large h results in only broad features becoming visible, an effect called over-smoothing (Silverman, 1998; Fleming *et al.*, 2015). The sensitivity to bandwidth selection is a major limitation of the KDE method (Downs, 2010).

A lot of research has been conducted into the adequate bandwidth choice and several methods have been developed to optimise the KDE bandwidth (Fleming *et al.*, 2015). Bandwidth can be chosen either by algorithms (e.g., plug-in and cross-validation approaches) or arbitrarily (Downs, 2010; Fleming *et al.*, 2015). The reference bandwidth is a common and easy approach for the bandwidth selection, using the standard deviation of the data to estimate the bandwidth (Silverman, 1998). In the *R* package 'adehabitatHR', often used to calculate home ranges in ecology, the reference bandwidth is calculated based on the bivariate normal distribution (Calenge, 2019). A second method to select the bandwidth is least-squares cross-validation (LSCV), which is completely automatic (Silverman, 1998). For this method, an estimate from the data is constructed and the mean integrated square error (MISE) of the bandwidth is minimised (Worton, 1989). Furthermore, the bandwidth can

be chosen subjectively by looking at several resulting estimates. Thereby the parameter that best suits the research question is selected. The process of testing multiple bandwidths may give first insights into the data, but the choice remains a challenge (Silverman, 1998; Calenge, 2019). Figure 2.3 shows the application of a kernel density estimation to a spatial point pattern using a bivariate normal kernel and a fixed bandwidth.

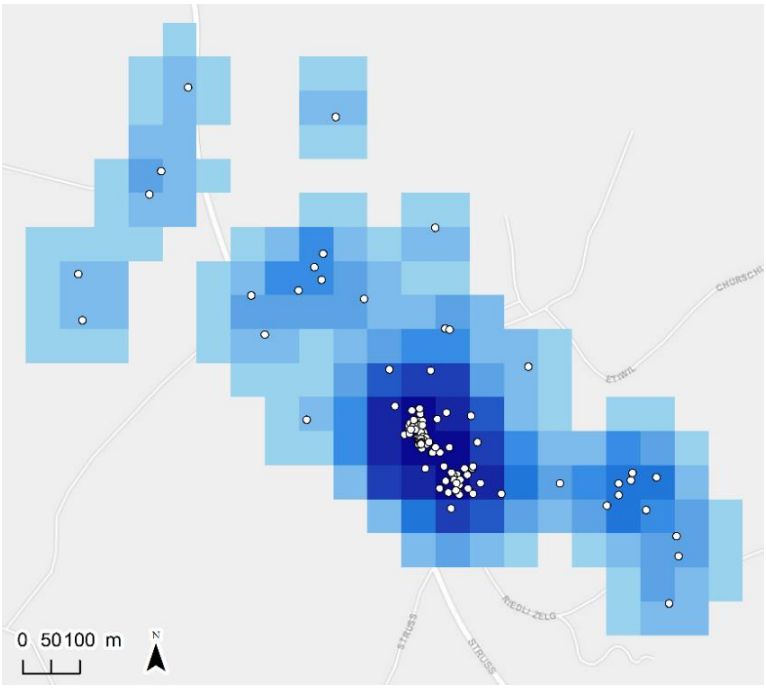


Figure 2.3. KDE applied to the spatial point pattern of a red kite's GPS locations. Base map: Esri, HERE, Garmin, OpenStreetMap contributors and the GIS user community.

2.4.2 Revisitation Analysis

As mentioned above in ecology (Section 2.3.2), recursive movement can be observed in a variety of species (Berger-Tal and Bar-David, 2015). The purpose of revisitation analysis is to obtain information on the return of an animal to specific areas throughout its lifetime (Berger-Tal and Bar-David, 2015; Mcgrady *et al.*, 2018). Revisitation takes place on different time scales such as daily, seasonal or annual, and across years depending on the species, as well as on different spatial scales (Berger-Tal and Bar-David, 2015; Bracis, Bildstein and Mueller, 2018). Various methods are available to study revisitation patterns (Bracis, Bildstein and Mueller, 2018). Here, the focus lies on the revisitation approach proposed by Bracis, Bildstein and Mueller (2018), which is implemented in the 'recurse' package in the *R* environment.

Bracis, Bildstein, and Mueller (2018) developed a method that focuses on point-based or visit-based analysis for temporal patterns of revisitation. The algorithm calculates revisits to specific locations in a movement trajectory or to arbitrary locations (Bracis, Bildstein and Mueller, 2018). By specifying locations, the analysis can be restricted to pre-identified sites (e.g., ground truth data). This makes the method applicable to conduct revisitation analysis to already identified locations that show certain ecological features. Moreover, the method can be used to quantitatively identify frequently used

sites along a movement trajectory or to investigate patterns of revisitation for single individuals and several individuals in order to uncover locations that are revisited across the population (Bracis, Bildstein and Mueller, 2018).

The method calculates different revisitation measures using a circle with a user-defined radius that moves along the trajectory or user-specified locations (Figure 2.4). For each location, the number of segments of the trajectory passing through the circle is counted to assess the number of revisitations (Bracis, Bildstein and Mueller, 2018). Instead of choosing a radius around locations, it is also possible to calculate revisits to a polygon where the number of trajectory segments entering and exiting the polygon is counted to evaluate the number of revisits. Additional measures, such as time between visits, time spent at each visit and residence time, which is the sum of time spent during all visits, are automatically calculated.

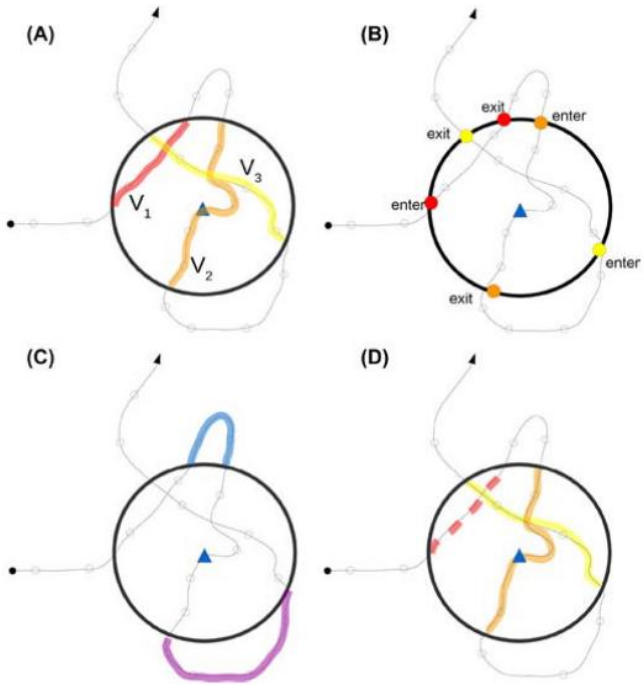


Figure 2.4. Graphical representation of revisitation calculations by the 'recurse' R package (Bracis, Bildstein and Mueller, 2018, p. 3).

The revisitation calculations when using user-specified radii along the movement trajectory are visualised in Figure 2.4. **(A)** Three visits (V_1, V_2, V_3) are counted to the specified circle. **(B)** For each visit entrance and exit times are calculated by linear interpolation between locations inside and outside the circle. With this information the time spent within the circle is determined. **(C)** For the revisits (V_2, V_3) the time since the last visit is calculated. **(D)** If trajectory segments only cross the circle (no observations inside) no visit is counted. Therefore, if data are not sampled finely enough, visits could remain undetected (Bracis, Bildstein and Mueller, 2018).

The *R* package 'recurse' accepts data supplied as data frame or as 'Move' or 'MoveStack' objects. According to Bracis, Bildstein and Mueller (2018) irregular sampling or a few gaps in the data should not lead to problems but nevertheless regularly sampled data are favoured. To obtain more accurate

calculations of the number of revisits or the visit duration it is preferable to have more than one trajectory point inside the radius. Therefore, it is recommended that the radius is not considerably smaller than the step length. Furthermore, the radius size should be larger than GPS measurement errors. However, in this master's thesis, the function calculating revisitations to a predefined polygon is of key interest.

2.4.3 Generalised Linear Mixed Model (GLMM)

To make predictions about locations of anthropogenic feeding sites a suitable prediction model is required. In ecology, basic statistical methods assuming normally distributed data are often not appropriate. Many applications in ecology deal with binary data (e.g., presence or absence of a species at a site or of a disease in an individual), proportions (e.g., infection rates) or counts (e.g., clutch sizes of storks) (Vicente *et al.*, 2006; Bolker *et al.*, 2009; Zuur *et al.*, 2009). Ecological problems often contain grouped data (e.g., individuals, blocks in studies) and therefore hidden error structures. To account for that random effects are included in the model. Instead of trying to transform non-normal data to fit the use of classical statistical methods, Bolker *et al.* (2009) suggest the use of methods matching the non-normal and random effect characteristics of the data. A generalised linear mixed model (GLMM) offers this possibility. GLMMs are a combination of linear mixed models (LMM) and generalised linear models (GLM) which are both commonly used in ecology and evolution (Zuur *et al.*, 2009). Linear mixed models (LMM) are used to incorporate random effects into linear models, whereas generalised linear models (GLM) provide an approach for non-normal data by using link functions and distributions of the exponential family (e.g., normal, Poisson, binomial). But GLMs assume independent data, which does not always hold in ecological applications. Here, GLMMs which allow for the correlation between observations (e.g., temporal auto-correlation, a bird has several observations or polygons) or nested data structures, offer a solution (Zuur *et al.*, 2009).

Although GLMMs show new opportunities for analysis in ecology and evolution, they also present challenges. Bolker *et al.* (2009) reviewed 537 ecology and evolution papers using GLMM analyses and found that in 58% of the cases GLMMs were inappropriately used. Furthermore, Zuur *et al.* (2009) point out that GLMMs 'are on the frontier of statistical research' (Zuur *et al.*, 2009, p. 323), suggesting that available information about GLMMs for ecologists is rare and somewhat technical.

GLMMs consist of fixed-effect parameters (effects of covariates and interactions, e.g., revisits per polygon) and random-effect parameters (standard deviations of random effects, e.g., variation in revisits per polygon across individual birds). GLMM estimation fits these parameters by maximum likelihood (ML) which is computationally intensive. Therefore, in research approximations of the likelihood are used to estimate GLMM parameters (Bolker *et al.*, 2009). Pseudo- and penalized quasi-likelihood (PQL) is the most widely used approximation but is also considered the least accurate. When standard deviations of random effects are large, it is known to show biased parameter estimates. Laplace approximation allows for more accurate approximations but is still not extremely computationally intensive (Bolker *et al.*, 2009). The choice of an appropriate estimation technique depends on expectations regarding computation time, availability of software packages and on the complexity of the model at hand (Bolker *et al.*, 2009).

According to Bolker *et al.* (2009) there are three different types of inference that can be performed after estimating parameter values with a GLMM: (1) hypothesis testing, (2) model comparison and (3) Bayesian approaches. In hypothesis testing test statistics (e.g., *F* statistics) are used to determine whether or not the null hypothesis can be rejected by comparing the distribution of estimates to the expected distribution under the null hypothesis. In model selection, the fit of a set of candidate models is compared. Either hypothesis tests or information theoretic approaches are used to establish a ranking of the models based on their expected predictive power.

In this master's thesis (2) model comparisons by information theoretic approaches will be described in detail. Information theoretic approaches provide a solution for the selection of a 'best' approximating model and a ranking from 'best' to 'worst' model for a set of predefined models based on the data. Since full reality cannot be captured in a model it is crucial to look for a model that approximates the effects given by the empirical data (Burnham and Anderson, 2002). For this reason, Akaike (1973) developed a methodology for the selection of the 'best' or parsimonious model applicable on the analysis of empirical data (Burnham and Anderson, 2002). He proposed the Akaike's information criterion (AIC), which is based on the relationship between the Kullback-Leibler distance and Fisher's maximised log-likelihood function (Burnham and Anderson, 2002). The Akaike's information criterion expresses 'an estimate of the expected, relative distance between the fitted model and the unknown true mechanism that actually generated the observed data' (Burnham and Anderson, 2002, p. 61). The AIC can be mathematically expressed as (Zuur *et al.*, 2009, p. 121):

$$AIC = -2 \times L(\theta) + 2 \times p \tag{2.2}$$

In Equation (2.2) L is the likelihood and p the number of parameters in θ. In other words, AIC is 'twice the difference between the value of the likelihood L (measure of fit) and the number of parameters p (penalty for model complexity) in θ' (Zuur *et al.*, 2009, p. 121). In conclusion, the AIC contains a term for the measure of fitness of the model and the complexity of the model (Zuur *et al.*, 2009). For more technical and mathematical information, see Burnham and Anderson (2002) or Akaike (1973). In practical applications, the AIC is calculated for each candidate model and the model with the minimum AIC value is considered the 'best' fitted model. The differences Δ_i in AIC values between the candidate models are of interest and not the absolute value which is influenced by sample size and constants (Burnham and Anderson, 2002). Therefore, the AIC cannot be used to compare models of different data sets. In some cases, no single best model can be established but rather a group of similarly appropriate models and another of inappropriate models. Thus, the AIC often leads to the identification of models with similar suitability (i.e. within $1 - 2$ AIC units of the minimum AIC) and, conversely, ones with poor explanatory power for the data (Burnham and Anderson, 2002). Applying the model selection to a new data set likely changes the ranking of candidate models (Burnham and Anderson, 2002). However, Akaike's Information Criterion is a respected choice for model selection in linear modelling and in linear mixed modelling (Zuur *et al.*, 2009).

In linear mixed effects modelling, a good approach to establish candidate models is to use a top-down strategy and then carry out model selection based on the AIC. The top-down strategy suggests starting with as many fixed effects (explanatory variables) as possible and their interactions included in the model. In a next step, the interaction with the highest, non-significant p-value can be omitted and the model is refitted. This step is repeated and finally, the AIC values of all candidate models are analysed (Zuur *et al.*, 2009).

2.4.4 Performance Measures and Classification

In ecology, presence-absence models are often used to predict the presence or absence of species based on habitat information (Fielding and Bell, 1997). This master's thesis deals with a model predicting presence-absence data and therefore, similar methods and performance measures are regarded suitable. To summarise the performance of a presence-absence model a 2 x 2 confusion matrix, also known as an error matrix, can be created indicating instances correctly predicted or confused by the classifier (Fielding and Bell, 1997; Tharwat, 2018). Table 2.1 shows an exemplary confusion matrix with the four possible outcomes: true positives (TP), true negatives (TN), indicating correctly classified instances, and false negatives (FN) and false positives (FP), representing incorrectly classified instances. False positives (FP) or Type I errors arise if the instance is classified as positive when the actual class is negative. False negatives (FN) or Type II errors occur if the instance is classified as negative when the actual class is positive (Fielding and Bell, 1997; Powers, 2011; Tharwat, 2018). In the case of the presented research, the true positives (TP) are polygons with an anthropogenic feeding site within their extent that were correctly predicted by the model. True negatives (TN) indicate polygons where no anthropogenic feeding site is reported, and the model predicted the absence of anthropogenic feeding. False positives (FP) stand for polygons where the presence of an anthropogenic feeding site was predicted but the ground truth information suggested the absence of anthropogenic feeding. False negatives (FN) are areas where the model predicted the absence of anthropogenic feeding while the ground truth stated the presence of anthropogenic feeding.

Table 2.1. Illustrative example of a confusion matrix from Powers (2011, p. 38); adapted by N. Heiniger. Colour indicates correct (green) and incorrect (red) counts.

		Actual	
		Positive (P)	**Negative (N)**
Predicted	**True (T)**	TP	FP
	False (F)	FN	TN

Imbalanced data sets introduce challenges when creating predictive models due to the underrepresentation of the minority class. However, there are various strategies to deal with imbalanced data sets. The data distribution can be changed during pre-processing to obtain a more balanced distribution, or the imbalance can be dealt with in the prediction post-processing (Branco, Torgo and

Ribeiro, 2016). An example for the latter case is to choose a pre-defined threshold to discriminate between classes in the classification process (Hernández-Orallo, Flach and Ferri, 2012). This is frequently done to evaluate results of the logistic regression where a classification threshold must be defined in order to dichotomise the prediction values. Often a default threshold of 0.5 is used for the validation of model performance to map the continuous prediction values to a binary category. A value equal or greater than the threshold indicates one class (presence, positive event) and smaller values the other class (absence, negative event). The threshold 0.5 works properly for well-balanced data but is unsuitable if the data is skewed. Therefore, the threshold value should be assessed, especially in logistic regression where predictions are biased towards the larger class (Fielding and Bell, 1997). When choosing a classification threshold there is a trade-off between the Type I (FP) and Type II (FN) errors. Therefore a threshold (cutoff) should be chosen based on context and costs associated (Fielding and Bell, 1997; Baldi *et al.*, 2000). Questions could be whether tolerance is greater for false negatives or false positives. Is it more important to have few false negatives but it does not matter whether some false positives occur? One must assess what is most important for the analysis or which errors would lead to worse consequences (Fielding and Bell, 1997).

A change in the classification threshold results in different values for the four fields of the confusion matrix and therefore most performance measures are threshold-dependent (Fielding and Bell, 1997; Powers, 2011; Tharwat, 2018). Some common measures derived from the confusion matrix (Table 2.1) are accuracy, recall (sensitivity), precision, F-measure, the Receiver Operating Characteristic (ROC) curve and its Area Under Curve (AUC) (Maratea, Petrosino and Manzo, 2014).

ACCURACY is often used to assess the probability of success but leads to misinterpretation of a model's performance with highly imbalanced data. If a model misclassifies most incidents of the minority class, but can predict the majority class incidents well, the accuracy value could be very high (Maratea, Petrosino and Manzo, 2014). The accuracy is calculated as (Maratea, Petrosino and Manzo, 2014, p. 334):

$$Accuracy = \frac{TP + TN}{TP + FN + FP + FN} \tag{2.3}$$

RECALL, also called sensitivity or true positive rate, is the proportion of correctly predicted real positives (Powers, 2011). Recall indicates which proportion of the actual positive cases a classifier can correctly identify (Powers, 2011). In the case of anthropogenic feeding sites, it would reflect the proportion of correctly predicted feeding site polygons at the locations of actual anthropogenic feeding sites (ground truth). Recall is defined as (Powers, 2011, p. 38):

$$Recall = \frac{TP}{TP + FN} \tag{2.4}$$

PRECISION reflects the proportion of predicted positive cases that are actual positives. Precision is a common and popular measure as it delivers a simple and intuitive interpretation. A precision of 0.5 translates to 50% correct predictions among all positive predictions (Saito and Rehmsmeier, 2015). In other words, 50% of the actual anthropogenic feeding sites are correctly predicted. Precision is defined as (Powers, 2011, p. 38):

$$Precision = \frac{TP}{TP + FP} \tag{2.5}$$

F-MEASURE is the harmonic mean of recall and precision and thus allows to integrate the two measures. It can be expressed by the general equation (Maratea, Petrosino and Manzo, 2014, p. 335):

$$F - Measure = (1 + \beta^2) \times \frac{Precision \times Recall}{(\beta^2 \times Precision) + Recall} \tag{2.6}$$

β can be adjusted to give different weights to precision or recall. The higher β the more weight is attributed to recall in comparison to precision (Maratea, Petrosino and Manzo, 2014).

Recall, precision and F-measure emphasise the positive incidents, predictions and error rates, whereas no information is given on how well the model handles true negatives (Powers, 2011). Only three cells of the confusion matrix are used for these metrics (Maratea, Petrosino and Manzo, 2014). Like the accuracy, the F-measure is sensitive to data imbalance (Maratea, Petrosino and Manzo, 2014; Boughorbel, Jarray and El-Anbari, 2017).

MATTHEWS CORRELATION COEFFICIENT (MCC) and AUC are more robust measures in case of data imbalance according to Boughorbel, Jarray and El-Anbari (2017). Furthermore, the MCC uses all four quantities of the confusion matrix and therefore provides a better summary of the performance than the F-measure (Saito and Rehmsmeier, 2015; Boughorbel, Jarray and El-Anbari, 2017). The MCC is calculated with the following equation based on the confusion matrix (Boughorbel, Jarray and El-Anbari, 2017, p. 5):

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(FP + FN)(TN + FP)(TN + FN)}} \tag{2.7}$$

The range of MCC is between -1 and $+1$. A value of $+1$ stands for complete agreement between prediction and observation, -1 for complete disagreement and 0 for random predictions (Baldi *et al.*, 2000).

All of the above-mentioned measures to evaluate a classifier's performance are bound to a defined classification threshold (Saito and Rehmsmeier, 2015).

The **ROC CURVE** (Figure 2.5) illustrates a model's performance over all thresholds instead of a single subjectively chosen one (Fielding and Bell, 1997; Lobo, Jiménez-Valverde and Real, 2008). The ROC curve delivers a good overview of the model's performance by showing the trade-off between the true positive rate (proportion of actual positives correctly identified; 'hit rate') and the false positive rate (proportion of actual negatives wrongly classified as positives; 'false alarm rate') (Baldi *et al.*, 2000; Saito and Rehmsmeier, 2015).

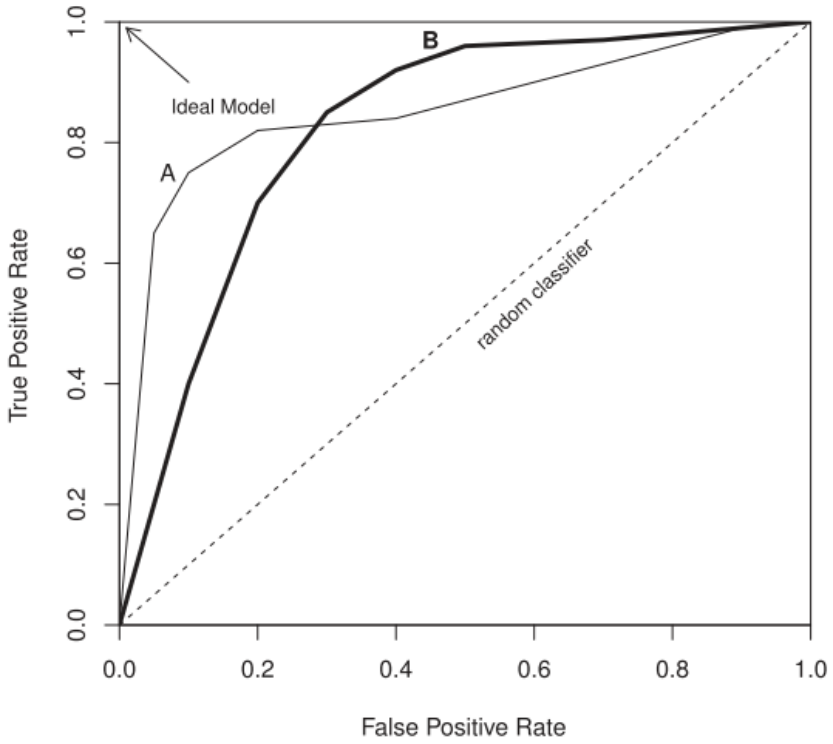


Figure 2.5. ROC curves show the performance of different classifiers at all classification thresholds (Branco, Torgo and Ribeiro, 2016, p. 12). The ideal prediction model would result in a point in the upper left corner, showing perfect classification. Models with curves above the diagonal line show better performance than random classifiers and those below show worse performance.

AUC, the area under the curve of the receiver operating characteristic (ROC), is a threshold-independent measure that is often used to assess the accuracy of probability prediction models (Lobo, Jiménez-Valverde and Real, 2008; Saito and Rehmsmeier, 2015). The AUC summarises the ROC curve into a single measure for the overall accuracy. Its values range from 0.5 – 1, where 0.5 means that the accuracy is no better than chance in case of balanced data (diagonal line in Figure 2.5). It can be used to quantify how accurately the model can discriminate between the positive and negative class (Fielding and Bell, 1997). For binary classification of imbalanced data, the AUC is often referred to as a valid performance measure (Zou *et al.*, 2016).

The **PRECISION-RECALL PLOT**, however, is more informative in case of imbalanced data sets than the ROC plot according to Saito and Rehmsmeier (2015). They state that the precision-recall plot (Figure 2.6), should be used in addition to the ROC. Both are visual representations of the performance of a model. For highly imbalanced data precision-recall plots are recommended as ROC would produce optimistic results about the classifier's performance (Davis and Goadrich, 2006; Branco, Torgo and Ribeiro, 2016).

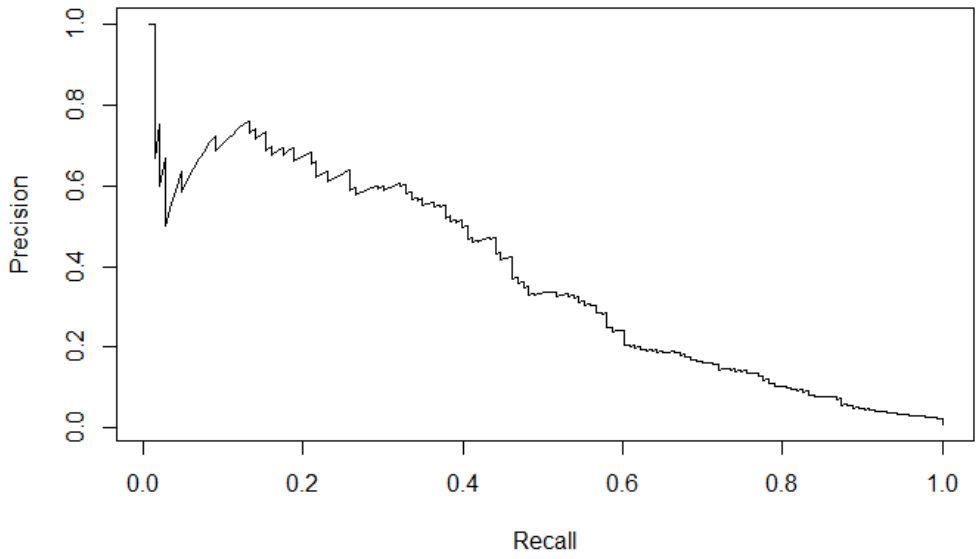


Figure 2.6. Precision-Recall curve.

Chapter 3

3 Study Area and Data

3.1 Study Area

The red kites project of the Swiss Ornithological Institute started in 2015 and is ongoing. The field work is focussed on a study area located in western Switzerland, more precisely in the Sense district of the canton of Fribourg and extending into the Schwarzenburgerland in the canton of Bern (Figure 3.1) with an area of 387.5 km^2. Red kites are tagged with GPS transmitters, feeding supplementation experiments take place and GPS movements are monitored to detect dead birds in time to retrieve carcasses and analyse the cause of death. During the breeding season, the work is focused on finding nests and monitor the breeding success of the red kites.

In this master's thesis, the spatial extent of the study area with a 5 km buffer is used as a basis for all analyses.

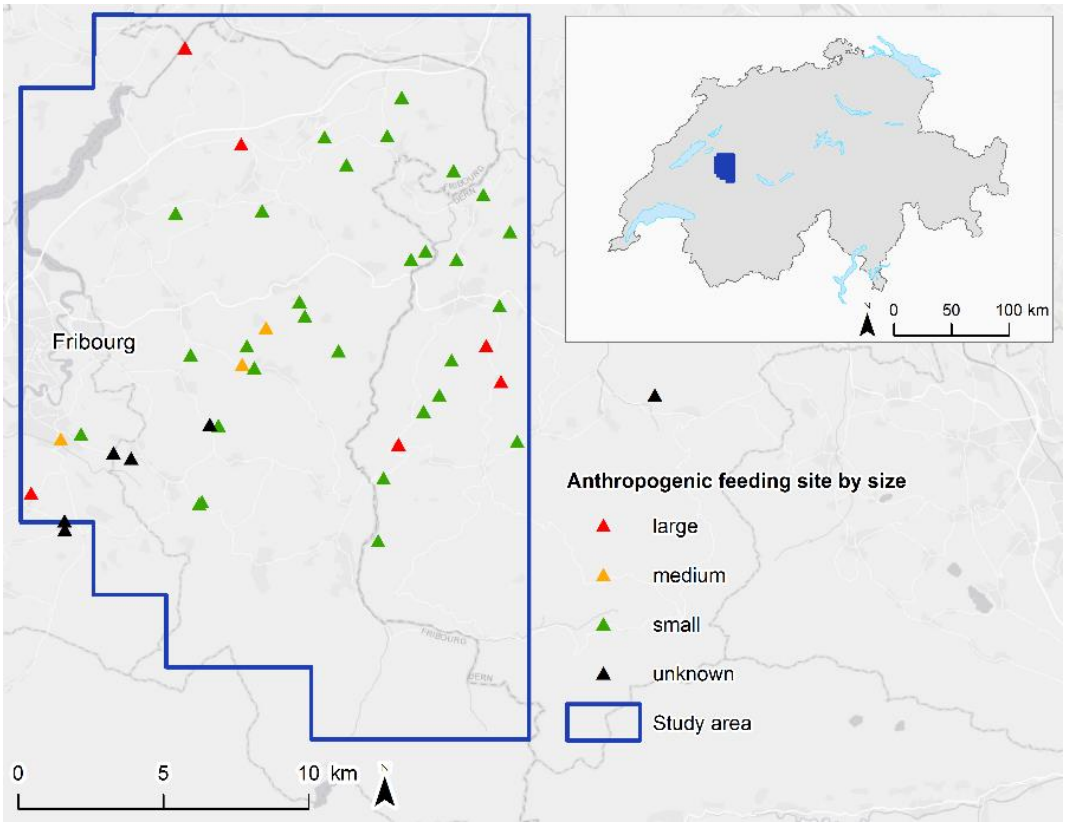


Figure 3.1. Study area with anthropogenic feeding sites (Swiss Ornithological Institute) and its location within Switzerland. Base map: Esri, HERE, Garmin, OpenStreetMap contributors and the GIS user community.

3.2 Data

3.2.1 Red Kite Trajectories

All movement data were collected by the Swiss Ornithological Institute in Sempach in terms of the red kite project. The data were available in Movebank, the online database for animal tracking data (Wikelski, Davidson and Kays, 2020). The data set contains the trajectories of 420 red kites equipped with solar-powered GPS transmitters (Figure 3.2) from two different models: Ecotone (SKUA/CREX type) and Milsar (M9/S9 type). Ecotone Telemetry GPS loggers have been deployed since 2015 and Milsar Telemetry loggers since 2017. So far, more than 5 million GPS localisations have been obtained during the project duration. The GPS transmitters provide hourly data, with some interruptions due to the limited battery life.

In this master's thesis the season 2018/2019, spanning from 01.04.2018 to 31.03.2019, was used for analysis. 412 red kites recorded GPS locations during this time period with 233 birds having GPS localisation within the study area of interest. For the study area a total of 1,077,374 GPS data points were available with 227 birds having sufficient data points (> 5 relocations every month) to be used for the kernel density estimation. Additional information for the individual red kites such as age class, sex, or territorial use was provided by the Swiss Ornithological Institute in Sempach.

Figure 3.2. Red kite in flight with a GPS transmitter attached on its back. Image credit: P. Scherler.

The 227 individuals that were used for the KDE analysis, were categorised into three age classes based on their age, occupation of a territory or breeding attempts: fledgling (first year of life), non-breeding birds, and breeding birds (Table 3.1). A rough division of the season into breeding (March – July) and non-breeding season (August – February) was performed and considered sufficient for this analysis even though in ecology a more detailed classification exists (Section 2.1).

Table 3.1. Age classes of red kites used for kernel density estimation.

Age class	Number of individuals
Breeding bird	80
Non-breeding bird	108
Fledgling	39
Total	227

3.2.2 Land Cover Information

VECTOR25 is the digital landscape model of Switzerland and is based on the national map 1:25'000 (Bundesamt für Landestopografie swisstopo, 2007). The location error is between 3 and 8 metres. VECTOR25 was used for the land cover data on the primary land cover class 'forest', specifically the two subcategories 'forest' and 'open forest'.

For the information on the anthropogenic environment, represented by buildings, the swissTLM^{3D} data set was obtained (Bundesamt für Landestopografie swisstopo, 2020b). Since no 3-D buildings were required for the analysis only the footprints of the buildings were used as a polygon layer. The geometric accuracy is 0.2 to 1.5 metres (Bundesamt für Landestopografie swisstopo, 2019). Furthermore, the swissTLM^{3D} object classes tree and shrub rows were used in the evaluation to determine the association of prediction polygons with these land cover classes.

SwissBOUNDARIES^{3D} was applied to limit the spatial extent of the data used in the revisitation analysis. It contains administrative units and the borders of Switzerland and the Principality of Liechtenstein (Bundesamt für Landestopografie swisstopo, 2020a).

3.2.3 Anthropogenic Feeding Sites

Ground truth data for anthropogenic feeding sites in Switzerland were provided by the Swiss Ornithological Institute. The data set was mostly established by Cereghetti *et al.* (2019) in a public survey of a sample of randomly selected households within the study area and supplemented by reports directed to the Swiss Ornithological Institute. Relevant for this master's thesis were 44 sites with active feeding practice in the season 2018/2019 and geographic coordinates available. Additional information about the frequency (e.g., daily, several times per week, weekly, 14 days, monthly, unknown, rarely) and the approximate amount of food (e.g., 100g, 300g, 500g, 1000g, 1500g, 2000g, 3000g, etc.) as well as the intentionality, the type of food (e.g., slaughter waste, meat, pet food, etc.) was available. Moreover, the Swiss Ornithological Institute provided a classification of the feeding sites in four classes: small, medium, large, unknown, based on a combination of the frequency and the amount of feeding. In addition, the data indicate whether a site carries out restricted feeding (i.e. only in winter, snow coverage) or all year round. Site characteristics (e.g., garden, meadow, farm) and the number of red kites observed at a particular site is recorded as is the time period during which the anthropogenic feeding has been taking place (e.g., 5 – 10 years, > 10 years, etc.).

Chapter 4

4 Methodology

The data cleaning, preparation, implementation, data analysis and evaluation were performed in the *R* language and environment (version 3.6.1; R Core Team, 2019) using the RStudio integrated development environment (IDE). The coordinates of all data sets were converted to the Swiss projected reference system (CH1903+/LV95).

The data for the time period 01.04.2018 – 31.03.2019 were directly downloaded as a 'MoveStack' object from www.movebank.org to the *R* environment. The data contained both live feed and manually uploaded data. For live feed data the timestamp indicates the minute when the position was taken, while the manually uploaded data have a timestamp with correct information about the second. For the detection of feeding sites, the second of recording a GPS localisation was considered irrelevant, so both data sets, live feed and manually uploaded data, were used for the analysis. The date and time columns of the data frame were transformed into a consistent format. The methodological workflow was developed as an individual-based approach due to a potential future application in areas with only a few tagged individuals. To obtain a first impression of the red kite trajectory data, some exploratory data analysis, including visualisations of exemplary trajectories, was conducted (Appendix A). A step-by-step flow chart with decision-making arguments of the complete methodology can be found in Appendix B.

4.1 Pre-processing

To achieve a clean and clearly arranged data set of the GPS points for the upcoming analyses only variables necessary for the task were retained and non-relevant ones discarded. Among the relevant variables were the local identifier, which identifies the individual red kite, the timestamp with date and time details, and coordinate information.

To remove erroneous GPS locations, the data set was analysed for spatial and velocity outliers. Locations outside Europe were discarded because the habitat of red kites is mainly Europe and all birds used in this study were tagged in the study area in Switzerland. After excluding spatial outliers, velocity for the remaining points was calculated based on step length and time lag of consecutive positions. Records with velocity values higher than 33.3 m/s (= 120 km/h) and their subsequent record were removed. Raptors rarely show higher velocities but might do so for short-term behaviour like hunting (Pennycuick, Åkesson and Hedenström, 2013; Vansteelant *et al.*, 2015; Hart *et al.*, 2018).

Relevant for this thesis was the spatial extent of Switzerland for an accurate revisitation analysis and the study area for a detailed analysis of the identification of anthropogenic feeding sites. Therefore, the swissBoundaries^{3D} vector data set with a 5 km buffer was used to filter the GPS data points for

the revisitation analysis (Bundesamt für Landestopografie swisstopo, 2020a). For the data used in the kernel density estimation, only GPS points within the spatial extent of the study area and a 5 km buffer were retained to ensure the identifiability of an anthropogenic feeding site at the border or just outside the study area.

To prepare the data optimally for the calculation of core areas by kernel density estimation and thus the task of identifying anthropogenic feeding sites, the number of GPS data points was further reduced. Based on ecological knowledge, GPS locations in forest areas were excluded. These locations mark resting places, nests or roosting sites, which are not relevant for this research focussing specifically on anthropogenic feeding. A 20 m buffer around forest areas was applied to take GPS errors into account and because the law on forests and natural hazards of the canton of Fribourg states that the construction of non-forest related structures at a distance of less than 20 metres from the edge of the forest are prohibited (Der Grosse Rat des Kantons Freiburg, 2019). As anthropogenic feeding is expected to take place in connection with human-made constructions such as houses this buffer seemed reasonable.

By calculating the daily period of daylight for each GPS location all points of darkness where excluded. GPS locations recorded between dawn and dusk were retained since red kites are diurnal birds and anthropogenic feeding is expected to take place during daytime due to human involvement (Aebischer, 2009; Yoda *et al.*, 2012). For the kernel density estimation more than 5 relocations per individual and month were needed, therefore this was another factor to be included.

4.2 Kernel Density Estimation

4.2.1 Generating Core Area Polygons

Through kernel density estimation the main areas of activity of red kites can be identified. In a first step kernel density estimation (KDE) was implemented to detect potential candidate regions for anthropogenic feeding sites. The polygons were then attributed with additional information from the revisitation analysis. The 'adehabitatHR' package in *R* was used to estimate the utilisation distribution in an individual-based approach (Calenge, 2019). KDEs for individual red kites for monthly snippets over the study period from 01.04.2018 to 31.03.2019 were calculated to obtain an individual-based KDE result. The monthly period seemed reasonable since the most impactful and therefore interesting anthropogenic feeding sites are those with a temporally high feeding frequency (i.e. daily, several times a week, weekly).

Most kernel functions deliver similar results, hence the frequently used bivariate normal kernel function was applied (Silverman, 1998; Downs, 2010; Fleming *et al.*, 2015). According to the literature the 50% home range area is commonly seen to represent the core area of activity indicating areas with a 50% probability of use (Harris *et al.*, 2012; Hötker, Krone and Nehls, 2017). This parametrisation was chosen under the assumption that anthropogenic feeding sites hold an important role in a red kite's use of geographic space and therefore should be located within their core areas of use.

Different parametrisations for the bandwidth h were tested on a random sample data set of 21 individuals and then applied to the full data set of 227 birds. Calenge (2019) stated that the reference bandwidth is often too large if an animal uses multiple activity centres and results in an over-smoothed probability surface, which was shown by the high values delivered by href (Table 4.1). Hence, the reference bandwidth was not suitable to capture events of small spatial extent, such as the anthropogenic feeding sites. Therefore, the LSCV method which computes the bandwidth h by a Least Square Cross Validation, was tested (Calenge, 2019). The calculated bandwidths were considerably lower compared to the href bandwidths but still large for the task (Table 4.1).

Therefore, a subjective, iterative visual choice of the bandwidth was considered (Silverman, 1998). Successive tests with bandwidths varying between 10 and 100 were conducted (Appendix C). Core area polygons resulting from bandwidth $h = 10$ were too small to include the known anthropogenic feeding sites in their extent, which is a sign of under-smoothing. Bandwidth $h = 100$ over-smoothed the result as it was too big for the spatial scale of anthropogenic feeding. However, the bandwidths $h = 30$ and $h = 50$ showed similarly suitable results. Considering the spatial scale of anthropogenic feeding (e.g., single backyard) and the distribution of GPS localisations of different individuals the bandwidth $h = 50$ was chosen (Table 4.1). Two underlying grid sizes were tested: *30m x 30m* and *60m x 60m*. The influence of the grid parameter was negligible, but the *60m x 60m* grid was selected for a better future application on larger areas, where computational effort is more demanding.

Table 4.1. Parametrisation of bandwidth h with reference bandwidth (href), LSCV and subjective choice method.

	Reference bandwidth	**LSCV**	**Subjective choice**
Bandwidth h	185 – 2500	9 – 199*	50

* The algorithm did not converge.

The KDE polygons represent the 50% home range area for each individual per month over the season 2018/2019 (Figure 4.1).

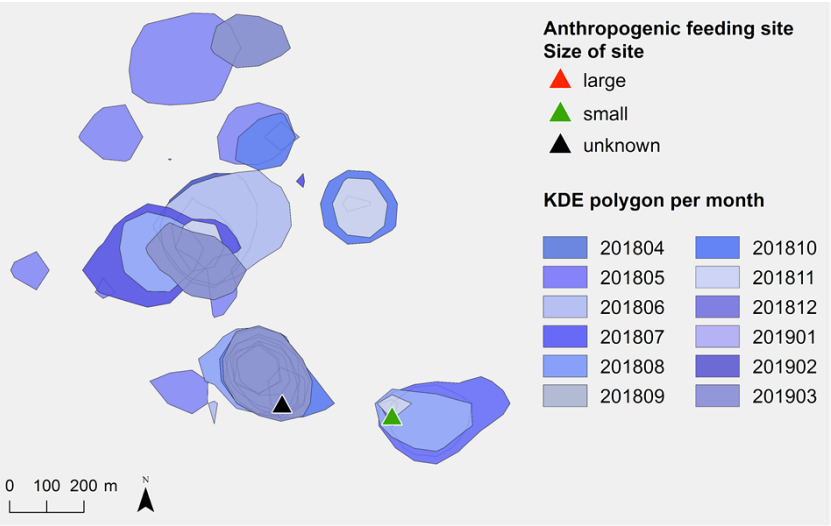


Figure 4.1. Exemplary 50% monthly home range polygons (core areas) of red kite ID013; evaluated parameters: h = 50, grid = 60 x 60m.

4.2.2 Anthropogenic Environment

As anthropogenic feeding is linked to human activity and human-made structures, the land cover information about settlement areas was included in the analysis using the swissTLM^{3D} data (Bundesamt für Landestopografie swisstopo, 2020b). The known anthropogenic feeding sites in the study area show location characteristics such as garden, orchard, meadow, compost, farm, house with garden, and are therefore closely associated with buildings. To include the aspect of inhabited buildings and exclude very small, potentially uninhabited buildings, a limit of 65 m^2 for the building floor area was set. This corresponds to the minimum design of a detached single-family house (Röthlisberger, 2019). Only 50% home range polygons that are in close proximity to such human activity, represented by the building layer, should be considered for further analysis of the identification of anthropogenic feeding sites. To find the best suitable buffer to include or exclude core area polygons different buffer sizes (25 m, 50 m, 75 m, 100 m) around buildings were analysed.

4.2.3 Ground Truth Information

The known anthropogenic feeding sites were classified into three groups: small or medium (called small), large, and unknown, subsequently referred to as the size of a feeding site, based on a combination of the frequency and the amount of feeding. All following analyses were based on this classification rather than only on the frequency of food provision suggested in RQ A.1.

To test the suitability of the kernel density estimation subsequently applied to detect anthropogenic feeding sites, it had to be determined whether a reported active anthropogenic feeding sites had been potentially visited by one of the red kites in the sample. 44 active anthropogenic feeding sites were reported for the season 2018/2019 in the study area (Section 3.2.3). However, active anthropogenic feeding sites without visits of an individual from the sample are impossible to detect with a method based on GPS tracking data and therefore would bias the evaluation process. Consequently, anthropogenic feeding sites were classified as 'potentially visited' if there was at least one at least one by at least GPS localisation of the pre-processed data set within a 50 m buffer around the site.

The remaining core area polygons were attributed with information about their connection to anthropogenic feeding. As the approach is individual-based, it had to be ensured that the GPS points which identify a feeding site as potentially visited and the 50% home range polygon belong to the same individual. Hence, a polygon was classified as 'feeding site' if the polygon and the GPS data point recording a visit to the feeding site belonged to the same individual. For example, if only red kite *no. 001* potentially visited anthropogenic feeding site *no. 1*, the 50% home range polygon of red kite *no. 001* was considered a 'feeding site' if the GPS point close to the site belonged to red kite *no. 001*.

4.2.4 Overlaps of KDE Polygons

Furthermore, it was tested whether intersection areas of a red kite's polygons over various months provides a possibility to reduce the number of polygons that do not detect an anthropogenic feeding site (false positives) but a natural feeding ground such as a field close to a building or tree rows

where red kites rest. For this reason, overlaps of the core areas of use for several months per individual red kite were calculated. However, this approach later proved to be unsuccessful and was therefore no longer pursued (Section 5.3).

4.3 Revisitation Analysis

The revisitation analysis was conducted to introduce additional information to the core area polygons. It was expected to allow for a differentiation between core area polygons correctly identifying anthropogenic feeding sites (true positives) and those with false indications (false positives).

For the revisitation analysis data of 224 individuals with suitable data available for the revisitation calculation were utilised. The *R* package 'recurse' from Bracis, Bildstein and Mueller (2018) was applied to calculate the revisitation metrics for the GPS trajectory data. The package has a built-in function for the calculation of revisits to a polygon (e.g., a protected area) (Bracis, 2018). Bracis (2018) states that polygons should have a convex shape to receive accurate results. For this reason, the above calculated 50% home range polygons representing core areas of use were transformed into convex polygons. The convex candidate polygons were used as areas of interest to calculate the revisitation measures. Since the function 'getRecursionsInPolygon' was developed for the application to a single polygon rather than multiple polygons it had to be modified to make it applicable for multiple polygons and multiple individuals.

The calculations were performed in an individual-based approach. For every polygon (representing a period of one month) of a specific individual, its recursion information was calculated for this particular month. As an example: Red kite *no. 001* has one polygon for the month of January 2019. So, based on all the GPS points from January 2019 it was calculated how many times the trajectory passes through this polygon. The algorithm automatically calculates revisits by counting the number of segments of a trajectory that cross the polygon. In addition, residence time or time since last visit is calculated assuming a linear movement between points inside and outside the polygon. If the value for revisits is 1, there were no revisits to the polygon, except the one initial visit (Bracis, 2018). The focus here lay on the two measures revisits to a polygon and residence time for each polygon. Since the polygons stand for the core areas used within one month, the revisits state the number of revisitations to this core area over the period of one month and the residence time indicates the total time spent within the core area over this time period.

The revisitation analysis was used as an intermediate step towards more information about the core area polygons to subsequently conduct predictions by the generalised linear mixed model rather than as a separate method.

4.4 Generalised Linear Mixed Model

To improve the detection of anthropogenic feeding sites and allow for the application to other regions the high number of core area polygons indicating anthropogenic feeding sites at locations where the ground truth indicates none (i.e. false positives) had to be reduced. The results of a GLMM informed by revisitation parameters were used for the prediction. The prediction values state the

probability of a polygon to be an anthropogenic feeding site (e.g., 0.2) and were later converted to a binary value of presence or absence (e.g., a polygon is an anthropogenic feeding site or not) by a classification threshold.

Every core area polygon contained information on the individual, the number of revisits, residence time, area of the polygon, age class, season, month and the average number of recorded GPS points the specific red kite showed in a month. Data of 224 birds were used for the predictive model (Table 4.2).

Table 4.2. Data used for the generalised linear mixed model grouped by age class.

Age class	Number of individuals	Number of polygons
Breeding bird	80	2,348
Non-breeding bird	108	11,634
Fledgling	36	718
Total	224	14,700

In a first, step numeric variables were scaled to make model convergence more likely. An *a priori* selection of feasible variables and interactions for the model was made based on ecological assumptions and knowledge (Dormann *et al.*, 2013). The number of revisits and residence time within a core area polygon were expected to differ between those associated with anthropogenic feeding activity and natural sites. During the previous steps of analysis differences between age classes could be observed; therefore, the inclusion into the model seemed reasonable. The following six fixed effects were included in the model: number of revisits, residence time, age class in 2018, season (breeding time, non-breeding time), the area of the polygon in hectares and GPS points per month (Appendix D, Table D.1).

Collinearity between the numerical fixed effects was tested by calculating a pairwise Pearson's correlation coefficient (r). Collinearity indicates that explanatory variables are linearly related, thus are not independent (Dormann *et al.*, 2013). When collinearity occurs the individual effect of the variables on the results cannot be separated and therefore the interpretation of the model is difficult (Dormann *et al.*, 2013). In pairwise correlation, the coefficient should not exceed a threshold of 0.5 – 0.7, values above that indicate high collinearity (Dormann *et al.*, 2013). The highest correlation coefficient occurred between the number of revisits and residence time (r = 0.68), hence all explanatory variables remained in the model.

Instead of resampling the GPS data to make sure all individuals have the same number of recorded points within a month, a control variable stating the number of points for a month was integrated into the model. The six fixed effects and all two-way interactions were included. In GLMMs, in addition to the fixed effect structure, an optimal structure for the random effects has to be found (Zuur *et al.*, 2009). The random effects were selected based on the data structure and study design. Random effects allow the modelling of differences between the individual red kites as well as the different months represented by the candidate polygons (Zuur *et al.*, 2009). Hence, two independent random

effects were integrated, one associated with each individual (BirdID) and another for each month, representing the repeated measures (Appendix D, Table D.2). The addition of a random term to the intercept resulted in a random intercept mixed effects model (Zuur *et al.*, 2009).

The choice of the error distribution for a GLMM depends on the response variable (Bolker *et al.*, 2009). Since the response variable 'feeding site' has either value 0 (polygon is not associated with an anthropogenic feeding) or 1 (polygon is a feeding site), a binomial GLMM is suitable (Zuur *et al.*, 2009). The link function defines the relationship between the expected value of the response variable and the predictor function. The logit link is the most common link function for binomially distributed data, and it achieves the goal of limiting the fitted values between 0 and 1 (Zuur *et al.*, 2009).

The above-mentioned steps resulted in a first model equation (4.1):

$$\begin{aligned} glmer(feedingsite \sim\ & revisits * residence\ time + revisits * season \\ & + revisits * age\ class + revisits * area + residence\ time \\ & * season + residence\ time * age\ class + residence\ time \\ & * area + points\ per\ month + (1\,|BirdID) + (1\,|\,month), \\ & data = data, \quad family = binomial(link = logit), \\ & nAGQ = 1) \end{aligned} \tag{4.1}$$

The expression $(1\,|BirdID)$ stands for the random effect where an intercept is calculated for each individual red kite. The operator '*' represents interaction terms (Zuur *et al.*, 2009).

The generalised linear mixed models were fitted by the glmer() function from the 'lme4' package in *R* (Bates *et al.*, 2015). As recommended by Zuur *et al.* (2009) a top-down strategy was used to create candidate models, starting with Equation (4.1), containing as many explanatory variables in the fixed component as possible, also known as 'beyond optimal model' (Zuur *et al.*, 2009, p. 121). Non-significant interactions were dropped one by one starting with the highest p-value, followed by dropping non-significant fixed terms. 13 candidate models were obtained (Appendix D, Table D.3). The 'best' of 13 models was chosen based on the information theoretic approach of the AIC.

The equation for the 'best' model is stated below (4.2):

$$\begin{aligned} glmer(feedingsite \sim\ & revisits * residence\ time + revisits * area \\ & + residence\ time * season + age\ class + points\ per\ month \\ & + (1\,|\,BirdID) + (1\,|\,month), \quad data = data, \\ & family = binomial(link = logit), \quad nAGQ = 1) \end{aligned} \tag{4.2}$$

The evaluation of the GLMM was conducted with the performance measures described in Section 2.4.4. The optimal classification threshold was based on the highest F-measure, giving precision and recall equal weight. Further information about land cover was integrated to analyse the origin of false positive results.

The prior classification of a polygon as an anthropogenic feeding site or not (class 1/0) is based on the location of a ground truth data point within the core area polygon's extent (Section 4.2.3). To study the influence of this classification on the best model chosen, the attribution of a polygon to an anthropogenic feeding site was repeated based on the condition that it must be located within 50 m from a ground truth data point. This approach offers a possibility to account for spatial uncertainty of ground truth data. The model selection process was rerun with the adapted classification and a model with the same fixed effects and interaction terms was found to be the optimal model based on its AIC. This is an indication that the model is robust.

The original data set shows great imbalance between the positive (1, anthropogenic feeding site) and the negative (0, no anthropogenic feeding site) events. Therefore, a test with balanced data was conducted. The balanced data set included all positive events (i.e. polygons associated with anthropogenic feeding) and an equal number of negative events (i.e. polygons not associated with anthropogenic feeding) generated by a random selection from all negative events. The model selection process was repeated for the balanced data set. The resulting 'best' model was similar to Equation (4.2), with the difference that neither an interaction term of revisits and residence time nor of revisits and area was included. However, all fixed terms remained in the equation. The imbalance in the data only slightly affected the model selection process and therefore the influence on the resulting prediction is not expected to be substantial. Based on those assessments the model of Equation (4.2) was used for all subsequent analyses and evaluations.

Chapter 5

5 Results

5.1 Pre-processing

During the pre-processing of the data, the amount of GPS data points was reduced based on context variables (Section 4.1). The main reduction occurred when GPS points within forest areas were excluded (Table 5.1). The pre-processing also resulted in a decrease of individuals suitable for the KDE analysis. 227 of 233 individuals still had sufficient data points (> 5 relocations per month) to be used.

Table 5.1. Pre-processing of GPS data points with the number of GPS points remaining after the exclusion of points based on context variables forest and daytime.

Total	Excl. forest areas	Excl. nighttime	Used
1,077,374	570,456	504,444	504,309

5.2 Use of Anthropogenic Feeding Sites

37 of 44 (84%) active anthropogenic feeding sites were potentially visited by tagged individuals (Table 5.2). 81% of small or medium sites, 100% of large sites and 83% of unknown-sized feeding sites were visited. Visits were indicated by the presence of at least one GPS localisation within a 50 m buffer around the site. Breeding birds and non-breeding birds made use of approximately the same number of different anthropogenic feeding sites, fledglings, however, show visits to a smaller number of sites.

Table 5.2. Potentially visited anthropogenic feeding sites by size and age class of red kite. (potentially visited feeding sites / active feeding sites).

Age class	Total	Small or medium	Large	unknown
Breeding bird	24 / 44	16 / 32	5 / 6	3 / 6
Non-breeding bird	28 / 44	20 / 32	5 / 6	3 / 6
Fledgling	13 / 44	8 / 32	2 / 6	3 / 6
All birds	37 / 44	26 / 32	6 / 6	5 / 6

105 red kites had GPS localisations in close range to known anthropogenic feeding sites, which indicates that 46% of the sample potentially used anthropogenic feeding sites. While only 26% of fledglings potentially visited a feeding site, 45% of breeding birds and 56% of non-breeding birds did so (Table 5.3).

Table 5.3. Number of individuals with GPS points in a 50 m buffer around anthropogenic feeding sites and the number of red kites used for KDE analysis grouped by age classes.

Age class	In 50 m buffer of feeding site	Used for KDE
Breeding bird	36	80
Non-breeding bird	59	108
Fledgling	10	39
Total	105	227

Out of the red kites which potentially used anthropogenic feeding sites, most (90 of 105, i.e. 86%) potentially visited one or two sites, with only a few birds benefiting from more sites (Table 5.4).

Table 5.4. Number of potentially visited anthropogenic feeding sites by age classes.

No. of visited sites	All birds	Breeding bird	Non-breeding bird	Fledgling
0	122	44	49	29
1	51	22	23	6
2	39	13	23	3
3	10	1	8	1
4	4	-	4	-
5	1	-	1	-

51 red kites potentially visited one site from April 2018 to March 2019, thereof 22 were breeding birds, 23 non-breeding birds, and 6 fledglings (Table 5.4). Breeding individuals showed potential visits to a maximum of three sites, with most of the birds visiting one. Non-breeding birds showed a maximum of five visited sites, with the majority concentrated around one and two sites. 87% of the individuals visiting more than two feeding sites are non-breeding birds.

5.3 Kernel Density Estimation

The kernel density estimation resulted in 39,517 polygons representing monthly core areas of use of the individuals. By including the information about a connection to human-made structures with different buffer sizes around buildings, the number of polygons was reduced to 16,936 (- 57%) for the 227 red kites in the study area (Table 5.5).

Table 5.5. Number of KDE core area polygons intersecting with different buffer sizes around buildings.

Total	25 m	50 m	75 m	100 m
39,517	12,269	16,936	21,521	25,719

Depending on a red kite's movement the core areas resulted in distinct numbers of polygons per month with different sizes. The number of core area polygons per month for non-breeding bird showed generally higher values than for breeding bird data (Figure 5.1). On average non-breeding

birds had 18 polygons categorised as core areas per month, while breeding birds had 3. Non-breeding bird data showed a larger within-group variability for the number of polygons per month than breeding bird or fledgling data. Furthermore, there was a large variation in size of the core area polygons as they reached areas of up to 60 ha (0.6 km²), with the majority being below 20 ha (0.2 km²).

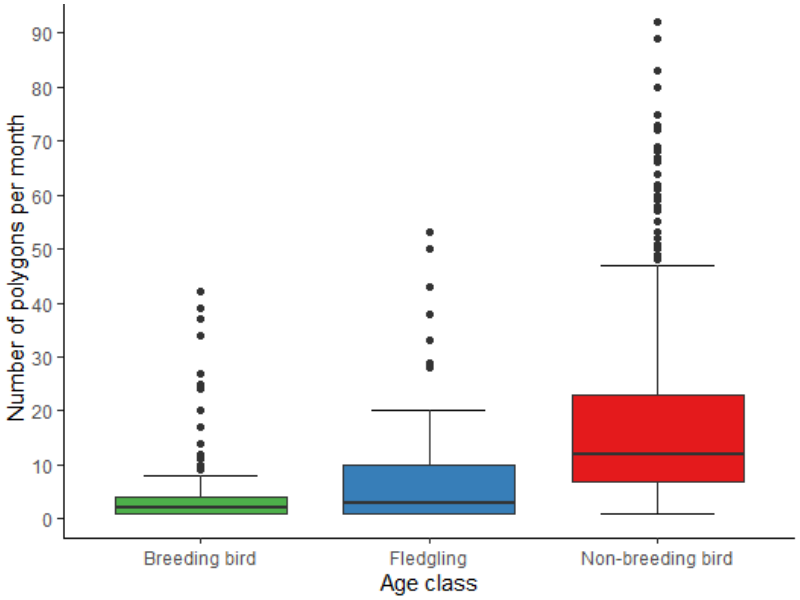


Figure 5.1. Number of polygons per month by individual, grouped by age class. The box plots show median (bold line), interquartile range (box), minimum and maximum value, and outliers (dots).

The detection success of KDE core areas was evaluated using the potentially visited anthropogenic feeding sites as ground truth. A group of 35 red kites benefiting from anthropogenic feeding potentially visited one site without detecting it (Figure 5.2). 17 red kites that potentially visited two feeding sites and detected one, resulting in a detection rate of 50%. With the KDE-based method one red kite has identified a maximum of four feeding sites, while potentially visiting five. Overall, 51 of 105 individuals (49%) successfully detected at least one feeding site that they potentially visited.

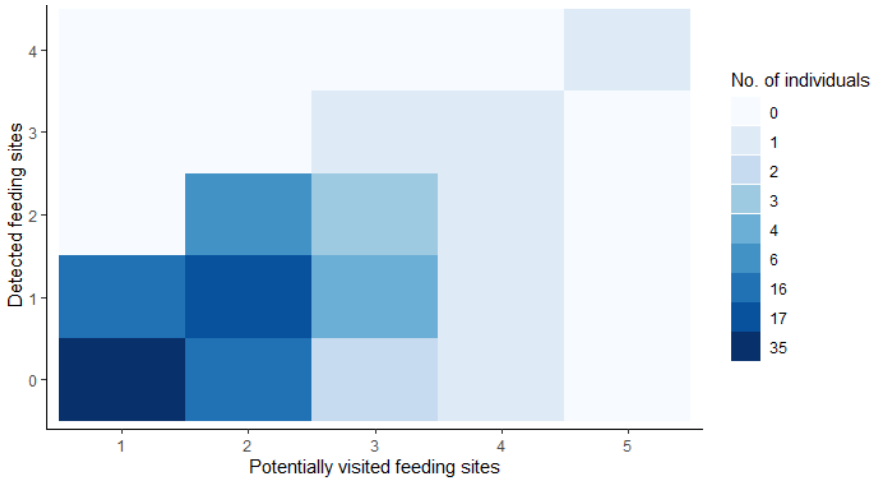


Figure 5.2. Number of potentially visited anthropogenic feeding sites in relation to the number of detected feeding sites. Reading example: 35 individuals potentially visited 1 site, detecting 0 sites.

When looking at the detection of individual feeding sites results showed that 26 out of 37 potentially visited feeding sites (70%) were positively identified by the KDE method (Table 5.6). Importantly, all six large anthropogenic feeding sites were located within the KDE 50% home range polygons. 17 of 26 (70%) small- and medium-sized as well as 3 of 5 (60%) sites of unknown size were successfully identified. Of the total 26 sites located within the core area polygons, six sites were used by both breeding and non-breeding individuals. Half of the large feeding sites were visited and detected by individuals from both age-classes, while small feeding sites were only detected by either a breeding or a non-breeding individual.

16,936 core area polygons were identified, but only 144 (0.85%) contained a known feeding site. The remaining 16,792 (99.15%) were not connected to a reported site. To reduce the number of polygons not connected to a known site, overlaps of the monthly core areas per individual bird were analysed. The assumption was that a red kite revisits the same anthropogenic feeding site in several months over the course of a season. The number of polygons incorrectly indicating an anthropogenic feeding site was reduced by 90.9% in this step. However, the condition of two monthly 50% home range polygons overlapping in season 2018/2019 reduced the number of detectable anthropogenic feeding sites by more than half (Table 5.6) and was therefore not pursued further.

Table 5.6. Number of detected anthropogenic feeding sites grouped by size located within KDE core areas and their intersection areas.

	Total	Small or medium	Large	Unknown
1-month core areas	26	17	6	3
2-months intersection	12	6	4	2
3-months intersection	9	5	2	2

5.4 Revisitation Analysis

The revisitation analysis was used as a step towards more information about the core area polygons to subsequently conduct predictions by the generalised linear mixed model rather than as a separate method.

Revisitation measures were calculated for the convex polygons of the original KDE core areas. There was a concern that the polygons would vastly increase in area and therefore change the outcome of the revisitation analysis by incorporating more data points. However, 78% of the polygons showed an area change smaller or equal to 10% and only 0.8% demonstrated an area change greater than 50% (Figure 5.3). The result is not surprising, as core area polygons derived from a smooth KDE surface inherently tend to be convex.

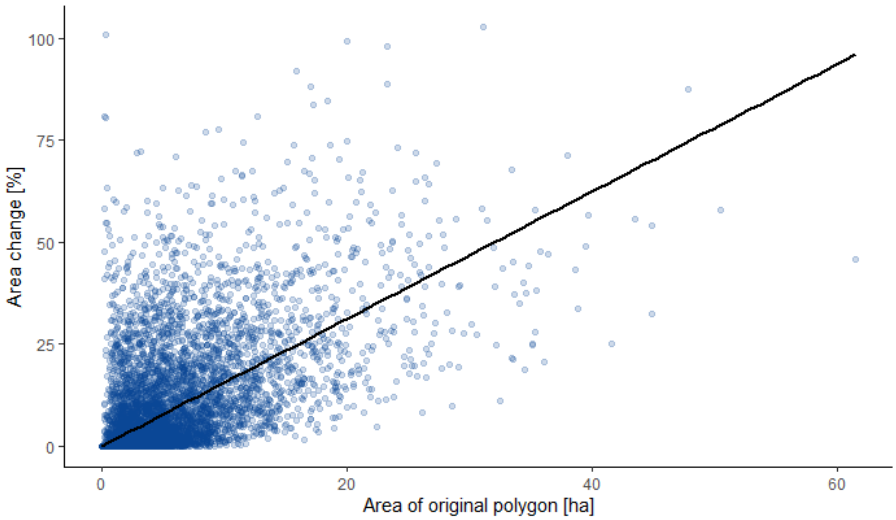


Figure 5.3. Area change [%] from original polygon created by 50% home range calculation to convex polygons used for revisitation analysis.

Revisitation rates were calculated for 143 polygons (out of 144, 99.3%) linked to feeding sites and 14,557 polygons (out of 16,792, 86.7%) not linked to anthropogenic feeding. Calculation for omitted polygons resulted in erroneous results. Most of the erroneous polygons showed very small areas resulting from the subjective parametrisation of the kernel density estimation. Since most of the error polygons were not associated with anthropogenic feeding, the revisitation calculation was continued with the 14,700 polygons where revisitation measures were successfully computed.

Table 5.7 shows a summary of the calculated revisitation measures. Generally, the mean number of revisitations to a core area was larger for polygons linked to anthropogenic feeding sites than those without connection to anthropogenic feeding across all age classes. The same was true for the average residence time within a polygon's boundary. Considering the average time per visit, the polygons associated with anthropogenic feeding sites showed higher values than such without connection to feeding activity.

Table 5.7. Average of revisitation measures with classification of polygons based on ground truth information about feeding sites grouped by age class.

Feeding site	Age class	Revisits		Residence time [h]		Time per visit [h]
		mean	SD	mean	SD	mean
No	Breeding bird	33.3	48.9	87.2	144.6	2.6
	Non-breeding bird	3.3	5.3	6.5	20.7	1.9
	Fledgling	15.7	38.9	35.6	106.4	2.3
	All birds	8.6	24.3	20.5	70.9	2.3
Yes	Breeding bird	48.1	41.4	165.0	188.5	3.4
	Non-breeding bird	9.4	11	25	48.3	2.7
	Fledgling	24.0	-	77.9	-	3.2
	All birds	30.1	36.5	99.8	157.2	3.3

At locations with and without feeding activity breeding bird data on average showed higher values for numbers of revisits, residence time and time spent per visit, than non-breeding bird data. However, the data has large standard deviations values and therefore a large amount of variation within the groups.

5.5 Generalised Linear Mixed Model

The data used for the prediction model were highly imbalanced, with 143 positive cases (actual anthropogenic feeding site polygons) and 14,557 negative cases (polygons without association to feeding). Based on the 'best' model described in Equation (4.2), probability predictions for each polygon for being linked to anthropogenic feeding or not were made.

The explanatory variables area of the polygon, number of revisits and residence time had a positive effect on the probability of a polygon being associated with anthropogenic feeding (Table 5.8). The area represented the strongest effect on the model. Of the interaction terms, interactions between revisits and residence time as well as revisits and area showed negative effects. For the random effects bird-to-bird variability made the greatest contribution, with a variance for the random intercept of 6.44, followed by the month-to-month variability, with a variance for the random intercept of 0.17 (Appendix D, Table D.4).

Table 5.8. Summary of the generalised linear mixed model.

Explanatory variable	Estimate	Std. Error	Z-value
Intercept	-6.88785	0.73041	-9.430***
Number of revisits	0.48391	0.16136	2.999**
Residence time	0.23391	0.11160	2.096*
Area [ha]	0.52588	0.07827	6.719***
Season: breeding time	0.52515	0.34080	1.541
Age class: fledgling	-2.02570	1.39290	-1.454
Age class: non-breeding bird	-0.38538	0.58339	-0.661
GPS points per month	-0.25130	0.21408	-1.174
Revisits*Residence time	-0.07640	0.03783	-2.020*
Revisits*Area [ha]	-0.08643	0.04301	-2.010*
Residence time*Season: breeding time	-0.25376	0.12965	-1.957

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The performance measures addressed in Section 2.4.4 were applied to evaluate the results of the generalised linear mixed model. The predictive model shows a high AUC of 0.952 close to the maximum value of 1, suggesting a good classification performance of the model (Figure 5.4).

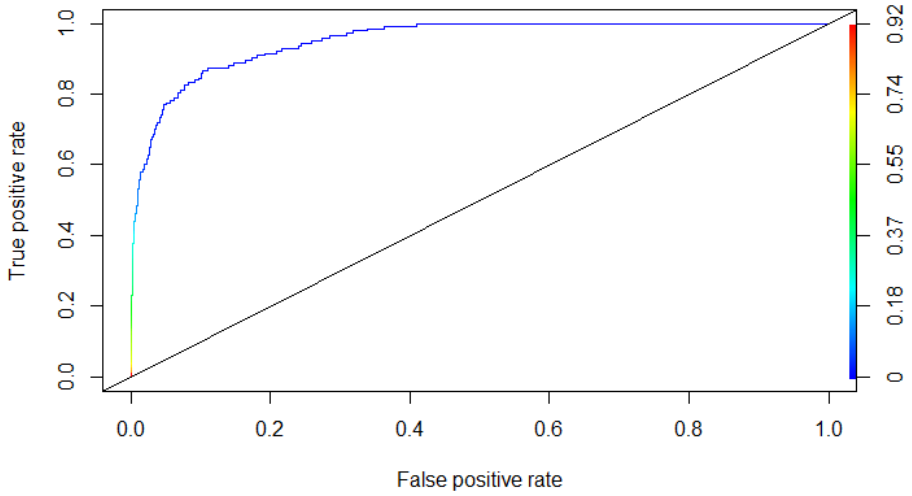


Figure 5.4. ROC curve colourised according to threshold, ROC AUC: 0.952.

However, the precision-recall curve recommended by Saito and Rehmsmeier (2015) for unbalanced data shows a less optimistic view of the model's classification performance (Figure 5.5). The curve shows a rapid decrease and an AUC of the precision-recall curve of 0.374. High values of recall were only achieved at low values of precision.

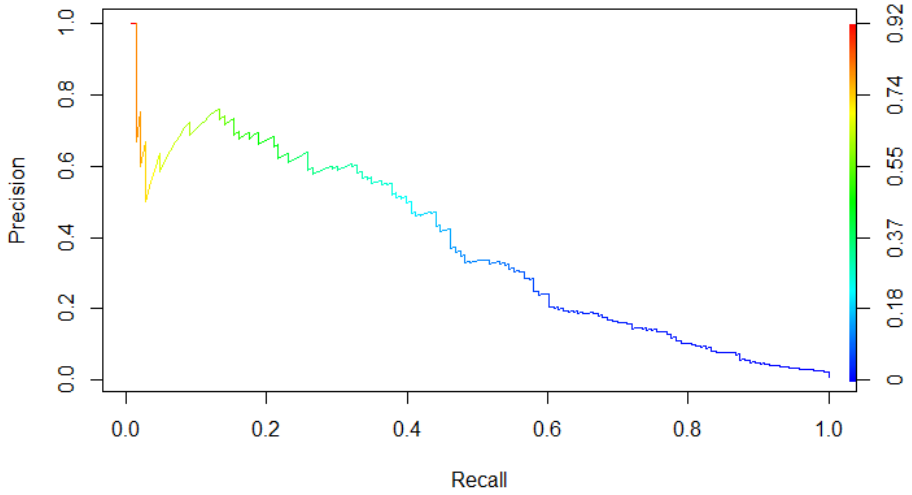


Figure 5.5. Precision-recall curve colourised according to threshold, PRC AUC: 0.374.

In Table 5.9 the confusion matrix and the related performance measures are illustrated for the default threshold of 0.5. However, as mentioned in Section 2.4.4 a threshold of 0.5 works well for balanced data sets but is unsuitable for imbalanced data such as the one at hand, which is indicated by the low values for the F-measure and MCC.

Table 5.9. Confusion matrix with classification threshold 0.5. Values indicate polygon counts and percentages.

		Actual: Feeding site	Actual: Non-feeding site
Predicted	**Feeding site**	TP 22 (0.15%)	FP 10 (0.07%)
	Non-feeding site	FN 121 (0.82%)	TN 14,547 (98.96%)

Precision	0.15
Recall	0.68
F-measure	0.25
MCC	0.34

The F-measure and MCC curves show similar courses, both displaying a maximum value of 0.45 at cutoffs of 0.188 and 0.266, respectively (Figure 5.6). Due to the simple calculation and easy understanding of the F-measure and the similar results of the F-measure and MCC, the threshold of 0.188 was chosen to accomplish a maximum value for the balanced F-measure, giving precision and recall equal weight.

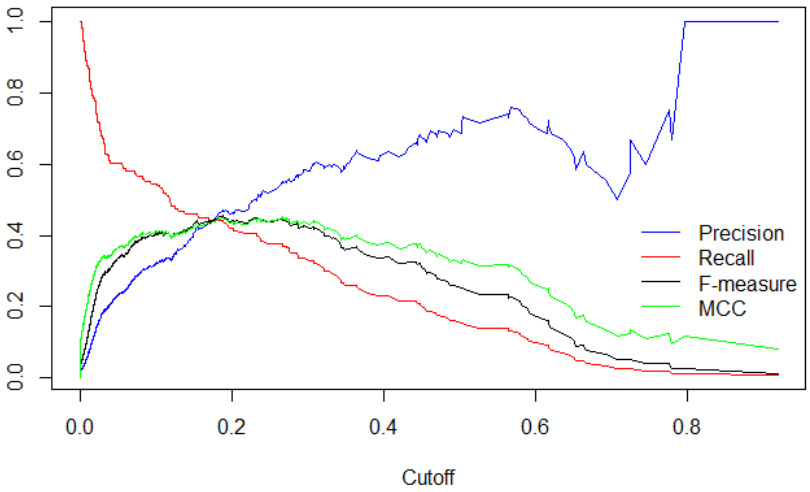


Figure 5.6. Precision, recall, F-measure and MCC curves for the overall GLMM.

Table 5.10 shows the confusion matrix and the performance measures for the optimal threshold 0.188 when evaluating the model's prediction among all birds in the sample.

Table 5.10. Confusion matrix with classification threshold 0.188. Values indicate polygon counts and percentages.

		Actual: Feeding site	Actual: Non-feeding site
Predicted	**Feeding site**	TP 63 (0.43%)	FP 71 (0.48%)
	Non-feeding site	FN 80 (0.54%)	TN 14,486 (98.54%)
	Precision	0.47	
	Recall	0.44	
	F-measure	0.45	
	MCC	0.45 (threshold: 0.266)	

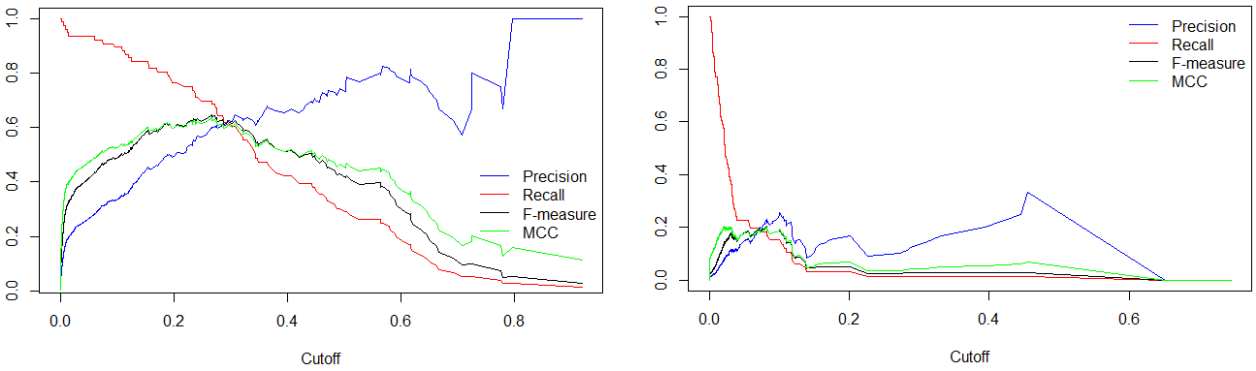
There is a probability of 47% that a predicted anthropogenic feeding site polygon indicates the location of an actual feeding site. The probability of the model to successfully recognize an anthropogenic feeding site polygon is 44%.

In more detail, the results show that data of only 15 out of 224 birds contributed to the true positive polygons. Furthermore, 13 of these are breeding birds and 2 non-breeding birds. 97% of the true positive polygons belong to breeding birds whereas 81% of the false negative polygons belong to non-breeding birds. However, also 86% of false positives belong to breeding birds. The contribution of fledglings to the identification of anthropogenic feeding sites was negligible as there was only one polygon attributed to a fledgling, which appears as a true positive at the chosen threshold.

Based on these first results and on ecological background knowledge that the age classes might show different behaviour the threshold choice and further evaluation was conducted separately for breeding birds and non-breeding birds.

Figure 5.7 shows the precision, recall, F-measure and MCC curves for the prediction model at different thresholds for (a) breeding bird and (b) non-breeding bird data, respectively. For breeding bird data, the optimal threshold based on maximum F-measure and MCC was at probabilities of 0.266, whereas for non-breeding bird data the values differed and lay at 0.082 and 0.021. The threshold choice based on F-measure and MCC both showed large differences indicating that the two age classes behave differently in the variables used for the model. The following model evaluation was conducted based on the maximum F-measure and the corresponding optimal threshold for each age class.

First, the evaluation of the results was done from a red kite's perspective, evaluating the predictions, and analysing the results per age class. Second, the focus was placed on the anthropogenic feeding sites, assessing the proportion of correctly identified sites and the influence of the size of a site and the age class of the red kite on the prediction success.



(a) Breeding bird data

(b) Non-breeding bird data

Figure 5.7. Precision, recall, F-measure and MCC curves for the GLMM by age class (breeding vs. non-breeding individuals).

The model performed better for core area polygons of breeding birds. All performance measures indicate better results for data on breeding than on non-breeding birds at the individually optimal threshold (Table 5.11). The F-measure and MCC with 0.65 and 0.64, as opposed to 0.2 and 0.21, were considerably higher for breeding bird data. When using breeding bird data, the model predicted 70% of the actual anthropogenic feeding site polygons correctly (recall) and was correct 60% of the time when predicting if a core area polygon is an anthropogenic feeding site (precision). For non-breeding bird data 18% of the actual anthropogenic feeding site polygons were correctly predicted (recall) and the model was correct 23% of the time when predicting if a core area polygon is an anthropogenic feeding site (precision).

Table 5.11. Comparison of breeding bird and non-breeding bird data evaluation of the GLMM. The two age classes show different optimal thresholds and varying values for the corresponding performance measures. Cell values indicate polygon counts and percentages.

Breeding birds (threshold: 0.266)

Predicted \ Actual	**Feeding site**	**Non-feeding site**
Feeding site	TP 53 (2.26%)	FP 35 (1.49%)
Non-feeding site	FN 23 (0.98%)	TN 2,237 (95.27%)

Precision	0.60
Recall	0.70
F-measure	0.65
MCC	0.64

Non-breeding birds (threshold: 0.082)

Predicted \ Actual	**Feeding site**	**Non-feeding site**
Feeding site	TP 12 (0.10%)	FP 40 (0.34%)
Non-feeding site	FN 54 (0.45%)	TN 11,528 (99.09%)

Precision	0.23
Recall	0.18
F-measure	0.20
MCC	0.21 (threshold: 0.021)

In general, only a small number of individuals accounted for the correctly predicted feeding site polygons. Just 10 of 80 breeding birds (13%) and 9 of 108 (8%) non-breeding birds contributed to the true positive polygons. In the case of false negative polygons 16 of 80 breeding birds (20%) and 30 of 108 (28%) non-breeding birds produced such results and 9 of 80 breeding birds (11%) and 23 of 108 non-breeding birds (21%) contributed to false positive results, respectively.

The false positive (FP) polygons are created by various underlying environmental effects. A visual assessment showed that the polygons were located primarily over tree rows and natural feeding grounds such as fields. Further instances included potentially unknown feeding sites and ones in close proximity to an actual anthropogenic feeding site surrounded by natural feeding grounds (e.g., farm house). False positive polygons were also located around known communal roosting sites (Figure 5.8).

(a) Breeding bird data **(b)** Non-breeding bird data

Figure 5.8. False positive core area polygons of breeding bird (left) and non-breeding bird data (right). Polygons are located above fields, near a farm, near a known feeding site, as well as tree rows and the area of a known communal roosting site. Base map: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS user community.

The analysis of the false positive polygons indicated that the model had difficulties distinguishing between the formerly mentioned sites and actual anthropogenic feeding sites. What stands out are the different origins of false positives by age class. For this evaluation, the swissTLM^{3D} object classes tree and shrub rows were included to determine the association to these land cover classes. False positive polygons originating from non-breeding bird data were more often associated with roosts and tree rows than such from breeding bird data (Table 5.12)

Table 5.12. Number of false positive polygons with their association to environmental factors such as communal roosting sites and tree/shrub rows. The evaluation was conducted at the individual optimal threshold for breeding and non-breeding bird data, respectively. Communal roosting sites incorporate tree/shrub rows; therefore, a polygon might be associated with both land cover classes.

Age class	Association with roosting site	Association with tree/shrub row
Breeding bird	1 / 35 (2.9%)	17 / 35 (48.6%)
Non-breeding bird	11 / 40 (27.5%)	37 / 40 (92.5%)

In a second step, the predictions were evaluated from the perspective of anthropogenic feeding sites. Table 5.13 shows the proportion of correctly predicted feeding sites and the potentially detectable number based on KDE. All results were evaluated by the individual optimal threshold chosen per age class. Overall, 50% of the possibly detectable anthropogenic feeding sites were correctly predicted by the model. For breeding bird data, the model correctly predicted 63% of the detectable anthropogenic feeding sites, whereas for non-breeding bird data 31% were predicted correctly.

Table 5.13. Correctly predicted anthropogenic feeding sites by size at individual optimal thresholds by age class. (correctly predicted feeding sites / potentially detectable feeding sites).

Age class	Total	Small or medium	Large	Unknown
All birds	13 / 26 (50%)	8 / 17 (47%)	4 / 6 (67%)	1 / 3 (33%)
Breeding bird	10 / 16 (63%)	6 / 10 (60%)	3 / 4 (75%)	1 / 2 (50%)
Non-breeding bird	5 / 16 (31%)	0 / 10 (0%)	4 / 5 (80%)	1/ 1 (100%)

For large anthropogenic feeding sites, the model showed similar prediction success for data of breeding birds and non-breeding birds with 75% and 80%, respectively. The detection success rate for small and medium-sized sites for breeding bird data was 60%, whereas the model delivered poor results for such sites used by non-breeding birds, where no site was correctly predicted.

An evaluation of the proportion of correctly predicted (true positives) and incorrectly predicted (false negatives) polygons at locations of known anthropogenic feeding sites was conducted (Appendix E). When evaluating at the optimal threshold for all red kites (breeding and non-breeding individuals) small feeding sites with feeding activity on a daily basis to several times per week showed the highest percentage of correctly predicted core area polygons, followed by large anthropogenic feeding sites with the same frequency of feeding. While most large feeding sites showed at least one correctly predicted polygon, for two sites all overlapping polygons were incorrectly predicted (false negatives). Small feeding sites with only weekly or unknown frequency of feeding activity often showed zero correctly predicted but a few incorrectly predicted core area polygons. Figure 5.9 shows true positive and false negative polygons over exemplary feeding sites.

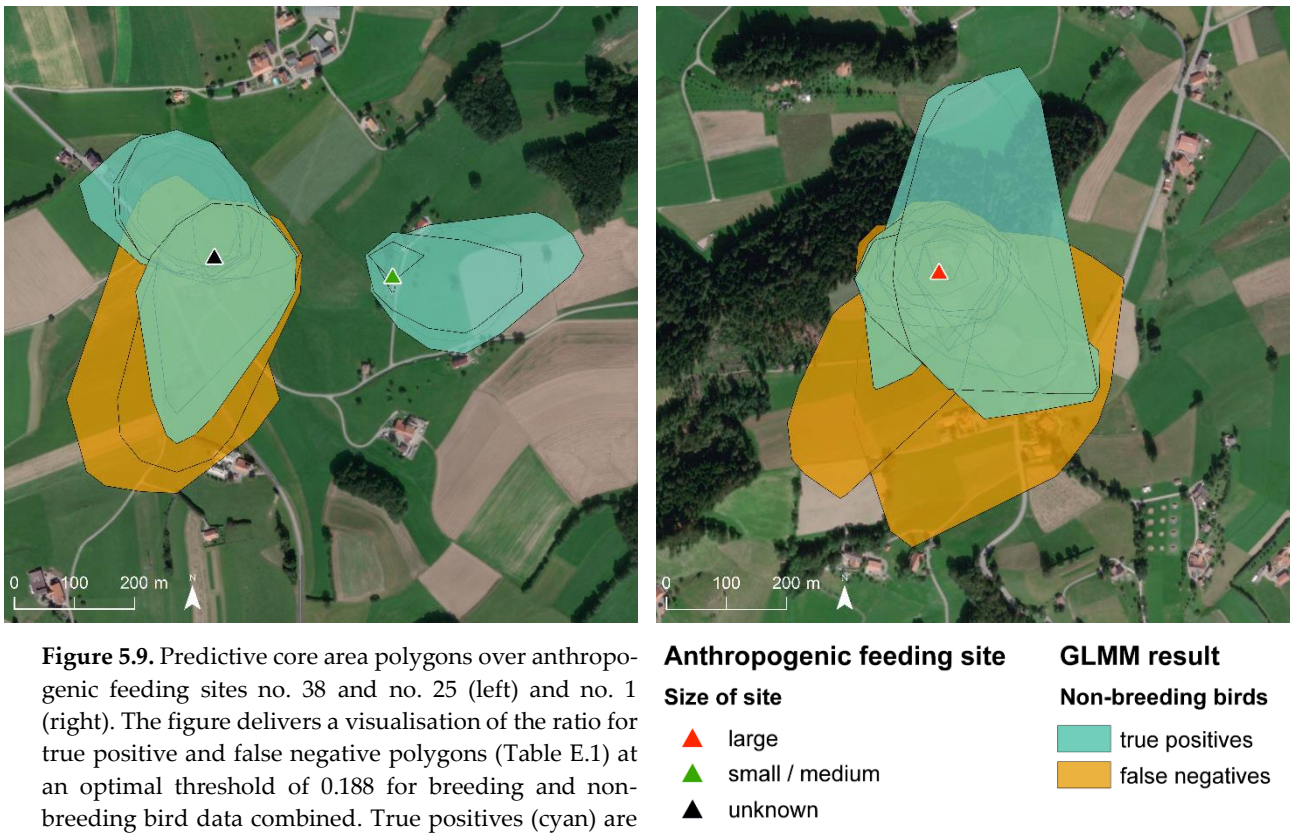


Figure 5.9. Predictive core area polygons over anthropogenic feeding sites no. 38 and no. 25 (left) and no. 1 (right). The figure delivers a visualisation of the ratio for true positive and false negative polygons (Table E.1) at an optimal threshold of 0.188 for breeding and non-breeding bird data combined. True positives (cyan) are correctly predicted polygons and false negatives (orange) are incorrectly predicted polygons, where the model predicts no feeding site but the actual value for the polygon is a feeding site. Base map: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS user community.

In general, the results demonstrate that the proportion of correctly and incorrectly predicted polygons per anthropogenic feeding site was larger when evaluated for breeding bird data than for non-breeding bird data (Appendix E). The data for breeding birds showed a range of 25 – 100% correctly predicted polygons at the feeding sites. Both small and large anthropogenic feeding sites showed high percentages of correctly predicted polygons (Appendix E, Table E.2). Some small feeding sites with activity only in winter and a large site with unknown frequency could not be correctly classified. The proportion of correctly classified polygons for non-breeding bird data was substantially lower ranging from 12.5 – 36.8% (Appendix E, Table E.3). Furthermore, the anthropogenic feeding sites with recall values above 0% (percentage of correctly predicted polygons at a feeding site) were all but one of large size. More detailed information on the proportion of correctly predicted polygons for the individual anthropogenic feeding sites with exemplary map visualisations for breeding and non-breeding bird data, respectively, can be found in Appendix E.

Chapter 6

6 Discussion

In this chapter, the results of the last chapter are discussed in further detail and the formulated research questions are answered. The results are examined from a methodological and a red kites' behavioural perspective.

6.1 Use of Anthropogenic Feeding Sites

Not all the active feeding sites within the study area were visited by a tagged red kite. However, a majority (37 of 44, 84%) showed visits of at least one tagged individual. The remaining sites were possibly used by untagged birds, undetectable due to a low temporal resolution of the data or not active anymore. Overall, 46% of the sample population showed visits to known feeding sites, indicating that anthropogenic feeding sites take an important role in foraging strategies. However, more than half of the tagged red kites did not use feeding sites and therefore seem to focus on natural food sources. Most red kites that potentially use anthropogenic feeding, benefitted from one or two sites, with only a few visiting more. Hence, red kites prefer to visit a small selection of feeding sites. In general, individuals of the classes breeding bird and fledgling visited a lower number of different sites than non-breeding birds. This implies behavioural differences in the space use of red kites depending on their age class. An explanation could be the territorial behaviour of breeding birds around their nests and the related coverage of smaller distances in search of food sources. In addition, breeding birds likely follow the central place foraging theory and therefore, at least during breeding season, possibly prefer the limited number of anthropogenic feeding sites that are in close range to their nest (Orians and Pearson, 1979). The lower number of visited anthropogenic feeding sites reported for fledglings was to be expected as those individuals leave the nest only in July and therefore have fewer data available for the analysis. Another factor could be the lack of knowledge about locations of feeding sites and the lack of experience in finding such sites. Non-breeding birds, however, are more experienced in finding sites, geographically more independent during the entire season and thus might locate new feeding sites. With this behaviour non-breeding birds extend their knowledge on the spatial and temporal distribution of different anthropogenic food sources.

6.2 RQ A – Methodological Approach

In the first research question, the selection of a methodological approach to detect anthropogenic feeding sites was key. Based on this research question the priorly discussed workflow combining information from kernel density estimation and revisitation analysis in a generalised linear mixed model was developed (Appendix B).

6.2.1 Kernel Density Estimation

A majority (70%) of the known anthropogenic feeding sites were within the extent of the calculated KDE core area polygons. Hence, the assumption that anthropogenic feeding sites take an important place within a red kite feeding routine and therefore lead to an accumulation of GPS points at such location, is assumed valid. Hence, the integration of kernel density estimation to define candidate polygons for further analysis was effective. The utilisation distribution of non-breeding birds showed a higher number of activity centres resulting in more core area polygons per month than breeding birds. This shows that breeding birds, while being bound to a nest for at least some part of the season, have a geographically less expanded area of use, or more precisely, use the same areas more frequently. Even after the breeding season, breeding birds continue to have less centres of activity than non-breeding birds. Non-breeding birds, however, seem to roam around in an exploratory manner using various places, resulting in a more scattered utilisation distribution and a higher number of activity centres. With just half of the individuals successfully detecting at least one visited feeding site with one of their core area polygons, still a majority of sites was detectable. Large feeding sites showed the highest percentage of locations within the core areas of use, indicating that those sites have a greater influence on the movement pattern of an individual than small, medium, and unknown sized sites. But also, some smaller sites have sufficient impact to be detectable by core areas of use.

Most individuals only detected a small number of feeding sites. This indicates that only a few selected sites have a large enough impact on the movement pattern to be detectable by core areas. While large feeding sites were detected by individuals of both age classes, small sites were identified by single individuals either breeding or non-breeding birds. This suggests that at large feeding sites multiple individuals gather, while small feeding sites are used only by a breeding pair or a single individual.

6.2.2 Revisitation Analysis

Since red kites respond to repeated placement of carcasses more strongly than other facultative scavengers and therefore exploit sites with high predictability of food occurrence a factor of revisitation was integrated (Welti, Scherler and Grüebler, 2019). The characteristic of recurrence was incorporated into the analysis by using the revisitation measures for each candidate polygon as fixed effects for the generalised linear mixed model. The assumption that red kites more frequently visit areas at anthropogenic feeding sites, based on the repetitive characteristic of food placement, and therefore show higher revisitation rates at such locations, was confirmed by the results. However, shorter residence times due to the temporally limited action of anthropogenic feeding could not be observed. In general, there was a large amount of variation for revisitation rates and residence time within all age classes, indicating large variabilities within the groups. The trajectory data showed a course, 1-hour sampling interval, which can cause insecurities in the revisitation analysis emerge. But these results could also be influenced by different behaviour of red kites at anthropogenic feeding sites. One reason could be that although the anthropogenic feeding takes place within a certain time frame, red kites behaviour differs from the assumption that they perceive the placed food and

simply grab it without hesitation. The longer residence time rather suggests that red kites circle above the anthropogenic food source cautiously before taking the food in the same way as they do when searching for natural food. This behaviour is potentially dependent on the number of red kites competing at the same anthropogenic feeding site and whether a red kite regularly uses the site and hence already trusts the source. Breeding birds showed higher revisitation rates than non-breeding birds which implies that breeding birds frequently use the same feeding site close to their nest. Hence, they have memorised the location which results in a frequent return pattern to check for food and consequently in a high number of revisits. However, non-breeding birds probably do not return to the same site as frequently because they are not bound to a nest.

6.2.3 Integration of Revisitation Parameters

Through the integration of revisitation parameters in a generalised linear mixed model and by classifying the probability predictions with an optimal threshold, the number of false positive polygons was reduced. The KDE analysis delivered 16,936 polygons which were potential feeding sites, with 144 true positives and 16,792 false positives. With the predictions of the GLMM the number of false positive polygons was reduced to 71, 35, or 40 depending on the optimal evaluation threshold for all birds, breeding, and non-breeding individuals, respectively. This increased the applicability of the approach by reducing the number of incorrectly predicted feeding sites and therefore improved the methodology. However, the absolute number of feeding sites that were identifiable, was reduced from 26 (KDE) to 13 (GLMM) when evaluating at an optimal threshold for all birds. With the age-dependent evaluation thresholds a total of 15 sites were identifiable, 10 detected by breeding bird and 5 by non-breeding bird data. Even if a smaller number of feeding sites was detectable, up to 80% of the large feeding sites were correctly predicted depending on the age-dependent threshold. By using age-dependent probability thresholds potential differences between age classes were accounted for. With breeding bird data 63% of potentially visited feeding sites were detected, with 60% and 75% detection rates for small and large sites, respectively. However, non-breeding bird data, while performing well for large-sized anthropogenic feeding sites with a detection rate of 80%, only reached a detection rate of 31% overall. These results show that while breeding bird data and non-breeding bird data are equally qualified to identify large anthropogenic feeding sites, only breeding birds delivered satisfactory results for small or medium size sites. Therefore, if only large-sized feeding sites are of interest the use of non-breeding bird data can be sufficient. But one should be aware that while the data shows prediction success for the detection of large feeding sites it does not for small anthropogenic feeding sites. This suggests that the methodological approach based on the idea of revisitation and anthropogenic feeding on a regular basis corresponds more to the behaviour of breeding birds than non-breeding birds. Non-breeding birds potentially show more exploratory behaviour using larger geographic areas and not the expected regularity in repetitive use of the same anthropogenic feeding site but rather use several different sites. In case those sites are not visited at least in a monthly pattern they would go undetected.

Furthermore, the number of individuals contributing to false positive and false negative results was higher for non-breeding birds than for breeding birds, even at the optimal individual threshold for the age classes, respectively. This indicates that the developed classification works better with data

of breeding birds. The flexibility of choosing different evaluation thresholds based on precision and recall makes the method a versatile tool as it is adjustable to individual project needs.

Overall, the results showed that not only the red kite age class but also the size of the anthropogenic feeding site, a combination of frequency and amount of feeding, had an influence on the identifiability of an anthropogenic feeding site. In general, a larger proportion of large anthropogenic feeding sites than small sites were correctly predicted, especially sites with feeding taking place daily, several times a week or weekly were detectable.

In RQ A the application of a population-based approach in addition to the individual-based approach was discussed. However, the implementation of a second, population-based approach went beyond the scope of this master's thesis. A population-based approach could potentially only deliver results for the detection of large anthropogenic feeding sites, since these are locations where multiple individuals gather. As the individual-based approach already delivers satisfactory and interesting results considering the detection of large and small anthropogenic feeding sites, a population-based approach was not expected to improve the methodology. Furthermore, a population-based approach is only applicable if sufficient data are available i.e. in regions where a lot of red kites are tagged. Hence, the here developed, individual-based approach enables broader applicability by being suitable for densely and scarcely sampled regions.

6.2.4 Uncertainties and Limitations

The examined methodological approach showed success in detecting small and large anthropogenic feeding sites, however, the step-by-step methodology involves a long decision-making process and each decision affects the results and the conclusions drawn from them. Therefore, their implications and limitations are discussed here.

While kernel density estimation (KDE) is a widely used method for point pattern analysis it might produce biased results when used for point patterns of moving objects. The analysis of point patterns resulting from animals with movement over time might require a method that includes both spatial and temporal elements (Downs, 2010). KDE assumes independent data, animal tracking data, however, shows autocorrelation and therefore violates this assumption (Fleming *et al.*, 2015). The impact of autocorrelation on KDEs performance increases when the temporal resolution of trajectories improves. Hence, the more finely the data are resolved, the worse is the KDE result (Fleming *et al.*, 2015). Since the data used were sampled at an hourly rate and a large number of consecutive points were removed when excluding all points within forests, autocorrelation was not considered a problem. Furthermore, red kites can cover large distances during the sampling interval. In this thesis, the temporal aspect of a red kite's movement data was incorporated into the methodology by the revisitation measures for the core areas. However, there is a variety of other methods to account for autocorrelation such as a new approach where kernel density estimation is combined with methods of time geography, and geo-ellipses are used to compute density estimates (Downs, 2010). Another option developed by Fleming *et al.* (2015) is the autocorrelated KDE (AKDE) which is suitable for movement data. Still, in future research, the issue of autocorrelation in animals' movement data should be considered to ensure the performance of point pattern analysis.

Furthermore, there are behavioural explanations for the limitation of the kernel density estimation to define candidate polygons. The large variety of different red kite movement behaviour within the sample proved to be difficult in the selection of appropriate parameters. While the subjectively chosen bandwidth for the kernel density analysis might be suitable for the location data of some red kites, it is potentially less suitable for others. This has been shown by the varying number and size of core area polygons depending on the red kite's movement behaviour. Furthermore, some polygons sizes show values up to 0.6 km², which is rather large considering that anthropogenic feeding takes place within a few square metres in a backyard, on a feeding platform, or a farm's compost heap. This leads to limitations in the application of the method. Within densely populated regions a polygon indicating an anthropogenic feeding site could easily contain several buildings. Hence, pinpointing the anthropogenic feeding to one house is not possible or would have to be investigated by a door-to-door survey in this area. However, for an anthropogenic feeding site on a farm standing alone between fields, the size of the polygon should not cause problems.

Even under the additional conditions of excluding GPS points in forests and during nighttime as well as the included proximity to anthropogenic building structures, a large number of core area polygons occurred at locations where no association with actual anthropogenic feeding sites could be established. For these locations a red kite's behaviour could be a key explanation because polygons occurred at locations of natural feeding sites, such as agricultural fields in close range to a farmhouse, a flight corridor over a settlement area that was frequently used to access a nest or a communal roosting site. Overall, the application of KDE with adjusted parametrisation for the small spatial scale of anthropogenic feeding sites was not sufficiently reliable for the detection of anthropogenic feeding sites. Therefore, the revisitation parameters were included in a generalised linear mixed model to improve the methodology.

When selecting and evaluating the generalised linear mixed model (GLMM) the same data was used for finding the optimal model and for the model validation. This method called resubstitution generally results in optimistic measures for prediction success. To obtain a more robust measure of prediction success independent data should be used by splitting data into training and test data or by applying the method to a completely new sample (Fielding and Bell, 1997). Furthermore, depending on the aim of the research the classification thresholds for the evaluation of the GLMM should be adapted.

The classification results of the generalised linear mixed model show that false positives polygons were associated with known roosting sites and natural feeding grounds. This indicates that polygons at natural feeding grounds, tree rows and roosts as well as potentially unreported anthropogenic feeding sites and the actual anthropogenic feeding sites show similar revisitation rates and residence times. Even with this difficulty of distinguishing between the formerly mentioned sites and actual anthropogenic feeding sites, the method can be useful to limit the locations of potential anthropogenic feeding sites.

6.3 RQ B – Environmental Context Variables

Research question B dealt with the context variables necessary to limit the results to an anthropogenic environment and lead to a confident identification of anthropogenic feeding sites.

Several context variables, used to define the anthropogenic environment of the feeding sites and exclude some natural sites, were considered in the development of the step-by-step workflow. Certain context variables were already integrated into the pre-processing of the data set. The main reduction of GPS points occurred when locations in forest areas were excluded, which accounted for about half of all points. This supports the assumption that red kites spend a lot of time in or close to forest areas. The number of GPS points was only slightly reduced by the exclusion of night points. Possibly because red kites are diurnal animals and most of the time spent in such areas is between dusk and dawn when they rest. Hence, a lot of nighttime points were already excluded by the prior step. Nevertheless, since feeding activities occur during daytime the omission was considered reasonable, to only retain areas of high GPS density potentially connected to foraging activities.

The kernel density estimation resulted in 39,517 polygons representing core areas of use of the individuals. The different buffer sizes tested to include information about the proximity to buildings, indicated that buffers larger than 50 m included many polygons without a link to anthropogenic feeding, such as areas of agricultural fields and other natural feeding grounds. However, both the 25 m and the 50 m buffer seemed suitable. As a conservative measure, a 50 m buffer was selected to retained polygons potentially linked to feeding based on their proximity to buildings, while still reducing the number of predicted polygons by more than half. Additionally, a 50 m walk (1 minute) was considered likely to be carried out by a person to place food on a neighbouring field to observe red kites. The large reduction of polygons indicates that many core areas, even with forest areas already excluded, have no association with the anthropogenic environment. But it also shows that the proximity to buildings helps to improve the method by excluding a large number of false positives. Hence, despite the relative simplicity of the integrated information, the context variables contributed to the improvement of the method. However, because of the high number of remaining candidate polygons unrelated to anthropogenic feeding the additional methodological steps, revisitation analysis and the inclusion of its information into a generalised linear mixed model, were required.

6.3.1 Uncertainties and Limitations

The issue of core areas predicted to be anthropogenic feeding sites which are in fact connected to natural feeding sites, roosting sites, tree rows or single trees remained even with the application of the generalised linear mixed model.

Hence, to further increase the method's success more detailed information about the land cover could be integrated. For example, single trees and tree rows which are not included in the used data set for forest areas could be excluded to avoid core area polygons and predicted anthropogenic feeding sites over such locations. Furthermore, information about agricultural fields or meadows that

serve as natural feeding grounds could be added to core area polygons. However, by a more restrictive selection of core area polygons, areas potentially associated with anthropogenic feeding (e.g., representing a flight corridor to a site), would be dismissed. As shown in the results of the GLMM prediction, locations of known roosting sites were present in the false positive polygons. Therefore, an intent for better predictions could be the inclusion of information about the proximity to a roosting site as an exploratory variable in the GLMM.

The order of integrating the context variables into the workflow influences the outcome of the method. GPS data points in forest areas were excluded in the first step, as it was expected that due to a high accumulation of points in those areas the occasional visits to feeding sites could go undetected by the kernel density estimation or only very frequently visited sites could be detected. However, a different implementation order was not tested, as it would have gone beyond the scope of this thesis.

Chapter 7

7 Conclusion

7.1 Summary

The aim of this master's thesis was to learn whether locations of anthropogenic feeding can be identified by an approach based on individual GPS trajectories of red kites and therefore set a basis for the quantification of anthropogenic feeding sites in Switzerland. Consequently, with an exploratory approach, a methodological workflow was developed and potentially relevant characteristics such as density of GPS localisation, anthropogenic environment and recurring, predictable feeding spots were included.

The results suggest that anthropogenic feeding sites used by red kites in the study area in western Switzerland can be detected with an individual-based approach using GPS tracking data. However, the detectability depends on (1) the amount and frequency of the food provision at the site and (2) the age class of the red kite, which provides the GPS tracking data.

7.2 Contributions

The findings of this master's thesis suggest that a method based on density measures and the inclusion of revisitation information into a generalised linear mixed model delivers a possibility to identify the locations of anthropogenic feeding sites with the data of individual red kite trajectories. Hence, it shows how individual animals can be used as an instrument to detect landscape features based on associated behaviours that are reflected in their movement patterns. On the example of red kites, it was demonstrated, how in particular anthropogenic feeding activity was detectable with GPS tracking data.

The method offers a new, less time-consuming way of detecting potential anthropogenic feeding sites than survey-based approaches and can possibly be applied to larger geographic areas. The reliable detection rate of large feeding sites makes the method especially useful, as those sites could have a substantial effect on the behaviour of red kites. Even smaller feeding sites, when in the proximity of breeding individuals can be successfully detected. While locations of feeding sites can be determined by the method, more detailed information about the frequency, amount and type of food would still need to be collected by survey-based door-to-door interviews.

Although the methodology shows some uncertainties when predicting anthropogenic feeding sites, it became evident that the use of trajectories of breeding birds shows higher prediction success. Depending on the research objective the data of a specific group of red kites can be used or the data can be evaluated based on different thresholds to account for differences between the age classes.

When applying the methodology to non-breeding bird data, one should be aware that while the approach shows success in identifying large feeding sites it is less successful for small anthropogenic feeding sites. By varying evaluation thresholds, the method can be adjusted based on research context and the expectations for precision and recall, respectively. Overall, the results indicate a focus on breeding bird data to detect small anthropogenic feeding sites.

7.3 Limitations

A major challenge was the temporal resolution of the GPS tracking data. Anthropogenic feeding takes place on a very small spatial and temporal scale, hence potential visits to anthropogenic feeding sites and therefore their positions might go undetected because of the low temporal resolution. The aim of the methodological approach, however, was the applicability to the low temporal resolution data set, as the Swiss Ornithological Institute has most data recorded at this interval and only fewer with a higher temporal resolution.

Another uncertain factor was the completeness of the ground truth data for the anthropogenic feeding sites within the study area. Even though there is a high level of confidence that the information about anthropogenic feeding sites in the study area is complete, most of the data is based on a survey and observational data from field assistants. The answers about frequency, amount of feeding and its continuation might depend on whether respondents believe that feeding of red kites is socially respected or frowned upon, respectively. This could influence the categorisation into small-/medium-sized and large-sized anthropogenic feeding sites.

The underlying data of this thesis was the season 2018/2019. However, annual variation of the use of anthropogenic feeding sites cannot be ruled out and was not accounted for in this thesis. The availability of natural food sources could impact the frequency of use of human-related food sources as red kites alternate between both sources. In case of an abundance of natural food, red kites possibly make less use of feeding sites which would result in fewer GPS localisations at those sites and hence affect their identifiability.

There was only little research and scientific literature available about anthropogenic feeding of birds of prey and particularly of red kites for Switzerland. Much is still unclear about the impact and influence of anthropogenic feeding within a red kite's life history and the behaviour of red kites around anthropogenic feeding sites, which made the development of a method challenging.

7.4 Future Research

The context variables included in this thesis were relatively simple and hence an interesting approach would be to test whether the methodology can be significantly improved by including extra land cover information such as locations of tree rows, roosting sites and places of natural food abundance to the generalised linear mixed model. Those additional variables could possibly further reduce the number of polygons incorrectly predicting anthropogenic feeding.

Since this thesis has developed a step-by-step workflow, the single steps can be easily replaced by different methods. Other methods to calculate candidate polygons for anthropogenic feeding sites, such as AKDE for autocorrelated data (Fleming *et al.*, 2015) should be studied to see whether results differ. Furthermore, the methodology discussed in this thesis has been developed based on the low sampling interval and therefore the low density of GPS locations and the revisitation to certain areas. For data with a higher temporal resolution, stop detection algorithms such as the one developed by Montoliu *et al.* (2013) could be tested to detect anthropogenic feeding sites shown by stops in red kite trajectories.

The conclusions drawn in this thesis are based on the red kite GPS tracking data and the anthropogenic feeding sites within the study area in the Sense district and the Schwarzenburgerland in western Switzerland. Hence, the application of the developed workflow to other regions within Switzerland should be tested in future research. As large feeding sites and even some small sites can be confidently identified by the method, its application could lead to better insight into the quantity of anthropogenic feeding taking place within different regions and all of Switzerland. By quantifying anthropogenic feeding in Switzerland and analysing regional differences the impact on red kites' foraging strategies, varying movement patterns based on the origin of the food source (natural vs. anthropogenic) and the potential influence on the population increase could be studied in further detail.

References

Aebischer, A. (2009) *Der Rotmilan: ein faszinierender Greifvogel*. Bern: Haupt.

Akaike, H. (1973) 'Information Theory and an Extension of the Maximum Likelihood Principle', in Petrov, B. N. and Csaki, F. (eds) *2nd International Symposium on Information Theory*. Budapest: Akadémiai Kiadó, pp. 267–281.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. F. and Nielsen, H. (2000) 'Assessing the accuracy of prediction algorithms for classification: an overview', *Bioinformatics*, 16(5), pp. 412–424. doi: 10.1093/bioinformatics/16.5.412.

Bar-David, S., Bar-David, I., Cross, P. C., Ryan, S. J., Knechtel, C. U. and Getz, W. M. (2009) 'Methods for assessing movement path recursion with application to African buffalo in South Africa', *Ecology*, 90(9), pp. 2467–2479. doi: 10.1890/08-1532.1.

Bartumeus, F., Giuggioli, L., Louzao, M., Bretagnolle, V., Oro, D. and Levin, S. A. (2010) 'Fishery Discards Impact on Seabird Movement Patterns at Regional Scales', *Current Biology*. Elsevier Ltd, 20(3), pp. 215–222. doi: 10.1016/j.cub.2009.11.073.

Bates, D., Mächler, M., Bolker, B. M. and Walker, S. C. (2015) 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software*, 67(1). doi: 10.18637/jss.v067.i01.

Benhamou, S. and Riotte-Lambert, L. (2012) 'Beyond the Utilization Distribution: Identifying home range areas that are intensively exploited or repeatedly visited', *Ecological Modelling*. Elsevier B.V., 227, pp. 112–116. doi: 10.1016/j.ecolmodel.2011.12.015.

Berger-Tal, O. and Bar-David, S. (2015) 'Recursive movement patterns: review and synthesis across species', *Ecosphere*, 6(9). doi: 10.1890/ES15-00106.1.

Blanco, G. and Montoya, R. (2004) 'Landscape, "muladares", or poisoning?: using GIS to assess factors related to the decline of breeding red kites (Milvus milvus) in Spain', in Rodríguez-Estrella, R. and Bojórquez Tapia, L. A. (eds) *Spatial analysis in raptor ecology and conservation*. Mexiko, pp. 133–152.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. and White, J.-S. S. (2009) 'Generalized linear mixed models: a practical guide for ecology and evolution', *Trends in Ecology and Evolution*, 24(3), pp. 127–135. doi: 10.1016/j.tree.2008.10.008.

Boughorbel, S., Jarray, F. and El-Anbari, M. (2017) 'Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric', *PLoS ONE*, 12(6), pp. 1–17. doi: 10.1371/journal.pone.0177678.

Bracis, C. (2018) *Package 'recurse'*. Available at: https://cran.r-project.org/web/packages/recurse/recurse.pdf.

Bracis, C., Bildstein, K. L. and Mueller, T. (2018) 'Software note: Revisitation analysis uncovers spatio-temporal patterns in animal movement data', *Ecography*, 41(11), pp. 1–11. doi: 10.1111/ecog.03618.

Branco, P., Torgo, L. and Ribeiro, R. P. (2016) 'A survey of predictive modeling on imbalanced domains', *ACM Computing Surveys*, 49(2), pp. 1–50. doi: 10.1145/2907070.

Buchin, M., Dodge, S. and Speckmann, B. (2012) 'Context-Aware Similarity of Trajectories', in Xiao, N. et al. (eds) *GIScience 2012, LNCS 7478*. Heidelberg: Springer-Verlag Berlin, pp. 43–56. doi: 10.1007/978-3-642-33024-7.

Bundesamt für Landestopografie swisstopo (2007) *Produkteinformation VECTOR25*.

Bundesamt für Landestopografie swisstopo (2019) *swissTLM3D*. Available at: https://shop.swisstopo.admin.ch/de/products/landscape/tlm3D (Accessed: 13 December 2019).

Bundesamt für Landestopografie swisstopo (2020a) *swissBOUNDARIES 3D: Grenzen schweizweit in 3D*.

Bundesamt für Landestopografie swisstopo (2020b) *swissTLM3d Version 1.8*.

Burnham, K. P. and Anderson, D. R. (2002) *Model selection and multilevel inference: a practical information-theoretic approach*. 2nd ed. Springer. doi: 10.1007/978-0-387-22456-5_7.

Bustamante, J. (1993) 'Post-fledging dependence period and development of flight and hunting behaviour in the Red Kite Milvus milvus', *Bird Study*, 40(3), pp. 181–188. doi: 10.1080/00063659309477181.

Cagnacci, F., Boitani, L., Powell, R. A. and Boyce, M. S. (2010) 'Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550), pp. 2157–2162. doi: 10.1098/rstb.2010.0107.

Calenge, C. (2019) *Home Range Estimation in R: the adehabitatHR Package*. Available at: https://cran.r-project.org/web/packages/adehabitatHR/vignettes/adehabitatHR.pdf.

Carter, I. (2001) *The Red Kite*. Chelmsford, Essex: Arlequin Press.

Carter, I. and Powell, D. (2019) *The Red Kite's Year*. Exeter: Pelagic Publishing.

Cereghetti, E., Scherler, P., Fattebert, J. and Grüebler, M. U. (2019) 'Quantification of anthropogenic food subsidies to an avian facultative scavenger in urban and rural habitats', *Landscape and Urban Planning*. Elsevier, 190. doi: 10.1016/j.landurbplan.2019.103606.

Cox, D. T. C. and Gaston, K. J. (2016) 'Urban bird feeding: Connecting people with nature', *PLoS ONE*, 11(7), pp. 1–13. doi: 10.1371/journal.pone.0158717.

Damiani, M. L., Issa, H. and Cagnacci, F. (2014) 'Extracting stay regions with uncertain boundaries from GPS trajectories: A case study in animal ecology', in *Proceedings of the 22nd ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*, pp. 253–262. doi: 10.1145/2666310.2666417.

Davis, J. and Goadrich, M. (2006) 'The Relationship Between Precision-Recall and ROC Curves', in *Proceedings of the 23rd International Conference on Machine Learning*. Pittsburg, PA, pp. 233–240. doi: 10.1145/1143844.1143874.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. and Weibel, R. (2015) 'Analysis and visualisation of movement: An interdisciplinary review', *Movement Ecology*, 3(5), pp. 1–24. doi: 10.1186/s40462-015-0032-y.

Dodge, S., Weibel, R. and Lautenschütz, A. K. (2008) 'Towards a taxonomy of movement patterns', *Information Visualization*, 7(3–4), pp. 240–252. doi: 10.1057/palgrave.ivs.9500182.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz García, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D. and Lautenbach, S. (2013) 'Collinearity: A review of methods to deal with it and a simulation study evaluating their performance', *Ecography*, 36(1), pp. 27–46. doi: 10.1111/j.1600-0587.2012.07348.x.

Downs, J. A. (2010) 'Time-Geographic Density Estimation for Moving Point Objects', in Fabrikant, S. I. et al. (eds) *GIScience 2010, LNCS 6292*. Zurich, Switzerland: Springer Berlin Heidelberg, pp. 16–26.

Fielding, A. H. and Bell, J. F. (1997) 'A review of methods for the assessment of prediction errors in conservation presence/absence models', *Environmental Conservation*, 24(1), pp. 38–49. doi: 10.1017/S0376892997000088.

Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P. and Calabrese, J. M. (2015) 'Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator', *Ecology*, 96(5), pp. 1182–1188. doi: 10.1890/14-2010.1.

Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P. and Calabrese, J. M. (2016) 'Estimating where and how animals travel: an optimal framework for path reconstruction from autocorrelated tracking data', *Ecology*, 97(3), pp. 576–582. doi: 10.1890/15-1607.1.

Garrison, J. S. E. and Gass, C. L. (1999) 'Response of a traplining hummingbird to changes in nectar availability', *Behavioral Ecology*, 10(6), pp. 714–725. doi: 10.1093/beheco/10.6.714.

Der Grosse Rat des Kantons Freiburg (2019) *Gesetz über den Wald und den Schutz vor Naturereignissen (WSG)*.

Grüebler, M. U., Schuler, H., Müller, M., Spaar, R., Horch, P. and Naef-Daenzer, B. (2008) 'Female biased mortality caused by anthropogenic nest loss contributes to population decline and adult sex ratio of a meadow bird', *Biological Conservation*, 141(12), pp. 3040–3049. doi: 10.1016/j.biocon.2008.09.008.

Gudmundsson, J., Laube, P. and Wolle, T. (2017) 'Movement Patterns in Spatio-Temporal Data', in Shekhar, S., Xiong, H., and Zhou, X. (eds) *Encyclopedia of GIS*. Springer, Cham. doi: 10.1007/978-3-319-17885-1.

Harris, M. P., Bogdanova, M. I., Daunt, F. and Wanless, S. (2012) 'Using GPS technology to assess feeding areas of Atlantic Puffins Fratercula arctica', *Ringing and Migration*, 27(1), pp. 43–49. doi: 10.1080/03078698.2012.691247.

Hart, L. A., Wreford, E. P., Brown, M. and Downs, C. T. (2018) 'Hunting flight speeds of five southern African raptors', *Ostrich*, 89(3), pp. 251–258. doi: 10.2989/00306525.2018.1455754.

Hernández-Orallo, J., Flach, P. and Ferri, C. (2012) 'A Unified View of Performance Metrics : Translating Threshold Choice into Expected Classification Loss', *Journal of Machine Learning Research*, 13(1), pp. 2813–2869.

Heuck, C., Brandl, R., Albrecht, J. and Gottschalk, T. K. (2013) 'The potential distribution of the Red Kite in Germany', *Journal of Ornithology*, 154, pp. 911–921. doi: 10.1007/s10336-013-0955-2.

Hötker, H., Krone, O. and Nehls, G. (2017) *Birds of Prey and Wind Farms: Analysis of Problems and Possible Solutions*. Springer, Cham. doi: 10.1007/978-3-319-53402-2.

Kays, R., Crofoot, M. C., Jetz, W. and Wikelski, M. (2015) 'ECOLOGY. Terrestrial animal tracking as an eye on life and planet', *Science*, 348(6240). doi: 10.1126/science.aaa2478.

Knaus, P., Antoniazza, S., Wechsler, S., Guélat, J., Strebel, N. and Sattler, T. (2018) *Schweizer Brutvogelatlas 2013 - 2016. Verbreitung und Bestandesentwicklung der Vögel in der Schweiz und im Fürstentum Liechtenstein*. Schweizerische Vogelwarte, Sempach.

Knaus, P., Müller, C., Sattler, T., Schmid, H., Strebel, N. and Volet, B. (2019) *The State of Birds in Switzerland: Report 2019*. Swiss Ornithological Institute, Sempach. Available at: www.vogelwarte.ch/state.

Laundré, J. W. (2010) 'Behavioral response races, predator–prey shell games, ecology of fear, and patch use of pumas and their ungulate prey', *Ecology*, 91(10), pp. 2995–3007. doi: 10.1890/08-2345.1.

Lobo, J. M., Jiménez-Valverde, A. and Real, R. (2008) 'AUC: A misleading measure of the performance of predictive distribution models', *Global Ecology and Biogeography*, 17, pp. 145–151. doi: 10.1111/j.1466-8238.2007.00358.x.

Lyons, A. J., Turner, W. C. and Getz, W. M. (2013) 'Home range plus: a space-time characterization of movement over real landscapes', *Movement Ecology*, 1(2), pp. 1–14. doi: 10.1186/2051-3933-1-2.

Maratea, A., Petrosino, A. and Manzo, M. (2014) 'Adjusted F-measure and kernel scaling for imbalanced data learning', *Information Sciences*. Elsevier Inc., 257, pp. 331–341. doi: 10.1016/j.ins.2013.04.016.

Mcgrady, M. J., Karelus, D. L., Rayaleh, H. A., Sarrouf Willson, M., Meyburg, B.-U., Oli, M. K. and Bildstein, K. (2018) 'Home ranges and movements of Egyptian Vultures Neophron percnopterus in relation to rubbish dumps in Oman and the Horn of Africa', *Bird Study*. Taylor & Francis, 65(4), pp. 544–556. doi: 10.1080/00063657.2018.1561648.

Menz, M. H. M., Mosimann-Kampe, P. and Arlettaz, R. (2009) 'Foraging Habitat Selection in the Last Ortolan Bunting Emberiza hortulana Population in Switzerland: Final Lessons before Extinction', *Ardea*, 97(3), pp. 323–333. doi: 10.5253/078.097.0308.

Millspaugh, J. J. and Marzluff, J. M. (2001) 'Radio-Tracking and Animal Populations: Past Trends and Future Needs', in Millspaugh, J. J. and Marzluff, J. M. (eds) *Radio Tracking and Animal Populations*. Academic Press, pp. 383–393. doi: 10.1016/B978-012497781-5/50016-5.

Monsarrat, S., Benhamou, S., Sarrazin, F., Bessa-Gomes, C., Bouten, W. and Duriez, O. (2013) 'How Predictability of Feeding Patches Affects Home Range and Foraging Habitat Selection in Avian Social Scavengers?', *PLoS ONE*, 8(1), pp. 1–11. doi: 10.1371/journal.pone.0053077.

Montoliu, R., Blom, J. and Gatica-Perez, D. (2013) 'Discovering places of interest in everyday life from smartphone data', *Multimedia Tools and Applications*, 62(1), pp. 179–207. doi: 10.1007/s11042-011-0982-z.

Mougeot, F., Garcia, J. T. and Viñuela, J. (2011) 'Breeding biology, behaviour, diet and conservation of the red kite (Milvus milvus), with particular emphasis on Mediterranean populations', in Zuberogoitia, I. and Martínez, J. E. (eds) *Ecology and conservation of European dwelling forest raptors and owls*. Bilbao, Spain: Diputación Foral de Vizcaya, pp. 190–204. Available at: http://www.eeza.csic.es/Documentos/Publicaciones/2011-Red kite book chapter.pdf.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. and Smouse, P. E. (2008) 'A movement ecology paradigm for unifying organismal movement research', *Proceedings of the National Academy of Sciences of the United States of America*, 105(49), pp. 19052–19059. doi: 10.1073/pnas.0800375105.

Navarro, J., Grémillet, D., Afán, I., Ramírez, F., Bouten, W. and Forero, M. G. (2016) 'Feathered Detectives: Real-Time GPS Tracking of Scavenging Gulls Pinpoints Illegal Waste Dumping', *PLoS ONE*, 11(7), pp. 1–9. doi: 10.1371/journal.pone.0159974.

Orians, G. H. and Pearson, N. E. (1979) 'On the theory of central place foraging', in Horn, D. J., Mitchell, R. D., and Stairs, G. R. (eds) *Analysis of Ecological Systems*. Columbus: The Ohio State University Press, pp. 154–177.

Orros, M. E. and Fellowes, M. D. E. (2015) 'Widespread supplementary feeding in domestic gardens explains the return of reintroduced Red Kites Milvus milvus to an urban area', *Ibis*, 157, pp. 230–238. doi: 10.1111/ibi.12237.

Pennycuick, C. J., Åkesson, S. and Hedenström, A. (2013) 'Air speeds of migrating birds observed by ornithodolite and compared with predictions from flight theory', *Journal of the Royal Society Interface*, 10(86). doi: 10.1098/rsif.2013.0419.

Pfeiffer, T. and Meyburg, B. U. (2015) 'GPS tracking of Red Kites (Milvus milvus) reveals fledgling number is negatively correlated with home range size', *Journal of Ornithology*. Springer Berlin Heidelberg, 156, pp. 963–975. doi: 10.1007/s10336-015-1230-5.

Plummer, K. E., Siriwardena, G. M., Conway, G. J., Risely, K. and Toms, M. P. (2015) 'Is supplementary feeding in gardens a driver of evolutionary change in a migratory bird species?', *Global Change Biology*, 21(12), pp. 4353–4363. doi: 10.1111/gcb.13070.

Plummer, K. E., Risely, K., Toms, M. P. and Siriwardena, G. M. (2019) 'The composition of British bird communities is associated with long-term garden bird feeding', *Nature Communications*, 10(2088), pp. 1–8. doi: 10.1038/s41467-019-10111-5.

Powers, D. M. W. (2011) 'Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation', *Journal of Machine Learning Technologies*, 2(1), pp. 37–63. doi: 10.9735/2229-3981.

R Core Team (2019) 'R: A Language and Environment for Statistical Computing'. Vienna, Austria. Available at: http://www.r-project.org/.

Reynolds, S. J., Galbraith, J. A., Smith, J. A. and Jones, D. N. (2017) 'Garden bird feeding: Insights and prospects from a north-south comparison of this global urban phenomenon', *Frontiers in Ecology and Evolution*, 5(24), pp. 1–15. doi: 10.3389/fevo.2017.00024.

Röthlisberger, H. (2019) *Günstiger bauen*, *Die Planungsphase*. Available at: https://buch-guenstiger-bauen.com/ (Accessed: 26 February 2020).

Saito, T. and Rehmsmeier, M. (2015) 'The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets', *PLoS ONE*, 10(3), pp. 1–21. doi: 10.1371/journal.pone.0118432.

Silverman, B. W. (1998) *Density Estimation for Statistics and Data Analysis*. New York: Routledge. doi: 10.1201/9781315140919.

Slingsby, A. and van Loon, E. (2016) 'Exploratory Visual Analysis for Animal Movement Ecology', *Computer Graphics Forum*, 35(3), pp. 471–480. doi: 10.1111/cgf.12923.

Swiss Ornithological Institute (2019) *Red Kite Milvus Milvus*, *Birds of Switzerland*. Available at: https://www.vogelwarte.ch/en/birds/birds-of-switzerland/red-kite (Accessed: 4 November 2019).

Swiss Ornithological Institute (2020) *Mechanisms of population dynamics in Red Kites*, *Ecological research*. Available at: https://www.vogelwarte.ch/en/projects/ecological-research/mechanisms-of-population-dynamics-in-red-kites (Accessed: 19 September 2020).

Tharwat, A. (2018) 'Classification assessment methods', *Applied Computing and Informatics*. doi: 10.1016/j.aci.2018.08.003.

Vansteelant, W. M. G., Bouten, W., Klaassen, R. H. G., Koks, B. J., Schlaich, A. E., van Diermen, J., van Loon, E. E. and Shamoun-Baranes, J. (2015) 'Regional and seasonal flight speeds of soaring migrants and the role of weather conditions at hourly and daily scales', *Journal of Avian Biology*, 46(1), pp. 25–39. doi: 10.1111/jav.00457.

Vicente, J., Höfle, U., Garrido, J. M., Fernández-De-Mera, I. G., Juste, R., Barral, M. and Gortazar, C. (2006) 'Wild boar and red deer display high prevalences of tuberculosis-like lesions in Spain', *Veterinary Research*, 37(1), pp. 107–119. doi: 10.1051/vetres.

Weimerskirch, H., Filippi, D. P., Collet, J., Waugh, S. M. and Patrick, S. C. (2018) 'Use of radar detectors to track attendance of albatrosses at fishing vessels', *Conservation Biology*, 32(1), pp. 240–245. doi: 10.1111/cobi.12965.

Welti, N., Scherler, P. and Grüebler, M. U. (2019) 'Carcass predictability but not domestic pet introduction affects functional response of scavenger assemblage in urbanized habitats', *Functional Ecology*, (00), pp. 1–11. doi: 10.1111/1365-2435.13469.

Wikelski, M., Davidson, S. and Kays, R. (2020) *Movebank: archive, analysis and sharing of animal movement data. Hosted by the Max Planck Institute of Animal Behavior.* Available at: https://www.movebank.org/cms/movebank-main (Accessed: 3 March 2020).

Williams, N. M. and Thomson, J. D. (1998) 'Trapline foraging by bumble bees: III. Temporal patterns of visitation and foraging success at single plants', *Behavioral Ecology*, 9(6), pp. 612–621. doi: 10.1093/beheco/9.6.612.

Van Winkle, W. (1975) 'Comparison of Several Probabilistic Home-Range Models', *The Journal of Wildlife Management*, 39(1), pp. 118–123. doi: 10.2307/3800474.

Worton, B. J. (1989) 'Kernel Methods for Estimating the Utilization Distribution in Home-Range Studies', *Ecology*, 70(1), pp. 164–168. doi: 10.2307/1938423.

Yoda, K., Tomita, N., Mizutani, Y., Narita, A. and Niizuma, Y. (2012) 'Spatio-temporal responses of black-tailed gulls to natural and anthropogenic food resources', *Marine Ecology Progress Series*, 466, pp. 249–259. doi: 10.3354/meps09939.

Zheng, Y. (2015) 'Trajectory data mining: An overview', *ACM Transactions on Intelligent Systems and Technology*, 6(3), p. 41. doi: 10.1145/2743025.

Zou, Q., Xie, S., Lin, Z., Wu, M. and Ju, Y. (2016) 'Finding the Best Classification Threshold in Imbalanced Classification', *Big Data Research*, 5, pp. 2–8. doi: 10.1016/j.bdr.2015.12.001.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. and Smith, G. M. (2009) *Mixed Effects Models and Extensions in Ecology with R, Statistics for Biology and Health*. New York, NY: Springer. doi: 10.1007/978-0-387-87458-6.

Appendix A

Exploratory Data Analysis

To obtain a first impression of the red kite trajectory data, visualisations of the data for all of Europe and within Switzerland was conducted.

A first finding was the large heterogeneity in the red kite's data set and the strongly varying behaviour between individuals visible in the GPS trajectories. While some individuals showed migrating patterns to Spain or France in winter others remained in Switzerland during the whole year.

When focusing on the patterns of GPS localisations in Switzerland it could furthermore be distinguished that many GPS points were located in forest areas. This was not unexpected, but nevertheless important for the development of an appropriate methodology. The pattern can be explained by red kites' behaviour of spending time sitting on tree branches, in their nest or returning to these areas. A variety of behavioural patterns, such as age-dependent movement patterns, were visible in the GPS trajectories. Two sample trajectories are illustrated in Figure A.1 and Figure A.2. Figure A.1 displays the trajectory of a migrating, non-breeding red kite, spending summer in Switzerland and Germany, showing exploratory behaviour and winter months in Spain. Figure A.2 shows the trajectory of a sedentary, breeding red kite spending 12 months in the same area in Switzerland using approximately 50 km².

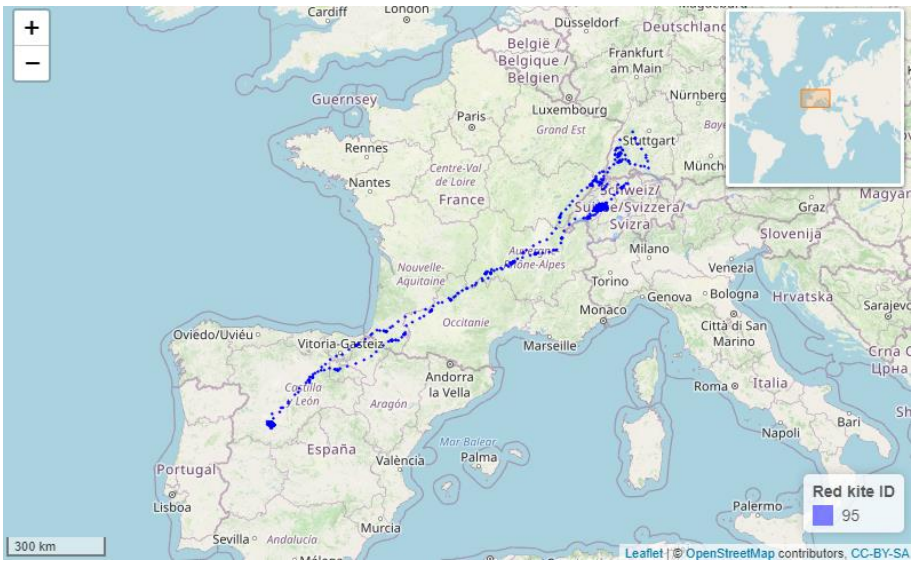


Figure A.1. Trajectory of a migrating red kite (season 2018/2019). Base map: OpenStreetMap contributors (2020).

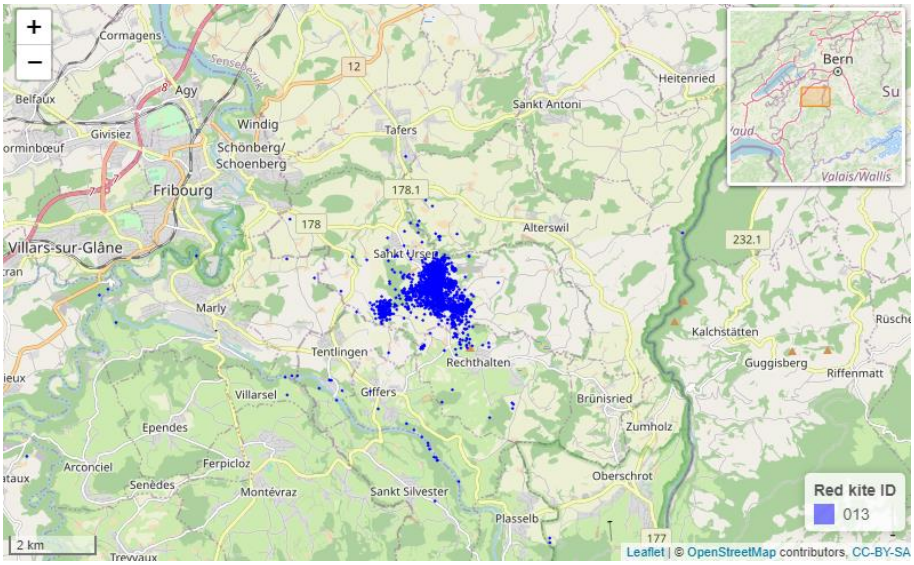


Figure A.2. Trajectory of a sedentary red kite (season 2018/2019). Base map: OpenStreetMap contributors (2020).

The data availability of red kites by age class and month was analysed to gain insight into the data structure. For the study area breeding birds showed data availability for more months over the course of the year than non-breeding birds or fledglings (Figure A.3). On average breeding birds spent ten months within the study area, non-breeding birds seven and fledglings three. As fledglings leave their nests only in July there were consequently fewer data available relevant for this thesis.

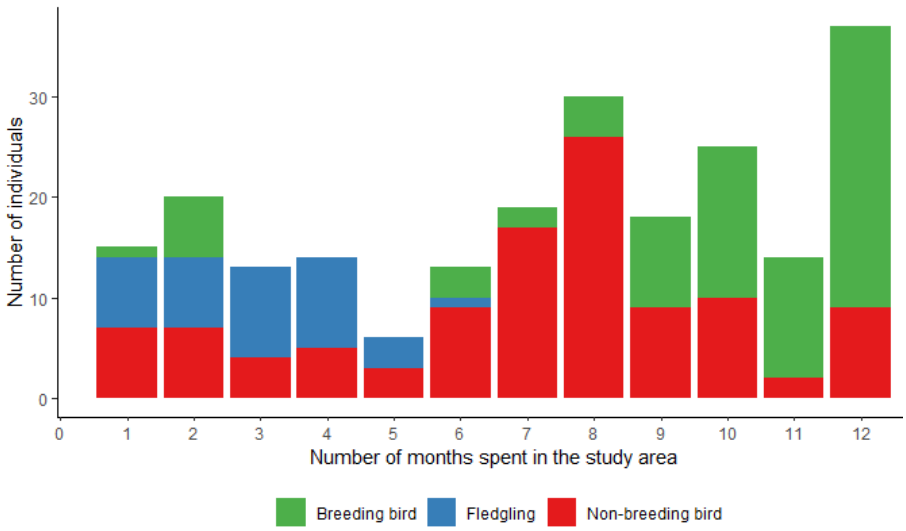
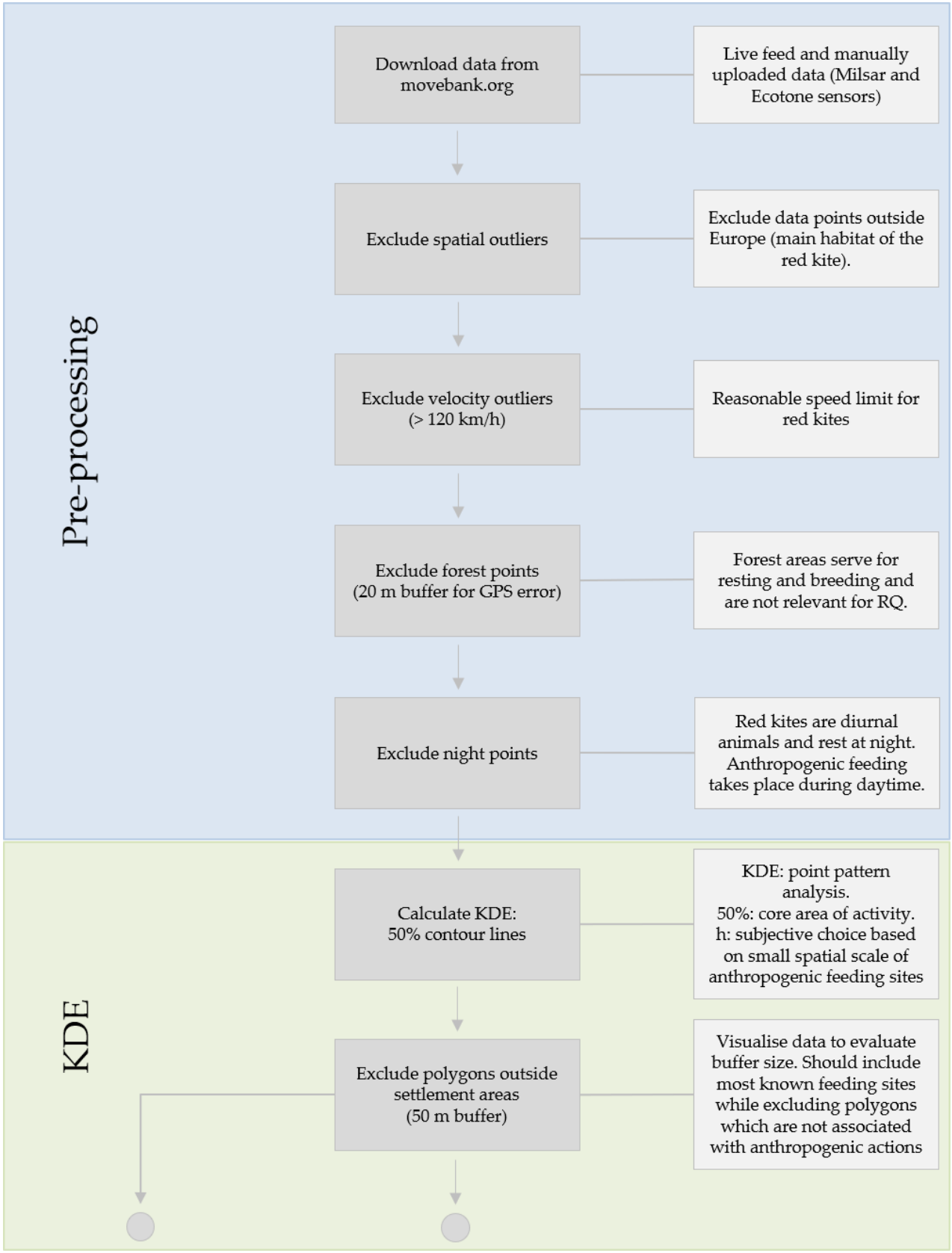


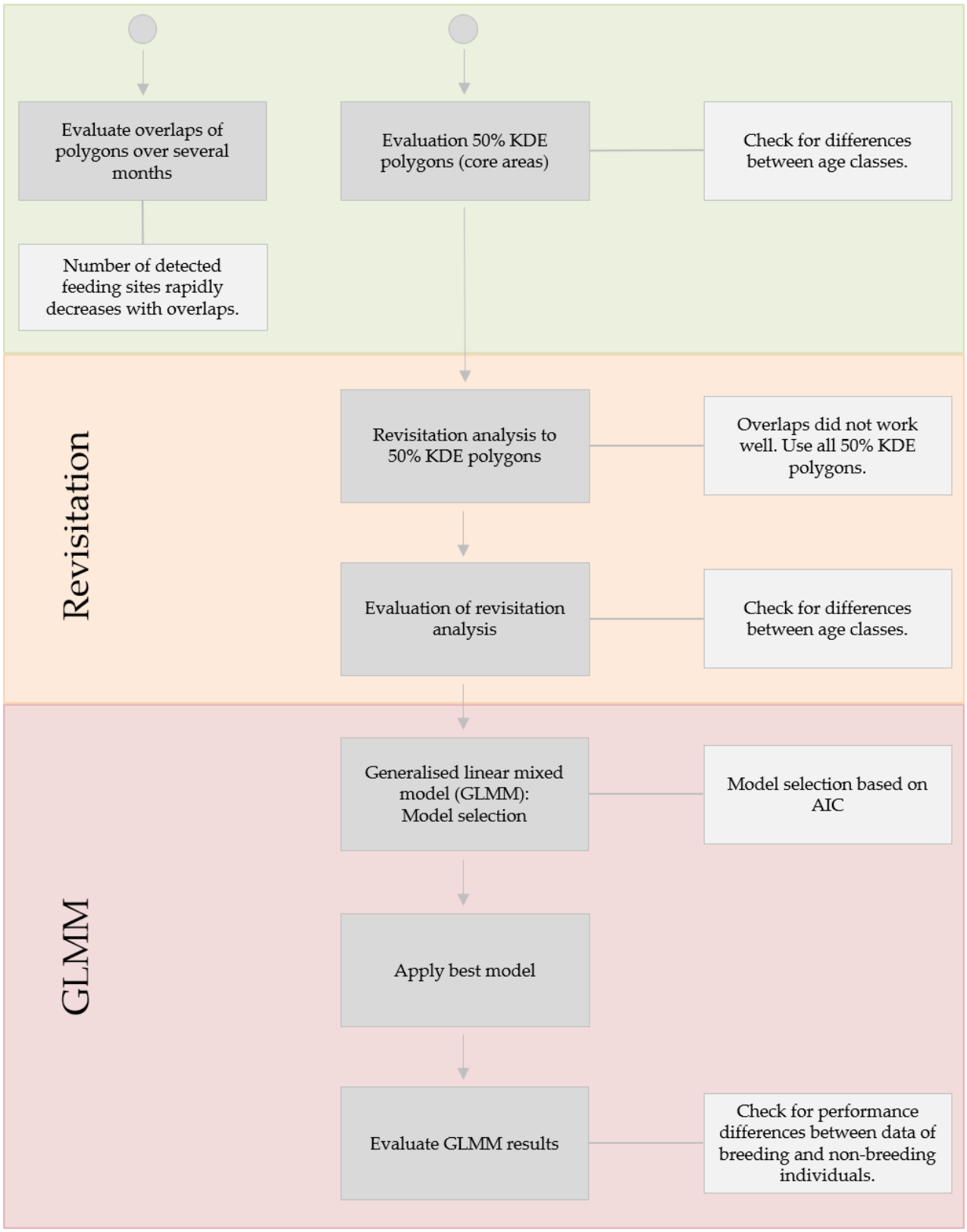
Figure A.3. Number of months with available data for the study area by age class.

Appendix B

Methodological Workflow

Figure B.1. Workflow with tasks in grey and decision-making arguments or descriptions in light grey.

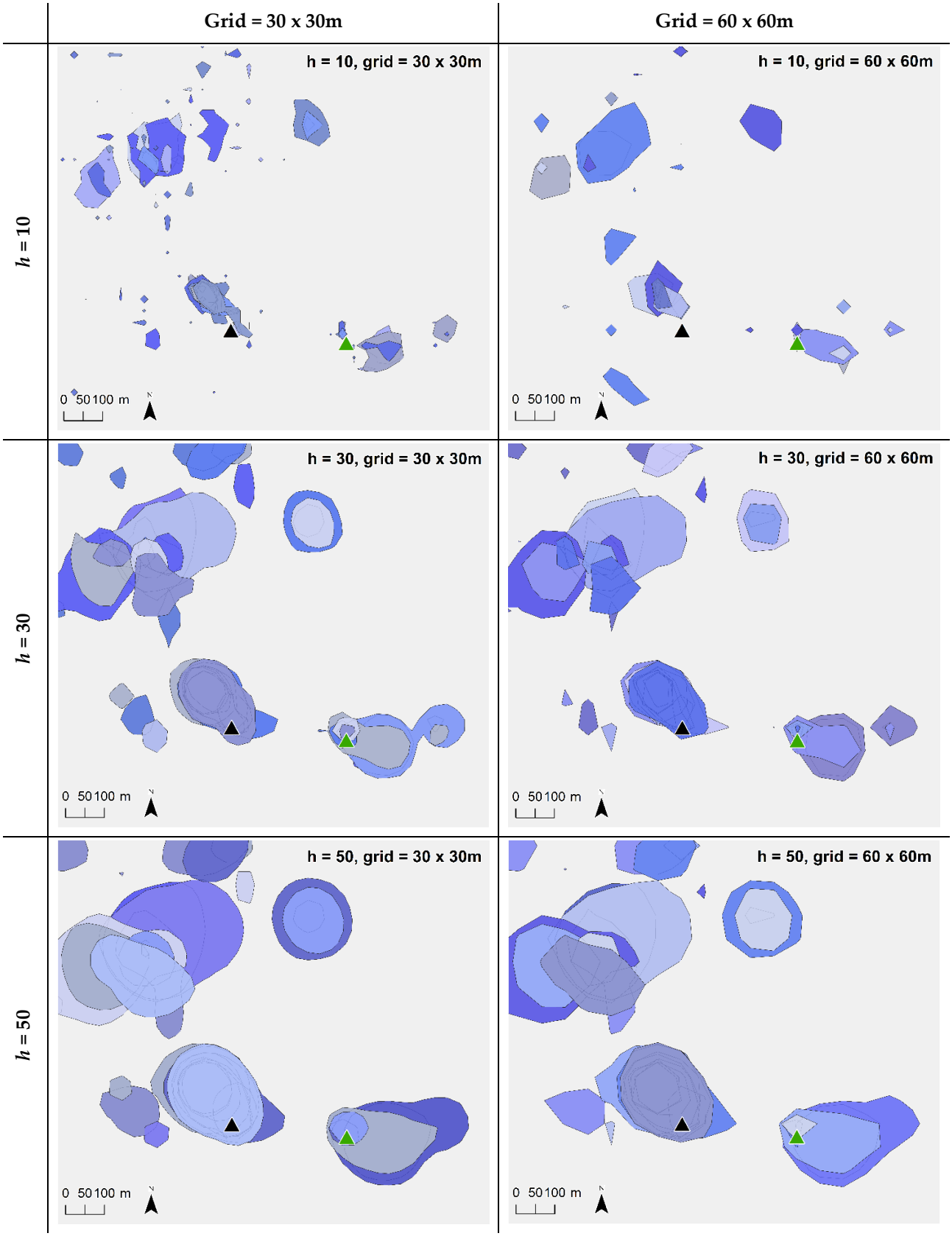


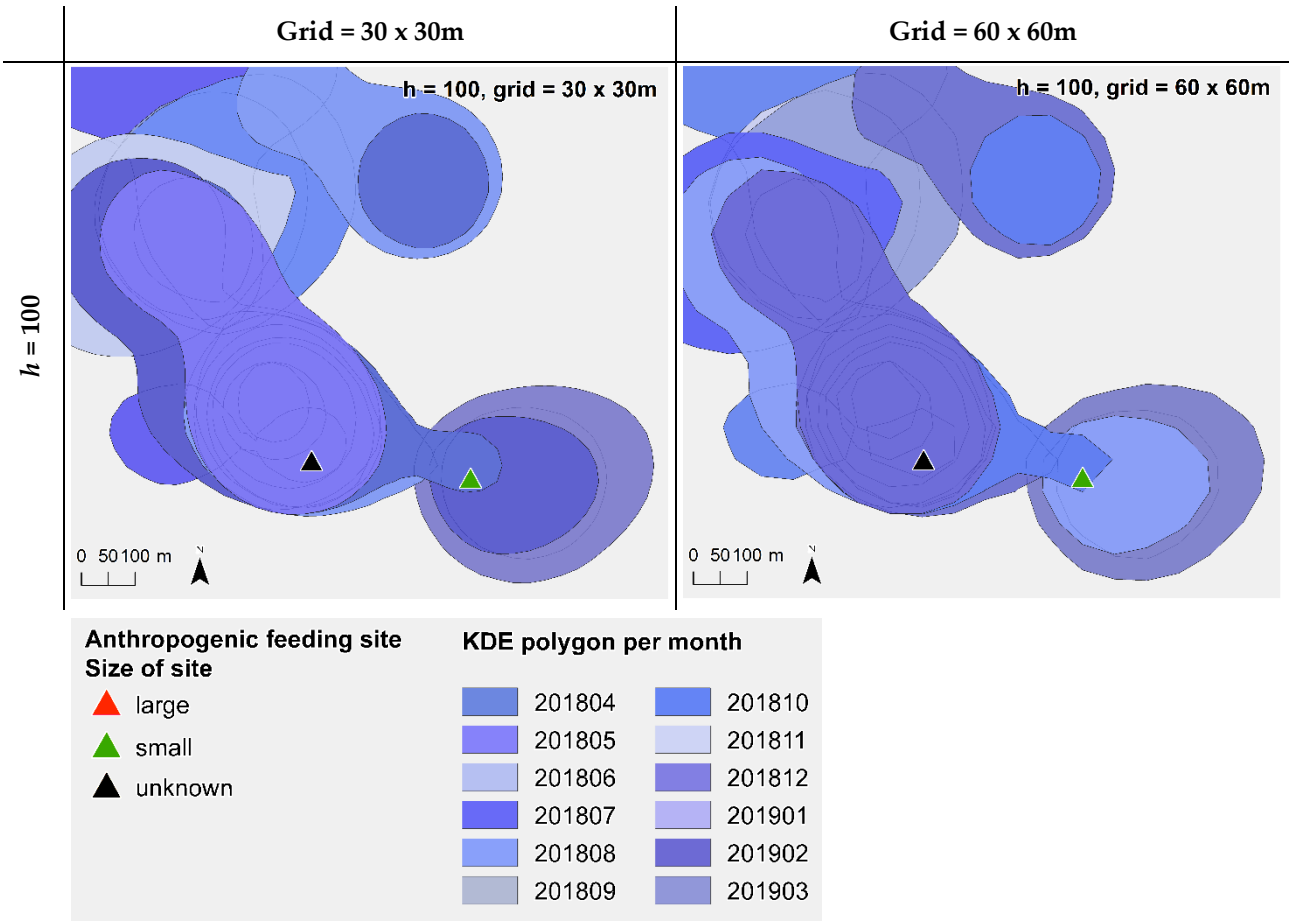

Evaluate overlaps of polygons over several months
Number of detected feeding sites rapidly decreases with overlaps.
Evaluation 50% KDE polygons (core areas)
Check for differences between age classes.
Revisitation
Revisitation analysis to 50% KDE polygons
Overlaps did not work well. Use all 50% KDE polygons.
Evaluation of revisitation analysis
Check for differences between age classes.
GLMM
Generalised linear mixed model (GLMM): Model selection
Model selection based on AIC
Apply best model
Evaluate GLMM results
Check for performance differences between data of breeding and non-breeding individuals.

Appendix C

Kernel Density Estimation

Table C.1. Parameter choice for kernel density estimation: example visualised for red kite no. 013.




Grid = 30 x 30m
Grid = 60 x 60m
h = 100
h = 100, grid = 30 x 30m
h = 100, grid = 60 x 60m
0 50100 m
0 50100 m
Anthropogenic feeding site
Size of site
large
small
unknown
KDE polygon per month
201804
201805
201806
201807
201808
201809
201810
201811
201812
201901
201902
201903

Appendix D

Generalised Linear Mixed Model

Table D.1. Overview of fixed effects included in the generalised linear mixed model.

Variable	Description	Type
revisits.z	Number of revisits to a polygon during a month (scaled).	Numeric
residencetime.z	Total time [hours] spent during all visits to polygon in one month (scaled).	Numeric
season	Information about season during which polygon was produced 2 levels: breeding time, non-breeding time.	Categorical
age_class_2018	Information about age class of red kite. 3 levels: fledgling, non-breeding bird, breeding bird	Categorical
area_ha.z	Area of polygon in hectares (scaled).	Numeric
points_month	Amount of GPS points recorded in the month of the KDE polygon for 1 individual. E.g., if 1 individual has several KDE-polygons in a single month, the amount of GPS points attributed to those polygons is the same.	Numeric

Table D.2. Overview of random effects included in the generalised linear mixed model.

Variable	Description	Type
BirdID	Number identifying the individual red kite creating a polygon.	Factor
month	Month attributed to a polygon.	Factor

Table D.3. GLMM candidate models with AIC and AUC values. Model no. 5 was selected based on the lowest AIC value.

Model number	Model equation	AIC	AUC
1	feedingsite ~ revisits.z * residencetime.z + revisits.z * season + revisits.z * age_class_2018 + revisits.z * area_ha.z + residencetime.z * season + residencetime.z * age_class_2018 + residencetime.z * area_ha.z + points_month.z + (1 | BirdID) + (1 | month)	1216.9	0.954
2	feedingsite ~ revisits.z * residencetime.z + revisits.z * season + revisits.z * age_class_2018 + revisits.z * area_ha.z + residencetime.z * season + residencetime.z * area_ha.z + points_month.z + (1 | BirdID) + (1 | month)	1215.8	0.953
3	feedingsite ~ revisits.z * residencetime.z + revisits.z * season + revisits.z * area_ha.z + residencetime.z * season + residencetime.z * area_ha.z + age_class_2018 + points_month.z + (1 | BirdID) + (1 | month)	1212.5	0.953
4	feedingsite ~ revisits.z * residencetime.z + revisits.z * season + revisits.z * area_ha.z + residencetime.z * season + age_class_2018 + points_month.z + (1 | BirdID) + (1 | month)	1210.7	0.953
5	**feedingsite ~ revisits.z * residencetime.z + revisits.z * area_ha.z + residencetime.z * season + age_class_2018 + points_month.z + (1 | BirdID) + (1 | month)**	**1209.4**	**0.952**
6	feedingsite ~ revisits.z * residencetime.z + revisits.z * area_ha.z + season + age_class_2018 + points_month.z + (1 | BirdID) + (1 | month)	1211.5	0.954
7	feedingsite ~ revisits.z + residencetime.z + revisits.z * area_ha.z + season + age_class_2018 + points_month.z + (1 | BirdID) + (1 | month)	1212.5	0.955
8	feedingsite ~ revisits.z + residencetime.z + season + age_class_2018 + area_ha + points_month.z + (1 | BirdID) + (1 | month)	1218.2	0.954
9	feedingsite ~ residencetime.z + season + age_class_2018 + area_ha + points_month.z + (1 | BirdID) + (1 | month)	1216.6	0.954
10	feedingsite ~ residencetime.z + age_class_2018 + area_ha + points_month.z + (1 | BirdID) + (1 | month)	1215.6	0.953
11	feedingsite ~ residencetime.z + age_class_2018 + area_ha + (1 | BirdID) + (1 | month)	1214.9	0.954
12	feedingsite ~ residencetime.z + area_ha + (1 | BirdID) + (1 | month)	1215.2	0.955
13	feedingsite ~ area_ha + (1 | BirdID) + (1 | month)	1215.3	0.955

Table D.4. Output and diagnostics of the 'best' model from Table D.3.

Output & diagnostics of the 'best' model

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod
Family: binomial  ( logit )
Formula: feedingsite ~ revisits.z * residencetime.z + revisits.z * area_ha.z +
    residencetime.z * season + season + age_class_2018 + points_month.z +
    (1 | BirdID) + (1 | month)
   Data: poly_kde50_2018_revisit
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))

     AIC      BIC   logLik deviance df.resid
  1209.4   1308.1   -591.7   1183.4    14687

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8770 -0.0699 -0.0282 -0.0202 24.1407

Random effects:
 Groups Name        Variance Std.Dev.
 BirdID (Intercept) 6.4418   2.5381
 month  (Intercept) 0.1673   0.4091
Number of obs: 14700, groups:  BirdID, 224; month, 12

Fixed effects:
                                     Estimate Std. Error z value Pr(>|z|)
(Intercept)                          -6.88785    0.73041  -9.430  < 2e-16 ***
revisits.z                            0.48391    0.16136   2.999  0.00271 **
residencetime.z                       0.23391    0.11160   2.096  0.03609 *
area_ha.z                             0.52588    0.07827   6.719 1.83e-11 ***
seasonnon_breeding_time               0.52515    0.34080   1.541  0.12333
age_class_2018fledgling              -2.02570    1.39290  -1.454  0.14586
age_class_2018non_breeding_bird      -0.38538    0.58339  -0.661  0.50887
points_month.z                       -0.25130    0.21408  -1.174  0.24045
revisits.z:residencetime.z           -0.07640    0.03783  -2.020  0.04339 *
revisits.z:area_ha.z                 -0.08643    0.04301  -2.010  0.04448 *
residencetime.z:seasonnon_breeding_time -0.25376    0.12965  -1.957  0.05031 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) rvsts. rsdnc. ar_h.z ssnn__ ag__2018 a__2018_ pnts_. rvs.:. rv.:_.
revisits.z  -0.188
residnctm.z -0.201  0.170
area_ha.z   -0.006 -0.097 -0.362
ssnnn_brdn_ -0.336  0.118  0.180  0.000
ag_cls_2018 -0.196  0.070  0.030 -0.048 -0.020
ag_c_2018__ -0.644  0.249  0.225 -0.109  0.136  0.208
pnts_mnth.z -0.053 -0.371  0.116  0.034  0.086 -0.158    0.089
rvsts.z:rs.  0.156 -0.623 -0.669  0.379 -0.075 -0.053   -0.208    0.084
rvsts.z:r_. -0.056 -0.257  0.136 -0.618 -0.008  0.039    0.036    0.037 -0.187
rsdnctm.:__  0.102 -0.221 -0.404 -0.037 -0.268  0.019   -0.058    0.018  0.275  0.059
```

Appendix E

Evaluation of Breeding Bird and Non-breeding Bird Data combined

Table E.1. Evaluation per anthropogenic feeding site with optimal threshold 0.188 for breeding and non-breeding bird data combined.

ID	Size	Frequency	Amount [g]	Restricted feeding	TP polygons	FN polygons	Ratio	Recall*
25	small	daily	100	year round	5	0	5	100
28	small	several times per week	700	wintering	1	0	1	100
40	small	unknown	unknown	unknown	1	0	1	100
31	small	several times per week	100	year round	10	1	10	90.9
7	small	several times per week	100	year round	11	2	5.5	84.6
1	large	several times per week	300	year round	11	8	1.4	57.9
38	unknown	unknown	unknown	unknown	8	6	1.3	57.1
20	small	several times per week	100	year round	1	1	1	50
24	small	weekly	100	year round	2	3	0.7	40
2	large	several times per week	3000	year round	9	18	0.5	33.3
13	small	14d	100	year round	1	3	0.3	25
36	large	weekly	ca. 5000	year round	2	6	0.3	25
33	large	daily	3000	year round	1	5	0.2	16.7
3	large	daily	500	year round	0	7	0	0
5	medium	weekly	ca.5000	wintering	0	1	0	0
8	small	weekly	1500	wintering	0	1	0	0
15	small	14d	1500	wintering	0	2	0	0
21	small	daily	100	year round	0	1	0	0
26	small	weekly	1500	wintering	0	1	0	0
29	small	weekly	1500	year round	0	1	0	0
32	small	daily	100	year round	0	1	0	0
34	unknown	unknown	>2000	unknown	0	8	0	0
35	large	unknown	ca. 5000	unknown	0	1	0	0
37	small	unknown	unknown	unknown	0	1	0	0
39	small	unknown	unknown	unknown	0	1	0	0
41	unknown	unknown		unknown	0	1	0	0

TP: true positives
FN: false negatives
*Recall: The model correctly identifies xx% of the polygons of an anthropogenic feeding site.

Evaluation of Breeding Bird Data

Table E.2. Evaluation per anthropogenic feeding site with optimal threshold 0.266 for breeding bird data.

ID	Size	Frequency	Amount [g]	Restricted feeding	TP polygons	FN polygons	Ratio	Recall
1	large	several times per week	300	year round	11	0	11 / 0	100
25	small	daily	100	year round	5	0	5 / 0	100
2	large	several times per week	3000	year round	7	1	7	87.5
31	small	several times per week	100	year round	9	2	4.5	81.8
7	small	several times per week	100	year round	10	3	3.33	76.9
38	unknown	unknown	unknown	unknown	7	3	2.33	70
20	small	several times per week	100	year round	1	1	1	50
24	small	weekly	100	year round	1	1	1	50
13	small	14d	100	year round	1	3	0.33	25
36	large	weekly	ca.5000	year round	1	3	0.33	25
21	small	daily	100	year round	0	1	0	0
26	small	weekly	1500	wintering	0	1	0	0
28	small	several times per week	700	wintering	0	1	0	0
34	unknown	unknown	>2000	unknown	0	1	0	0
35	large	unknown	ca.5000	unknown	0	1	0	0
40	small	unknown	unknown	unknown	0	1	0	0

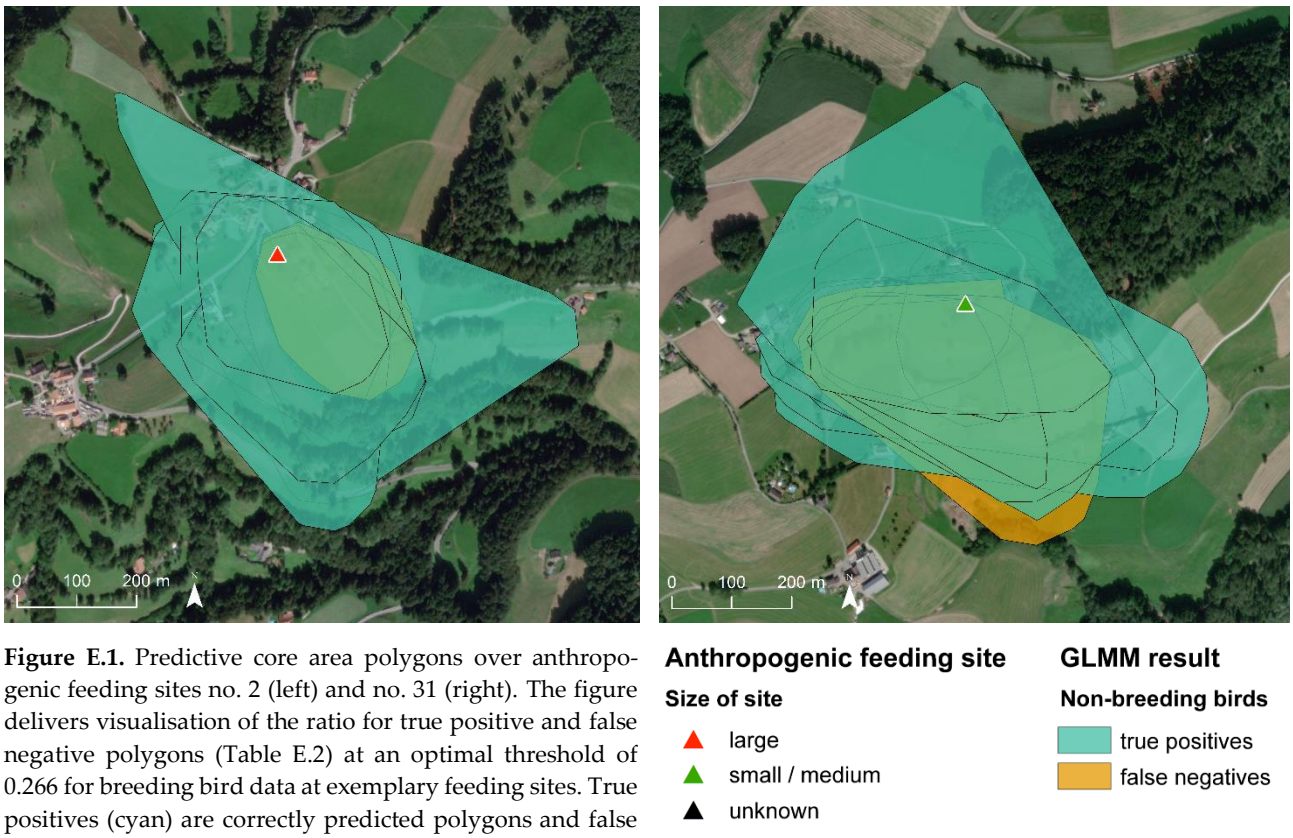


Figure E.1. Predictive core area polygons over anthropogenic feeding sites no. 2 (left) and no. 31 (right). The figure delivers visualisation of the ratio for true positive and false negative polygons (Table E.2) at an optimal threshold of 0.266 for breeding bird data at exemplary feeding sites. True positives (cyan) are correctly predicted polygons and false negatives (orange) are incorrectly predicted polygons, where the model predicts no feeding site but the actual value for the polygon is a feeding site. Base map: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS user community.

Evaluation of Non-breeding Bird Data

Table E.3. Evaluation per anthropogenic feeding site with optimal threshold 0.082 for non-breeding bird data.

ID	Size	Frequency	Amount [g]	Restricted feeding	TP polygons	FN polygons	Ratio	Recall
2	large	several times per week	3000	year round	7	12	0.6	36.8
33	large	daily	3000	year round	2	4	0.5	33.3
3	large	daily	500	year round	1	6	0.2	14.3
34	unknown	unknown	>2000	unknown	1	6	0.2	14.3
1	large	several times per week	300	year round	1	7	0.1	12.5
5	medium	weekly	ca.5000	wintering	0	1	0	0
8	small	weekly	1500	wintering	0	1	0	0
15	small	14d	1500	wintering	0	2	0	0
24	small	weekly	100	year round	0	3	0	0
29	small	weekly	1500	year round	0	1	0	0
32	small	daily	100	year round	0	1	0	0
36	large	weekly	ca.5000	year round	0	3	0	0
37	small	unknown	unknown	unknown	0	1	0	0
38	unknown	unknown	unknown	unknown	0	4	0	0
39	small	unknown	unknown	unknown	0	1	0	0
41	unknown	unknown	unknown	unknown	0	1	0	0

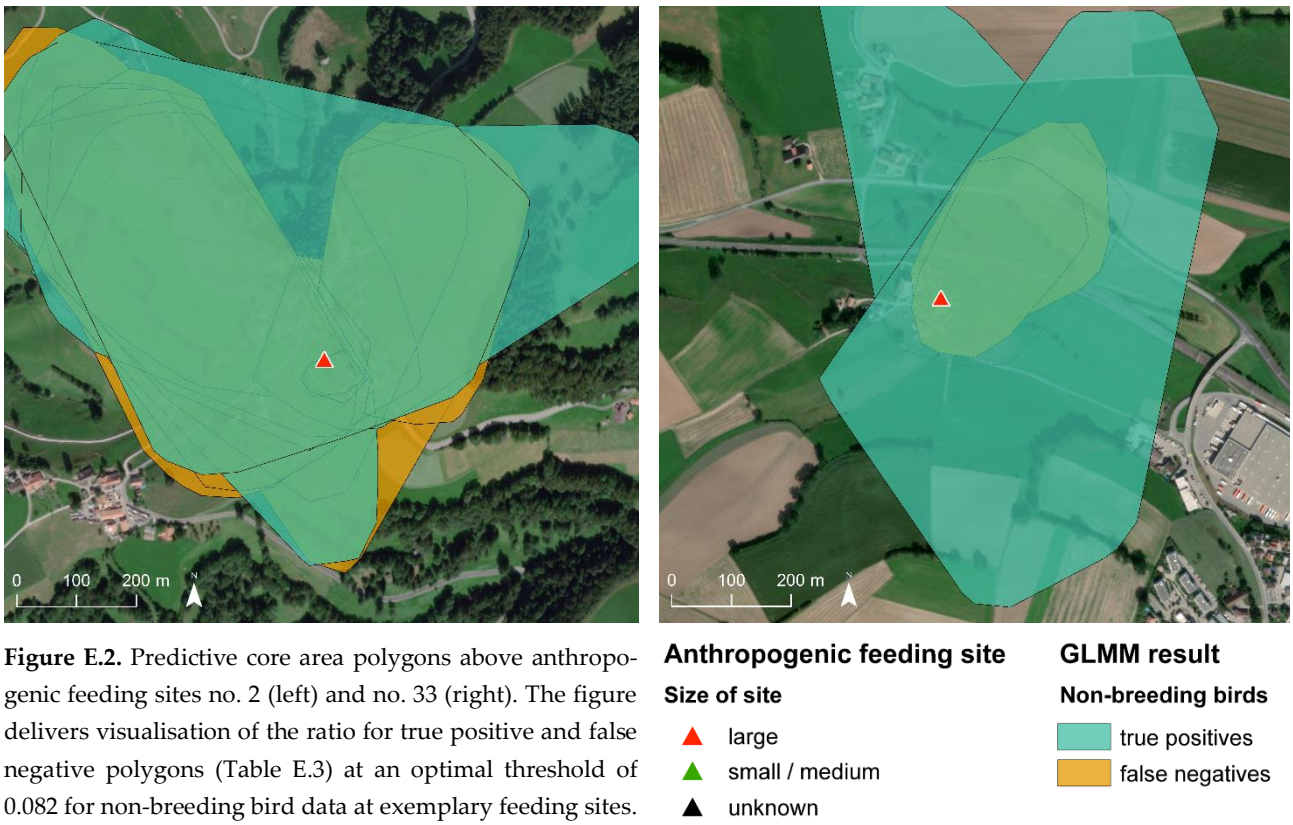


Figure E.2. Predictive core area polygons above anthropogenic feeding sites no. 2 (left) and no. 33 (right). The figure delivers visualisation of the ratio for true positive and false negative polygons (Table E.3) at an optimal threshold of 0.082 for non-breeding bird data at exemplary feeding sites. True positives (cyan) are correctly predicted polygons and false negatives (orange) are incorrectly predicted polygons, where the model predicts no feeding site but the actual value for the polygon is a feeding site. Base map: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS user community.

Declaration of Authorship

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Date	Signature of the author
30 September 2020