Analysis of geo referenced contact and GPS data of wild ranging animals

Master thesis GEO 511
Geographic Information Science GIS
Department of Geography, University of Zurich

Submission date
30 September 2018

Submitted by
Susanne Täschler
susanne.taeschler@uzh.ch
13-738-687

Supervisor and faculty representative
Prof. Dr. Robert Weibel, Dept. of Geography, University of Zurich

Co-Supervisors
Dr. Francesca Cagnacci & Dr. Federico Ossi, Applied Ecology Group,
Fondazione Edmund Mach/ CRI

Abstract

Improving the knowledge about interactions between animals is of interest for ecologists as it helps to better understand social behaviour and dynamics, mating system and the spread and transmission of diseases. The research of monitoring wild ranging animals deals either with collecting location data of the animals or detecting proximity between animals. modelling spatial and temporal explicit contact The WildScope system integrates both in one system and thus enables the detection of geo referenced proximity data. But in ecology, the vast majority of tracking data collected today is GPS-only data. This thesis shows different approaches to detect interactions between animals from location data only and mapping them spatially explicit. Three methods are applied, namely Euclidean distance buffer, dynamic Brownian Bridge Movement Model (dBBMM) and joint Potential Path Area (jPPA) in combination with a developed procedure to estimate contacts from telemetry data. The data of the WildScope sensor serves as a ground truth for comparing the different methods and their performance. All methods identified potential contact area between two animals and between an animal and fixed nodes. In accordance with the recall value the jPPA and the BBMM perform better. Whereas, the linear buffer method returns the most detailed contact areas. The results depend on the parameterization of the methods.

Content

Figures

Tables

1 Introduction

1.1 Context and Motivation

In animal ecology, the use of global positioning system (GPS) sensors is nowadays a well-established way to track movements of large animals (Cagnacci et al. 2010). Movement ecology tries to find answers to the questions of why, how, when and where organisms move and how this movement is related to environmental factors (Nathan et al. 2008). Spatio-temporal data, as collected by GPS devices, is the basic information for dealing with these questions (Demšar et al. 2015). The modelling of the movement path based on the positional data is a fundamental task. More accurate statements on resource use and behaviour of animals can be made by the integration of additional information obtained from other sensors (Cagnacci et al. 2010).

In parallel, in the last decade the animal ecologists' community saw the development and deployment of new sensors to detect occurrence and duration of proximity patterns between individuals (Ossi et al. 2016) via contact detection between dyads of loggers fitted on a collar and then attached to the animals (see e.g. Boyland et al. 2013). Among several proximity logger models, those that work in a Wireless Sensor Network (WSNs, Picco et al. 2015) are particularly suitable for investigating this issue. Usually, the proximity data gained from these loggers is spatially implicit, however, thus not providing any information about where a given contact occurred, which still is crucial for a full ecological understanding of the proximity pattern under investigation. An elegant way to overcome this limitation is through an integration of the two technologies into one system, achieving so-called geo-referenced proximity detection (Picco et al. 2015; Ossi et al. 2016)This allows to detect when and where a contact between two animals may have occurred. Moreover, the proximity sensors can also be deployed as fixed loggers that are placed at locations of particular ecological or behavioural meaning that should be monitored (e.g. feeding stations, water holes). The contacts detected between a given animal and a fixed logger provide the duration of the visit of an animal to that site, thus informing about the use of these specific resources in a given habitat, overcoming the limitations imposed by the periodic nature of GPS data (see Ossi et al. 2016 for details).

1.2 Research gaps

In spite of these new advancements in the biologging and biotelemetry fields, the vast majority of tracking data collected today is GPS-only data. Hence, being able to infer proximity relations and potential contacts from GPS-only data, in the absence of proximity sensors, can importantly contribute to improve ecologists' knowledge on relevant ecological issues, such as e.g. predation-prey events, maternal care, territory defence, resource use or hierarchy establishment. Different methods exist to infer the distance path between GPS locations, allowing to estimate if a contact between animals possibly might have happened (e.g. Buchin et al. 2015; Long et al. 2015). However, typically GPS data are collected at coarse temporal sampling intervals, thus rendering this task non-trivial to accomplish.

Starting from this motivation, this MSc project aims at comparing different methods for inferring proximity patterns from GPS data, possibly enriching the detected proximity pattern by the addition of contextual (e.g. habitat) variables. This work relies on a fundamental premise, that is, that contact loggers perform better than other methodologies (GPS, camera traps) to assess proximity patterns. Building on this assumption, it will then be possible to accurately investigate the performance of different GPS-based methodologies to infer proximity patterns, using known occurred contacts as ground truth.

1.3 Research objectives and questions

Obj. 1: Validation of GPS-based interpolated trajectories via contact detection

1a. Is there any difference in the performance between the tested methods?

1b. Is there any influence of the properties of the data sets that will be used (e.g. concerning the sampling interval or the positional accuracy of the GPS)?

1.4 Data

1.4.1 Telemetry data – GPS

Telemetry is a widely used tool in animal ecology for remotely tracking the locations and the movement of animals. With the technological development of the last years GPS-based radiotelemetry is more and more applied in movement ecology (Cagnacci et al. 2010). GPS is a satellite-based system that measures spatiotemporal information. GPS-based radiotelemetry enables high resolution spatiotemporal tracking and mapping of animal movements (Krause et al. 2013).

Limitations are the battery lifetime, the memory and that the performance of GPS tracking is limited in densely vegetated areas (Krause et al. 2013). As a consequence of these limitations a trade-off between the weight of the battery synonymous for the lifetime of the sensor and the costs has to be found. But also a trade-off between the weight of battery and the sampling frequency. (Cagnacci et al. 2010)

The measured GPS locations trough time form the basic unit to rebuild the trajectory of an animal (Demšar et al. 2015).

1.4.2 Proximity sensors

In biologging small sensors are attached on animals to log and eventually transmit data about behaviour, physiology or environments of animals. Proximity logging is a biologging system that allows the dynamic mapping and quantifying patterns of animal-to-animal contacts (Krause et al. 2013; Drewe et al. 2012).

Proximity sensors give the information about the time and duration of a contact and which animals are involved in a contact. There is no reference to 'where?' a contact occurs. The information gained from these sensors is geospatially implicit (Ossi et al. 2016). The sensor is a transceiver, being a transponder and receiver of radio signals at the same time. The detection range of the sensor is determined by the power level of the signal. It ranges between 5-40 m depending on the power level (Ossi et al. 2016). The epoch time determines how often a signal is transmitted. When a sensor receives a signal from another sensor a contact is started. If the sensor does not receive a signal over a certain time (separation time) the contact is closed. Proximity sensors work reciprocal, meaning that every participating sensor records a contact independently (Krause et al. 2013).

1.4.3 WildScope

The independent development and application of these two systems has its limitations, as proximity logger produce spatially implicit information and GPS data does not have information about encounters (Picco et al. 2015; Ossi et al. 2016). To overcome this limitations WildScope was developed. WildScope is a system that integrates the two technologies of GPS tracking and proximity logging in a single system and thus enabling to detect geo-referenced proximity detection (Picco et al. 2015). With this system it can be answered 'Where?' and 'When?' a contact occurred (Picco et al. 2015; Ossi et al. 2016).

The system consists of three components: a radio unit, GPS sensor and a GSM/GPRS modem. The technology of wireless sensor networks is used for the collection and exchange of data from the fixed and animal-borne sensors to a main data collection system through radio pulses (Picco et al. 2015). It consists of mobile nodes, sensors attached to the animals, and fixed nodes and a base station (Ossi et al. 2016). Beside detecting contacts with the mobile nodes, the fixed nodes allow to monitor when and how long an animal visits certain places, as the fixed nodes are installed at landmarks such as feeding stations/places or resting areas (Picco et al. 2015).

This system enables two kinds of GPS measurements: periodic GPS measurement controlled by the sampling rate and triggered GPS measurement activated by a contact detection. This enables monitoring the location during a contact. Not only to know where a contact occurs, but also being able to set it into context of the trajectory is of interest, therefore the entire path of animal is tracked with the periodic GPS measurements (Picco et al. 2015).

There are two options for the data offloading: in-field collection with fixed nodes or remotely with cellular modems (Picco et al. 2015).

1.4.4 Foxes, UK (2015)

The data initially used in the intended study is from 5 foxes in the urban area of Brighton, UK. The data set is provided by Bryony Tolhurst from the University of Brighton. The foxes were tagged with collars with a GPS sensor and proximity detection sensor using low-power wireless technology. GPS location is recorded periodically every 30 min and also proximity triggered. Besides the individuals (mobile nodes) 15 fixed nodes were set up in the area at 6 locations. The contacts which are used for the analysis are mainly those which are detected between an individual and a fixed node (i.e. mobile-fixed contacts). The collection of the data took place between the 15 July 2015 and the 15 October 2015.

1.4.5 White tailed deer, Canada (2015)

Eight white tailed deer were fitted with collars and 23 fixed sensors were installed at the University of Saskatchewan's Goodale farm outside of Saskatoon in Canada. Data owner is Evelyn Merrill from the Department Biological Sciences at University of Alberta. The animals were fenced into three main paddocks, and the goal was to calibrate the system for future deployment in the wild (still going on). The periodic GPS fixes were taken every 2 hours, while the triggered GPS were taken only at the beginning of a contact. Data were collected from the 25 March 2015 until the 21 May 2015. The proximity loggers were set with two different power levels (3 and 7) (Pettitt et al. 2015).

2 Theoretical background

2.1 Movement ecology

A fundamental issue in ecology is understanding the organization of animals in space and time (Kernohan et al. 2001).The field of movement ecology is a part of ecology. This research field deals with understanding patterns, drivers, physiology and consequences of animal movement such as seasonal migration, dispersal and foraging (Hays et al. 2016). To achieve this, Nathan et al. (2008) propose four fundamental questions about the movement of organisms:

- 'Why?'
- 'How?'
- 'When and where?'
- 'How is the relation of the organism influenced by the environment?'

The challenge is to understand the processes that cause and influence the movement of the animals (Nathan et al. 2008). Movement is a spatial as well as a temporal phenomenon. Linking space and time together into spatiotemporal information can help understanding geographic processes (Holloway & Miller 2017). To answer questions in ecology it is essential to model the movement of an animal. The elemental unit of an animal's trajectory is its position. Dodge et al. (2016) consider the position to be the key for study and understanding movement. It shows where individuals interact with the surrounding environment (Cagnacci et al. 2010).

Technological advances enabled the transformation of movement ecology from a data-poor into a data-rich discipline (Demšar et al. 2015). The amount and quality of available data facilitates to work on new research questions or find new answers to existing questions (Holloway & Miller 2017).

Dodge et al. (2016) regards the study of movement as a continuum of research. The movement research is shaped two linked parts that benefit from each other. One part is the understanding of movement, where knowledge about the behaviour of moving individuals is gained which then can be used for modelling and prediction of movement (Dodge et al. 2016). The understanding of movement comprises of quantifying movement by developing appropriate methods, analysis of the movement and the interacting context and computational movement analysis for detection of movement patterns. Whereas modelling movement entails the creation of models, simulation development and making predictions (Dodge et al. 2016).

2.2 Contact

Information about interactions between animals is of interest as it helps to better understand social behaviour and dynamics, mating system and the spread and transmission of diseases (Ji et al. 2005).If the information about encounters are known for all animals of a group or population, further issues are of interest, such as formation of social hierarchies, cooperation between individuals, information flow, and responses to ecological event (Krause et al. 2013). But movement and thus interactions are also influenced by exogenous factors such as landscape and climatic changes (Long et al. 2015).

Several papers (Tosa et al. 2015; Rushmore et al. 2013; Ji et al. 2005; Walrath et al. 2012; Schauber et al. 2007) emphasize the importance of recording contacts (frequency and duration) for the transmission and spread of diseases and bacteria or the transmission of information, as they are spread by social interactions. Therefore, the study of disease-transmission requires an initial knowledge and understanding of behavioural dynamics (Krause et al. 2013)

The frequency and the duration of contacts varies according to social group characteristics. It is influenced by the characteristics such as the age, sex, kinship, social positon and patterns of space-use of an individual (Walrath et al. 2012; Tosa et al. 2015).

Interaction among animals is usually linked with spatial proximity. The *"Gambit of the group"* is a technique assuming that spatial proximity implies social interaction between individuals (Whitehead & Dufault 1999). This assumption is built on the realistic expectation that among nonhuman animals, close physical proximity is necessary for interaction. But it does not consider non-social spatial factor that influence animal movement and co-habitation, meaning the co-occurrence of individuals at the same location without social interactions happening (Whitehead & Dufault 1999).

This leads to the assumption that the detection of contact is driven by two main parameters; spatial proximity and the duration of a contact. Therefore, it has to be determined at what distance between two individuals a contact occurs. Further it can be determined a minimum duration for being counted as a contact. Even though this does not allow to say something about the kind of contact.

When measuring interaction, it can be distinguished between static and dynamic interactions. Static interaction is specified as the joint space use, not concerning temporal information, which traditionally has been focused on (Holloway & Miller 2017). The dynamic interaction covers how the movement of two individuals is interdependent and related concerning spatial as well as temporal overlap (Long et al. 2014).

There are two main conceptual approaches for mapping encounters. Contacts can either be recorded directly or indirectly. Direct encounter mapping uses technology that allow animal-to-animal data exchange, as for example proximity loggers. Indirect encounter mapping uses the recorded spatiotemporal positions of the individual animals, to infer contacts from the spatiotemporal information of two animals by combining them (Bettaney et al. 2015; Krause et al. 2013).
WildScope, as proximity loggers are a part of it, is an example for direct encounter mapping. The work done here is part of the approach of indirect encounter mapping.
For both approaches, the major limitation is that it is only possible to record spatial proximity of the animals, but no further information about the context and the kind of contact (active/passive) is known (Krause et al. 2013). If contacts are measured based on the distance between animals, it cannot be detected whether a contact occurs from attraction between individuals or an essential resource attracting individuals not concerning sociality (Spiegel et al. 2016). It is still an issue to reliably distinguish between real contacts and spatiotemporal proximity only (Holloway & Miller 2017).

2.3 State of research

The data from proximity loggers has so far often been used to apply for social network analysis (SNA). SNA enables to answer a variety of fundamental behavioural, ecological and conservation led questions (Haddadi et al. 2011). It allows to associate individual behaviour patterns with the population-level social structure and reversed (Croft et al. 2008). In a SNA the individuals are represented as nodes and the social interactions as edges (Wey et al. 2008). The edges can further be directional and/or weighted by the frequency of interactions (Pinter-Wollman et al. 2014). SNA can also be applied with telemetry data where the distances between individuals is inferred from the location (Haddadi et al. 2011).

In the paper of Tosa et al. (2015) three methods to estimate contact patterns are compared. It is compared between space-use overlap, direct contact rate measured with simultaneous GPS and direct contact rate measured with proximity loggers. Contact rates are estimated within and between social groups. Based on the estimated contact patterns, predictions for disease transmission dynamics are inferred (Tosa et al. 2015).
The volume of intersection, as a measure for joint space-use, is inferred from the overlap of the fixed-kernel use distributions. For the GPS, a contact is defined as simultaneous if the distance between them is <10 m. If >30 s passes between a proximity logger contact, it is defined as a separate contact. The contact rate for GPS and proximity logger is computed as number of contacts per week (Tosa et al. 2015). The methods show different interaction patterns as the methods have varying ranging distances and this leads to different management strategies (Tosa et al. 2015).

The research about interactions among animals using tracking data follow typically two conceptual approaches for joint movement processes: static interaction and inter-individual interaction. Static interaction, as mentioned earlier, explores the joint space use of two individuals indicating if animals are sharing their habitats. This is done by calculating the overlap of the two home ranges (Long et al. 2015). The area that an animal uses over a year is called the home range (Burt 1943). The limitations of this approach is that it does not incorporate temporal information about the joint space use (Long et al. 2015).

Different methods are known for home range estimation as for example the frequently used kernel density estimation (KDE). KDE calculates a utilization distribution. It represents the space use as a probability distribution quantifying the intensity of space use (Byrne et al. 2014). However, a weakness of the method is that it does not consider temporal autocorrelation, assuming independence between sample points (K Buchin et al. 2012). The Brownian bridge movement model (BBMM) estimates the movement path and the intensity of space use (Buchin et al. 2015). BBMM as well as the KDE are used with low sampling rates so that independent movement between the fixes can be expected. Assuming linear movement is reasonable for higher sampling rates(Buchin et al. 2015; Kevin Buchin et al. 2012). The BBMM assumes random movement (Brownian motion) and takes into account the characteristics of the movement path by concerning the temporal structure of the telemetry data (Buchin et al. 2015, Byrne et al. 2014, Kranstauber et al. 2012).

Long et al. (2015) introduce a new method for mapping areas of spatio-temporal overlap named the joint potential path area (jPPA). It is based on the approach of time geography (Hägerstrand 1970). It should overcome the issue of home range overlap being only spatial by incorporating timing. As an intermediate step the potential path area (PPA), depicting the home range, is calculated. In a further step a spatio-temporal overlap of the PPA is applied. The jPPA delineates the area where it is possible for two animals to interact in space and time (Long et al. 2015).

There exist several indices to test for occurrence of interaction between individuals from telemetry data. Inter-individual interactions are detected by these methods using spatial and temporal thresholds looking for simultaneous occurrence of the fixes in space and time (Long et al. 2015). Long et al. (2014) reviews in his paper several indices of dynamic interaction. These indices follow two approaches. A distinction is made between point- and path-based approaches, which has a conceptual impact on the different calculation and interpretation. Point-base indices reveal attraction and avoidance behaviour. Path-based indices, in contrast, allow to state about the cohesiveness of the movement behaviour (Long et al. 2014).

Examples for a point-based indices are the proximity analysis (Bertrand et al. 1996), the coefficient of association (De Almeida Jácomo et al. 2009; Karlin & Chadwick 2011) or the half-weight association index (Atwood & Weeks 2003; Bromley & Gese 2001). The indices are calculated with ratio between the observed number of contacts and the expected number of contacts. These indices range between 0 and 1, where 0 means no association or avoidance and 1 means attraction. The result of these indices depend on the distance and/or the time threshold which is required to define when or at what distance it is assigned as a contact.

The dynamic interaction index from (Long & Nelson 2013) uses the path-based approach. The corresponding path segments are compared to the means of the entire path by using correlation statistics. This index ranges between -1 (opposing movement) and 1 (cohesive movement) and 0 represents random movement (Long et al. 2014).

But all these indices have in common that they lack the spatial context. Similar to the proximity logger it is not possible to infer spatial explicit information. It is only a measure of how connected animals are but not were the social encounter occurs within the environment. Furthermore, the jPPA enables detecting infrequent and random contacts between animals whereas the indices of interaction, using formal statistical test, are not able to identify contacts of this type as they are not statistically significant. Detecting infrequent and random interactions among individuals is of interest since they have impact biological processes (Long et al. 2015).

Only recently a new R-package, called “spatsoc”, was introduced by Robitaille et al. (2018). It facilitates detecting potential interactions between individuals by analysing GPS data in space and time. The objective of this package is equal to the aim of this thesis. The available functions allow to analyse the data at different spatial and temporal scales. Spatial scales in terms of point-, line- or polygon-based overlap and the temporal scale range between minutes and years (Robitaille et al. 2018). Basically the functions work spatially but can be extended to also enable spatiotemporal interaction detection. If the data was temporally grouped in advance, the location data can be analysed in space and time. **group_lines()** function assumes linear movement and additionally a distance threshold for a buffer can be added (Robitaille et al. 2018).

2.4 Methods

The first two methods are not able to detect contacts yet as they are only able to model the movement path of an animal. The methods applied here should enable producing outputs that can be mapped. This requires the development of a procedure where geometric methods are applied to detect contacts from location data.

2.4.1 Euclidean distance

The Euclidean distance is the simplest distance measure and will be used as a baseline method. It corresponds to the length of the shortest and most direct connection between two points based on the theorem of Pythagoras.

This method is not dependent on any further parameters. A prerequisite for this method is that the telemetry data is sorted by time. As an expansion to this method, to every pair of sequential fixes that form a line segment together a width value is assigned to form a buffer around. The buffer is based on the epsilon band model which is used for modelling positional uncertainty of lines. The uncertainty is represented as a buffer zone on each side of the line with fixed width (Li et al. 2017). Here, the width can be the same single value for all segments or an individual value for every path segment. The width value of the buffer can be adjusted to the research purpose. In terms of the proximity loggers the detection distance is a possible value for the width. Another possibility is the location error regarding the measurement accuracy of the GPS device.

2.4.2 Brownian Bridge Movement Model

Brownian bridges were first introduced by Bullard (1999). It is based on the concept of Brownian Motion. This method assumes that the movement between relocations is random conditional on the starting and ending location. The temporal structure of the data is included trough the order of the locations and the time passing between them (Horne et al. 2007). Beside this, location error is incorporated as well. With increasing time interval, the conditional random movement between sequential locations becomes less likely (Horne et al. 2007).

The estimation of the positions a random walk of an animal from position *a* to *b* using a normal distribution (Horne et al. 2007) is described as:

$$\varphi(z;\mu,\sigma^2) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(z-\mu)^2}{2\sigma^2}\right)$$

where $\mu(t) = a + \frac{t}{T}(b-a)$ and $\sigma^2(t) = \frac{t(T-t)}{T}\sigma_m^2$

Z represents any position in $\mathbb{R}^2$.The mean values is denoted as μ and the Brownian motion variance is denoted as σ^2. T is the time passed between the positions a and b and with $t \in [0,T]$.

The density function of the BBMM (Horne et al. 2007) for a series of space-time observations at position z, where T_{toal} is the total duration and n is the total number of locations, is:

$$h(z) = \frac{1}{T_{total}} \sum_{i=0}^{n-1} \left\{ \int_0^{T_i} \varphi\left(z;\ \mu_i(t), \sigma_i^2(t)\right) dt \right\}$$

The BBMM described by Horne et al. (2007) has a constant Brownian motion variance σ_m^2 along the trajectory. It is a measure of how irregular the path of an animal is between consecutive locations and as such the driving parameter (Byrne et al. 2014)The dynamic Brownian Bridge Movement Model (dBBMM) is an advancement of the BBMM (Kranstauber et al. 2012). The dBBMM is able to modify the Brownian motion variance along the path according to behavioural changes in the movement path of an animal (Kranstauber et al. 2012). The dynamic variance is estimated with a moving window involving a series of locations along the path (Byrne et al. 2014)

The result of this method is a utilization distribution in form of a raster with the probability values of an animal being in a raster cell and a utilization distribution bounding (volume isopleth/contours in %). A contour level of 50% contains the half of all raster cells with the highest probability values.

2.4.3 Joint Potential Path Area

The method of joint Potential Path Area (jPPA), introduced by Long et al. (2015), relies on the framework of time geography (Hägerstrand 1970). The framework serves to study how individual movement is influence by spatial-temporal processes. The space-time prism is a conceptual component of time geography that represents all potential possibilities in space and time that are open to an animal considering all restrictions. It consists of three dimensions, the x and y coordinates and the time.

As a base for the jPPA the potential path area (PPA) is calculated. The PPA is an approach to estimate animal home range including the temporal information of animal movement (Long & Nelson 2012). It shows all locations an animal can access knowing the starting and ending location and the animal's maximum velocity and corresponds to the projection of the space-time prism on to the geographical plane.

For any point in time t within the time interval $[t_i, t_{i+1}]$ of two sequential fixes $\{a_i, a_{i+1}\}$ a disc $D_{i,\tau}$ centered on the first fix with radius $r_{i,\tau}$ is generated from:

$$r_{i,\tau} = v_{max} * (\tau - t_i)$$

v_{max} is a measure for an animal's mobility and can be inferred from the data according to Van der Watt's approach (1980). Another disc $D_{i+1,\tau}$ is centred on the second fix generated with the radius $r_{i+1,\tau}$ and v_{max}.

The accessibility space G_τ^A for an animal at time τ is the intersection of both discs $D_{i,\tau}$ and $D_{i+1,\tau}$. The PPA ellipse is the union of all G_τ^A for $[t_i < \tau < t_{i+1}]$. In a set of n locations this is done for every pair of sequential fixes resulting in *n-1* PPA ellipses. The union of all *n-1* ellipses produces the PPA home range (Long & Nelson 2012).

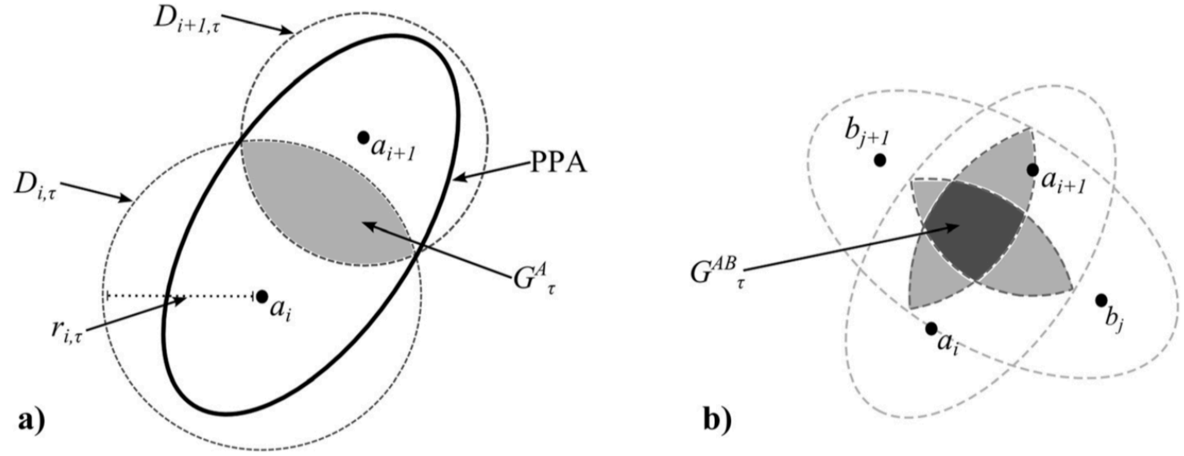


Figure 1: a) The PPA is the union of all G_τ^A for any τ. b) The joint accessibility space of two animals, A and B, is the intersection of G_τ^A and G_τ^B. (Long et al. 2015)

Intersecting the PPA of two individuals produces the overlapping accessibility spaces. The jPPA is then the spatial union of all overlapping accessibility spaces of two animals for every point in time within the time interval $[t_i, t_{i+1}]$. The area of the jPPA is always a part of the intersected PPA's of the two animals (Long et al. 2015).

The advantage of this method lies in its simple and uncomplicated implementation and interpretation. The result of this method is a polygon containing all regions where it is spatially and temporally possible for two animals to be in contact with each other. The driving parameters of this method are the available time budget and the maximum speed of an animal v_{max} (Long et al. 2015).

The utilization distribution resulting from a BBMM is a surface with uneven probabilities. Whereas the PPA, the utilization distribution of the jPPA, has all over the surface the same discrete probability (Long et al. 2015).

3 Methodology

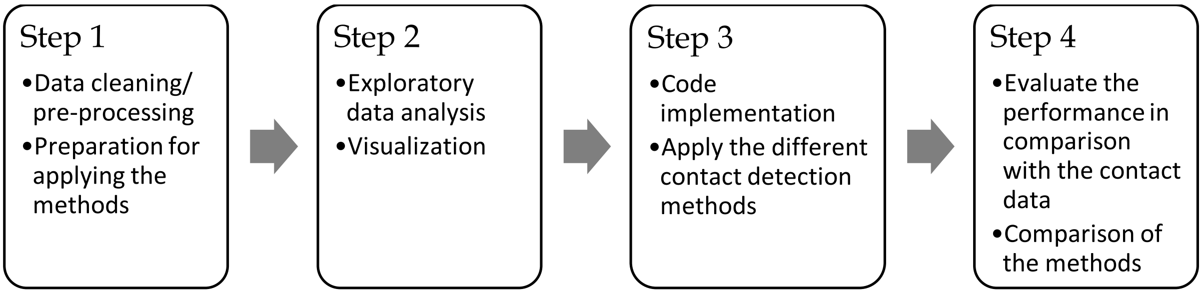


Figure 2: Work flow.

The methodology follows the steps of the work flow (Figure 2). Steps 1 and 2 of the work flow deals with the pre-processing and the analysis of the data sets. The methods have to be implemented and adjusted to the purpose as described by step 3 of the work flow. The concept and the implementation of the contact detection procedure is described above. Step 4 of the work flow considers the evaluation of the performance ant the comparison of the methods. The entire data analysis is conducted with R (R Core Team 2013).

3.1 Explorative data analysis

Pre-processing

For the data pre-processing it is necessary to remove all missing values within the data. The buffer method requires chronologically sorted data; therefore, the data is ordered by time. As all animals are within the same data frame a subset for every single animal is done. After removing the outliers periodic GPS measurements are subset as the contact estimation is based only on the periodic GPS data. Data has to be removed from proximity logger data for nodes without location, as they cannot be used for the evaluation of the methods.

As the data of the proximity logger only contains start and end time of a contact and the id of the animal/fixed node transmitting and the peer id, the data has to be interpolated by based on the GPS data. Although it is known whether a GPS measurement is periodic or triggered, it cannot be linked with the contact data. A reason for this is the GPS limit time. It is the period after a periodic GPS measurement where no new location is recorded, if a contact is detected within the GPS limit. The contact data is interpolated linearly.

One of the foxes cannot be used for the animal to animal contact detection due to a later and non-overlapping recording time with the other individuals.

Data analysis

This section is about the data analysis and data cleaning. The aim is to develop an understanding of the used data by visualizing the data, calculating and evaluating statistical plots and values, and removing inconsistencies and outliers due to measurements uncertainties.

First of all, the data is visualized to see how the data is distributed in space and how the individuals are connected with each other spatially. For every animal a centre point is determined by calculating the mean longitude and the mean latitude of all GPS measurement of every animal. Various statistical measurements of the step length, the distance between two consecutive locations, are computed including the minimum, the maximum, the mean and the standard deviation of the step length. In a histogram the distribution of the step lengths is shown.

Outlier recognition and removal

For the detection of outliers, the distance of all GPS measurements from every animal to its centre point is calculated and the according mean and standard deviation. An outlier is defined as (Knight & Wang 2009):

$$\mathrm{Outlier} = \sum_{i=1}^{n} d_i - \overline{D} \geq 3 * SD$$

Three standard deviations are equal to 99.73 %. All the values that fulfil this condition are removed from the data. Beside this, the percentage of outliers in all data is determined.

After the detection and removal of the outliers the analysis steps of before are conducted again to see the influence of the outliers and if the statistical measurement adjusted.

The same procedure of data analysis and data cleaning is applied for the foxes and the white tailed deer.

3.2 Methods

3.2.1 Euclidean distance

To get a buffer for every pair of sequential locations from the GPS data a function called **oneWidthBuffer()** is created. As input parameters the function requires a data frame with the location data, the Universal Transverse Mercator (UTM) time zone and the buffer width. The buffer width is the main parameter of the function which stands for the location error of the GPS device. As a first step the coordinates are transformed to UTM projection. The coordinates of every i-th and i+1-th element are extracted and inserted into a 2*2 matrix. All matrixes are stored in a list. The list of matrixes can then be converted into spatial lines. The spatial lines are needed for the buffer creation which is the last step. The function puts out spatial polygons where a single polygon represents a pair of sequential locations.

With a data subset the influence of different buffer widths on the resulting contact rate was tested. Some of the width values are not realistic as location error, but are still interesting to see how close contacts occur. The width values range from 2,3,5,10,15 to 20 meters. The number of contacts increases until a certain point at 10 m where it remains nearly static for the rest of values.

Once the location error of 10 m is applied as buffer width. For the deer data the buffer width is set to 15 m and 40 m according to the power level settings of the proximity loggers for every animal. For the white tailed deer UTM time zone 13 is used and for the foxes the UTM time zone 30.

3.2.2 Brownian Bridge Movement Model

An adapted version of the function **move.contour()** from the moveud R package (Collier 2016) is used to obtain Brownian bridges for every time step.

In a first step a dBBMM model is fitted to the entire path of an animal using the R package move (Kranstauber et al. 2017). This is then used as an input into the **move.contour()** function. Other required parameters are the location error and the contour level. The function returns the utilization distributions, in form of polygons bounded by the contour level, for every pair of two successive fixes based on the Brownian motion variance from the fitted dBBMM.

The size of the resulting polygons can be influenced by varying the contour level. With a lower contour level only the higher probabilities are chosen.

3.2.3 Joint Potential Path Area

The function **jppa()** from Long et al. (2015) is used. Required function parameters are the trajectories of two animals, t.int, tol and ePoints. The trajectories of the two animals need to be an object of class ltraj. T.int is a time parameter that defines the frequency of time slices which is applied to delineate the joint activity space. Tol, also a time parameter, is used for filtering out the segments with too long time intervals between two locations. The parameter ePoints determines the number of corner points that are used for constructing the PPA ellipse. The function returns a polygon corresponding to the joint accessibility space of two animals.

As mentioned earlier, the size of the PPA depends on the v_{max} parameter. This parameter can be estimated or can be defined by the researcher manually. With the estimated v_{max} large PPAs are generated. To test the effect of the parameter different v_{max} values are applied. When applying the mean speed inferred from the data as v_{max} no change is apparent.

3.2.4 Contact detection / estimation

There is a difference between animal-to-animal contact estimation and animal-to-fixed node contact estimation. The contact detection with fixed nodes is a simpler case of inter-individual contact as the time component does not have to be considered. Only the spatial overlap has to be checked. As the location of the fixed stations is known, they are represented as points. The animal's track is represented with different polygons, depending on the applied method. Therefore, the check whether an interaction occurred, can be estimated with a simple point-in-polygon test with the **FixedNodesContact()** function. As parameters it requires the fixed node locations and the path of the animal as polygons. It returns the contacts as matching pairs with the fixed node and the path segment which belong together.

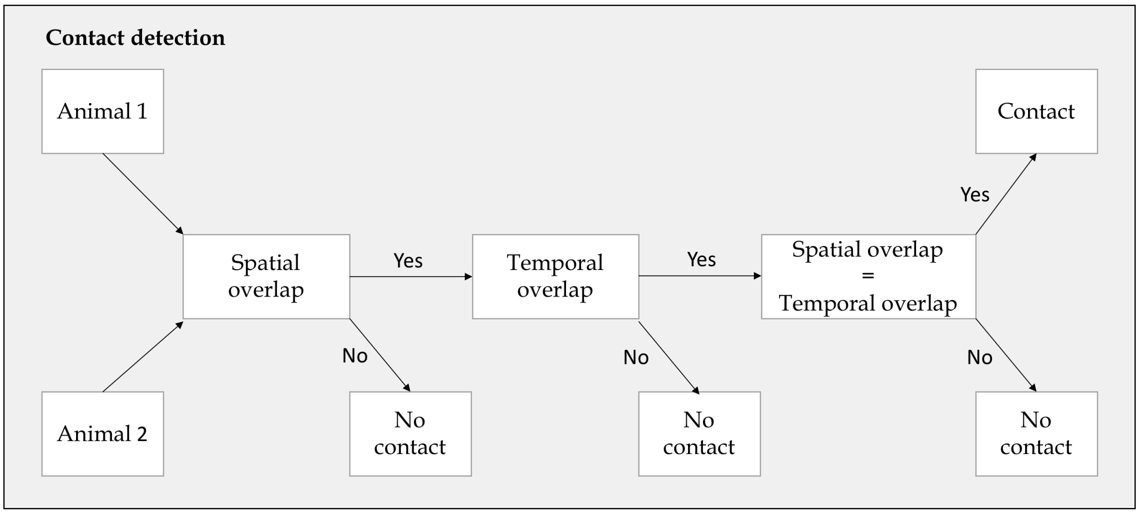


Figure 3: Three-step procedure for contact detection.

As mentioned earlier, the Euclidean distance and the BBMM are two different methods to estimate the path or areas with high probability of an animal based on known location points. To be able to make an estimation whether any interactions between animals is possible a three-step procedure is developed. This procedure allows to detect contacts, based on GPS location data.

In the first step the paths of two animals are tested if their paths intersect in some place. The created function is called **SpatialPairs()**. If there is no spatial overlap, no encounter between the animals occurred. If there is spatial overlap, the indices of the involved segments of both animals are extracted. The spatial indices give the information about where the spatial overlap is within the animal track. Returning the matching path segments pairwise.

The second step checks for temporal overlap with the **TemporalOverlap()** function. To do this, time intervals are created where start and end of the interval is equal to the time at the starting and ending location of a segment. The comparison of the time intervals of two animals checks whether the same time can be found in the intervals of both animals. The output is a data frame with two columns containing the pairwise indices of the temporal overlapping path segments.

In the third step the results from the test for spatial and temporal overlap are compared for coincidence. The results are intersected with a SQL-statement in the **ComparSpaceTime()** function. Resulting from the final step are the pairwise indices from both animals that fulfilled the constraint. These indices indicate where it is spatially and temporally possible for two animals to meet.

These three steps are joined in a single function, called **FindAnimalContacts(),** for detecting potential contacts. Input parameters for the function are the polygons, created with a method explained before (Euclidean distance or dBBMM), and the data frame with the temporal information for each of the two animals. If in the first or second step of the function no contact is detected, the function breaks.

The result of this procedure does not give exact locations of contact but returns path segments, as polygons, within which one or more contact occurred. This means that not an absolute number of contacts and an explicit location (a point) can be derived. The accuracy of the result depends on the temporal resolution of the data, as the path segments decrease with decreasing sampling rate.

3.3 Evaluation criteria

The confusion matrix is used to test the applied method for its performance in comparison to the ground truth (Fawcett 2006). The ground truth consists of the data recorded by the proximity loggers knowing where a contact occurred. Based on these evaluation criteria the methods can be compared regarding the measures of relevance.

A confusion matrix sums up the performance of a classification model and is commonly used in information retrieval (Fawcett 2006; Lewis 1991). In this case the classification models are the methods to estimate contacts. The result of the methods is the prediction. When compared to the ground truth four outcomes are possible which are represented in a 2x2 matrix. If a contact is predicted and the loggers detected a contact as well, these are the true positives. If no contact is predicted but the loggers actually detected a contact, these are the false negatives (Fawcett 2006). If no contact is predicted and the loggers detected contact either, these are the true negatives. If a contact is predicted and the loggers did not detect a contact, these are the false positives (Fawcett 2006). From the data used here only the true positives and false negatives can be obtained, because there is not an absolute number of no contacts known.

These four values are the basis to calculate further measures (Fawcett 2006). Here, it is only possible to calculate the recall but not the precision, as the true negatives and false positives are not known. The recall, also known as sensitivity or true positive rate, shows the proportion of predicted contacts that were predicted correctly and is calculated as:

$$recall = \frac{true\ positives}{total\ positives}$$

where the total positives are the sum of the true positives and false negatives (Fawcett 2006; Lewis 1991). It ranges between 0 and 1 whereby 0 means that no contact was predicted correctly and 1 means all true contacts were detected by the prediction.

The true positives and false negatives can be obtained with a point-in-polygon test where the prediction or estimated contacts are a polygon and the detected contacts are points. The number of TRUE values of this test corresponds to the true positives and the number of FALSE values corresponds to the false negatives. The ground truth for the animal to fixed node contacts is the contact data of the animal recording a contact with a fixed node.

4 Results

4.1 Explorative data analysis

4.1.1 Foxes

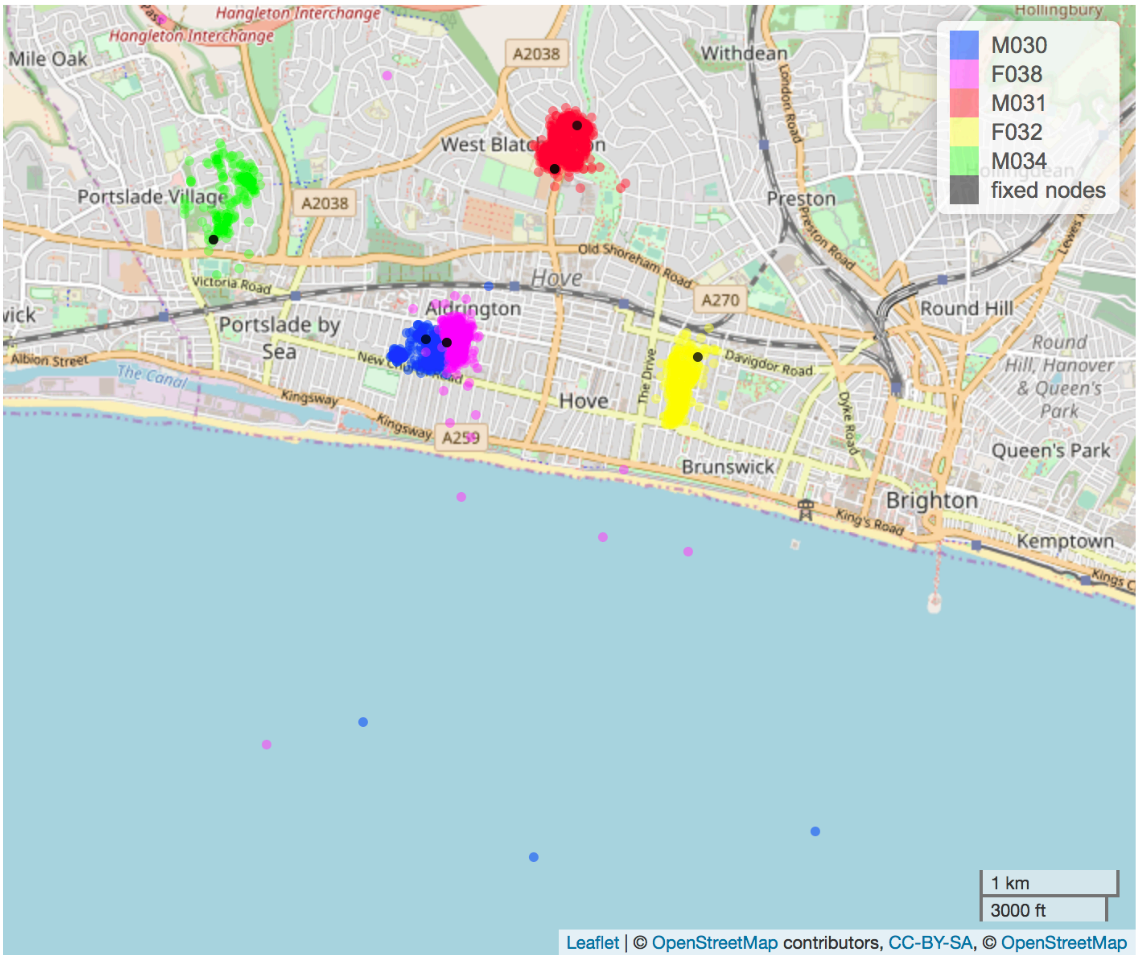


Figure 4 : Fox locations and fixed stations in Brighton, UK.

To get a first impression of how the data of the foxes look like, a map with all data is created. Figure 4 shows that all foxes move within their own territories. Only the animals M030 and F038 live in close proximity. These two animals are the only ones that detected animal-to-animal contacts. The rest of the contact data consists of contacts between mobile and fixed nodes. The data shown in the map indicates that there are some outliers or odd measurements that have to be considered. There are some points that are located in the sea which seems to be unrealistic. Some other points are far off the majority of the points as it seems unreal to reach these locations considering the temporal resolution (30min) and the speed of a fox.

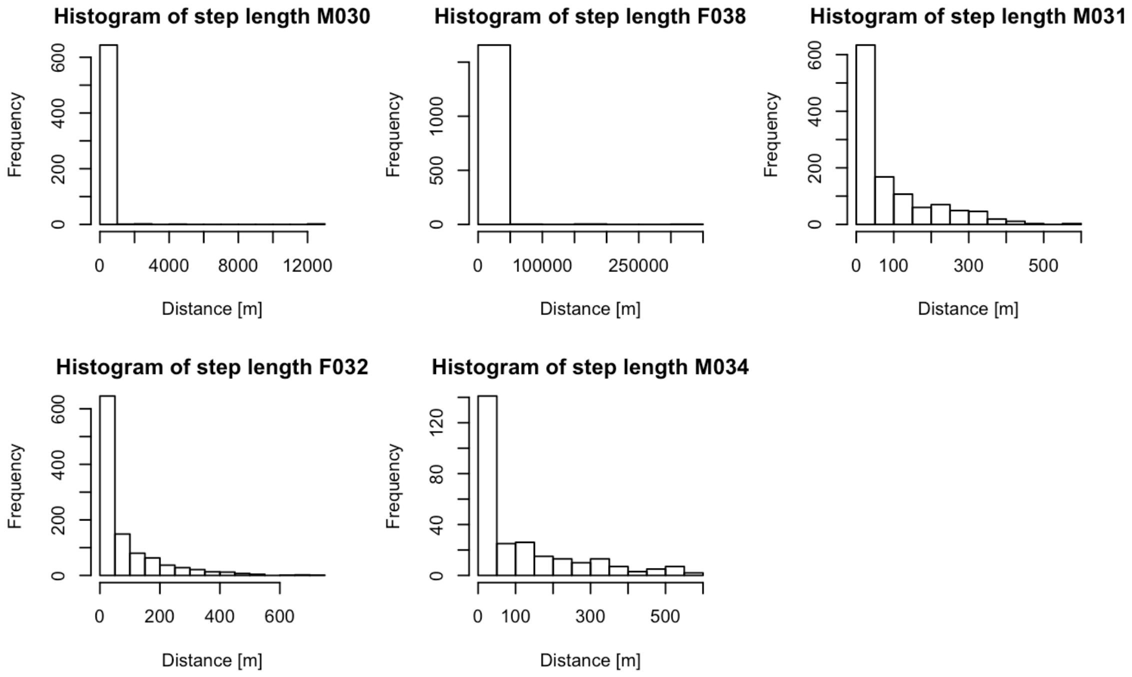


Figure 5: Histograms with step length distances of the foxes.

The histograms in Figure 5 show a single sided distribution with a lot short distances ranging between 0-100 m and few long distances (>100 m). The scale of the x-axis varies from animal to animal. The maximum, the mean and the standard deviation distance (SDD) decreased after the data cleaning. Especially for the animals M030 and F038 which have extreme outliers, the maximum decreased from 12 km and 345 km to 0.8 km and 1.0 km referring to Table 1.

Table 1: Statistical measures of the step length in [m] before and after data cleaning.

Fox	Original data			Cleaned data		
	Max	**Mean**	**SDD**	**Max**	**Mean**	**SDD**
M030	12887.7	114.1	81.5	795.0	58.0	81.5
F038	345991.1	1017.1	77.7	1032.0	53.2	77.7
M031	598.9	92.1	101.6	489.4	89.7	101.6
F032	737.9	80.4	100.1	538.7	77.4	100.1
M034	588.5	117.5	141.1	588.5	114.0	141.1

The percentage of outliers of all data ranges between 0.3-1.4 % for all foxes. For one animal (F038) outliers located in the sea could not be removed with this approach. These points are removed manually.

4.1.2 White tailed deer

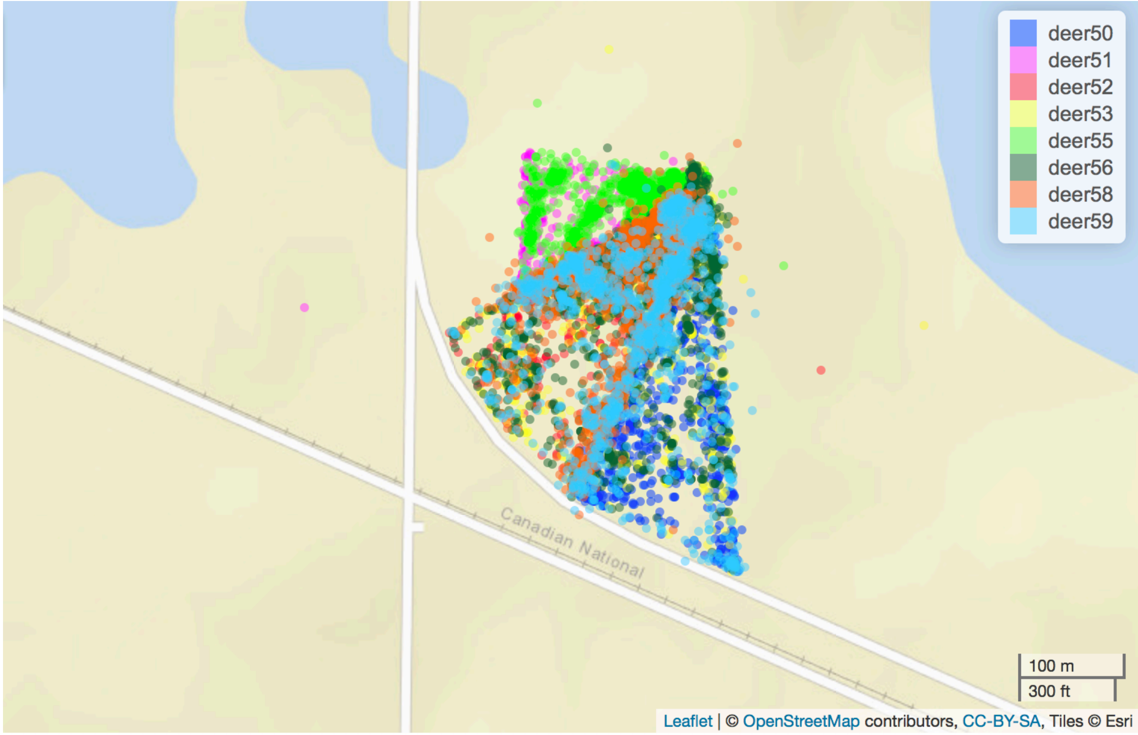


Figure 6: Deer location fixes in Canada.

The Figure 6 shows how the deer are distributed in space. It shows only an extent with the majority of the data, indicating that there are some outliers or odd measurements that have to be considered. Some of these points are far off the majority of the points which is unreal to reach these locations considering the temporal resolution (2h) and the average speed of a deer. Further it can be seen that the deer were divided into groups which were in separated enclosures. Even though the animals are grouped into separate enclosures, they live in immediate neighbourhood. This implies that also contacts between animals of different enclosures are detected by the proximity loggers.

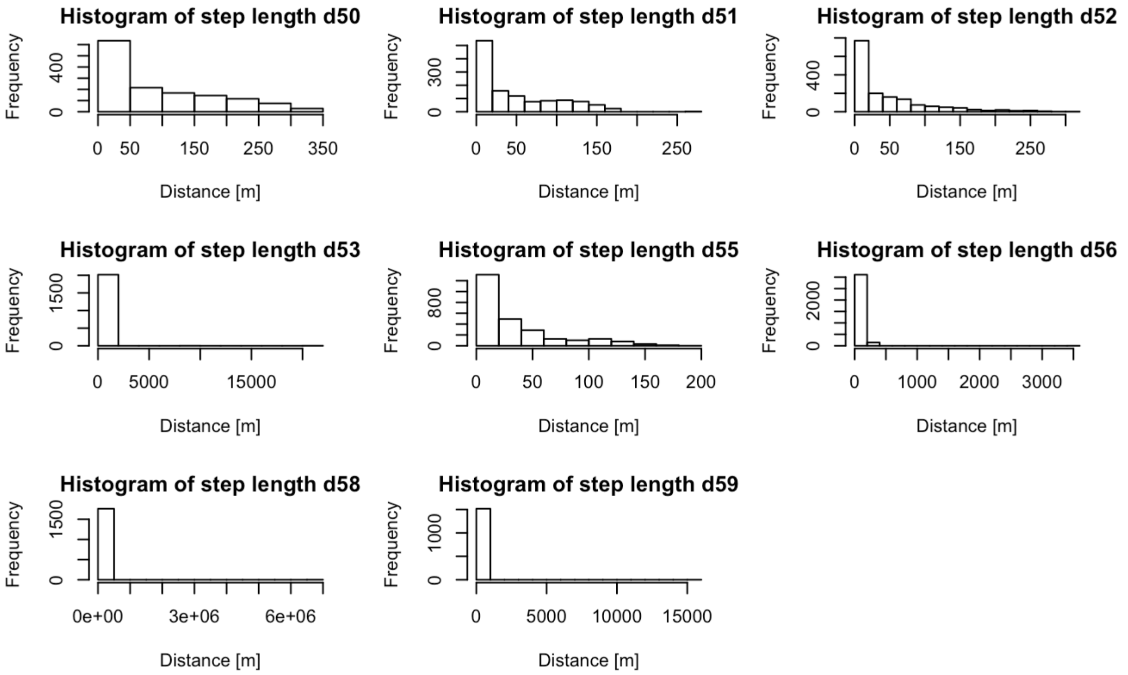


Figure 7: Histogram with step length distances for all white tailed deer.

The frequency of the displacement for each animal are shown in Figure 7. The distribution is right-skewed indicating high frequency of short distances (0-50 m) and low frequency of long distances (>50 m). The maximum, the mean and the standard deviation distance decreased after the data cleaning. Especially for the deer 53,56,58 and 59 which have extreme outliers. For example, comparing to Table 2 the maximum of deer 53 decreased from 20 km to 0.3 km.

For all deer the percentage of outliers of all data ranges between 0.0 and 1.6 %.

Table 2: Statistical measures of the step length in [m] before and after data cleaning.

Deer	**Original data**			**Cleaned data**		
	Max	**Mean**	**SDD**	**Max**	**Mean**	**SDD**
d50	346.0	92.6	89.2	346.0	92.6	89.2
d51	268.6	48.4	50.0	178.5	48.0	49.2
d52	315.2	45.9	57.7	266.6	43.7	53.9
d53	20’364.6	81.6	691.9	330.7	53.5	66.1
d55	187.8	33.7	38.0	179.1	33.2	37.2
d56	3’499.8	48.3	105.1	338.0	46.1	61.0
d58	6’564’851.9	7’494.5	220’984.6	314.7	51.5	63.0
d59	15’127.8	65.8	393.4	349.8	55.7	71.4

4.2 Methods

4.2.1 Euclidean distance

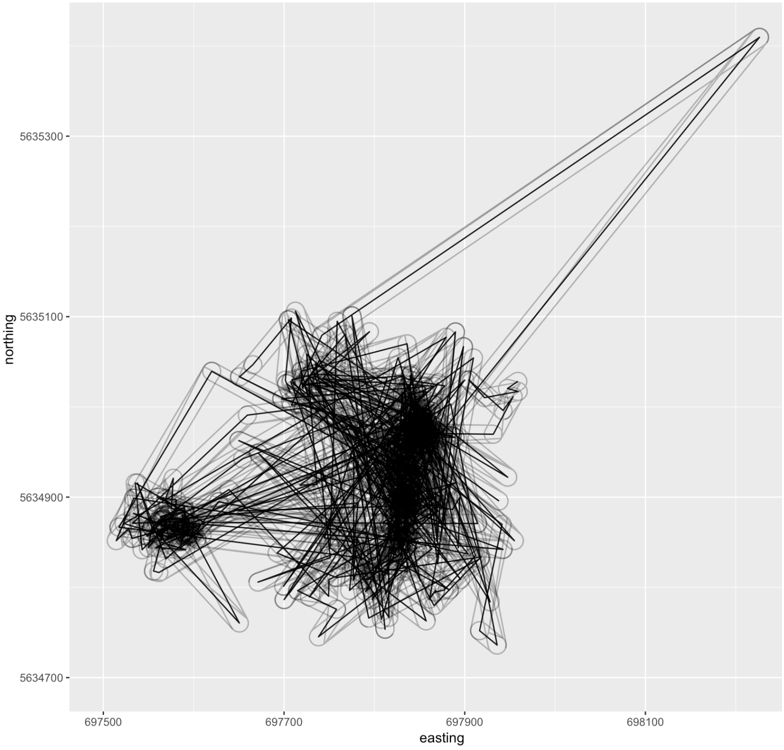


Figure 8: Linear movement path and buffer of fox M030.

Shown in the Figure 8 is the modelled linear movement path of the animal and the linear buffers with a width of 10 m. The width of 10 m represents the location error. For every path segment consisting of two sequential location fixes a buffer is computed.

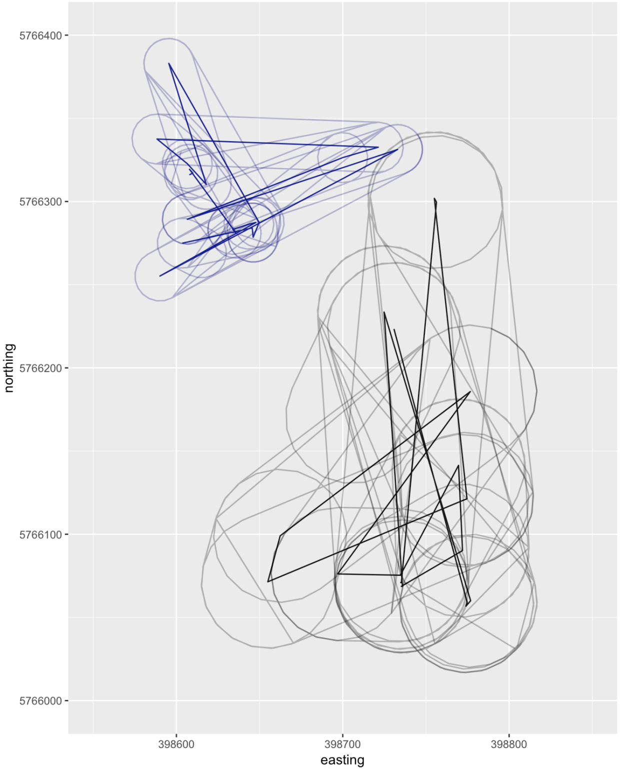


Figure 9: Linear movement path and buffer of deer 50 (black) and deer 51 (blue).

Figure 9 shows again the linear movement path of two animals and the buffer with different widths. The width of 15 m and 40 m are inferred from the distance range of different power level of the proximity loggers. It shows that the movement path only does not spatially intersect but the buffers intersect.

For every animal the buffer method was applied with 10 m width and for the deer data additionally the power level ranges were applied.

4.2.2 BBMM

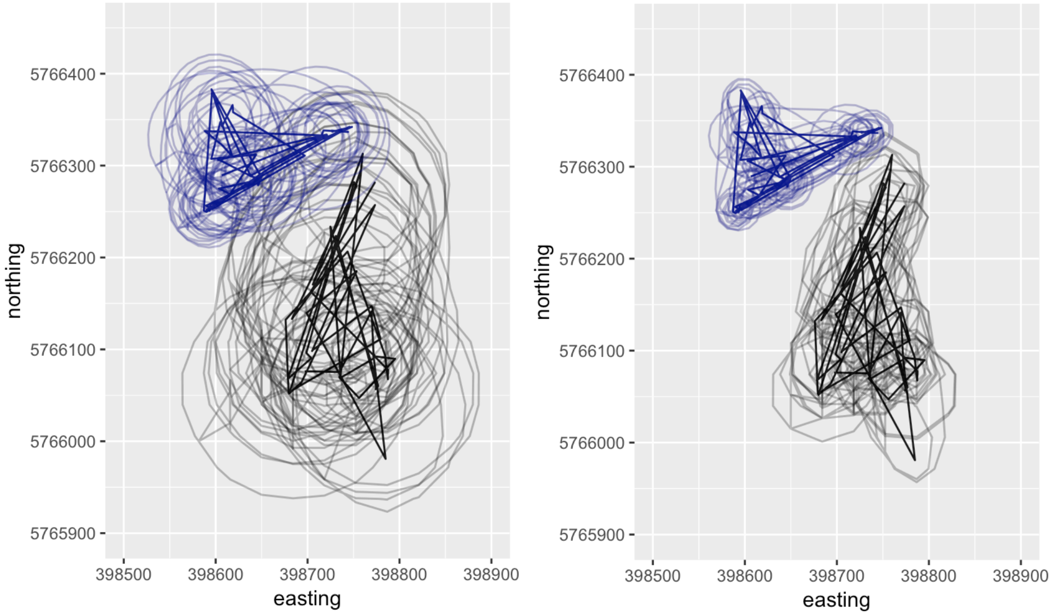


Figure 10: Linear movement path and BBMM with contour level 90% and 50% for deer 50 and 51.

The buffers represent the dynamic Brownian bridges at a certain contour level. As for the linear buffers, for every pair of two successive locations fixes there is Brownian bridge. The movement paths of the animal do not necessarily overlap. However, with the Brownian bridges buffers a spatial overlap results. In Figure 10 the difference in the area size is shown depending on the chosen contour level. With 90% contour level there is much more spatial overlap between the two animals than there is with a 50% contour level.

4.2.3 jPPA

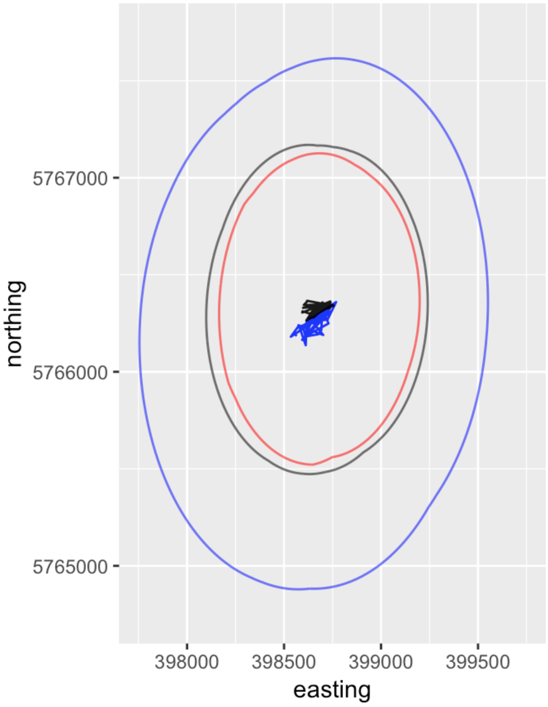


Figure 11: jPPA (red), PPAs and trajectories of deer 52 (blue) and 55 (black).

For all deer dyads the results of the jPPA are equal to the one shown in Figure 11. The area within the red ellipse is the jPPA where interaction is spatially and temporally possible. The jPPA is smaller than the overlapping zone, which is the intersection of both PPAs. Here, the overlapping zone is equal to the PPA of the deer 55. The plots of all dyads can be found in the appendix.

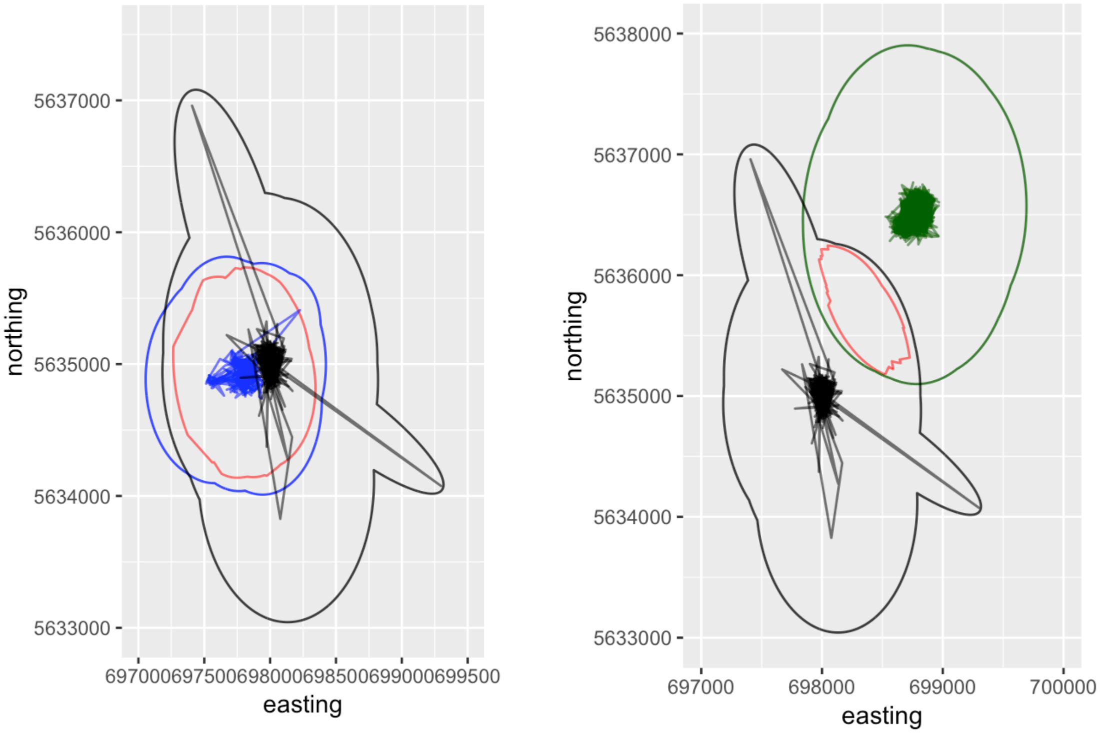


Figure 12: jPPA (red), PPAs and trajectories of the foxes F038 (black), M030 (blue) and M031 (green).

As mentioned earlier M030 and F038 are the only foxes that live in close neighbourhood. Nevertheless, the jPPA method can be computed for foxes where the trajectories do not spatially overlap but the PPAs do. This is shown in Figure 12. In this case the jPPA is an area between the two animals where neither of them has actually been or no location is recorded there, respectively. Even so, when applying the jPPA on the foxes M030 and M031 the absence of a joint area is determined. The PPAs of the two animals have a small overlapping area but the joint use of this overlapping area is spatially and temporally not possible for the both foxes.

4.2.4 Contact detection / estimation

Animal to fixed node contacts

The most contacts between a deer and the fixed nodes are estimated by the BBMM method with the 90% contour level (Table 3). Only for three deer the buffer method with the width from the power level estimates more contacts than the BBMM 90%. Apart from two deer, with the buffer method, with 10 m width, the fewest number of contacts are estimated. All methods have in common that only few contacts are estimated in comparison to the number of real detected contacts with the proximity loggers. The number of contacts detected by the logger ranges between 3119 and 13410. For the buffer with 10 m it ranges between 98 and 703 and for the buffer with power level as width it ranges from 125 to 826. For the BBMM with 90% it ranges between 203 and 676 and with 50% between 136 and 480.

Table 3: Absolute contact rates between an animal and fixed nodes for deer data with the different methods.

Deer	Buffer 10 m	Buffer (15 and 40m)	BBMM 90%	BBMM 50%	Proximity logger
50	206	427	521	321	4910
51	368	457	541	419	6058
52	310	495	478	369	5827
53	449	559	676	480	11769
55	337	468	633	420	8265
56	703	826	595	414	13410
58	177	292	203	136	7959
59	98	125	247	161	3119

It is striking in Table 4 that for F032 no contact is detected with the logger, but all methods estimated contact with fixed loggers for this fox. Beside this fact, it is similar to the deer data. The most contacts are estimated with the BBM 90% method and the fewest with the buffer method with 10 m of width. The methods do not estimate the most contacts for the same fox as the logger. The methods computed the most contacts for the fox M031, where the logger counted the most for F038.

Table 4: Absolute contact rates between an animal and fixed nodes for fox data with the different methods.

Fox	Buffer 10 m	BBMM 90%	BBMM 50%	Proximity logger
M030	9	106	26	110
F038	5	44	32	1424
M031	27	310	93	368
F032	4	72	18	0
M034	1	23	3	48

Animal to animal contacts

For the fox data only M030 and F038 are possible to have contact, as these are the only foxes that live in direct neighbourhood and though have spatial overlap.

Unlike the fox data, with the deer data is possible to conduct contact estimation with all combinations of animals, as all animals live in direct proximity. Both methods were able to estimate contacts between all deer pairs except from dyad of deer 50 and 51 with the buffer method with width of 10 m.

The BBMM 90% estimates the most contacts for all pairs of two individuals.

Table 5: Absolute contact rates with buffer, BBMM and measured contacts of the deer data.

Deer dyad	Buffer (10m)	Buffer (15 and 40m)	BBMM 90%	BBMM 50%	Proximity logger
d5051	0	95	422	82	29
d5052	76	463	641	295	1324
d5053	258	623	1042	570	728
d5055	6	142	497	81	20
d5056	307	704	1115	538	395
d5058	114	369	525	237	409
d5059	273	442	528	367	335
d5152	192	552	805	435	492
d5153	129	215	834	332	139
d5155	549	653	931	732	487
d5156	154	262	788	275	136
d5158	140	304	620	280	553
d5159	84	98	244	116	86
d5352	491	728	844	544	3533
d5355	221	340	1069	424	348
d5356	1313	1477	1498	1065	6915
d5358	271	516	682	336	1176
d5359	234	289	623	360	734
d5552	175	510	691	342	379
d5556	228	380	876	302	236
d5558	120	319	544	249	236
d5559	52	85	350	130	29
d5652	590	887	911	499	2945
d5658	344	590	730	304	1299
d5659	303	369	580	293	640
d5852	375	625	607	390	1206
d5859	166	222	267	217	1118
d5952	57	82	85	59	1144

The buffer method with 10 m width estimates the fewest contact, except for the dyads of d5356, d5652, 5658 and d5659. For these pairs the BBMM 50 % computes the fewest contacts. For some pairs the proximity logger detects more contacts (e.g. 5052, d5352) but for some pairs the methods estimate more contacts (e.g. d5051, d5559) than the proximity loggers.

Table 6.: Absolute contact rates with buffer, BBMM and measured contacts of the fox data

Deer dyad	Buffer (10m)	Buffer (15m)	Buffer (40m)	BBMM 90%	BBMM 50%	Proximity logger
M030-F038	0	0	14	20	1	5

The proximity loggers only detected contact between the foxes M030 and F038 and even the logger recorded only a small number of 8 contacts (Table 6). The buffer method with a width of 10 m and 15 m did not compute a contact between the animals. With a width of 40 m for the buffer 14 contacts are estimated. The BBMM method estimated contact between the foxes with both contour levels, but with 50% only a single contact. The interaction between these two foxes occurs over longer distances as the results of the methods show.

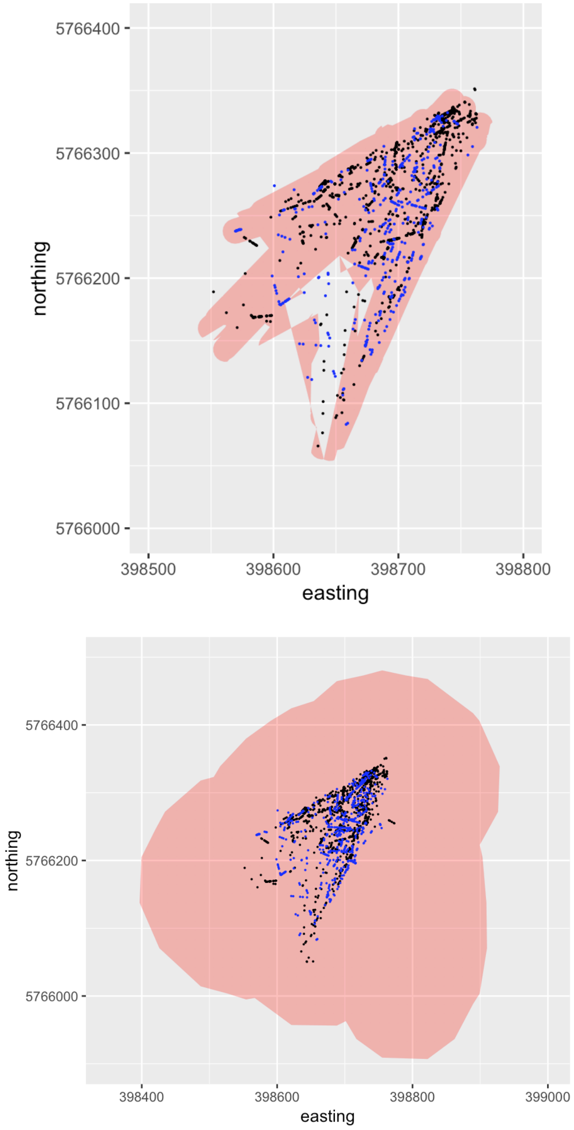


Figure 13: Estimated contact area (red) and measured contacts (black, blue) between deer 56 and 58 with linear buffer 10m (upper) and BBMM 90% (lower).

The red area in Figure 13 is the spatial intersection of all buffers that were part of a contact. Together with the measured contact data it gives a first estimate of how good the method worked.

With the deer data the jPPA contains all measured contacts within the area. The area is even by far exceeding the extent of the contact data.

For the fox data where only contacts were detected for two of five individuals. The method is also applicable for the animals without contact and returns a potential contact area.

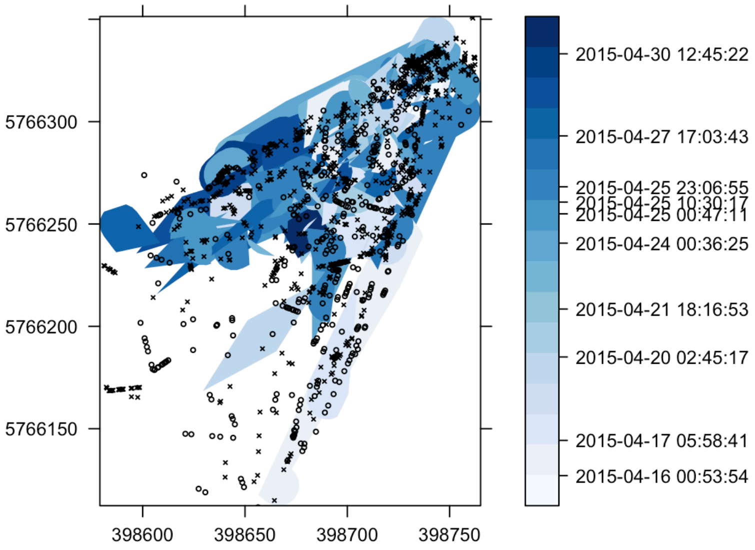


Figure 14: Estimated (blue) and measured (black) contacts between deer 56 and 58.

In contrast to Figure 13, Figure 14 shows the contact area not as a whole but the single contact events are coloured in blue according to the point in time when a contact occurs. The time range lasts over a month (April) and starts with light blue and gets darker with time passing. The measured contacts are the circles and the crosses where the circles are the contacts measured by the proximity logger deer 58 and the crosses by the logger of deer 56.

4.3 Evaluation

Animal to fixed nodes contacts

The BBMM method performed well nearly all contacts were identified correctly with a recall of >0.99 for all animals. The buffer method shows a different result. For the foxes M030 and F038 a recall of 0.7 and 0.9 could be achieved. For the other foxes only few contacts were correctly identified with recalls of <0.25.

Table 7: Confusion matrix of animal to fixed node contacts with fox data.

	Buffer 10 m			**BBMM 90%**			**BBMM 50%**		
Fox	**True positives**	**False negatives**	**Re-call**	**True positives**	**False negatives**	**Re-call**	**True positives**	**False negatives**	**Re-call**
M030	70	30	0.7	11	0	1	11	0	1
F038	320	34	0.9	354	0	1	354	0	1
M031	26	83	0.24	108	0	1	108	0	1
F032	2	46	0.04	48	0	1	48	0	1
M034	283	1137	0.2	1420	0	1	1408	12	0.99

The BBMM method shows the same performance for the deer data, actually for every deer a recall of 1 is achieved (Table 9). The buffer method has higher recall values with the deer data than with the fox data. With buffer 10m the recall ranges between 0.75 and 0.97 and with buffer power level it ranges from 0.79 to 1.

Animal to animal contacts

The performance of the methods with different parameterization for estimating animal to animal contacts for the two foxes M030 and F038 is shown in Table 8. The methods estimated only few contacts between the animals and from this few contacts only a small number was identified correctly. Buffer 40m with 1 true positive and BBMM 90% with 3 true positives which lead to a recall value of 0.13 and 0.35, respectively. The methods with smaller parameterization were not able to estimate contacts and though have a recall of 0. Only the jPPA for this two animals contained all ground truth contacts and has a recall of 1. The jPPA was applied on other foxes, but performance could not be tested as no ground truth data exists.

The

Table 10 shows the confusion matrix for the linear buffer and the BBMM method with different parameterisation applied on the deer data set. The confusion matrix consists of three values: the true positives, false negatives and the recall. Figure 13 above visualizes the true positives and false negatives. All points inside the red area are true positives whereas points lying outside this area are false negatives.

The BBMM method 90% estimates ≥ 0.95 of the contacts correctly for all dyads. For most of the dyads even 100% are reached. Such a high recall is achieved with 0 or a small number of false negatives as in the case of BBMM 90%. More than half of the pairs reached a recall of 1 with the BBMM 50%, but there are also some poor values ranging between 0.00 and 0.33 indicating more false negatives were estimated than true positives.

For the dyad d5051 the buffer method with power level all estimated contacts were predicted at the wrong place which is denoted with 0 true positives. Beside this and one other all dyads reached a recall of > 0.5 and five even are 1.00 rounded.

The recall for buffer method with width of 10 m for the pair d5051 is nearly 0 similar to the other buffer method. The highest recall value with the buffer method 10 m is 0.99.

For the jPPA method the recall is 1 one for all deer and the foxes M030 and F038. For all other foxes it cannot be determined as there is not even ground truth to test for the performance.

Table 8: Confusion matrix of animal to animal contacts with foxes M030 and F038.

M030-F038	true positives	false negatives	recall
Buffer 10m	0	8	0
Buffer 15m	0	8	0
Buffer 40m	1	7	0.13
BBMM 90%	3	5	0.35
BBMM 50%	0	8	0

Table 9: Confusion matrices for animal to fixed node cotacts of the deer data.

Deer dyad	Buffer 10 m			Buffer power level			BBMM 90%			BBMM 50%		
	True positives	False negatives	Recall	True positives	False negatives	Recall	True positives	False negatives	Recall	True positives	False negatives	Recall
50	14896	2063	0.88	16950	9	1.00	5653	0	1	5653	0	1
51	21236	1528	0.93	22214	550	0.98	7588	0	1	7588	0	1
52	14487	1339	0.92	15708	118	0.99	7913	0	1	7913	0	1
53	30304	9992	0.75	31988	8308	0.79	13432	0	1	13432	0	1
55	28474	743	0.97	29017	200	0.99	9739	0	1	9739	0	1
56	38064	9774	0.80	39568	8270	0.83	15946	0	1	15946	0	1
58	23199	1005	0.96	24150	0	1.00	8050	0	1	8050	0	1
59	5855	777	0.88	6221	411	0.94	3316	0	1	3316	0	1

Table 10: Confusion matrices for animal to animal contacts of the deer data.

Deer dyad	Buffer 10 m			Buffer power level			BBMM 90%			BBMM 50%		
	True positives	False negatives	Recall	True positives	False negatives	Recall	True positives	False negatives	Recall	True positives	False negatives	Recall
d5051	1	18	0.05	0	19	0	19	0	1	19	0	1
d5052	206	149	0.58	307	48	0.86	1324	0	1	1169	155	0.83
d5053	654	61	0.91	713	2	1.00	715	0	1	715	0	1
d5055	1	2	0.33	1	2	0.33	3	0	1	2	1	0.67
d5056	349	39	0.90	387	1	1.00	388	0	1	388	0	1
d5058	260	149	0.64	361	48	0.88	409	0	1	409	0	1
d5059	275	37	0.88	284	28	0.91	312	0	1	1	311	0.00
d5152	327	126	0.72	437	16	0.96	492	0	1	492	0	1

Deer dyad	Buffer 10 m			Buffer power level			BBMM 90%			BBMM 50%		
	True positives	False negatives	Recall	True positives	False negatives	Recall	True positives	False negatives	Recall	True positives	False negatives	Recall
d5153	104	34	0.75	115	23	0.83	138	0	1	133	5	0.96
d5155	465	20	0.96	470	15	0.97	485	0	1	473	12	0.98
d5156	90	46	0.66	102	34	0.75	136	0	1	136	0	1
d5158	347	206	0.63	40484	57	1.00	553	0	1	553	0	1
d5159	67	19	0.78	68	18	0.79	86	0	1	16	70	0.19
d5352	2264	136	0.94	2362	38	0.98	3533	0	1	3508	25	0.99
d5355	153	172	0.47	192	133	0.59	312	13	0.96	276	49	0.85
d5356	6832	76	0.99	6881	27	1.00	6908	0	1	6908	0	1
d5358	1129	47	0.96	1165	11	0.99	1176	0	1	1176	0	1
d5359	693	40	0.95	727	6	0.99	733	0	1	733	0	1
d5552	124	190	0.39	303	11	0.96	376	0	1	376	0	1
d5556	107	123	0.47	146	84	0.63	230	0	1	230	0	1
d5558	70	166	0.30	221	15	0.94	236	0	1	236	0	1
d5559	18	6	0.75	18	6	0.75	24	0	1	8	16	0.33
d5652	2230	145	0.94	2333	42	0.98	2945	0	1	2941	4	1.00
d5658	1266	33	0.97	1289	10	0.99	1299	0	1	1299	0	1
d5659	590	42	0.93	614	18	0.97	632	0	1	632	0	1
d5852	886	30	0.97	913	3	1.00	1100	0	1	1100	0	1
d5859	930	132	0.88	1025	37	0.97	1062	0	1	1017	45	0.96
d5952	199	105	0.65	282	22	0.93	78	4	0.95	4	78	0.05

5 Discussion

The explorative analysis of the data shows that data quality is good and the number of outliers is relative small. The statistical measures of the step length improved after the data cleaning and are more realistic, especially for certain animals.

As the results show, it is possible to compute possible contacts between individuals with the methods and the newly introduced FindAnimalContacts–procedure from location data only. Modelling contacts between animals and fixed nodes is possible as well. The comparison of the methods for animal to animal contacts based on the fox data is unfeasible as the ground truth is very small.

The jPPA and the BBMM 90% are performing best regarding the recall values. When the resulting contact area sizes are visually explored it is apparent that the linear buffer method generates the smallest and the most angular areas (Figure 13). Whereas the dBBMM and especially the jPPA generates coarse contact areas. This pattern goes along with the recall values. The coarser the area is; the better recall values are achieved. This means that there are big parts of the area where actually no contact was detected by the loggers and so estimating large areas of omission. Thus, this cannot be quantified as there is no ground truth data about where no contact occurred available. The BBMM 50% achieves high recall values as well, but differently from the BBMM 90% the area is smaller. The difference in the area size is recognisable in Figure 10.

The recall seems not to be an appropriate measure for the performance or at least not if being the only value to compare the performance of the methods. The methods should also be compared based on a ratio of the area.

The same conclusions can be drawn for the detection of contacts between an animal and the fixed nodes. The jPPA and the BBMM 90% have the best recall values.

The result of the three step procedure is list of pairwise ids of which path segments, represented as polygons, have contact. The polygons represent a part of the movement path of an animal in form of a segment area. In this way an interaction is determined by the spatial extent and the time interval of the polygon. In contrast to that the jPPA returns an area which is not part of the movement path but the size of the area is inferred from the location data. The joint area cannot be considered in detail, meaning no separate interaction event can be detected. Instead an area of potential interaction is computed.

The developed procedure for estimating contact enables to state where and when a contact happened. Nothing is known about the kind of contact and the associated behaviour. Combining the analysis for contact detection with other techniques as for example the different interaction indices, described by Long et al. (2014), could give further information about what behaviour is linked to the contacts.

The confusion matrices show that the results depend on the parameterization of the methods. Every method has a parameter which has impact on the result. For the linear buffer it is the width. For the dBBMM the choice of the contour level is crucially. For the jPPA the v_{max} and the available time budget are the important parameters.
With increasing area per path segment the potential to find contacts between them increases. The aim is not to estimate as many contacts as possible but to estimate the most probable contacts in terms of the ecology.

The probability information of the utilization distribution of the dBBMM method gets lost when applying the contouring as done here. Being able to keep the probability information for the estimation of contacts would result in contact probabilities. A more detailed result could be achieved which allows further interpretations.

The jPPA method applied on the deer data show that the PPA and the jPPA are large. The reason for this is the lower sampling rate than for the foxes, because with lower sampling rate the potential for movement between fixes is considerable and resulting in large PPAs. This leads large jPPA and thus to an over-estimation of interactions and uncertainty of the result (Long et al. 2015). Due to this large jPPAs the method achieves good recall values for the deer data set, but at the same time the mapping is not very detailed due to the large dimension of the area.
The jPPA detected a potential interaction area for the foxes even when trajectories lie over 1 km apart from each other. But how probable is it that the foxes have actually been in this joined path area if no spatial overlap of the location fixes exists and no contact is detected by the logger, either. But as the example of M030 and M031 show not in any case where the PPAs overlap a jPPA is detected.

The jPPA method is not suitable to estimate contacts with fixed nodes. It could be tested whether the fixed node lies within the PPA but the result does not reveal additional information and therefore is not very useful.

The contact data had to be interpolated to serve as ground truth. The interpolation was done based on the GPS locations. Thereby errors due to the interpolation appeared, in fact some contacts were located at unrealistic positions. An unrealistic position is if the contact is located outside of the ranging distance defined by the power level.

Regarding the number of contacts between an animal and fixed nodes the number of contacts by the proximity logger is many times larger than the number of estimated contacts. The reason for this is that within a buffer segment several measured contacts can be located. This shows the issue of the methods. The methods enable to map contacts but the result is not a single location but rather an area of interaction. This fact also complicates the interpretation of the results.

Depending on the sampling rate of tracking data assuming linear movement is not realistic. Therefore, the application of the linear buffer method is no appropriate for all kind of data. Due to its simplicity, the method might still be useful for a first impression to see in what relationship the animals are. The advantage of the method is that it has only one parameter, the width, which can be defined exactly in accordance with the research of different animals and if the need of different spatial scales is necessary. The advantage of the BBMM is that it works with lower sampling rates and the location error of the GPS measurement is included.
In contrast to space use overlaps where only spatial overlap is relevant, for the methods applied here beside the spatial also the temporal coincidence is requested. On the other hand, the joint space use of two individuals is a prerequisite for successfully applying the methods.

As the used data sets were collected in complex environments, it is difficult assess how reasonable the results are. The deer were fenced in three paddock adjoin each other. The methods do not consider the given boundary; therefore, contacts might be estimated outside this area which does not make sense.

The fox data was collected in the urban area of Brighton. Therefore, an additional layer with buildings should be added to see the results in the context of the environment. Also other habitats it might be useful to add environmental data to the results.

It was expected that the data set with higher sampling rate performs better than the data set with lower sampling rate. Due to fact that for the fox data the methods could only be applied on one dyad and the ground truth is very small, a comparison between the data sets is not reasonable. The possible effect of the sampling rate on the result requires further research. A possible approach is using a very high sampled data set from which several data subsets with lowering sampling rates are derived and then applied on the methods.

An advantage of the introduced procedure to detect contacts is that temporal coincidence is considered as well as spatial overlap. The temporal information about 'when' a contact occurs is also stored in the results as start and end point of the path segment. However, an adequate visualization of the spatial and temporal contact information is challenging. Figure 14 shows a static attempt of visualizing contacts by shading them according to the time. Several temporal layers lie on the top of each other, thus not all data is visible. A better approach would be a dynamic visualization where time could be controlled by a slide bar. The temporal information enables to detect temporal patterns on different time scales.

The package "spatsoc" seems to follow the same idea of detecting potential contact between individuals but with different implementation. Therefore, a comparison between the methods applied in this thesis and the functions provided by the package would be useful to see if the results are equivalent. Especially, the group lines()-function which corresponds to the linear buffer method applied here, could be compared by using the same buffer width.

6 Conclusion

6.1 Major findings

The thesis shows an explorative study of different methods to map inter-individual contact spatially explicit from location data only. Three different methods, namely Euclidean distance, dBBMM and jPPA, were applied, where for the first two an additional procedure was developed to detect possible contacts among two individuals. In combination of the methods and the introduced procedure it was possible to estimate contacts between animals from GPS data only. Additionally, it was also possible to compute whether certain locations (fixed nodes) were visited by an individual. Due to the fact that the methods underlie different assumptions the outcome varies. The evaluation of the results shows the methods perform differently and the outcome depends on the parameterization. According to the confusion matrix the linear buffer method performs the worst, even though the resulting interaction area is the most detailed. Unlike the other two methods, the jPPA method lacks a segment-wise consideration.

6.2 Limitations

This thesis carried out an explorative study of the methods. To be able to make reliable statements about the performance of the methods the available data volume is insufficient considering that, more data is required. Further research is necessary to better understand the opportunities but also the limitations of the methods. But based on the gained knowledge the jPPA method is not appropriate for detailed mapping of contacts.

The modelled contacts were only compared with the ground truth based on the spatial accordance, but it is not checked whether temporal coincidence is present. To be able to make reliable statements about the performance of the methods the available data volume is insufficient considering that, more data is required.

6.3 Future work

A systematic research of the methods with is essential to obtain a comprehension of an appropriate parameterization for the methods and in respect to the species the data is about. To better understand how the method and the result is influenced by the sampling rate of the data, very high sampled telemetry data is required to systematically test the methods. Also in terms of the parameterization a systematic research is required for a better understanding the impact of the parameter setting on the results. The optimal parameterization computes good results and the results are reasonable for the species.

By adding additional layers with environmental data the informative value of the results could be enhanced and the interactions can be set into the context of the animals’ habitat.

For a simpler, clearer but also more detailed interpretation of the results a more appealing visualization is required. An interactive visualization is a good alternative to static representations. This would allow to show the results more detailed as different points in time can be seen in the same animation.

7 References

De Almeida Jácomo, A.T., Kashivakura, C.K., Ferro, C., Malzoni, M., et al., 2009. Home range and spatial organization of maned wolves in the Brazilian grasslands. *Journal of Mammalogy*, 90, pp.150–157.

Atwood, T.C. & Weeks, H.P., 2003. Spatial home-range overlap and temporal interaction in eastern coyotes: the influence of pair types and fragmentation. *Canadaian Journal of Zoology*, 81, pp.1589–1597.

Bertrand, M.R., DeNicola, A.J., Beissinger, S.R. & Swihart, R.K., 1996. Effects of parturition on home ranges and social affiliations of female white-tailed deer. *J Wildl Manage*, 60.

Bettaney, E.M., James, R., St Clair, J.J.H. & Rutz, C., 2015. Processing and visualising association data from animal-borne proximity loggers. *Animal Biotelemetry*, 3(1), pp.1–11.

Boyland, N.K., James, R., Mlynski, D.T., Madden, J.R., et al., 2013. Spatial proximity loggers for recording animal social networks: Consequences of inter-logger variation in performance. *Behavioral Ecology and Sociobiology*, 67(11), pp.1877–1890.

Bromley, C. & Gese, E.M., 2001. Effects of sterilization on territory fidelity and maintenance, pair bonds, and survival rates of free-ranging coyotes. *Canadaian Journal of Zoology*, 79, pp.386–392.

Buchin, K., Sijben, S., Loon, E.E. Van, Sapir, N., et al., 2015. Deriving movement properties and the effect of the environment from the Brownian bridge movement model in monkeys and birds. *Movement Ecology*, 3(18), pp.1–11.

Buchin, K., Sijben, S., Willems, E.P. & Arseneau, T.J.M., 2012. Detecting Movement Patterns Using Brownian Bridges BT - ACM SIGSPATIAL. In Redondo Beach, CA, USA: ACM Press.

Buchin, K., Willems, E.P., Arseneau, T.J.M. & Sijben, S., 2012. Detecting Movement Patterns using Brownian Bridges * Categories and Subject Descriptors. In *SIGSPATIAL '12 Proc. 20th International Conference on Advances in Geographic Information Systems*. pp. 119–128.

Bullard, F., 1999. *Estimating the home range of an animal: a Brownian bridge approach*. University of North Carolina.

Byrne, M.E., McCoy, J.C., Hinton, J.W., Chamberlain, M.J., et al., 2014. Using dynamic Brownian bridge movement modelling to measure temporal patterns of habitat selection. *Journal of Animal Ecology*, 83, pp.1234–1243.

Cagnacci, F., Boitani, L., Powell, R.A. & Boyce, M.S., 2010. Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philos Trans R*

Soc Lond B Biol Sci, 365.

Collier, B., 2016. *moveud: Dynamic brownian bridge movement model individual time-step utilization distributions*,

Croft, D.P., James, R. & Krause, J., 2008. *Exploring animal social networks*, Princeton University Press.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., et al., 2015. Analysis and visualisation of movement : an interdisciplinary review. *Movement Ecology*, 3(5), pp.1–24.

Dodge, S., Weibel, R., Ahearn, S.C., Buchin, M., et al., 2016. Analysis of movement data. *International Journal of Geographic Information Science*, 30(5), pp.825–834.

Drewe, J.A., N, W., Carter, S.P., Bearhop, S., et al., 2012. Performance of Proximity Loggers in Recording Intra- and Inter-Species Interactions: A Laboratory and Field-Based Validation Study. *PLoS One*, 7(6).

Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognition Letters*, 27, pp.861–874.

Haddadi, H., King, A.J., Wills, A.P., Fay, D., et al., 2011. Determining association networks in social animals: choosing spatial-temporal criteria and sampling rates. *Behavioral Ecology and Sociobiology*, 65, pp.1659–1668.

Hägerstrand, T., 1970. What about people in regional science? *Pap Reg Sci Assoc*, 24.

Hays, G.C., Ferreira, L.C., Sequeira, A.M.M., Meekan, M.G., et al., 2016. Key Questions in Marine Megafauna Movement Ecology. *Trends in Ecology & Evolution*, 31(6), pp.463–475.

Holloway, P. & Miller, J.A., 2017. Analysis and Modelling Movement in Comprehensive Geographic Information System. In *Comprehensice Geographic Information Systems*. Elsevier Inc., pp. 162–179.

Horne, J.S., Garton, E.O., Krone, S.E. & Lewis, J.S., 2007. Analyzing animal movements using Brownian bridges. *Ecology*, 88(9), pp.2354–2363.

Ji, W., White, P.C.L. & Clout, M.N., 2005. Contact rates between possums revealed by proximity data loggers. *Journal of Applied Ecology*, 42, pp.595–604.

Karlin, M. & Chadwick, J., 2011. Measures of space use and association of two unrelated male red wolves in a shared area. *Mammal Study*, 36, pp.147–153.

Kernohan, B.J., Gitzen, R.A. & Millspaugh, J.J., 2001. Analysis of Animal Space use and Movements BT - Radio Tracking and Animal Populations. In J. J. Millspaugh & J. M. Marzluff, eds. New York: Academic.

Knight, N.L. & Wang, J., 2009. A Comparison of Outlier Detection Procedures and Robust Estimation Methods in GPS Positioning. *Journal of Navigation*, 62(4), pp.699–709.

Kranstauber, B., Kays, R., LaPoint, S.D., Wikelski, M., et al., 2012. A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, 81(4), pp.738–746.

Kranstauber, B., Smolla, M. & Scharf, A.K., 2017. *move: Visualizing and Analyzing Animal Track Data*,

Krause, J., Krause, S., Arlinghaus, R., Psorakis, I., et al., 2013. Reality mining of animal social systems. *Trends in Ecology and Evolution*, 28(9), pp.541–551.

Lewis, D., 1991. Evaluating text categorization. In Morgan Kaufmann, ed. *Proc. SPeech and Natural Language Workshop*. pp. 312–318.

Li, L., Ban, H. & Wechlser, S.P., 2017. Spatial Data Uncertainty. In *Comprehensice Geographic Information Systems*. Elsevier Inc., pp. 313–334.

Long, J.A. & Nelson, T.A., 2013. Measuring dynamic interaction in movement data. *Trans GIS*, 17.

Long, J.A. & Nelson, T.A., 2012. Time geography and wildlife home range delineation. *J Wildl Manage*, 76.

Long, J.A., Nelson, T.A., Webb, S.L. & Gee, K.L., 2014. A critical examination of indices of dynamic interaction for wildlife telemetry studies. *Journal of Animal Ecology*, 83, pp.1216–1233.

Long, J.A., Nelson, T.A., Webb, S.L. & Gee, K.L., 2014. A critical examination of indices of dynamic interaction for wildlife telemetry studies. *J Anim Ecol*, 83.

Long, J.A., Webb, S.L., Nelson, T.A. & Gee, K.L., 2015. Mapping areas of spatial-temporal overlap from wildlife tracking data. *Movement Ecology*, 3(1), pp.1–14.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., et al., 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.19052–9.

Ossi, F., Focardi, S., Picco, G. Pietro, Murphy, A., et al., 2016. Understanding and geo - referencing animal contacts : proximity sensor networks integrated with GPS - based telemetry. *Animal Biotelemetry*, 4(21), pp.1–14.

Pettitt, T., Lin, I., Lewis, M. & Merrill, E., 2015. *Functionality and error assessment of new GPS-porxmity collars on captive deer*,

Picco, G. Pietro, Molteni, D., Murphy, A.L., Ossi, F., et al., 2015. Geo-referenced Proximity Detection of Wildlife with WildScope : Design and Characterization.

Pinter-Wollman, N., Hobson, E.A., Smith, J.E., Edelman, A.J., et al., 2014. The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behavioral Ecology*, 25(2), pp.242–255.

R Core Team, 2013. *R: A language and environment for statistical computing*, Vienna, Austria.

Robitaille, A.L., Webber, Q. & Vander Wal, E., 2018. *spatsoc: Group Animal Relocation Data by Spatial and Temporal Relationship*,

Rushmore, J., Caillaud, D., Matamba, L., Stumpf, R.M., et al., 2013. Social network analysis of wild chimpanzees provides insights for predicting infectious disease risk. *Journal of Animal Ecology*, 82, pp.976–986.

Schauber, E., Storm, D. & Nielson, C., 2007. Effects of joint space use and group membership on contact rates among white tailed deer. *J Wildl Manage*, 71.

Spiegel, O., Leu, S.T., Sih, A. & Bull, C.M., 2016. Socially interacting or indfferent neighbours? Randomization of movement paths to tease apart social preference and spatial constraints. *Methods in Ecology and Evolution*, 7, pp.971–979.

Tosa, M.I., Schauber, E.M. & Nielsen, C.K., 2015. Familiarity breeds contempt: combining proximity loffers and GPS reveals female white-tailed deer (odocoileus virginianus) avoiding close contact with neighbours. *Journal of Wildlife Diseases*, 51(1), pp.79–88.

Walrath, R., Vercauteren, K.C. & Deer, W., 2012. Efficacy of Proximity Loggers for Detection of Contacts Between Maternal Pairs of White-Tailed Deer Efficacy of Proximity Loggers for Detection of Contacts Between Maternal Pairs of. *Wildlife Society Bulletin*, 35(4), pp.452–460.

Van der Watt, P.A., 1980. A note on estimation bounds of random variables. *Biometrika*, 97, pp.712–714.

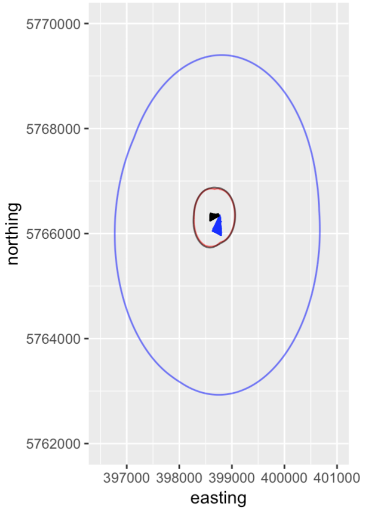
Wey, T., Blumstein, D.T., Shen, W. & Jordán, F., 2008. Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour*, 75, pp.333–344.

Whitehead, H. & Dufault, S., 1999. Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. *Adv Study Behav*, 28, pp.33–74.

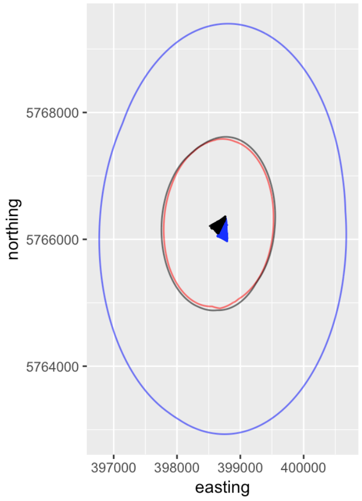
8 Appendix

A1 jPPA

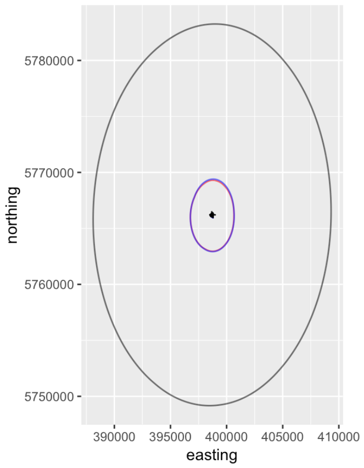
The red ellipse corresponds to the jPPA of the two animals and the first animal is in blue and the second in black.



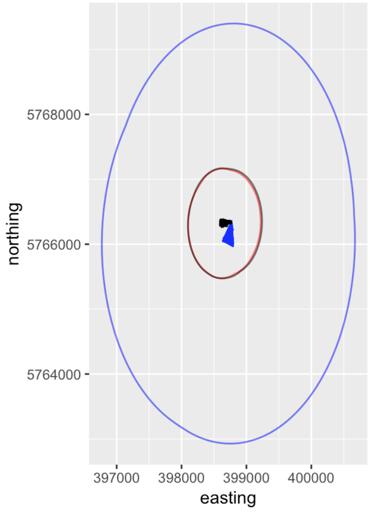
D5051



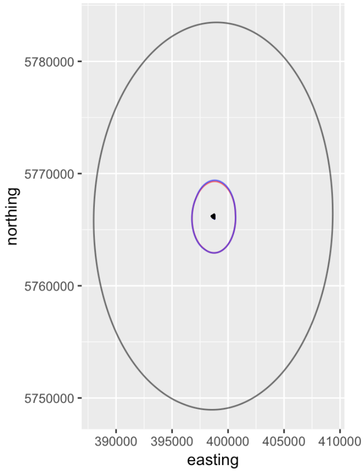
D5052



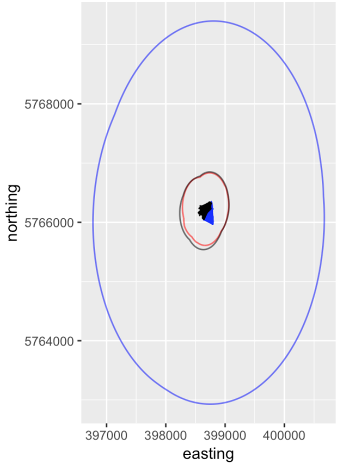
D5053



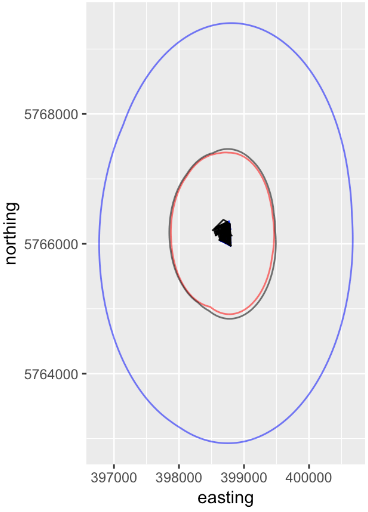
D5055



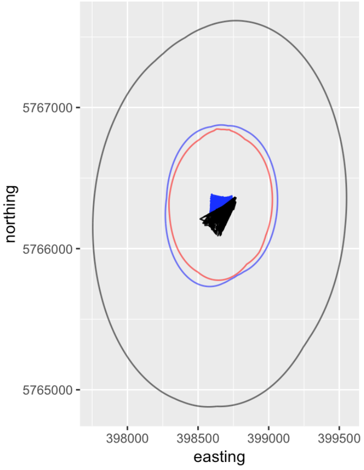
D5056



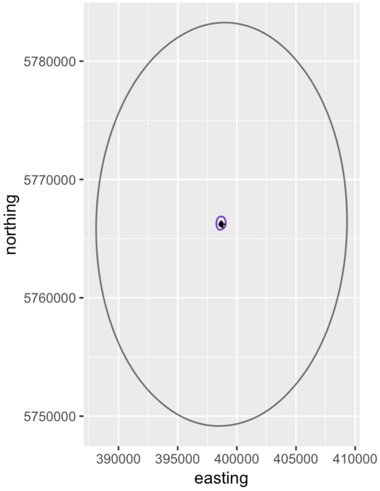
D5058



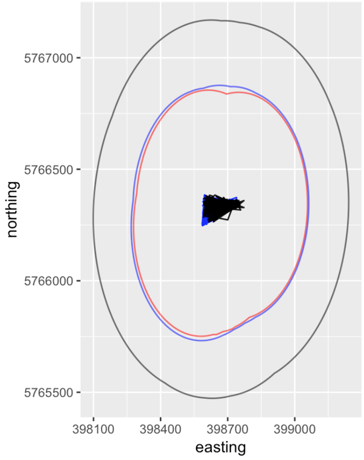
D5059



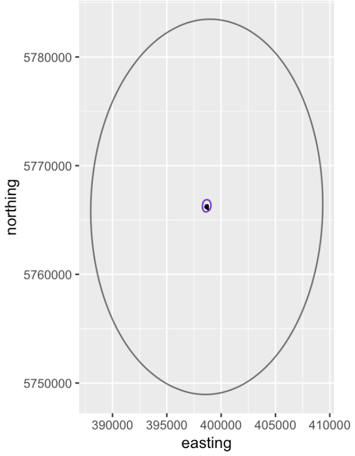
D5152



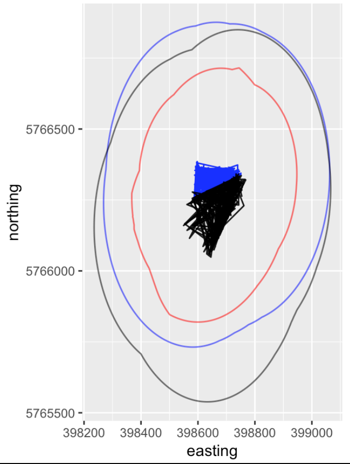
D5153



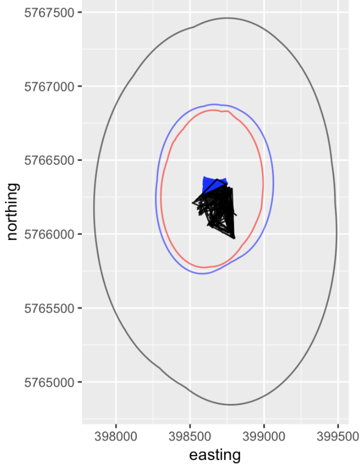
D5155



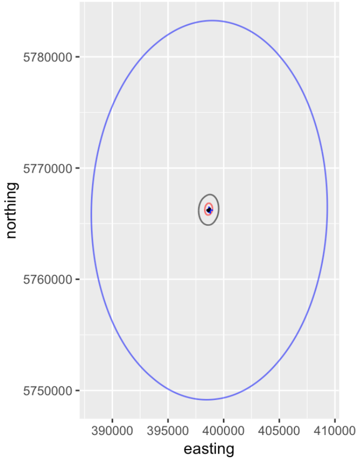
D5156



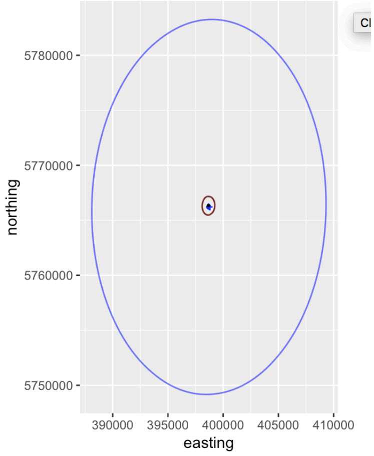
D5158



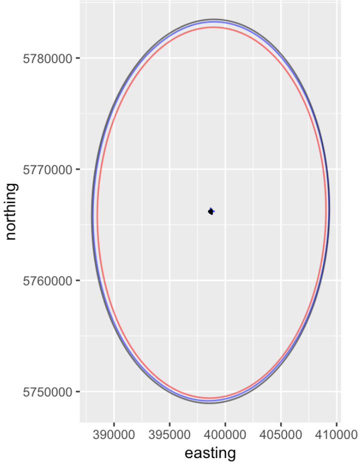
D5159



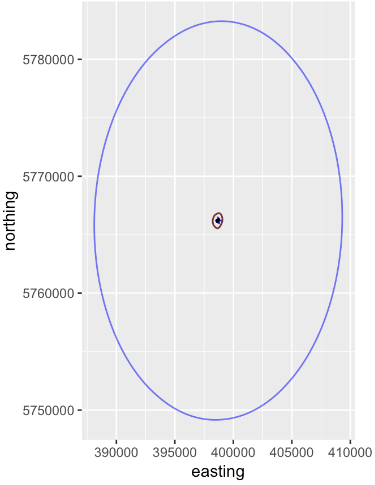
D5352



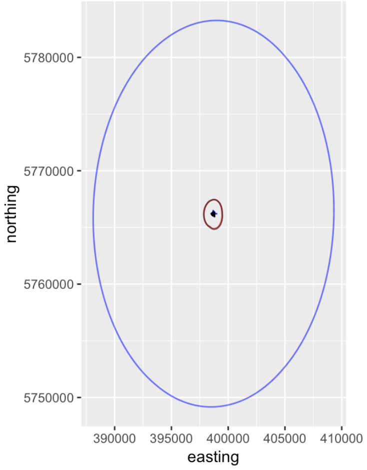
D5355



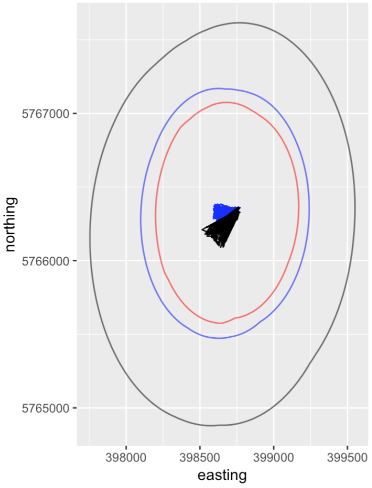
D5356



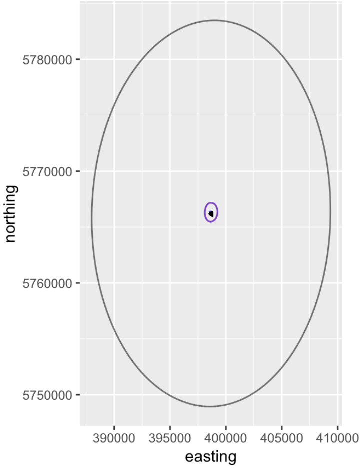
D5358



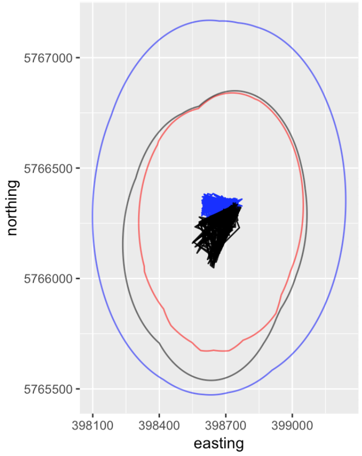
D5359



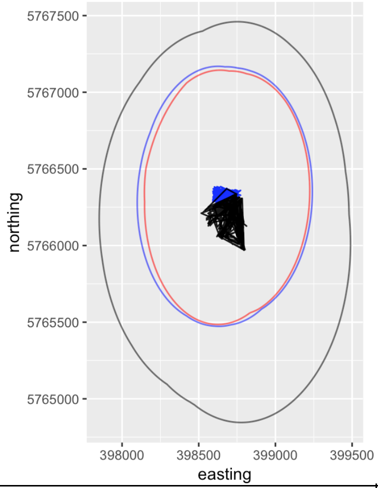
D5552



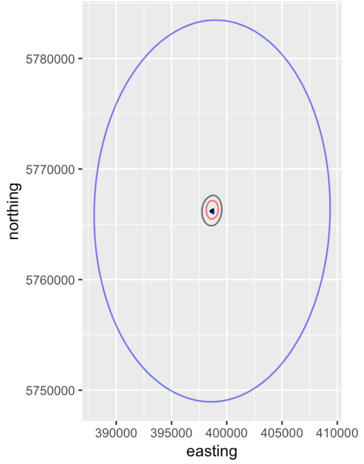
D5556



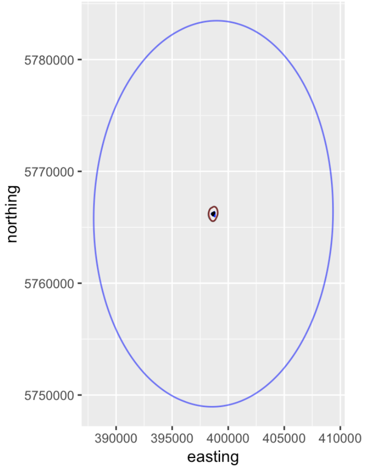
D5558



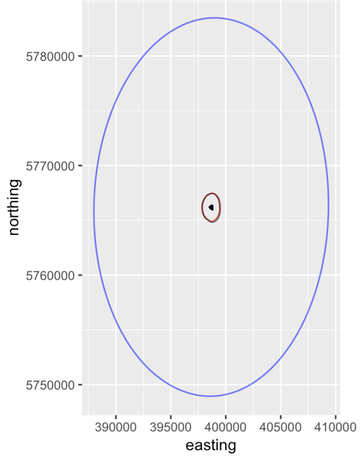
D5559



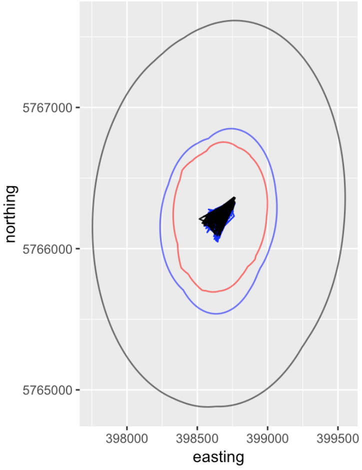
D5652



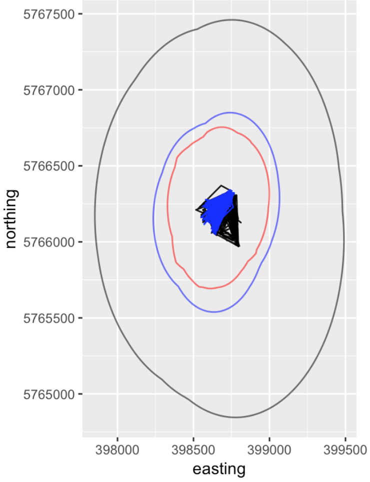
D5658



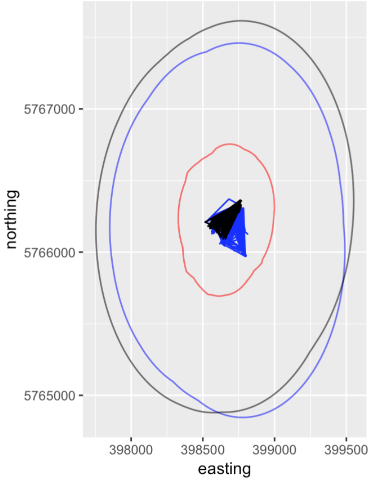
D5659



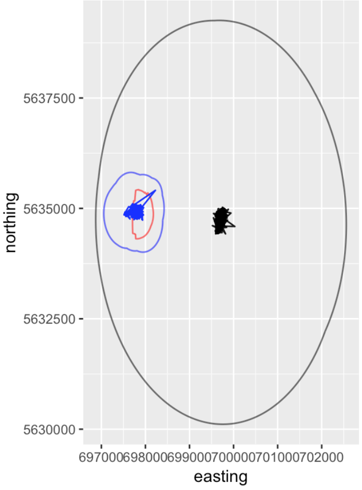
D5852



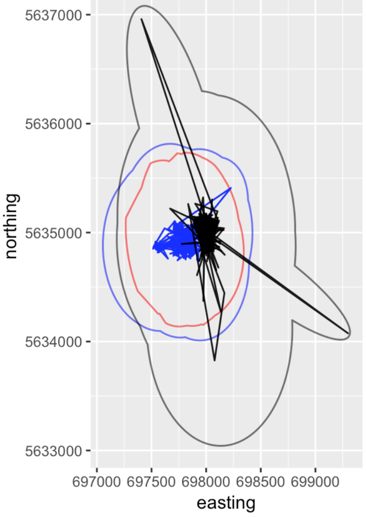
D5859



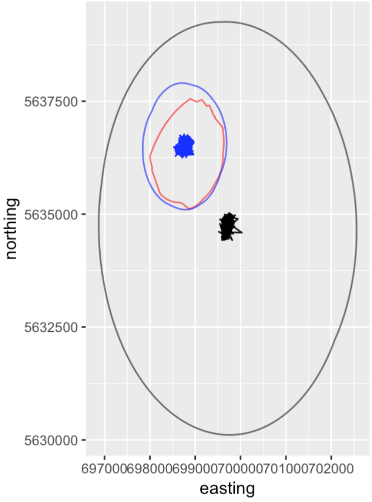
D5952



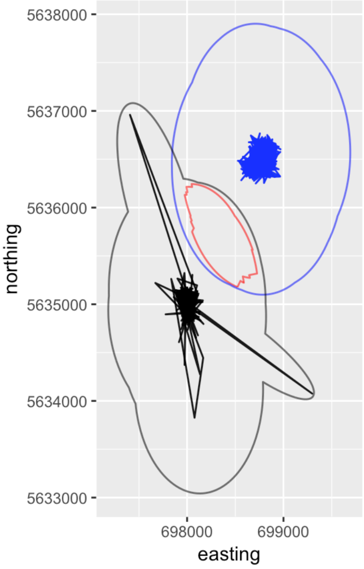
M030-F032



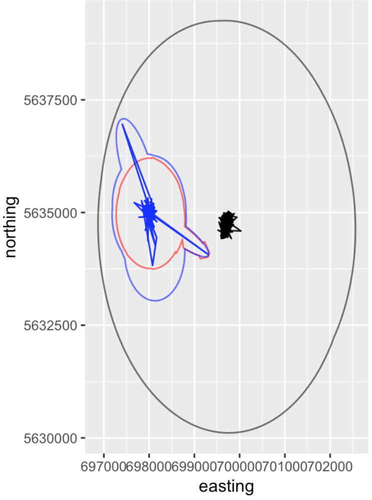
M030-F038



M031-F032



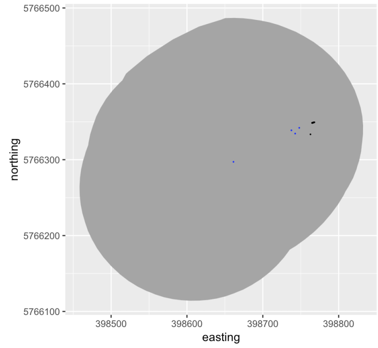
M031-F038



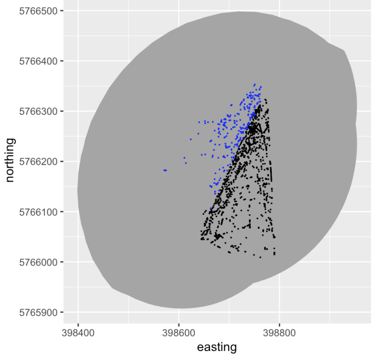
F038-F032

A2 dBBMM 90%

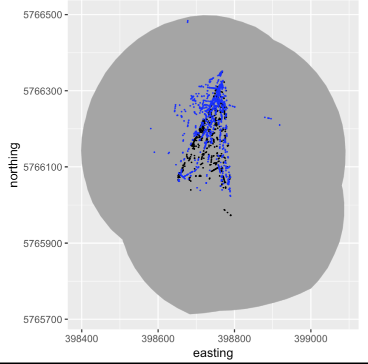
The grey area corresponds to the interaction area of the two animals and the measured contact data of first animal is in black and of the second in blue.



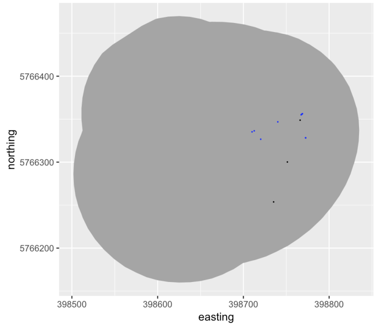
D5051



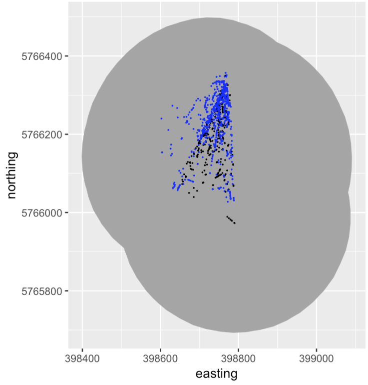
D5052



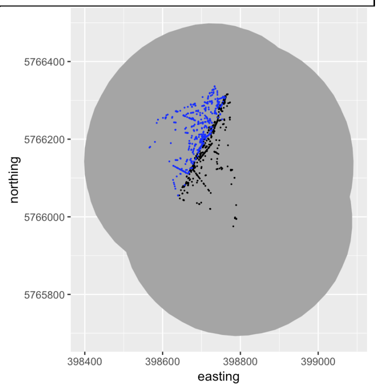
D5053



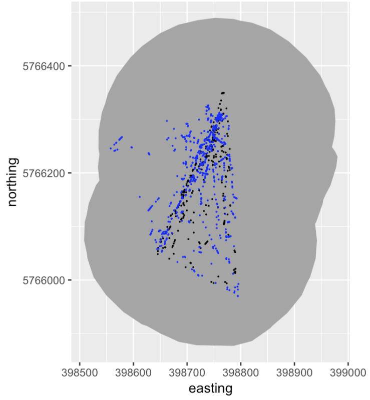
D5055



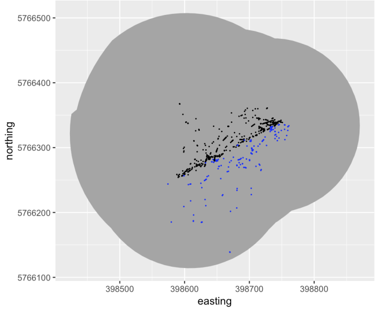
D5056



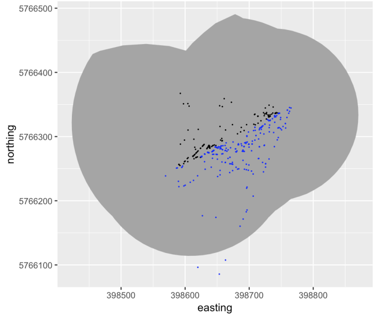
D5058



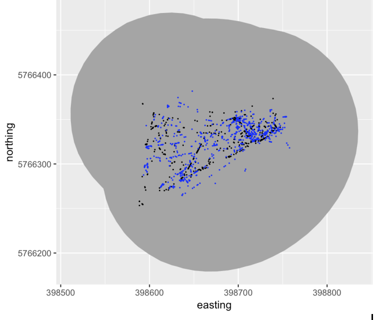
D5059

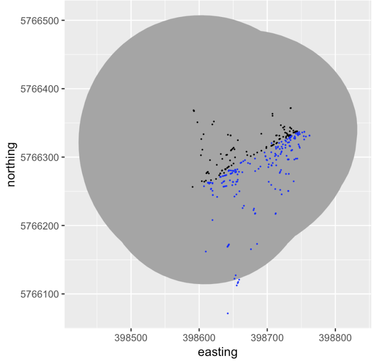


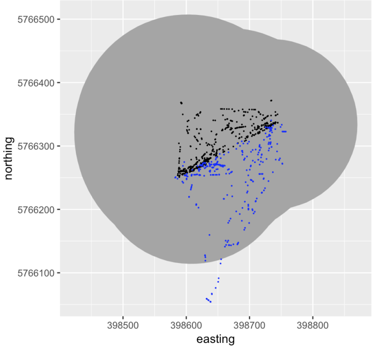
D5152



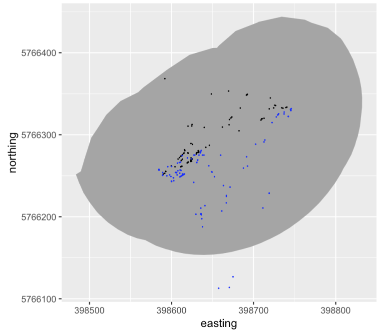
D5153

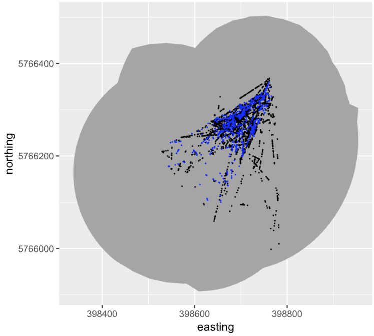


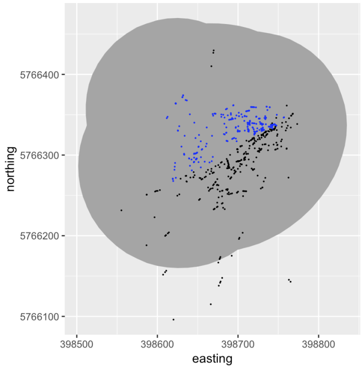




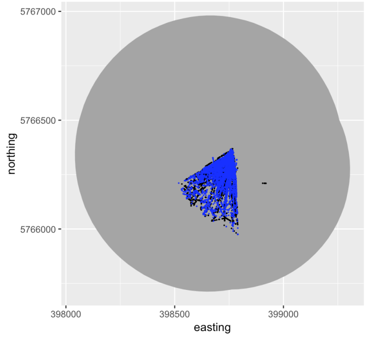
D5155 D5156 D5158

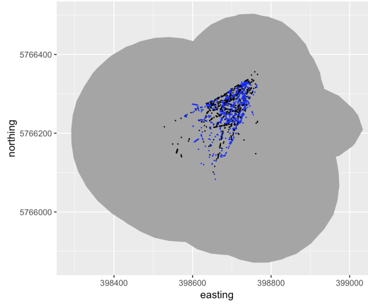


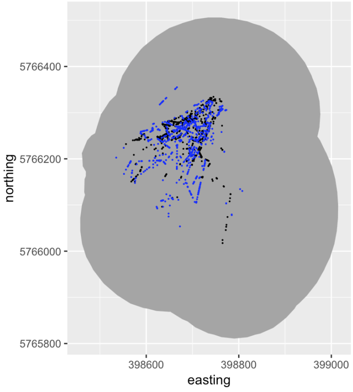




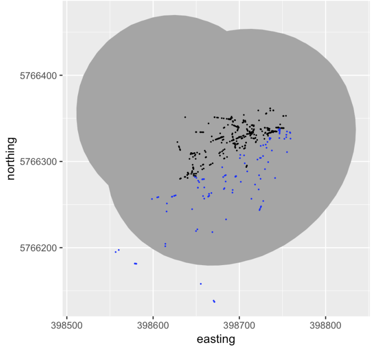
D5159 D5352 D5355

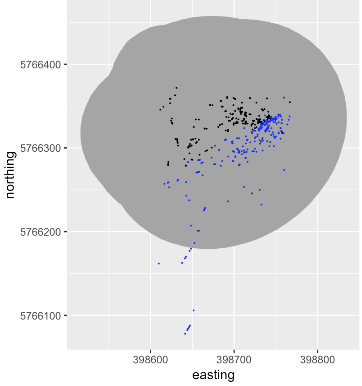


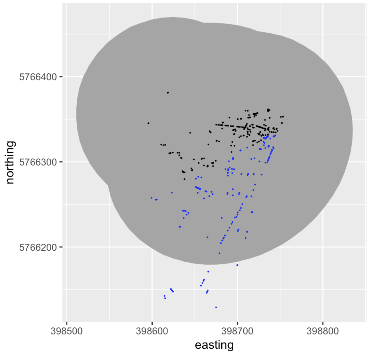




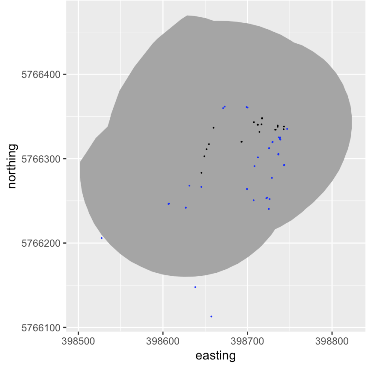
D5356 D5358 D5359

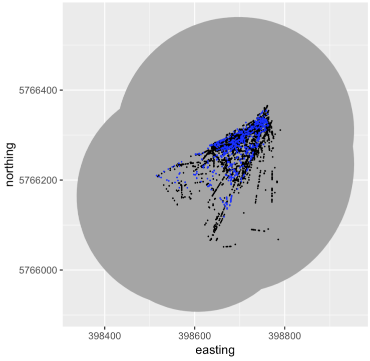


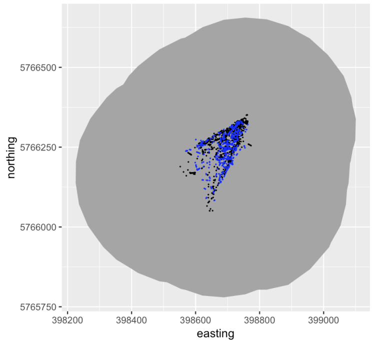




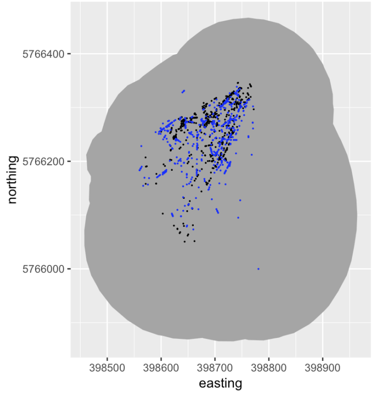
D5552 D5556 D5558



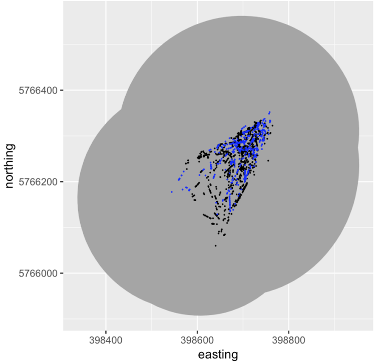




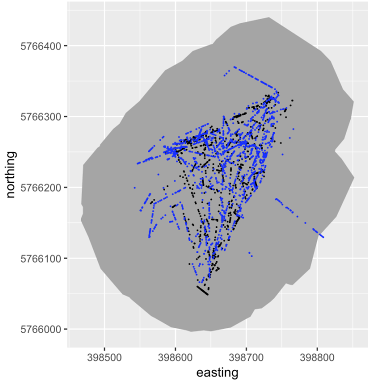
D5559 D5652 D5658



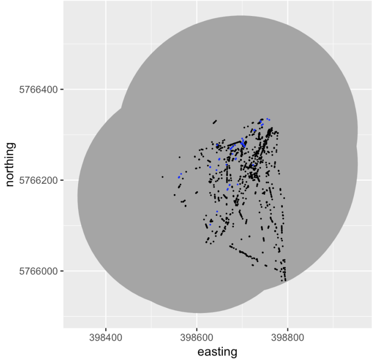
D5659



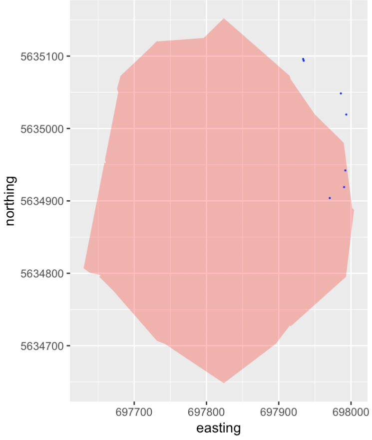
D5852



D5859



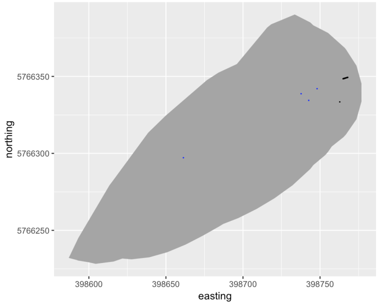
D5952



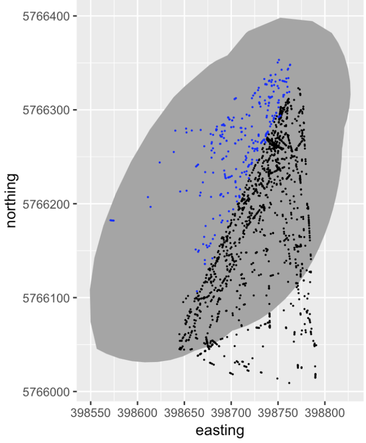
M030-F038

A3 dBBMM 50%

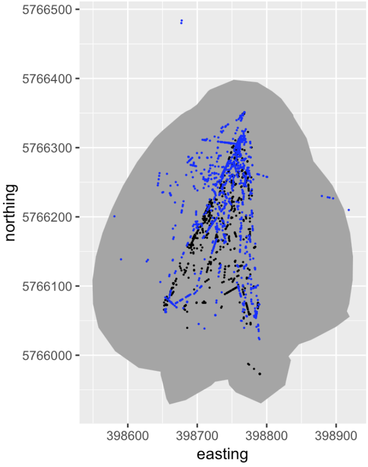
The grey area corresponds to the interaction area of the two animals and the measured contact data of first animal is in black and of the second in blue.



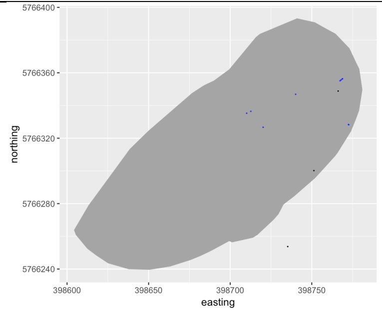
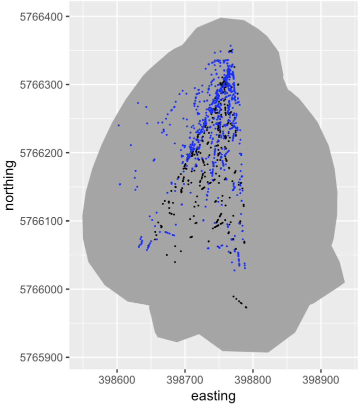
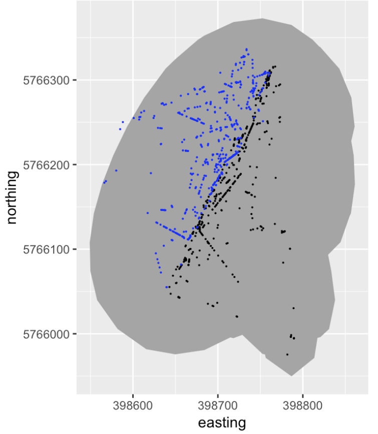
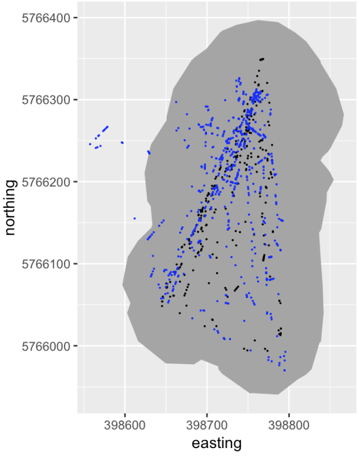
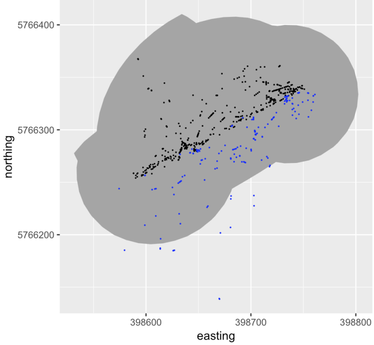
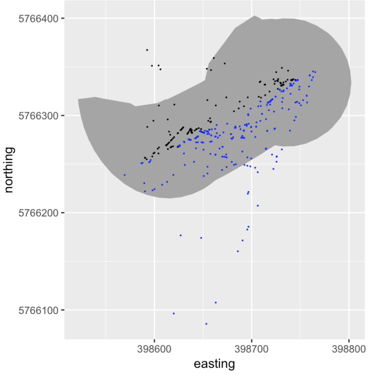
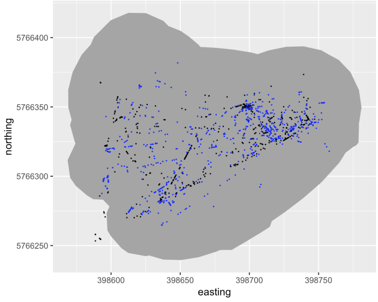
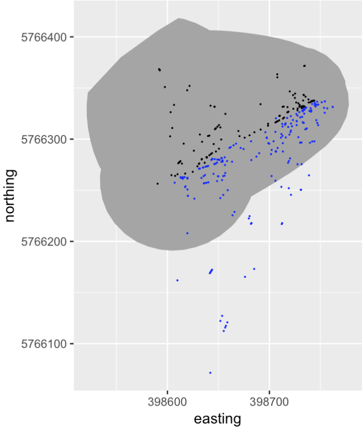
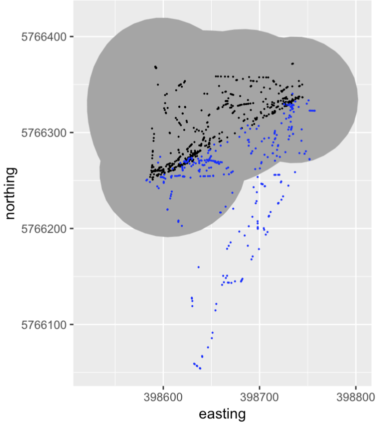
D5051

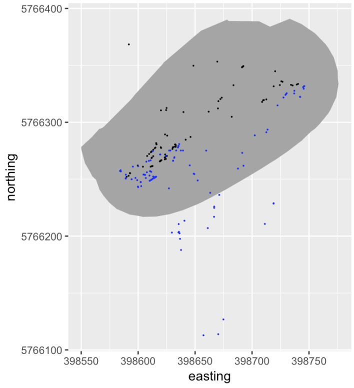


D5052

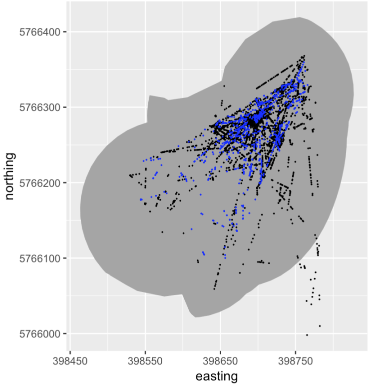


D5053

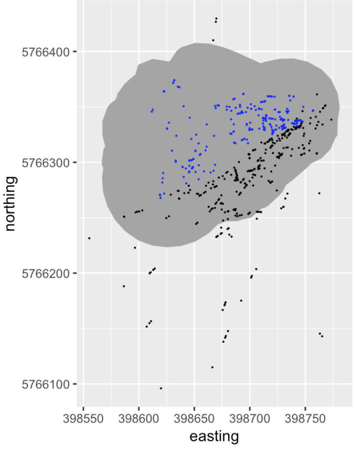
		
D5055	D5056	D5058
		
D5059	D5152	D5153
		
D5155	D5156	D5158



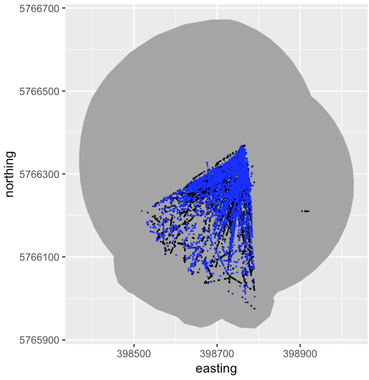
D5159



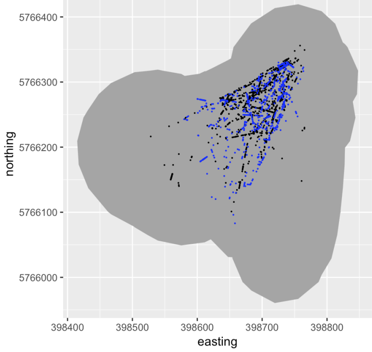
D5352



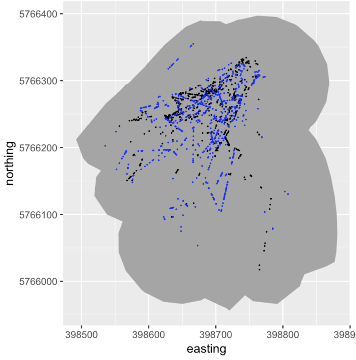
D5355



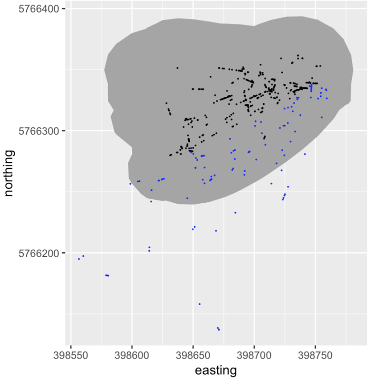
D5356



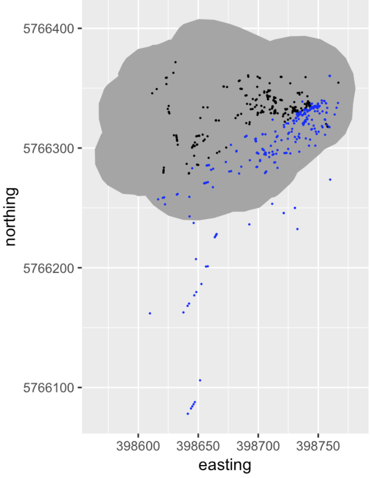
D5358



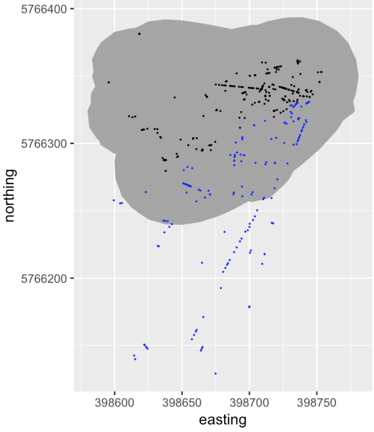
D5359



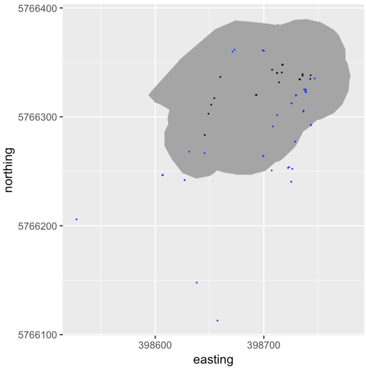
D5552



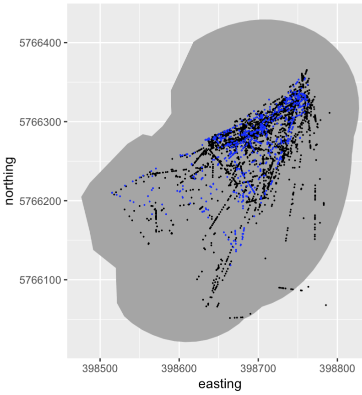
D5556



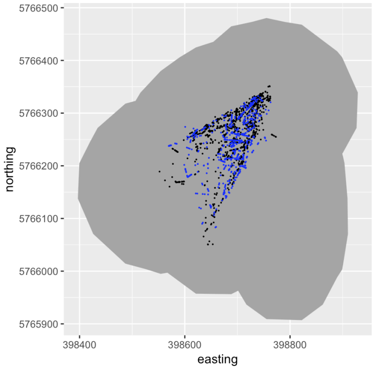
D5558



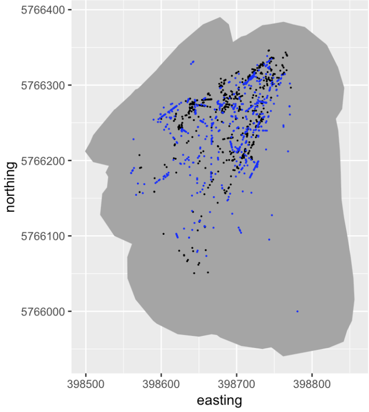
D5559



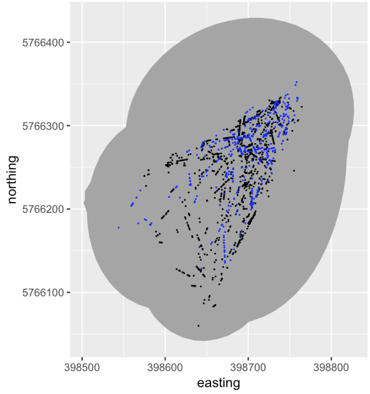
D5652



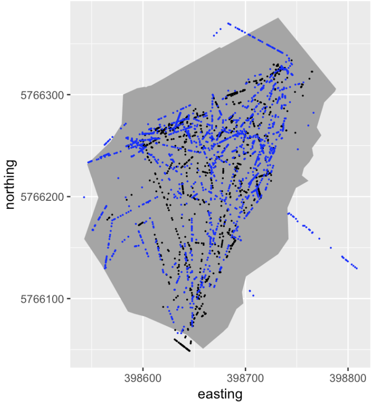
D5658



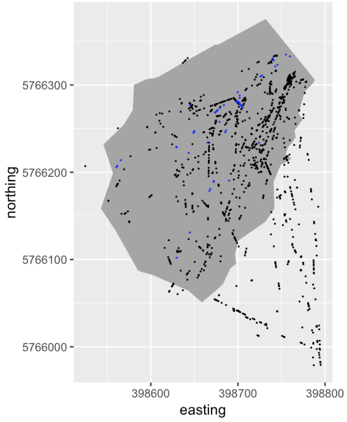
D5659



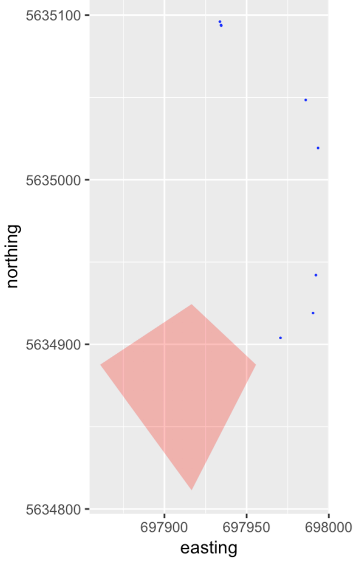
D5852



D5859



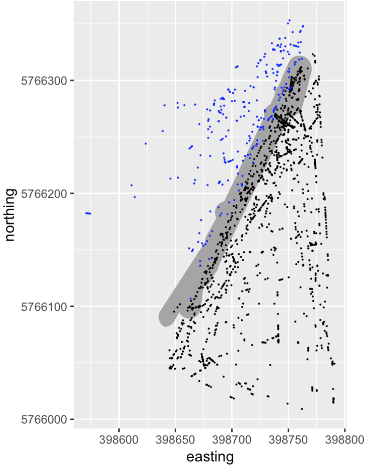
D5952



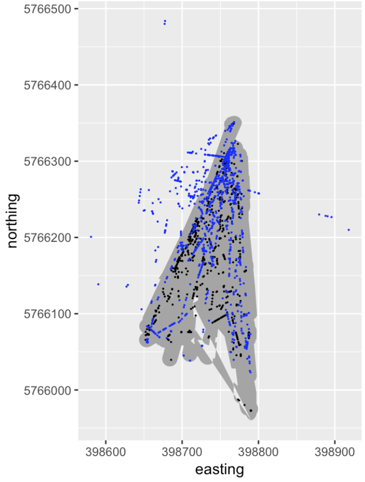
M030-F038

A4 Buffer 10m

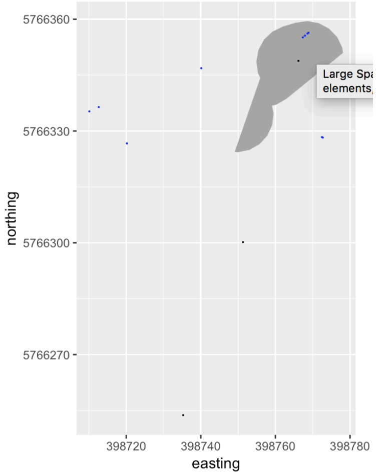
Buffer 10m
The grey area corresponds to the interaction area of the two animals and the measured contact data of first animal is in black and of the second in blue.



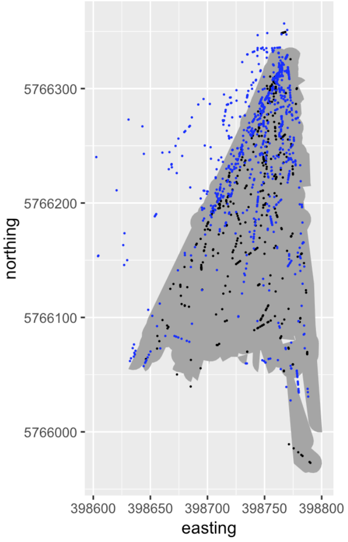
D5052



D5053

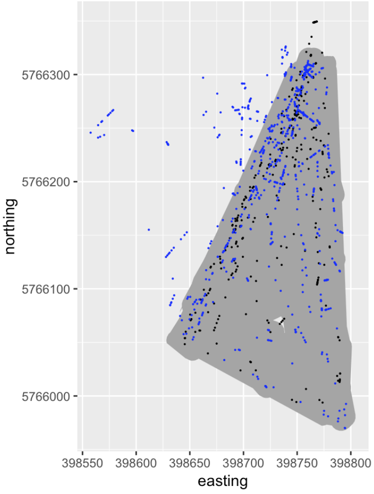


D5055

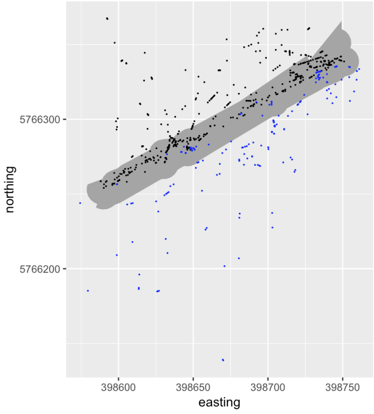


D5056

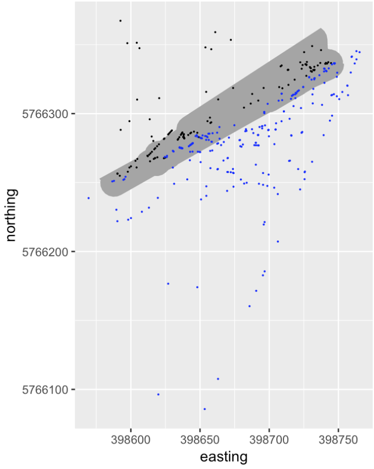
D5058



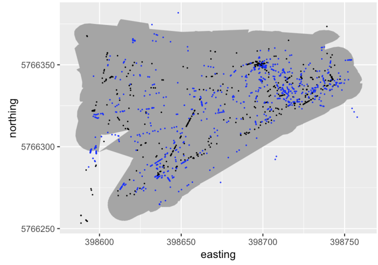
D5059



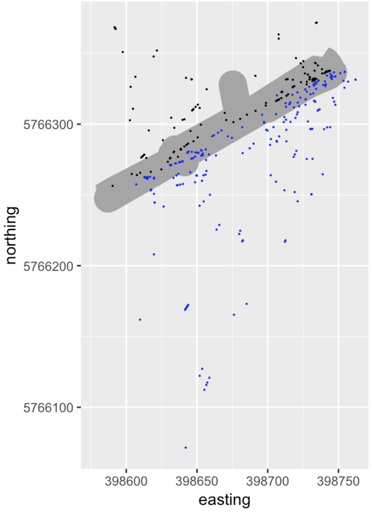
D5152



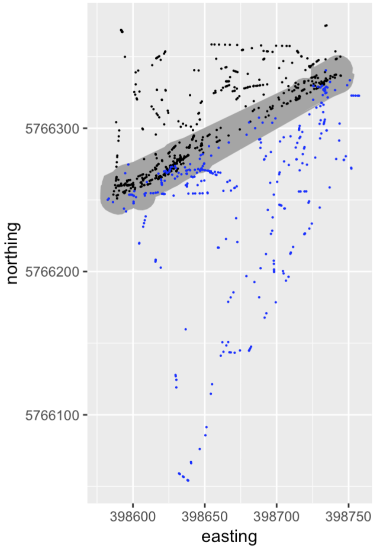
D5153



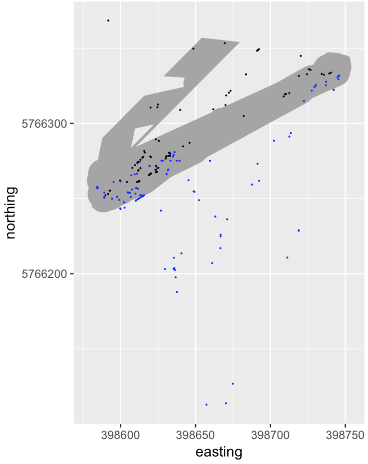
D5155



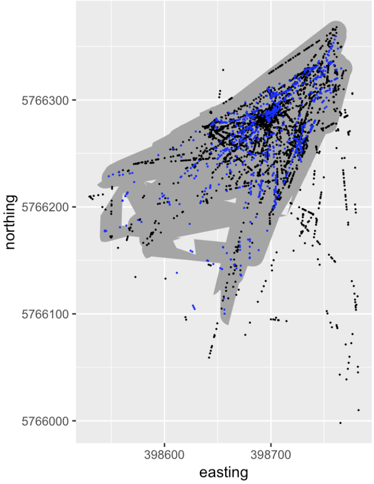
D5156



D5158

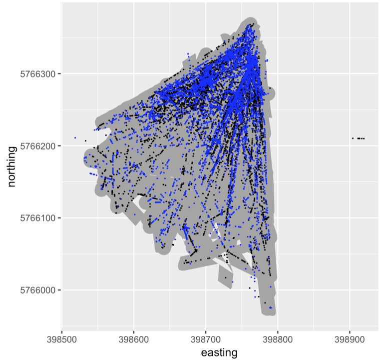


D5159

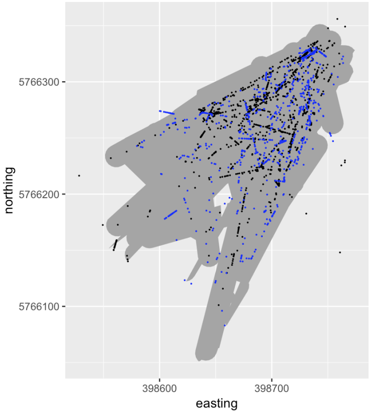


D5352

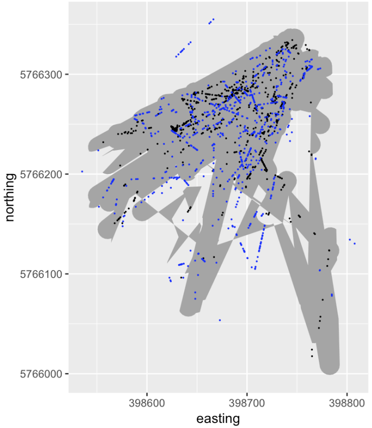
D5355



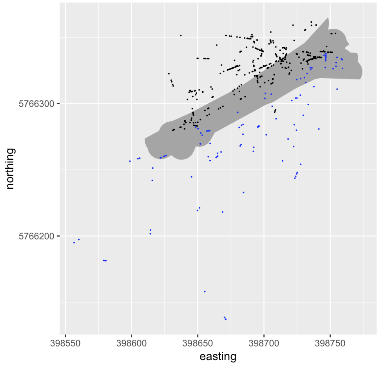
D5356



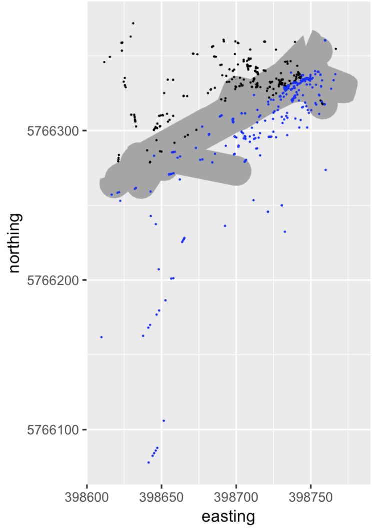
D5358



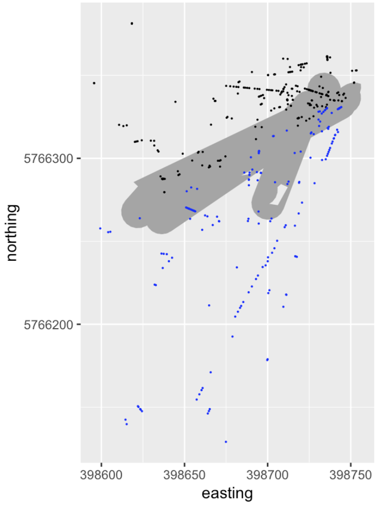
D5359



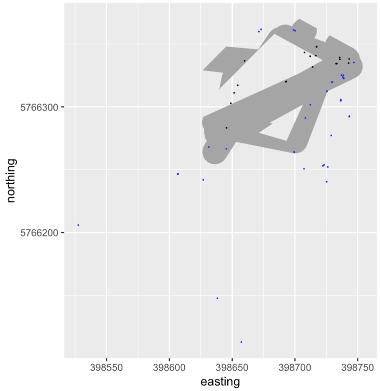
D5552



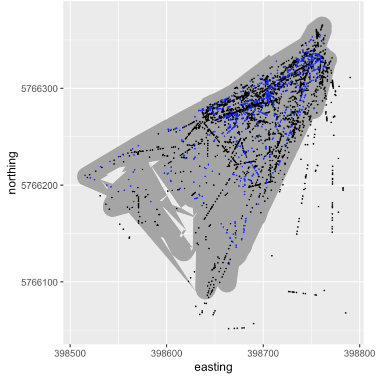
D5556



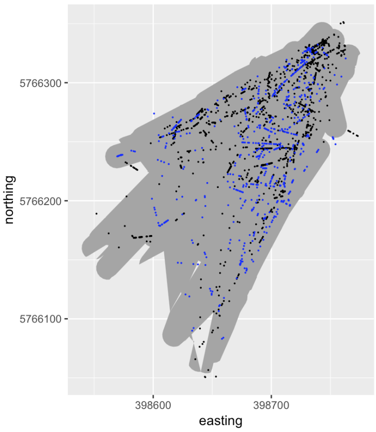
D5558



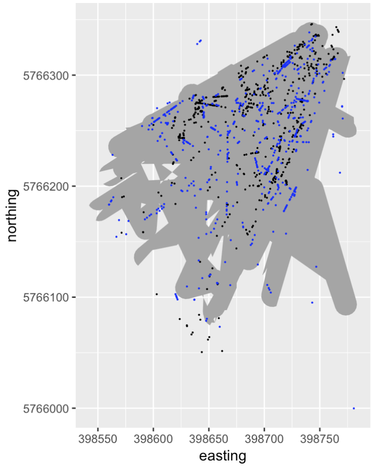
D5559



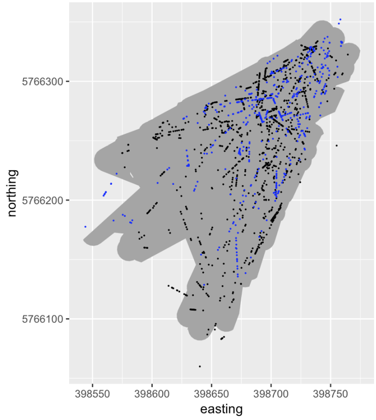
D5652



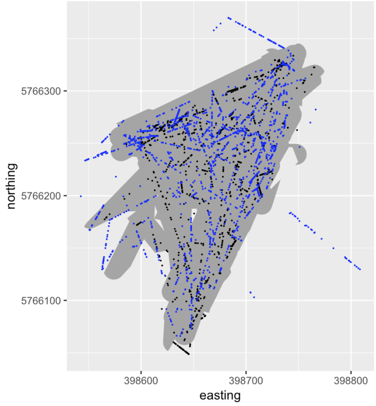
D5658



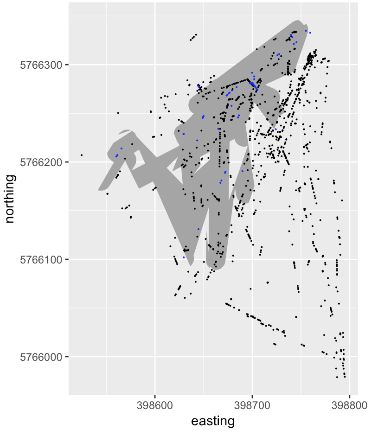
D5659



D5852



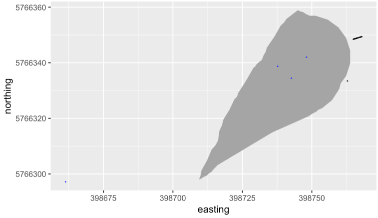
D5859

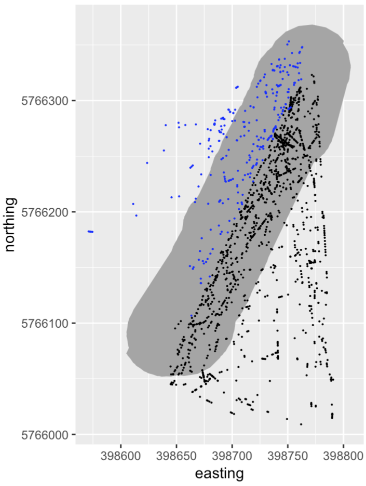


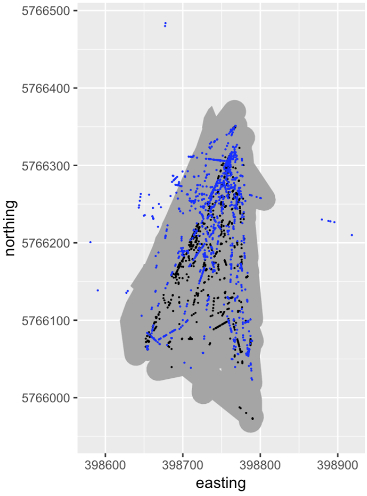
D5952	M030-F038

A5 Buffer 15 and 40m

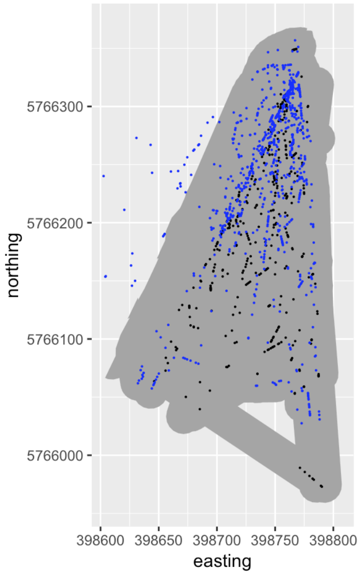
The grey area corresponds to the interaction area of the two animals and the measured contact data of first animal is in black and of the second in blue.

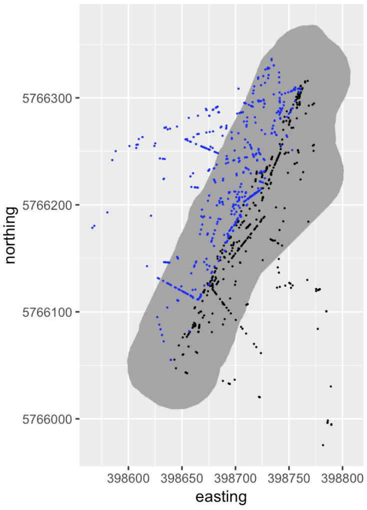




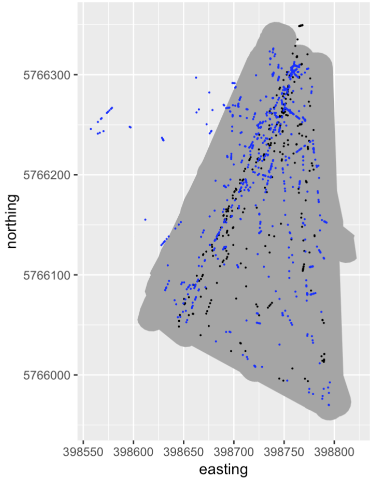


D5051	D5052	D5053

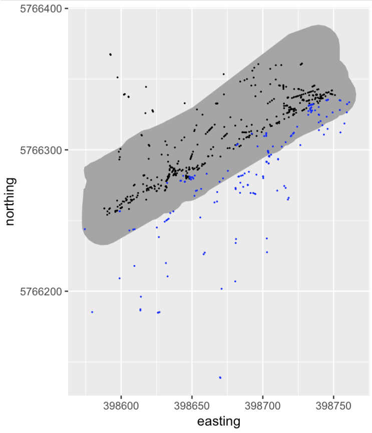




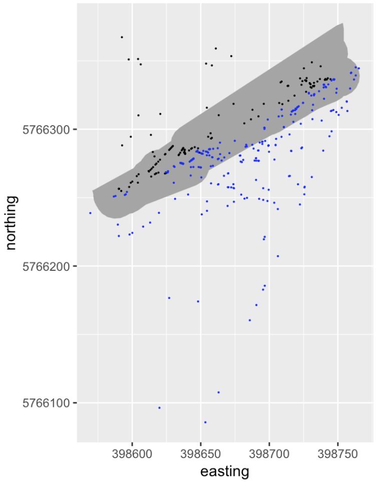
D5055



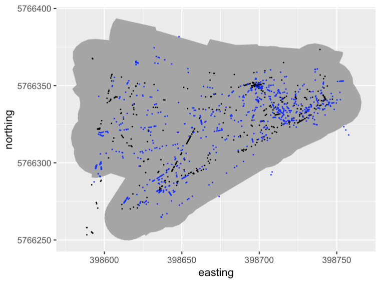
D5056



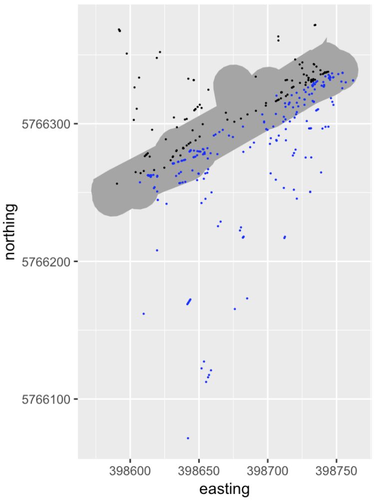
D5058



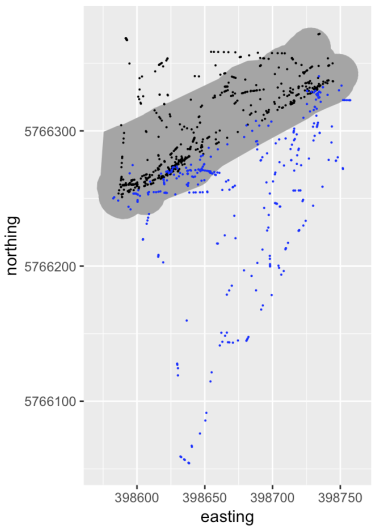
D5059



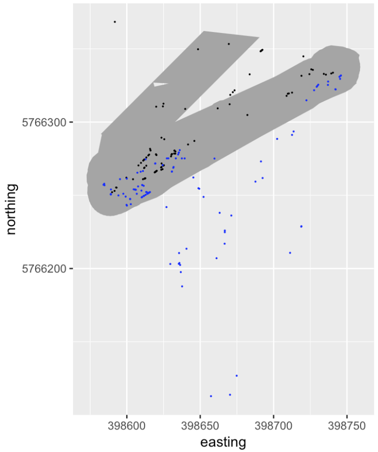
D5152



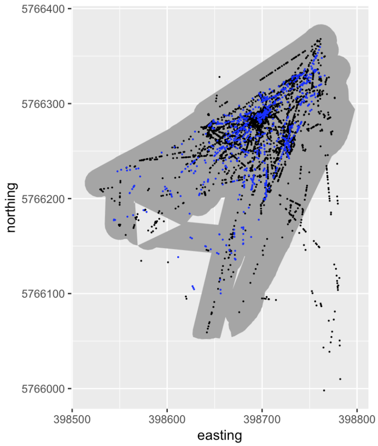
D5153



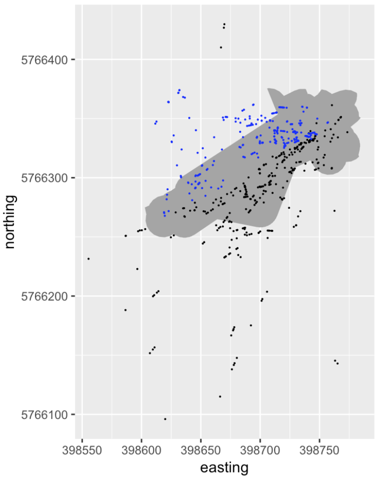
D5155



D5156



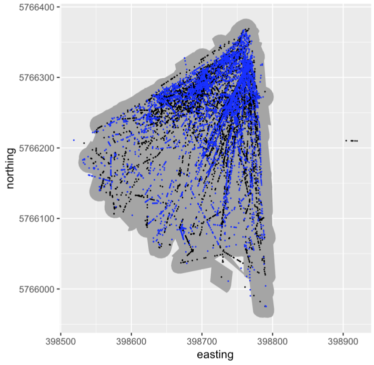
D5158



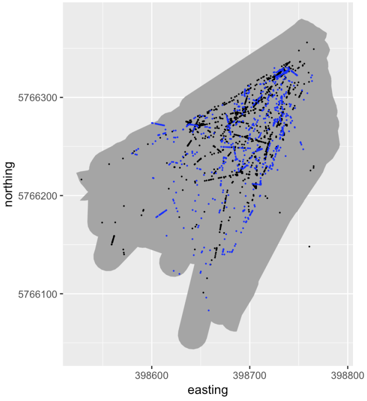
D5159

D5352

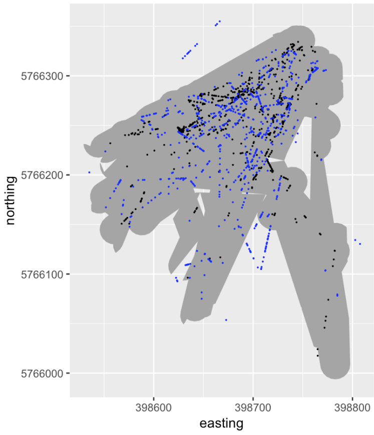
D5355



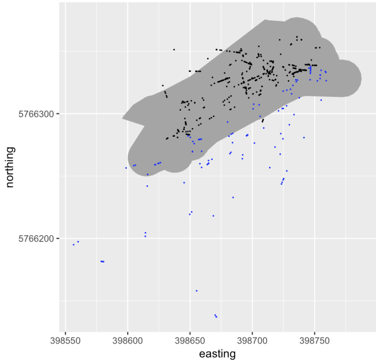
D5356



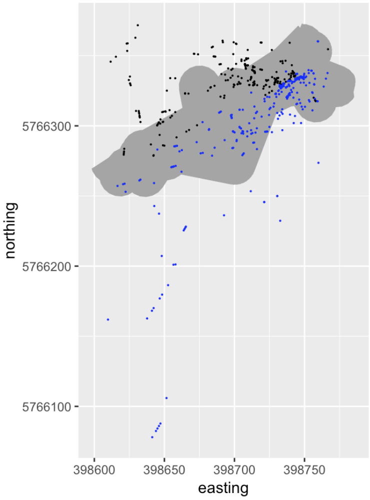
D5358



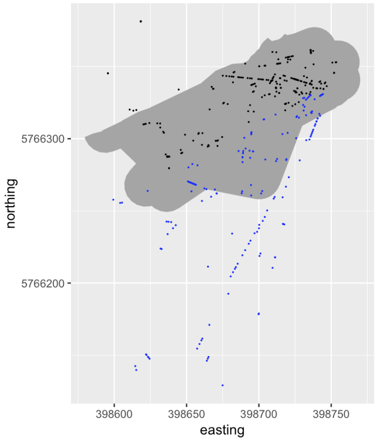
D5359



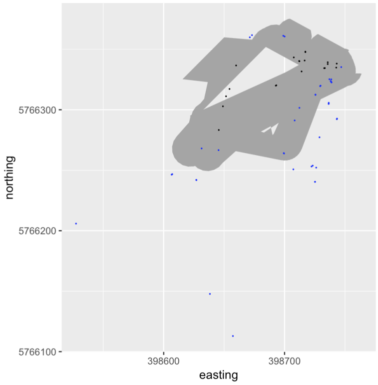
D5552



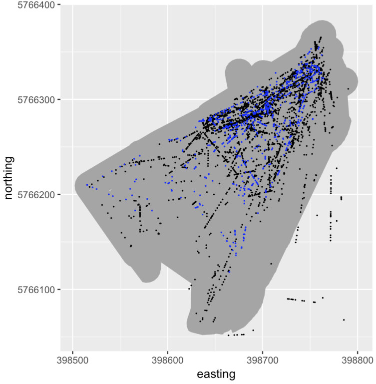
D5556



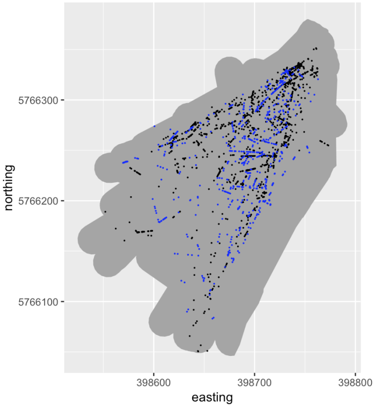
D5558



D5559

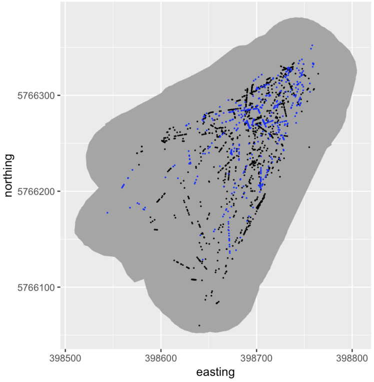


D5652

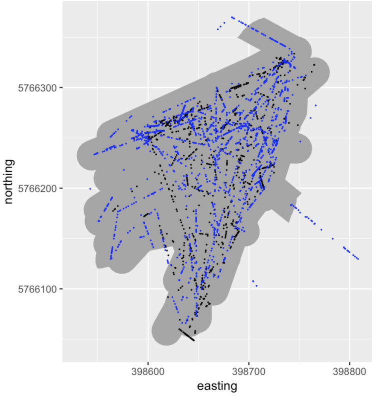


D5658

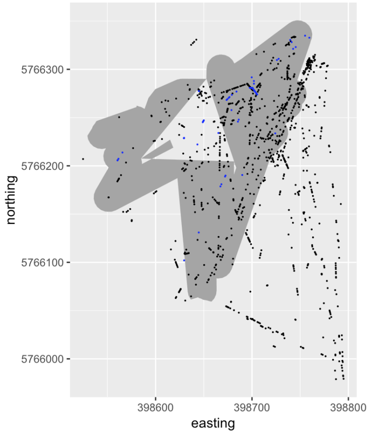
D5659



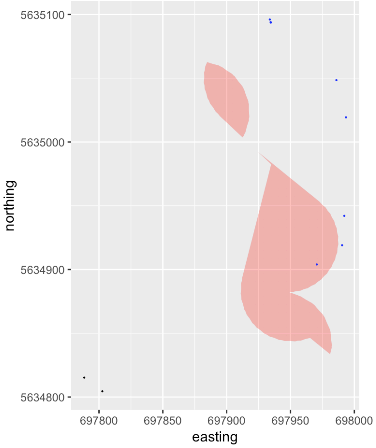
D5852



D5859



D5952



M030-F038 (40m)

Personal declaration

Personal declaration: "I hereby declare that the submitted thesis is the result of my own and independent work. All external sources are explicitly acknowledged in the thesis."

Place and date

St. Gallen, 30.09.2018

Signature

Susanne Täschler

S. Täschler