Spatial Statistical Methods in the Reconstruction of Badger Territories

Martín Alfredo Legarreta González

Submitted for the degree of PhD

School of Mathematics and Statistics

March 2018

Supervisor: Prof. Paul G. Blackwell

The University of Sheffield

To:

my dad and mom from Heaven.

My sons: Yhoshua, Luis and Christian, my wife Rosy,

my mom Martha, my brothers Pepe, Jorge, Jaime, Lupita and César

and specially to my dad José Luis †

Acknowledgements

My sponsors: Consejo Nacional de Ciencia y Tecnología, Universidad Autónoma de Chihuahua, Secretaría de Educación Pública.

My supervisor: Prof. Paul G. Blackwell.

University of Sheffield.

Food and Environmental Research Agency.

People who helped me and my family before and during my course: Paul and Jacqui, Cristina Legarreta, Mónica Legarreta, Jorge Gómez, Saúl Hernández and family, Javier Hermosillo, Guillermo Hermosillo, Francisco Márquez, Armando Segovia, Paniel Reyes, Sawaporn Siripanthana, Fernanda Vidal and Jesús Balcorta.

To my friends: Jesús González, Miguel Juárez and Jorge Puebla.

Contents

List of Tables

List of Figures

Abbreviations

AIC: Akaike Information Criterion

AOM: Adjusted Ordinal Model

ANOVA: Analysis of Variance

COPM: Conditional Outlier Prediction Model

CSnl: Conditional Sum of neighbouring latrines

CSDnl: Conditional Squared Difference of neighbouring latrines

CSRF: Corrected Scale Reduction Factor

EOPM: Elstub Outlier Prediction Model

FERA: Food and Environmental Research Agency

GLMs: Generalized Linear Models

gpclib: General polygon clipping routines for R based on Alan Murta's C library

GPS: Global Positioning System

Ha: Hectares

LRM: Logistic Regression Model

MASS: Functions and datasets to support Venables and Ripley, 'Modern Applied Statistics with S'

MCMC: Markov chain Monte Carlo

MCP: Minimum Convex Polygon

MDnl: Mean of Distances of Neighbouring Latrines

mgcv: Mixed GAM Computation Vehicle

MPSRF: Multivariate Potential Scale Reduction Factor

OPM: Outlier Prediction Model

PSRF: Potential Scale Reduction Factor

RDH: Resource Dispersion Hypothesis

UD: Utilization Distribution

UOPM: Unconditional Outlier Prediction Model

UOM: Unadjusted Ordinal Model

VHF: Very High Frequency

WP: Woodchester Park

Chapter 1

Introduction

For decades, Eurasian badgers *(Meles meles)* have been the object of several studies trying to explain their primitive social organization, feeding, territory and lately, their relationship with bovine tuberculosis which can cost £1bn over the next 10 years. Badgers spend the day sleeping in their setts and foraging during the night. They live in clans, sharing and defending a communal territory but foraging and feeding individually. Several attempts to explain what influences the size and shape of badgers' territories have been made, considering, for example whether they are determined by the dispersion of resources or by the location of the main sett which consists of several holes with large spoil heaps and obvious paths emanating from and between sett entrances. Since badgers use communal latrines to mark their territories, another approach is to use statistical methods based on this information to delineate their territories.

A common method employed to reconstruct badger territories from latrines is the Minimum Convex Polygon (MCP), another approach classifies the latrines as hinterland, boundary or extraterritorial excursions, based on elements surrounding them (fences, badgers paths, etc.). The use of extra information such as the presence of other latrines in the same direction from the main sett, can provide more robust models that can be used not only in a point estimation approach but in a sampling approach that uses the probability distribution fitted by a model and permits to quantify the uncertainty in the reconstruction of the territories.

This thesis consists of 7 chapters: Chapter 1 is this Introduction. Chapter 2 is the literature review which looks at badger ecological behaviour, techniques used to obtain information about the territories of the badgers and methods used to reconstruct them. Chapter 3 is the Unconditional Outlier Prediction Model (UOPM). It is an extension of an unpublished paper that uses a logistic regression to estimate the probability that a latrine is part of the territory or, alternatively, is an extraterritorial excursion. This information is used with the 100% MCP to make the reconstructions. Chapter 4 talks about the Conditional Outlier Prediction Model (COPM), which is an extension to the UOPM that uses Gibbs sampling in the reconstruction of the territories to allow for dependance between latrines. This model uses the 100% MCP of the sampled latrines which are not outliers, in order to map the reconstructed territory at each iteration. Chapter 5 presents the Unadjusted Ordinal Model (UOM). This model uses the original classification from the 2010 baitmarking Woodchester Park Badger Survey made by the Food and Environmental Research Agency (FERA), applying a cumulative ordinal model to estimate the probability distribution fitted by the model using a sampling approach to reconstruct the territories. All the previous chapters employ information from only the territory being reconstructed; Chapter 6 adjusts the probabilities obtained by a territory and the territories sharing at least one latrine with it, to reconstruct the territory using a sampling approach. The last chapter discusses the results of the methods proposed in this research.

Chapter 2

Literature Review

2.1 Introduction

This literature review introduces badger ecology and behaviour in the context of the ecological question of how to define territories and home ranges and the applied problem of bovine TB, which is thought to be transmitted by badgers and is a function of their habitat use.

2.2 Badgers

2.2.1 Group size

Badgers are nocturnal animals with an elusive life style, this characteristic and the difficulty of identifying individuals within a population, makes large scale direct counts impractical. Trying to count the number of badgers whilst they are emerging from their setts has the disadvantage of underestimating the population size, due to: the degree of error being different from season to season for reasons such as the surveyor and the shyness of the clan under observation (Macdonald et al., 1998), for example. Direct (Smal, 1993; Wilson et al., 1997) and indirect secondary indices of abundance, notably sett density (Macdonald et al., 1996; Thornton, 1988) and quantifying latrine use (Tuyttens et al., 2001) have been used to estimate the size of a clan. Sett surveys often rely on the differentiation of main setts from the other types of setts (Wilson et al., 1997): "*Annexe*, Normally less than 150m from main sett, comprising several holes. May not be in use all the time, even if main sett is very active; *Subsidiary*, Usually at least 50m from main sett with no obvious paths connecting to other setts. May only be used intermittently; and *Outlier* Little spoil outside holes. No obvious paths connecting to other setts and only used sporadically. May be used by foxes and rabbits" (Scottish Natural Heritage, 2003).

Badger abundance is then estimated as the number of main setts multiplied by the average size of badger social groups derived from trapping studies. This method assumes that there is one main sett per social group and that surveyors are able to find and classify correctly all main setts in an area. Both assumptions may be invalid (Tuyttens et al., 2000). Moreover, this method will work only in areas where the relationship between sett density and badger density is adequately known. This relationship may differ not only in space but also in time for reasons such as food availability or migration (usually by females).

An unnatural mechanism of territorial variation can be the culling of badgers, for example as part of the UK government's policy to control the transmission of bovine tuberculosis from badgers to cattle. This may substantially reduce badger abundance without altering the number of setts in the short to medium term (Tuyttens et al., 2000) which is a good example of unknown and unnatural variation of group size. Furthermore, given the considerable variation in the size of badger social groups (ranging from one individual to more than 30), this method is only useful for estimating badger abundance over large areas. These misgivings support the conclusion that there appears to be no reliable and cost-effective badger census technique at present that does not involve capture (Macdonald et al., 1998).

Indeed, the abundance of badgers in small intensively studied populations is usually estimated from mark-recapture studies (Rogers et al., 1997; Tuyttens et al., 1999). These methods are believed to be reliable if the underlying assumptions are valid, if variations in capture probability are adequately controlled, and if an adequate proportion of the badgers is trapped over a sufficient number of trapping occasions (Tuyttens et al., 2000; Tuyttens et al., 1999). However, there are some disadvantages for live trapping of badgers such as: costs, labour-intensive, in many countries a special permit is needed since it is illegal and potentially disturbing to the badger population.

Faecal counts have frequently been used as a measure of animal abundance (Krebs et al., 1987; Neff, 1968; Plumptre, 2000; Putman, 1984; Sutherland, 1996). Badgers use clusters of dung pits to defecate called 'latrines' which are not very difficult to find and their faeces can be distinguished from other species easily. Some people who work in this research area have realized that counting these latrines could be a good way to estimate badger density (Brown, 1993; Hutchings and Harris, 1999). Palphramand (Palphramand et al., 2007) used a combination of two direct methods: a) counts of the numbers of animals trapped at each sett and b) direct observations from consecutive nights and the Tuyttens indirect method which consists on counting the number of latrines, or preferably the number of separate dung pits, which were known from bait-marking to be used by members of a social group (Tuyttens et al., 2001). Thence, they adjusted their estimates to include radio-collared badgers (see below) that were known to be resting in the sett but not observed in that particular night.

2.2.2 Habitat distribution and quality within territories

The Resource Dispersion Hypothesis (RDH), introduced by MacDonald (Macdonald, 1983) and reviewed by Johnson (Johnson et al., 2002) is a hypothetical model which postulates that if unpredictable resources (food or mates) are distributed in discrete aggregations (patches) in space or time, an animal will have to increase its territory to encompass the variability of these resources and ensure that at least one resource aggregation will be available to satisfy its requirements. If these defended aggregations are also rich (high prey or mate density and large patch size), the territory is predicted to have an excess of resources some or all of the time. Large resource patchiness causes territories of multiple animals to overlap, and high patch richness simultaneously permits these additional residents to be supported. In other words, the RDH states that badger social groups can develop where resources are dispersed such that the smallest economically defensible territory to supply year-round needs for a pair of badgers can also sustain additional animals at no net cost to the original pair. The RDH thus favours group living without requiring active cooperation or relatedness within groups. Therefore, the RDH is a testable hypothesis for it makes very clear predictions (Macdonald, 1983; Johnson et al., 2002):

1 If the number of animals per territory is increased, the group size will be increased and the territory size will be the same but it will be increased if the change is extreme;

2 if the Total resource abundance is increased, the group size will be increased and the territory size will be decreased unless the resources are still dispersed;

3 territory size and group size increase as resource aggregations become more clumped in space and time;

4 group size increases with an overall increase in patch richness; and

5 if the variability in patch availability is increased (heterogeneity) the group size is increased (providing overall abundance is the same), although the territory size will be increased. (Johnson et al., 2002).

Palphramand tested the applicability of the RDH to a population of badgers, examining territory size and group size in relation to food patch quality and dispersion using graphical plots (Palphramand et al., 2007). They identified five habitat categories within the study area (coniferous woodland, broad leaved woodland, mixed woodland, scrub and grassland) and mapped the habitat composition of the study area in a GIS (ArcView 3.3: ESRI 1992). Patches were defined as areas of a specific habitat over 0.05 ha in size, which were spatially separated from patches of the same biotope by at least 20m, as these seem to be distinguishable by badgers when foraging. They calculated the number and size of habitat patches within the territories and compared these against the territory sizes and group sizes of the different social groups. Using the same method as Broseth (Broseth et al., 1997) to measure food patch dispersion, they determined the overall mean distance from the perimeter of each type of habitat patch to the perimeter of all other patches of the same habitat type within each territory. They also measured the mean distance from the main sett to patches of each habitat type within each territory. They found that seasonal patterns in homes ranges were consistent with the RDH, and food appeared to be more important than mating opportunities as a limiting resource.

2.2.3 Demarcation and Defence

Badgers defend the boundaries of their territories not only with direct aggression; they also use several types of visible signs to mark them. These are paths, latrines, and tussocks used for 'squat-marking'. These are very important in high density areas or where the main sett is close to a boundary (Kruuk, 1978b; Kruuk, 1989; Kruuk and Dekock, 1981; Kruuk et al., 1984). Badgers form paths simply by repeatedly using exactly the same route. These paths are often very conspicuous in areas where territory border crosses large tracts of lands (i.e. open woodlands or fields without landmarks such as fences or roads associated with boundaries). Paths occur on the border between ranges, but they are also associated with setts (Kruuk, 1978b; Kruuk, 1989; Kruuk and Dekock, 1981; Kruuk et al., 1984). Squat-marking is associated with territorial behaviour; however, the most important signs of badger territory boundaries are the latrines. These are aggregations of up to 60 defecations, in small pits dug by the badgers themselves. Pits are usually 5 to 10cm deep, though sometimes as much 30cm, and each contains one or several droppings, or often nothing. Usually there are signs of scratching of the ground around the latrine. Latrines are used at all times of the year, but especially in spring with a second lower peak in October. Many latrines are used by badgers for several years (Kruuk, 1978b; Kruuk, 1989; Kruuk and Dekock, 1981; Kruuk et al., 1984). Most latrines are on the territorial boundary and those that are more inside the territory are smaller. Also very striking is the position of the latrines close to some conspicuous landmark (a fence, a track, or a woodland edge). The landmarks reinforce the striking pattern of territorial boundaries, of home ranges which are defended by all possible means, and with a large input of energy, against the neighbours (Delahay et al., 2000; Delahay et al., 2006a; Delahay et al., 2006b; Palphramand et al., 2007; Palphramand and White, 2007).

2.2.4 Diet and Foraging

Diet and feeding strategies of the European badger show great variation across its range (Fischer et al., 2005; Kruuk and Parish, 1981; Neal and Cheeseman, 1996; Roper, 1994; Rosalino et al., 2005). Due to this diversity of feeding strategies, badgers have been able to colonize a wide range of habitats across Europe (Fischer et al., 2005; Kruuk, 1978a; Kruuk and Parish, 1981; Martín et al., 1995; Roper, 1994; Rosalino et al., 2005; Shepherdson et al., 1990). The badgers' diet in Europe is related to latitude: in northern latitudes badgers usually eat earthworms (mainly *Lumbricus terrestris*) and in southern latitudes fruits and insects (Kruuk, 1978a). According to dietary studies, badgers are generalist feeders (Neal and Cheeseman, 1996; Rosalino et al., 2005) and, depending on environmental conditions and habitat type, they have the capacity to survive on different resources (Roper, 1994).

Some studies in Continental Europe about badgers' strategies and foraging behaviour showed that they are generalist foragers (Pigozzi, 2010; Roper, 1994) and their diet fluctuated with seasonal availability of *Lamellicornia* (Fam. *Scarabaeidae*) larvae. Badgers have a preference for these larvae, consuming them when abundant (February to May) (Pigozzi, 2010). In Doñana National Park (Spain) the badgers' preferred prey were rabbit kittens (*Oryctolagus cuniculus*), but secondary prey were consumed in smaller amounts compensating for temporal fluctuation in the availability of the preferred prey (Martín et al., 1995). In southwest Portugal, badgers showed seasonal specialisation on olives and consumed other food types, such as arthropods and fruits, when olives were not available (Kruuk and Dekock, 1981; Rosalino et al., 2005).

However, in contrast to the view that badgers are generalists, Kruuk et al (Kruuk and Parish, 1981) showed that badgers can be thought of as 'earthworm specialists' as *Lumbricidae* dominate the diet of badgers over a wide range of habitats, from Scotland (Kruuk and Parish, 1981) through the forests and mountains of central Europe to supra-Mediterranean habitats in Madrid (Virgós et al., 2004). Badgers in these studies changed their foraging efforts to compensate for fluctuations in earthworm availability but intake of earthworms was largely independent of availability. Consequently, the proportion of earthworms in the diet was relatively constant, with little seasonal variation. Virgós et al reached a different conclusion and viewed badgers as 'facultative specialist', searching preferentially for earthworms, but taking other food resources as they were encountered. This had the effect that in times of low earthworm availability, for example during summer droughts, other foods predominated in the diet (Virgós et al., 2004).

Whatever the term used to describe the badgers' diet and foraging behaviours, a general observation is that their diet is variable and they appear to forage opportunistically on whatever food types are most abundant (Fischer et al., 2005; Neal and Cheeseman, 1996; Roper, 1994; Roper and Lups, 1995; Shepherdson et al., 1990). In England, *Lumbricidae* were found to be a large proportion of the diet in all seasons except summer, when they increased their consumption of cereals (Shepherdson et al., 1990). In Switzerland, *Lumbricidae* were taken at a high frequency but in low volumes throughout the year, but seasonally they consumed large volumes of wasps, cherries, plums and oats (Roper and Lups, 1995). In the Swiss Jura Mountains, mammals and cereals dominated the diet, while in mid-mountain and lowland areas, maize dominated the diet (Roper and Lups, 1995). Earthworms were only of secondary importance in the diet in mountainous areas and were negligible in the mid-elevation and lowland areas. Roper (Roper, 1994) suggested that the simplest explanation of the badgers' diet is that consumption of different foods is determined primarily by their relative profitability. That is, badgers consume whatever is more common or available.

2.3 Experimental techniques

2.3.1 Bait-marking

One of the techniques used for determining the territorial configuration of social groups of the European Badger is bait-marking, which has applications in both ecological research and applied wildlife management problems. The basic idea is to feed plastic markers to the badgers, and then record where the markers are recovered, after excretion. Badgers mark their territorial boundaries with communal latrines, so this method is particularly suited to them. Counting marked droppings may be of limited value in quantifying defecation rates and latrine use, as data from sequential visits to latrines, show substantial short-term variation in the number of marked droppings counted at a given latrines. However, bait-marking data do allow the estimation of territorial boundaries between badgers' groups. The description here is based on the review by Delahay (Delahay et al., 2000). Table 2.1, reproduced from Delahay et al (Delahay et al., 2000) shows the advantages and weaknesses of bait-marking. This technique consists of the following steps:

1 **Timing and planning**

Bait marking is generally conducted during the annual peak of territorial activity, February, March and April (Kruuk, 1978b). Before conducting this activity, the study area is thoroughly searched for badger setts (Harris, 1984). Only main setts should be selected for feeding, each with a unique type of plastic marker, during at least 10 consecutive days, completing the latrine survey as quickly as possible after this.

2 **The bait**

It usually consists of peanuts and plastic beads in a matrix of golden syrup. Some suggestions about characteristics of the beads are: food grade plastic, approximately 2mm size, avoiding dark colours (difficult to see in droppings, soil and leaf litter) and white ones (in calcareous or 'chalky' areas). It's important to ensure that beads fed to adjacent social groups are clearly distinguishable in appearance. The ratio of beads to peanuts should be approximately 15:1:2, peanuts:beads:syrup.

3 **Feeding the bait**

The active main setts to be fed are visited on a daily basis in late afternoon to limit exposure of the bait to other animals. On each visit, around 25-30 bait points are created, scattered around the set at distances of about 15-20m. Each bait point consists of about 50-100ml of bait, placed in a small depression made using the heel of the foot, covered with a large stone.

4 **Latrine survey**

The survey itself consists of visiting as many of the latrines as possible, and recording the numbers and types of beads found at each one. It is necessary to

Strength	Weaknesses
Can yield quick and accurate estimates of territorial configuration	Skilled staff required for best results
Provides objective data	Subjective judgment required for interpretation
Not labour intensive or time consuming compared to other methods that would yield similar quality data	Can be difficult at any time other than spring or autumn
Requires no specialist equipment and is therefore relatively inexpensive	Less effective at lower badger densities and where territorial systems are not clearly defined
Marked bait is harmless and the technique is non-invasive	Possible adverse effects of artificially feeding badgers
Technically simple in execution	Accurate delineation of a single social group territory also requires feeding bait at adjacent setts

Table 2.1: Strengths and weaknesses of the bait-marking technique for the determination of badger social group territories

rake up the ground and to break apart individual droppings, to ensure that beads are found if present.

5 **Mapping territories**

After the survey is finished, each latrine position and the origins of its coloured markers are displayed on a map of the study area, connecting each latrine with a line to the main sett at which those markers were fed. Latrines where more than one type of marker is found, known as shared latrines, are shown as linked to both the corresponding setts. These points are the ‘raw data’ for reconstructing the badgers’ territories; possible methods are described in section 2.3.3.

Woodchester Park Data

An example of the application of bait-marking technique, mention above, in the reconstruction of badger territories, is the data used to estimate the models on this thesis which consists of the year 2010 bait-marking results from Woodchester Park, where 481 latrines were surveyed from 23 baited setts, along with boundary reconstructions by experts at the Food and Environmental Research Agency (FERA) (Delahay et al., 2000) shown on Figure 2.1 where the stars represent setts, the red dots latrines and the lines boundaries between badger territories where convex and concave areas can be observed. Thus, for these particular latrines, it is known whether they are extraterritorial excursions (outliers), or part of the territory (boundary or hinterland latrines), allowing logistic or ordinal regression models to be fitted.

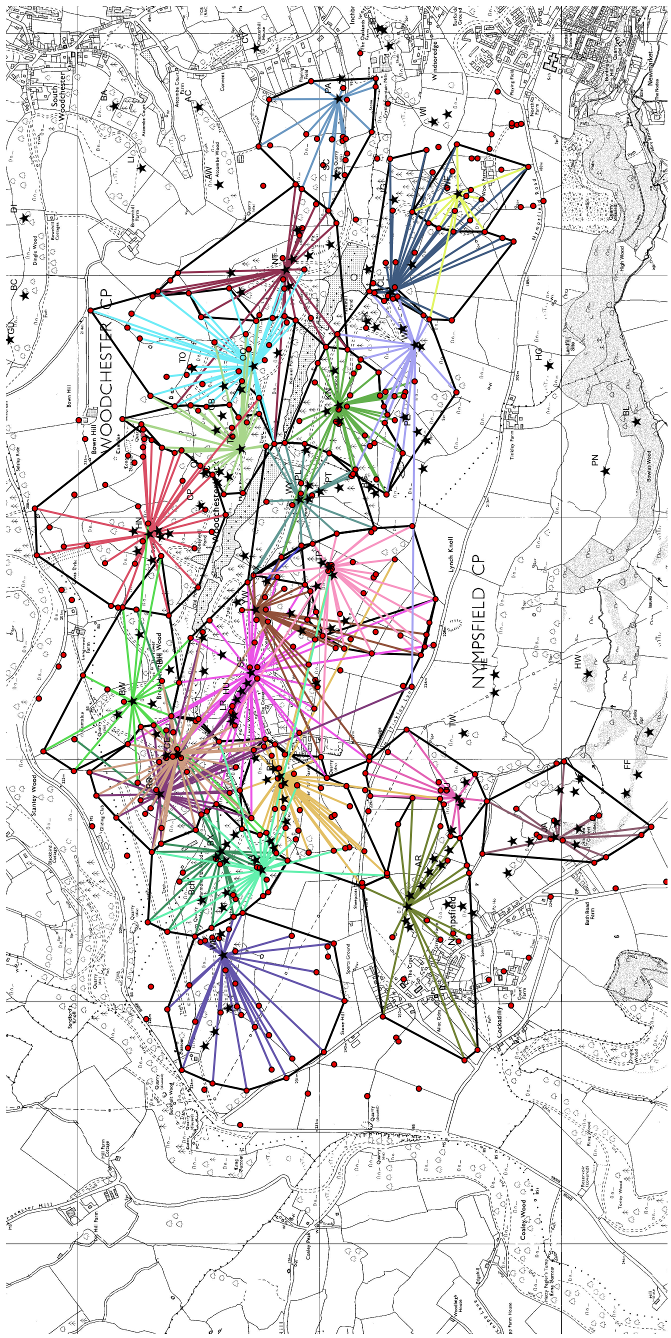


Figure 2.1: Map of Woodchester Park with the 2010 Bait-marking Survey Results. Stars represent baited setts, red dots latrines and lines boundaries between badger territories. Grid lines shown at 1 km intervals; region shown is approximately 5.5 km x 3 km.

2.3.2 Radio-tracking

Even though radio-tracking data were not used in this thesis, it is explained roughly since radio-tracking (or wildlife telemetry) is one of most widely used techniques to study wildlife. This technique consists of the utilization of radio signals to or from a device that is carried by an animal. This information can be sent or received through the atmosphere by radio waves (telemetry). Wildlife research has been benefited from this technique because it helps to identify individual animals and to locate them when desired. There are three types of radio-tracking systems:

1 Conventional Very High Frequency (VHF) radio-tracking systems consist of two parts, one attached to the animal (transmitting system) and a receiving one. Their range varies from 5-10km (ground to ground) to 15-25 km aerially.

2 The second type of radio-tracking (satellite telemetry) uses a package attached to the animal which sends a location signal of the animal to satellites and these send this signal back to receiving stations. Its greatest advantage is that far-ranging and elusive species can be tracked from the office of the researcher but it's less accurate than VHF and Global Positioning System (GPS) radio-tracking.

3 The third one is GPS radio-tracking which uses a package attached to the animal to receive information from a special set of satellites to record and calculate, at programmed intervals, the date, time and animal's location. This information can be stored in the telemetry unit for remote downloading or later retrieval. It has a potentially accuracy of 5m. However, these units are more expensive and they do not last as long as VHF units (Mech and Barber, 2002).

2.3.3 Delineating Badger Territories from Bait-Marking Data

Estimation of overall territories

One approach to delineating badger territories is considering several territories at once. One example of this is the work of Blackwell and McDonald (Blackwell and Macdonald, 2000) who delineated badger territories using a model in which the latrines follow the edges of a Dirichlet tessellation. They fitted their model through a standard maximisation routine in which the Dirichlet borders were calculated using the iterative algorithm of Lee and Schachter (Lee and Schachter, 1980). Badgers' territories surrounded by others were well estimated; however, due to the incomplete information on territories near the edges of the study, the boundaries were undetermined in these ones. They showed that considering the badger sett as the centre of the territory does not improve the model and suggested that identifying the type of latrine (boundaries and hinterland) would improve it.

Estimation of single territories

- **Minimum Convex Polygon (MCP)**

  There are a number of possible approaches to reconstructing a map of an animal's territory or home range from point data such as latrine locations. Here we concentrate on polygon methods, in which the peripheral locations of an animal are joined by some connecting rule to determine the home range. The simplest and most popular method uses the MCP (Bearder and Martin, 1980; Schoener, 1981; Tinkle et al., 1962; Worton, 1987). A convex set is one with the property that, given any two points in the set, all the other points in the line segment joining them must also be in the set. In the case of a polygon, that is equivalent to the condition that none of the polygon internal angles exceeds 180°. The MCP is called minimum because it is the smallest convex polygon that contains all location points. That is, any other convex polygon (in fact any other convex set) that covers all the data must contain the MCP within it.

  A key difficulty is that the MCP tends to increase with the number of data points observed: adding a point can either extend the MCP or leave unchanged, but never reduce it. Summaries such as the area will generally be highly correlated with sample size. Also the particular MCP obtained will be very sensitive to a few 'extreme' values.

  Various ways have been proposed to overcome these problems. The most common approach involves omitting some of the points. This is analogous to calculating a robust summary of a 1-dimensional data set but 'trimming' or omitting a number of the smallest and largest values. However, in the two-dimensional case, where the term 'peeling' is commonly used instead of trimming, it is not quite so clear which points should be excluded.

  Schoener (Schoener, 1981) suggests removing those points furthest away from the centre of the data (defined as the location with coordinates given by the mean x-coordinate and the y-coordinate from the data), for example the furthest 10% or 20% of the data. Green (Green, 1981) discusses other ways to choose the points to 'peel' concentrating on 'convex hull peeling' in which the points that lie on the MCP itself are removed and the MCP then recalculated.

  If only the area occupied is of interest, which is not generally the case in this thesis, then an alternative approach is simply to 'correct' the area of the MCP by a factor that depends on the sample size. Jennrich and Turner (Jennrich and Turner, 1969) provide numerical values for these sample size factors, assuming that observations come from a bivariate normal distribution; Worton compares these with factors obtained from the uniform distribution on a disk (Worton, 1995).

  Another limitation of the MCP, is that it does not give any information about the way in which an animal spends its time within its home range (Utilization Distribution) –it is not the aim of this thesis to get a Utilization Distribution from latrine data–. A realistic Utilization Distribution can be very complex, and the sample size factors used to 'correct' estimates of area do not generally reflect this complexity and variability as (Schoener and Schoener, 1982) found when they compared the home range and its utilization of –bigger– male lizards with the –smaller– home range of female lizards. In general, correction factors based on unimodal or uniform distributions will be misleading when applied to very multimodal Uniform Distributions.

Mohr (Mohr, 1947) originally considered the case where animals could be observed only by being captured in traps at specific locations, generally arranged in a grid. He referred to the MCP of the grid points at which a particular animal was trapped as the 'minimum home range' of the animal.

He looked at the how the calculated areas were affected by the resolution of the grid and the details of how home ranges were derived from the data, including the way the boundary was adjusted for grid size, and showed that such issues may explain differences in published estimates of home range size although it is not the aim of this thesis to get a Utilization Distribution from latrine data.

A limitation of this method is it can not produce reconstructions of home ranges or territories with concave areas, only convex ones, even if they exist, it not takes on consideration neighbouring territories and it not uses badger behaviour on feeding nor number of uses of a latrine nor if it is shared by different badger clans.

- **Modelling latrine locations**

  Given the possibility of reconstructing territories from latrine locations, it is of interest to think about the relationship in the other direction too, modelling the types of locations that seem to be preferred for latrines, given knowledge of territory boundaries that were not defined only from the latrine data. Badger territories can be obtained by other signs such as landmarks, badger paths and other features as explained below. This data helps to estimate badger territories. Telemetry data is also used to estimate badgers home range.

  A study was carried out at Woodchester Park, England, using the bait-marking technique for latrine identification (Delahay et al., 2007) to investigate, in a high density badger population, the distribution of badger latrines in relation to habitat composition. They found that the density and the density of badger latrines had variation according with land use, with linear landscape features such as hedges and stone walls, and woodland. Distance from linear features was an important factor in the number of the latrines which decreased significantly. Another result was that grasslands had the highest number of latrines, this is important since may offer opportunities to manage exposure of cattle to badger latrines by habitat manipulation in farmland.

- **Elstub Outlier Prediction Model**

  An unpublished project by Elstub, supervised by T. Etherington, and carried out at the FERA, (Elstub, 2006) attempts to predict whether a given latrine is an outlier (an extraterritorial excursion), based on covariates that include properties of the latrine itself and on the configuration of the other latrines associated with the same clan. The aim is to use those predictions to improve the reconstruction of territories. Developing this approach is important for the subject matter of this PhD Thesis, so the method is described in detail here. Variations and improvements are described in next chapter.

The data used consist of bait-marking results from a site known as Woodchester Park, along with boundary reconstructions by experts at the FERA. Thus, for these particular latrines, it is known whether they are outliers, allowing Elstub to fit a Logistic Regression Model (LRM) to estimate the probability that a given latrine is an outlier based only on the data.

The variables obtained directly from the 1990-1996 bait-marking Woodchester Park surveys are:

a) Identification of the social groups using the baited setts $(sett_i)$;
b) x and y co-ordinates of each social group's sett (sx_i, sy_i);
c) the x and y co-ordinates of the latrine (x_i, y_i);
d) whether the latrine is shared (S_i) or not; and
e) number of returns to the latrine (R_i).

From the previous variables, the following variables were calculated:

a) Distance (d_i) is calculated using the co-ordinates of the baited sett (sx_i, sy_i) and the co-ordinates of the latrines (x_i, y_i) where the plastic pellets were recovered.
b) The 'Angular fit' $(Angfit_i)$ which is the ratio that compares the distance of one latrine from the sett against the distance between the sett and the other latrines in a similar direction. The way to calculate it is to divide the distance from the latrine to the sett by the mean of the distances from the other latrines to the sett that are within a specific circular sector or 'swath' starting at the sett and with a centreline running through the latrine in question. We write 2θ for the angular width of the swath. Elstub takes $2\theta = 45°$. In the calculation, only latrines at a distance of more than some threshold m from the sett were included, because, if a latrine has only latrines that are close to the sett in a similar direction, this would otherwise result in a very high angular fit for the latrine and may lead to incorrectly identifying it as an outlier, whereas such latrines are best thought of as 'central' rather than implying a boundary that excludes the latrine in question. Elstub takes $m = 40m$. The $Angfit$ is given for a $latrine_i$ by

$$Angfit_i = \frac{d_i}{\frac{1}{n_i}\sum_{j \in Se_i} d_j} \tag{2.1}$$

where d_i is the distance of the i^{th} latrine from the sett (Se_i) of latrines j such that $d_j > m$ and

$$a_i - \theta \leq a_j \leq a_i + \theta, \tag{2.2}$$

where a_i is the bearing from the sett (Se_i) to the $latrine_i$, and $n_i = S_i$ is the number of latrine in Se_i. If $n_i = 0$, that is there are no other latrines within the i^{th} swath, the angular fit for latrine i is defined as 1.

If non-circular home ranges are under consideration, this comparison is very important as it captures the idea that the boundary is likely to be relatively smooth, and so a latrine much further out than its 'angular' neighbours is likely to be an outlier.

c) Elstub used the variable Returns (R) as a design variable, because in some bait-marking surveys, this variable was a description whilst others it was the actual number. A 3 level design variable was created as follow:

$$R = \begin{cases} 1 \quad if \quad \rho = 1 \quad or \quad \rho = \text{`}few' \\ 2 \quad if \quad \rho = 2 \leq \rho \leq 3 \quad or \quad \rho = \text{`}some' \\ 3 \quad if \quad \rho \geq 4 \quad or \quad \rho = \text{`}many' \end{cases}$$

where R is the number or amount of returns of pellets from the same baited sett.

As R is a nominal variable with levels 'few', 'some', 'many'. One way to analyze it is using dummy variables for the LRM. Therefore, since R has three possible values, two dummy variables are needed (table 2.2).

Return level	R_1	R_2
1	0	0
2	1	0
3	0	1

Table 2.2: Coding for the Dummy Variables of Return Levels

d) The model also contains the independent variable S, whether the latrine is shared ($S = 1$) or not ($S = 0$).

- **Model fitting and variable selection**

  The LRM was fitted, using a stepwise algorithm, selecting models on the basis of the Akaike Information Criterion (AIC) (Akaike, 1978). The final model selected was:

$$g(x) = -7.04 + 1.44 \times S - 0.62 \times R_1 - 1.92 \times R_2 + 0.01 \times Dist + 0.65 \times Angfit \quad (2.3)$$

- **Using the predictions**

  Elstub's model (2.3) gives a probability of each latrine being an outlier. She considers two ways in which this information could be used to delineate territories. Firstly, she tries a 'hard' classification of each latrine, by choosing a threshold probability intended to minimize both false positives and false negatives. However, this ‘optimal’ threshold is found to be very low (approximately 0.043), and excluding all latrines with probability above this level would make mapping territories impossible. Instead, since the optimum cut-off is strongly dependent on prevalence, she reduced the distance of a latrine to the sett by a proportion $\pi(x)$ which is the probability (from the model) that a latrine x_i is an outlier. She compared predicted territories using this method with reconstructed territories using the 95% MCP method and actual territories from Woodchester Park and to the convex version of the actual territories (created from the same returns) as the field based polygons but with no concave regions. She also investigated how many latrines are needed to have a good approximation of the badger territories using this method and she found that territories estimated from between 3 and 7 latrines are highly erroneous, territories estimated from between 8 and 10 latrines are erroneous, territories estimated from 11 latrines reasonable and territories estimated from 12 or more latrines will give good approximations when these reconstructions of the territories using this method are compared with the reconstructions of the territories made by the experts of Woodchester Park.

Limitations, from a biological point of view, of this method are that it can not detect concave areas of the territories, the way the probability is used (shrinking the distance form the main sett to the latrine), it fixes the minimum distance and the angle used to calculate the $Angularfit$, the use of a latrine contained in the same direction is independent of the other ones in the same direction and, finally, it not takes on consideration neighbouring territories.

2.4 Overview of the methods developed in this thesis

This thesis is about boundary definitions of badger territories. The data used, as mention above, are from Woodchester Park latrine survey 2010. From these data different variables were obtained: badger territories (according with experts opinion), coordinates of main setts and latrines (for each territory), and times each latrine was used by a badger clan. Moreover, other variables were calculated from the mentioned variables such as: distance from main setts of the badger territory and latrines used by the badger clan, direction of each latrine in relation to the main sett and latrines on the same direction in relation of the main sett.

The first method explored in this thesis, is an adaptation of the Elstub Outlier Prediction Model (EOPM), explained in detail above. The method was called ‘Unconditional Outlier Prediction Model (UOPM)’. The difference between both methods was that, in the calculation of the *Angular fit* (2.1), the minimum distance and the angle were fixed on the EOPM and in the UOPM, these were varied in the estimation for the calculation of the Angular fit use in the estimation of the best model. The model were used in two different ways: to produce one territory of a badger clan to be compared with the MCP and the EOPM. The second one was using the model with a sampling approach. This is done in the next chapter.

One of the limitations of the UOPM is that it not takes in consideration wether a latrine in the same direction is not part of the territory or not. A method that uses this information was proposed, the Conditional Outlier Prediction Model (COPM). This method is explained in Chapter 4.

Since the COPM uses the MCP to define boundaries of the badger territories, it produces only convex territories. Moreover, the UOPM and the COPM classifies the latrines as 'Part of the territory' and 'Outliers'. On Figure 2.1 some concave areas can be observed. To deal with the problem of concave areas in some territories, a new method was proposed: the Unadjusted Ordinal Model (UOM) (Chapter 5). For this, and ordinal classification was given to the latrines: 'Hinterland', 'Boundary' and 'Outlier'.

However, the three previous methods not take on considerations the neighbour territories. A final method was explored that used this information: the Adjusted Ordinal Model (AOM) which is explored in Chapter 6.

Chapter 3

Unconditional Outlier Prediction Model

3.1 Introduction

An unpublished project by Elstub (Elstub, 2006) —supervised by T. Etherington, and carried out at the Food and Environmental Research Agency (FERA)— explained in detail in the literature review on page 13, attempts to predict whether or not a given latrine is an outlier, based on covariates that include properties of the latrine itself and on the configuration of the other latrines associated with the same clan. The aim is to use those predictions to improve the reconstruction of territories. The Elstub Outlier Prediction Model (EOPM) was estimated with data from 1990 to 1996. The number of social groups in each year varied from 26 to 37 in this study. Variables are explained in detail in section 2.3.3.

In the present work, the EOPM was replicated as a starting point, then the criteria to calculate the *Angular fit* (minimum distance m and the angle 2θ) were allowed to vary, treating them as parameters to be estimated. This defines a more general model than the EOPM, which we refer to as the Unconditional Outlier Prediction Model (UOPM). The term 'Unconditional' here refers to the fact that when a latrine is being classified as an 'outlier' or not, we do not take into account the classification of other latrines in similar direction. It contrasts with a conditional version of the model, introduced in the next chapter, in which the classification of a particular latrine is allowed to depend on the classifications of others.

The data used to estimate this model consists of the year 2010 bait-marking results from the same site (Woodchester Park) in which 481 latrines were surveyed from 23 baited setts, along with boundary reconstructions by experts at the FERA (Delahay et al., 2000) shown on Figure 3.1 where the stars represent setts, the red dots latrines and lines boundaries between badger territories. Thus, for these particular latrines, it is known whether they are extraterritorial excursions (outliers), or part of the territory, allowing a logistic regression model to be fitted.

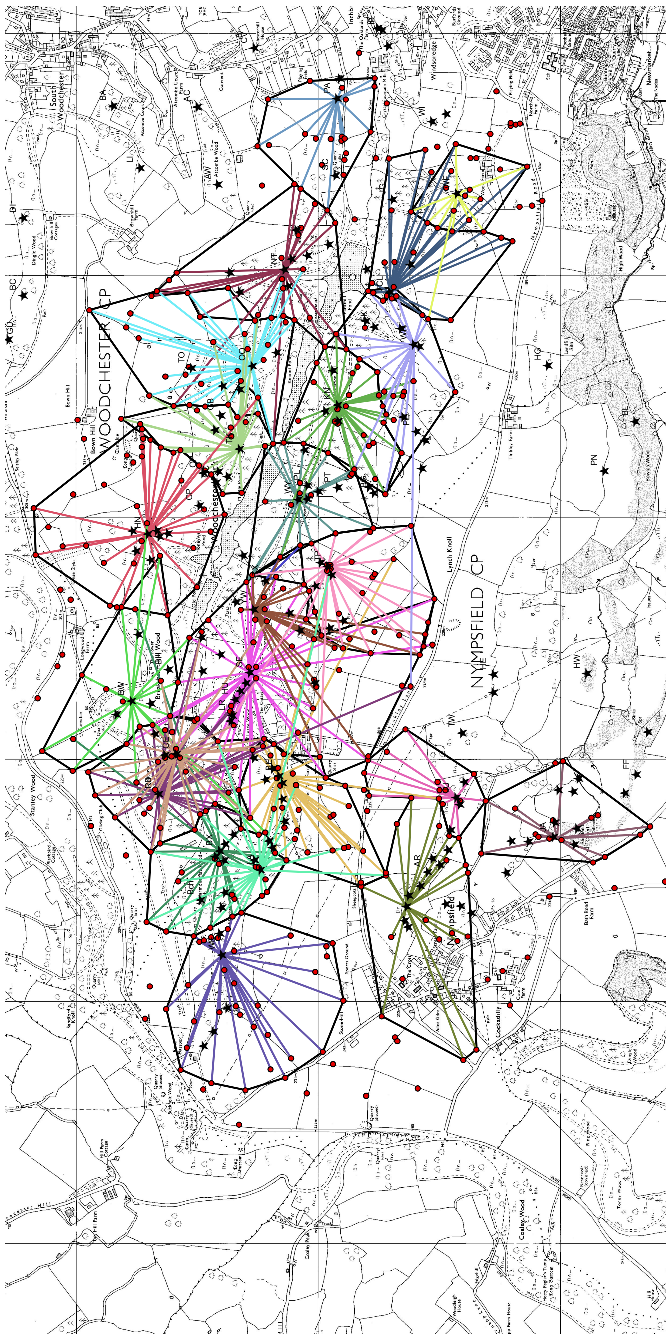


Figure 3.1: Map of Woodchester Park with the 2010 Bait-marking Survey Results. Stars represent baited setts, red dots latrines and lines boundaries between badger territories. Grid lines shown at 1 km intervals; region shown is approximately 5.5 km x 3 km.

The UOPM was used to estimate probabilities whether a latrine is part of the territory or not and then this information was used to reconstruct the badger territories firstly with a *Point Estimation Approach* (section 3.3.1) and secondly with a *Sampling Approach* (section 3.3.2). The reconstructed territories using the Point Estimation Approach were compared with the EOPM, the 95% Minimum Convex Polygon (MCP) and the territories generated by the researchers from Woodchester Park. For the Sampling Approach, two graphical representations of the territory reconstruction using the UOPM are presented: a) the territory reconstructions with the four with highest probabilities and b) the probability of inclusion that presents all the samples in one single graph.

3.2 Model Selection

As starting point for the modelling of outliers, a logistic regression was applied, using the *"glm"* function in R (R Core Team, 2014) in which the response variable was binary, representing the status of each latrine as an outlier or not. The EOPM (section 2.3.3) was re-fitted to 2010 Woodchester Park data set using firstly an identical approach to Elstub (same angle, $2\theta = 45°$ and minimum distance, m=40m, to calculate the *Angular fit*, and no interaction terms in model) the coefficients of this model are presented in table 3.3; secondly the angle was changed from 20° to 360° in steps of 1° and m from 1m to 500m in steps of 1m to calculate the *Angular fit* ($340 \times 501 = 170,340$ different combinations of meters and degrees) and using main effects and their interactions to estimate the best logistic regression model (27 different combinations, $27 \times 170,340 = 4'612,707$). This was done in a single run, saving the Akaike Information Criterions (AICs) values of the best model, for each combination of meters and degrees, in a "csv" file, to select the smallest one. Finally, four different logistic regression models were estimated for each combination of θ and m using the original values or transformation of Number of returns to the latrine by badgers of the same baited sett (R_i) and the distance from the baited sett to the latrine ($4'612,707 \times 4 = 18'450,828$) ($d_i$) as follows:

- R_i as a quantity (original value and its square root transformation both at once) with the original values of d_i.
- Using R_i as quantity (original value and its square root transformation both at once) with the logarithmic transformation of d_i.
- Using R_i as a factor (as explained in table 2.2) with the original values of d_i.
- Using R_i as a factor with the logarithmic transformation of d_i.

The best logistic regression models were chosen using the AIC (Akaike, 1978) and were obtained with a $2\theta = 46°$ and m=90m in the calculation of $AngFit_i$ which gave the best AICs for all six cases listed above. A summary of the best AICs is given in table 3.1. From this table, we can say that the variables that produced the best AIC are R_i as a Factor and d_i un-transformed. The estimate of the coefficients of the best

UOPM is given in table 3.2. This model was applied to reconstruct the territories of the badgers using the "Point Estimation" and "Sampling" approaches. It should be noted that using interactions and varying θ and m to estimate the UOPM gives a better AIC than the EOPM, even increasing it by 4 due to the two more parameters, m and θ estimated by the Outlier Prediction Models (OPMs). The estimated coefficients for the other models in table 3.1 are in Appendix A.1.

This approach using generalized linear modelling allows the model to be fitted very efficiently for each combination of m and θ. As an alternative, it would be possible to use a more general non-linear approach, and estimate all parameters simultaneously. This approach was not explored, as the method described performed satisfactorily.

Model	No. of Returns as	AIC	
		Distance	Log(Distance)
UOPM	Quantity	267.98	269.10
	Factor	265.39	267.10
Elstub	Factor	286.01	-

Table 3.1: AIC values for the different models (UOPM)

Parameter	Coefficient
Intercept	-11.887
angfit1	7.137
Dist	0.004
Shared(1)	-12.315
R2(1)	-3.623
Shared(1): R2(1)	27.183
angfit1:R2(1)	2.535
angfit1:shared(1)	13.152
Dist: R2(1)	0.005
angfit1:shared(1):R2(1)	-28.464
Degrees of Freedom:	630 Total (i.e. Null); 621 Residual
Null Deviance:	448
Residual Deviance:	245.4
AIC:	265.4

Table 3.2: Unconditional Outlier Prediction Model with variable *Number of Returns as an indicator variable and Distance Untransformed*

3.3 Reconstruction of the Territories

3.3.1 Unconditional Outlier Prediction Model: Point Estimation Approach

As mentioned above, to reconstruct the territories, the estimated UOPM coefficients shown in table 3.2 were used to calculate the probability of each latrine being an outlier, i.e. the fitted values in the model. Following (Elstub, 2006) as a new way to deal with extreme values as an alternative to the commonly used 95% MCP, the d_i of the $Latrine_i$ to the baited sett was multiplied by the probability to construct a new point for each latrine, applying, as a final step, a 100% MCP to this new data set to each one of the territories for its reconstruction.

In addition to this OPM variant that we called UOPM, the 23 badgers' territories that make up the 2010 Woodchester Park data set were delineated using three alternative methods:

- 95% MCP explained in detail in 2.3.3 (on page 12).
- Experts: They use a criterion to delineate the territories based on the presence of shared latrines, field signs (such as boundary ruts), distance and activity at a fed sett (active hole and presence of latrines around the main sett) (Delahay et al., 2000); they results are shown in Fig. 3.1.
- Elstub's method (EOPM): The reconstruction was made using the same steps as the UOPM but fixing the values she proposed to calculate the *Angular Fit* (explained in detail in 2.3.3 on page 13) with the estimated coefficients for the 2010 Woodchester Park survey shown in table 3.3.

Parameter	Coefficient
Intercept	-11.100
angfit1	5.780
Dist	0.006
Shared(1)	1.488
R2(1)	-1.280
Degrees of Freedom:	630 Total (i.e. Null); 625 Residual
Null Deviance:	448
Residual Deviance:	274
AIC:	286.013

Table 3.3: Elstub Outlier Prediction Model using $2\theta = 45°$ and minimum distance=40m

An example of the territories reconstruction using the four methods for the *Beech territory* is shown in Figure 3.2a in which the burgundy star represents the *'baited sett'* and the circles the latrines where the pellets from the baited sett were recovered. From Fig. 3.1, black circles are *'hinterland latrines'*, orange circles *'boundary latrines'* and red circles *'outlier latrines'* according to the 'expert's reconstruction'. The lines are the different versions of the reconstruction of the territory: red line is the *'Experts version'*, green line represents the *'UOPM version'*, magenta line shows the *'EOPM version'* and the blue one the *'95% MCP'*. The reconstruction of the other territories can be found in the appendix A.2. Later in this chapter, we show a better method of reconstruction based on the UOPM, which is illustrated for Beech and several other territories.

Four pairwise comparisons were made from the different reconstructions of the territories, using the 'Expert reconstruction' as a reference (Figure 3.2):

a) 95% MCP - Experts (Figure 3.2b).

b) UOPM - Experts (Figure 3.2c).

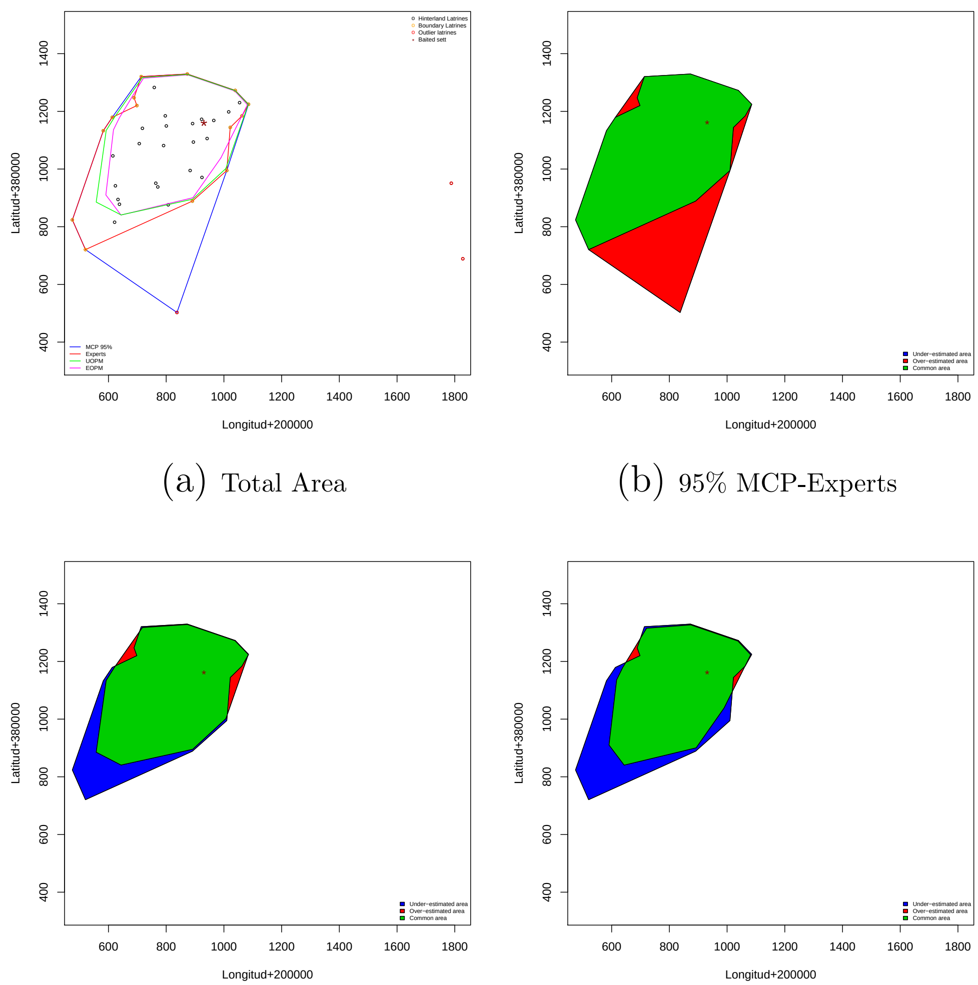
c) EOPM - Experts (Figure 3.2d).

A graphical representation with these comparisons of the Beech territory can be viewed in Fig. 3.2 in which the green zone represents *common areas* predicted by both methods, blue areas are zones that the first method is not predicting as part of the territory but the reference method is, i.e. *Under-estimated areas* and red areas represent zones that the first method is predicting are part of the territory but the reference method is not (*Over-Estimated Areas*).

Comparison of the predictions of the OPMs

The first comparison made was the estimation of the predicted probabilities of being an outlier for each $latrine_i$ for the UOPM and the EOPM. A scatter plot was produced using the log of these probabilities since most of the latrines are around or near the main sett making it difficult to visualize them and make and interpretation of the results. Figure 3.3 shows the comparison of 4 territories. Four territories have been chosen as examples to present and discuss the results of the Adjusted Ordinal Model (AOM): Beech, Junction, Parkmill and Woodrush. Beech was picked randomly since the beginning of the research and all the models and scripts were tested firstly in this territory; Junction was chosen since it is on the edge of Woodchester Park and it has several shared latrines; Parkmill is in the edge of the park, however it does not have too many shared latrines and Woodrush was selected because it was one of the territories that the Unadjusted Ordinal Model (UOM) (Chapter 5) was able to present valid configurations.

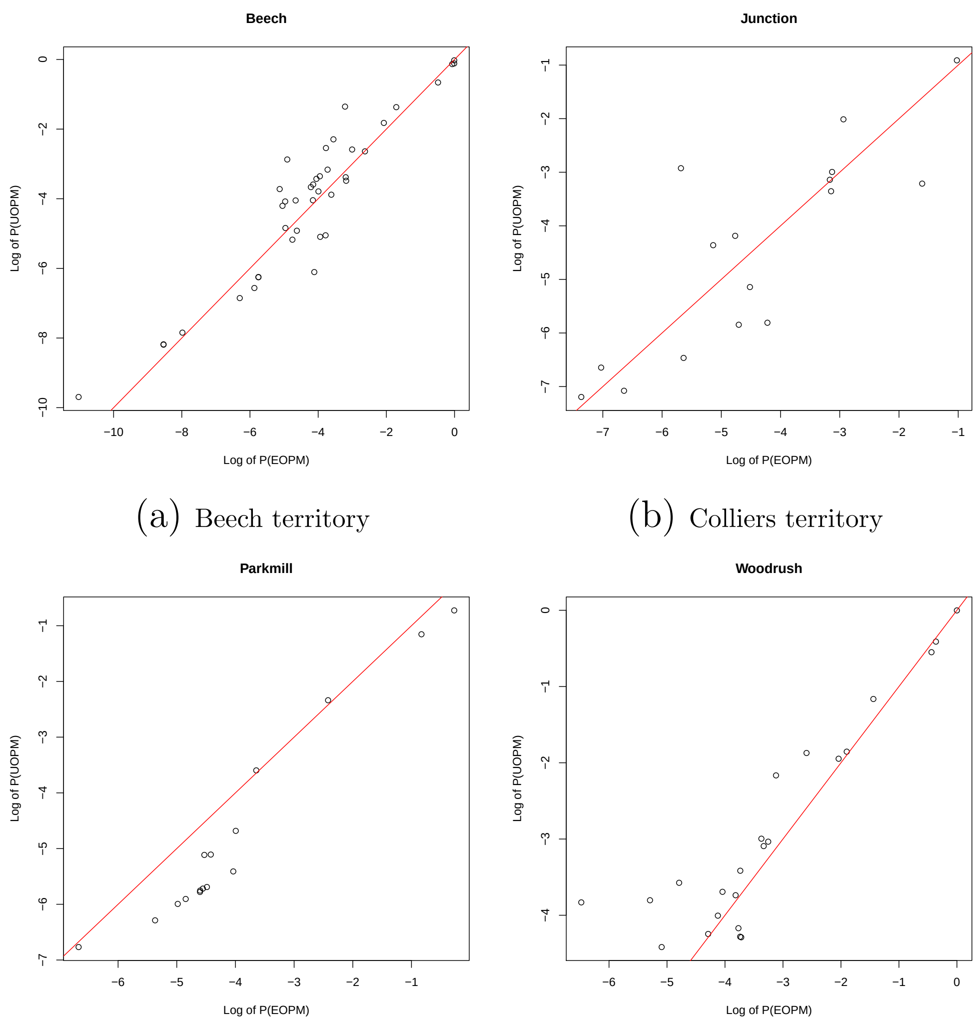
There is a clear association between the two sets of probabilities for each territory, but the strength of the relationship varies. In Beech and Honeywell, the relationships are quite strong, and the two sets of probabilities are quite similar. In Colliers the relationship is rather noisy; in Top it is fairly good except that some of the very low probabilities differ a lot between the two models.



(a) Total Area (b) 95% MCP-Experts

(c) UOPM-Experts (d) EOPM-Experts

Figure 3.2: Comparison of the reconstruction of the territory Beech using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



(a) Beech territory

(b) Colliers territory

(c) Honeywell territory

(d) Top territory

Figure 3.3: Comparisons of the log p(UOPM) and the log p(EOPM). The line $y = x$ is shown in red.

Graphical Comparisons

The different territories generated were compared visually by delineating them, plotting the latrines then plotting the territories in the same scale. We compared the reconstructions of the the methods mentioned in the previous section. The 'expert' territory reconstruction was used as a reference to compare over or under-estimated areas by the other methods that could be shaded for visual comparison as presented in Fig. 3.2 in which one of the territories (Beech) is shown. Green areas represent zones that both methods are predicting, red areas represent Over-estimated areas and blue ones Under-estimated areas. The 'star' represents the baited sett. In this figure, 4 comparisons are made. On the pairwise comparisons, the reconstruction taken from Fig. 3.1 was used as a reference: green areas represent areas from the territory predicted by both methods, blue areas are those that were proposed by the experts but the other methods are not predicting as part of the territory (under-estimated areas) and red zones represent areas that the reference method is not predicting as part of the territory but the other methods is doing it (over-estimated areas).

- **Total Area:** In Fig. 3.2a, the reconstruction of the territory using the four methods is used to make a visual comparison of the total area predicted for each method. The red line shows the territory reconstruction made from Fig. 3.1, the blue line represents the reconstruction using the MCP, the green line the territory predicted by the UOPM and the magenta line the EOPM; The burgundy star represents the baited sett; circles represent latrines, the colours are based again on Fig. 3.1 in which black ones are hinterland, orange ones are boundary and red ones are *'outliers'*.

- **95% MCP with Experts:** In Fig. 3.2b, there are three areas that the 95% MCP is over-estimating, the two smaller ones are due to concavities of the 'Expert version' and the bigger one is because, as mentioned in several reports, the MCP tends to over-estimate territories, in this case is identifying an *'outlier latrine'* as part of the territory.

- **UOPM with Experts:** Fig. 3.2c presents the comparison of the UOPM with the Experts version of the reconstruction. Two red areas representing over-estimated areas can be observed to the west and east of the reconstructions. A blue zone representing an under-estimated area not predicted by the UOPM can be observed to the south and southwest of the territory.

- **EOPM with Experts:** The comparison of the original OPM proposed by Elstub and the Experts version is made on Fig. 3.2d. Since the predicted territory using this method is smaller compared with those reconstructions made with the MCP and the UOPM, the over-predicted areas are smaller even when they are the same two. However the under-predicted area is not only bigger but there is another one on the east side of the territory.

Comparison of the Mean of the Areas of the Reconstruction

The different reconstructions can be compared numerically by considering the predicted areas of each one of the 23 territories for the four methods utilized to reconstruct the badger territories and their mean.

The R program (R Core Team, 2014) was used to compare the areas (total, under, over-estimated and under and over-estimated together) with an Analysis of Variance (ANOVA) and when significative differences were found, the Tukey's test (Tukey, 1949) was applied (library Functions and datasets to support Venables and Ripley, 'Modern Applied Statistics with S' (MASS) (Venables and Ripley, 2002)). The library Mixed GAM Computation Vehicle (mgcv) (Wood, 2011) was used to plot the boundary latrines — not to smooth them: Generalized Additive Models were not used to reconstruct the territories nor to plot the boundaries — and the library General polygon clipping routines for R based on Alan Murta's C library (gpclib) (Peng, 2013) to calculate territories areas and to plot them as in Fig. 3.2.

Elstub used only the total area of the reconstructed territories in comparisons. In this work, not only the *Total Areas* were compared since this could lead to the loss of important information about the quality of the reconstruction because the over and under-estimate areas could cancel each other out. For this reason, in addition, three types of possible errors in the reconstruction were calculated and analyzed: a) Over-estimated areas, b) Under-estimated areas, and c) the sum of these two areas (total error) defined in section 3.3.1 (on page 23).

1 **Total Area**

As mention above, the library mgcv was used to draw the reconstruction of the territories for each method (Fig. 3.2). Then, the library gpclib was utilized to calculate their *Total Area* in Hectares (Ha). The mean of each method was compared through an ANOVA running a *Tukey's post-hoc test* after finding significant differences $F(3, 88) = 4.45$ (table 3.5). The method that produced on average the bigger territories was the 95% MCP (mean=28.89, SD=11.97) followed by the Experts (mean=24.45, SD=10.72), the UOPM (mean=20.57, SD=7.59), and finally the EOPM (mean=19.4, SD=6.82) (table 3.4). The mean in Ha of the *Total Area* for the MCP was 4.43 Ha, 95% CI[-2.92, 11.78] bigger than the 'Experts', which was not statistically significant ($p = 0.40$); it is 8.32 Ha, 95% CI[0.97, 15.67] bigger than the UOPM ($p < 0.05$); and 9.14 Ha, 95% CI[1.80, 16.49] than the EOPM ($p < 0.05$). On average, the 'Experts reconstruction' produces a territory 3.88 Ha, 95%[-3.46, 11.23] bigger than the UOPM and 4.71 Ha, 95% CI[-2.64, 12.06] bigger than the EOPM, however this differences were not statistically significant ($p = 0.51$) and ($p = 0.34$) respectably. The mean of the the UOPM is 0.83 Ha, 95% CI[-6.52, 8.18] bigger than the EOPM although was not statistically significant ($p = 0.99$). This comparisons can be observed graphically in Figure 3.4a.

Sett	Experts	MCP	EOPM	UOPM
Jacks	18.74	14.21	16.10	17.20
Mead	35.57	35.64	27.31	28.24
Junction	10.98	12.83	11.57	11.25
West	40.01	35.55	30.76	33.23
Box	18.91	24.88	17.44	17.36
Beech	21.51	30.94	16.28	18.17
Septic Tank	55.79	62.68	32.00	36.85
Yew	30.47	25.46	11.76	17.58
Top	25.31	27.18	21.98	23.33
Wych Elm	17.44	18.16	13.75	13.93
Kennel	16.52	21.49	15.93	16.50
Woodrush	18.55	24.45	14.45	15.36
Colliers	20.11	27.38	22.07	22.03
Wood Farm	10.44	10.44	9.06	8.93
Parkmill	17.40	14.80	11.77	10.51
Nettle	28.58	31.40	23.98	25.38
Old Oak	31.44	33.72	22.58	21.48
Trackside	13.66	22.51	15.12	14.20
Honeywell	40.77	38.39	33.96	34.42
Boxwood	23.79	49.27	22.65	26.48
Rabbit	23.75	32.62	21.62	21.70
Cedar	23.35	32.57	21.83	23.34
Larch	19.32	37.82	20.09	15.64
Mean	24.45	28.89	19.74	20.57
SD	10.72	11.97	6.82	7.59

Table 3.4: Total Area in hectares for each reconstructed territory using the *'Point Estimation Approach'* methods.

2 **Under-Estimated area**

To calculate the Under-Estimated area, the three methods (MCP, UOPM and the EOPM) are compared with the 'Experts version' of the territory and both OPMs between them (for this comparison, the EOPM is used as the reference method). The *Under-Estimated Area* is the area that is predicted by the reference method and is not predicted by the other one. This areas are indicated in blue colour in Fig. 3.2 and in the Appendix A.2. If a '0.00' is obtained as a result, it means that that method is not predicting under-estimated zones for that territory, either because is the same predicted territory or is predicting a bigger territory that the reconstruction by the experts' territory.

A one-way ANOVA was conducted to determine if the Under-Estimated Areas were different for the 3 different methods used to reconstruct the badger territories. Data is presented as mean ± standard deviation in table 3.6. Under-

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Mean of Total Area	3	1209.82	403.27	4.45	0.0058
Residuals	88	7969.78	90.57		

Table 3.5: Anova for Total Area

estimated areas were statistically significantly different between the different methods, $F(2,66) = 0.95, p < 0.05$ (table 3.7). The OPMs tend to underestimate more when compare with the MCP: EOPM (mean=6.32, SD=5.80) and UOPM (mean=5.49, SD=4.80) whilst MCP (mean=2.06, SD=3.80). Tukey post-hoc analysis revealed that the under-estimation difference between the MCP and the UOPM was marginal, 3.43 Has, 95% CI[-0.10, 6.87], $p = 0.05$ with the EOPM and 4.26, 95% CI[0.82, 7.70] $p < 0.05$ within the OPMs was 0.83 Has, 95%[-2.61, 4.28] which was not significant $(p = 0.83)$. The differences between the two OPMs was 0.83 Has, 95% CI[-2.50, 4.16] which was not statistically significant $(p = 0.91)$. Fig. 3.4b shows graphically the post-hoc Tukey's test.

3 **Over-estimated area**

To calculate the Over-Estimated area, the three methods (MCP, UOPM and the EOPM) are compared with the 'Experts version' of the territory and then the over-estimated areas were compared between them. The *Over-Estimated Area* is the area that is not predicted by the reference method and is predicted by the other one. This areas are indicated in red colour in Fig. 3.2 and in the Appendix A.2.

The Over-Estimated Areas were compared using a one-way ANOVA to determine if there were differences in the means for the 3 different methods used to reconstruct the badger territories. Data is presented as mean $\pm$ standard deviation in table 3.8. Statistically significant differences were found on the Over-estimated areas between the different methods, $F(2,66) = 9.89, p < 0.05$ (table 3.9). The MCPs was the method with bigger overestimated area (mean=6.49, SD=6.99) followed by the EOPM (mean=1.61, SD=1.94) and the UOPM (mean=1.61, SD=1.17). Tukey post-hoc analysis showed that, when the 'Experts reconstruction' was the reference method, the over-estimation difference between the MCP and the EOPM was 4.88, Ha, 95% CI[1.84, 7.92], $p < 0.05$ and with the UOPM and the difference was 4.89 Ha, 95% CI[1.85, 7.93] $p < 0.05$. Fig. 3.4c shows graphically the post-hoc Tukey's test.

4 **Total Error areas**

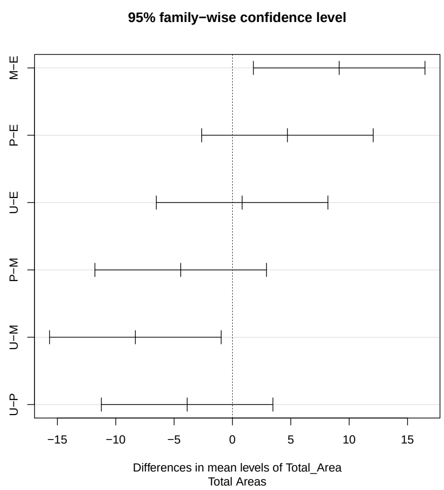
To do this comparison, the Over-Estimated and the Under-Estimated areas (sections 2 and 3) were summed since this is the *'total error'*, when two territories are compared. As in the previous sections, the three methods were compared with the 'Experts version'. Note that total error is zero only if the the two reconstructions agree complectly.

The method that on average has a smaller Total Error Area, when is compared

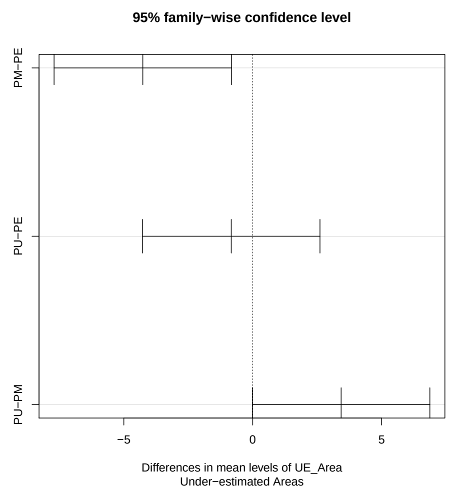
Sett	MCP	EOPM	UOPM
Jacks	4.53	2.64	1.53
Mead	1.23	9.10	8.22
Junction	0.31	1.41	1.75
West	4.46	9.24	6.77
Box	0.00	4.13	3.26
Beech	0.00	5.43	3.90
Septic Tank	14.50	25.33	19.87
Yew	11.66	20.02	17.15
Top	0.00	4.57	3.45
Wych Elm	0.60	3.73	3.56
Kennel	0.00	2.47	2.61
Woodrush	1.06	4.65	4.01
Colliers	0.00	2.82	2.44
Wood Farm	0.00	1.40	1.51
Parkmill	2.60	5.63	6.89
Nettle	0.00	5.18	4.08
Old Oak	3.16	10.32	10.60
Trackside	0.47	0.54	0.95
Honeywell	2.81	6.81	6.36
Boxwood	0.00	4.84	3.55
Rabbit	0.00	3.38	3.53
Cedar	0.00	4.25	2.33
Larch	0.00	7.56	7.95
Mean	2.06	6.32	5.49
SD	3.80	5.80	4.80

Table 3.6: Mean of Under-estimated Areas (UOPM) in hectares

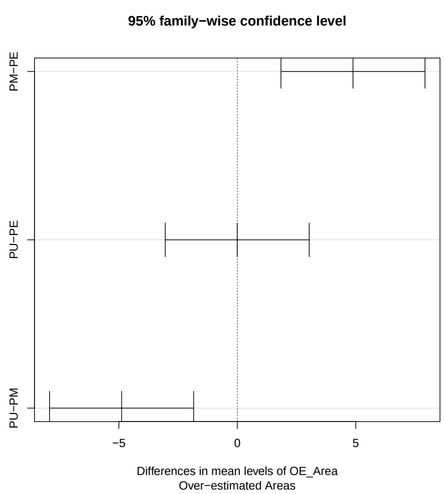
with the ‘Experts Version’, is the UOPM (mean=7.10, SD=5.20), followed by the EOPM (mean=7.94, SD=6.08) and the method the produces the bigger Total Error Area is the MCP (mean=8.56, SD=8.57) (table 3.10). However, after a one way ANOVA, statistically significant differences were not found on the Over-estimated areas between the different methods, $F(2, 66) = 0.268, p = 0.77$, (table 3.11) and Fig. 3.4d.



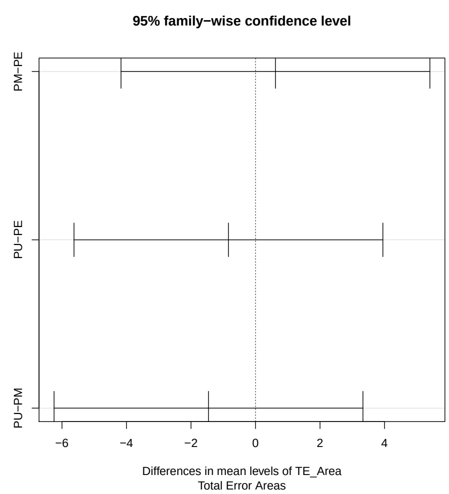
(a) Total Area



(b) Under-estimated Area



(c) Over-estimated Area



(d) Over and Under-estimated Area

Figure 3.4: Tukey HSD for the Outlier Prediction Model. The abbreviations are: UE UOPM-EOPM, PU 'Experts'-UOPM, PM 'Experts'-MCP and PE 'Experts'-EOPM.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Mean of Under-estimated Areas	2	234.90	117.43	4.952	0.0099
Residuals	66	1565.11	23.71		

Table 3.7: Anova Under-estimated Areas (UOPM)

Sett	MCP	EOPM	UOPM
Jacks	0.00	0.00	0.00
Mead	1.30	0.84	0.89
Junction	2.16	2.00	2.02
West	0.00	0.00	0.00
Box	5.97	2.66	1.71
Beech	9.44	0.20	0.56
Septic Tank	21.39	1.54	0.93
Yew	6.66	1.31	4.26
Top	1.87	1.24	1.46
Wych Elm	1.32	0.04	0.06
Kennel	4.97	1.88	2.59
Woodrush	6.96	0.55	0.82
Colliers	7.26	4.77	4.35
Wood Farm	0.00	0.02	0.00
Parkmill	0.00	0.00	0.00
Nettle	2.82	0.59	0.88
Old Oak	5.45	1.47	0.64
Trackside	9.33	2.01	1.49
Honeywell	0.43	0.00	0.02
Boxwood	25.47	3.69	6.24
Rabbit	8.86	1.25	1.48
Cedar	9.22	2.73	2.32
Larch	18.50	8.33	4.28
Mean	6.49	1.61	1.61
SD	6.99	1.94	1.71

Table 3.8: Mean of Over-estimated Areas (UOPM) in hectares

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Mean of Over-estimated Areas	2	365.7	182.84	9.89	0.0002
Residuals	66	1220.2	18.49		

Table 3.9: Mean of Over-estimated Areas (UOPM)

Sett	MCP	EOPM	UOPM
Jacks	4.53	2.64	1.53
Mead	2.53	9.95	9.11
Junction	2.47	3.41	3.77
West	4.46	9.24	6.77
Box	5.97	6.80	4.98
Beech	9.44	5.63	4.46
Septic Tank	35.90	26.87	20.79
Yew	18.32	21.33	21.41
Top	1.87	5.81	4.91
Wych Elm	1.92	3.77	3.62
Kennel	4.97	4.35	5.20
Woodrush	8.02	5.20	4.82
Colliers	7.26	7.59	6.79
Wood Farm	0.00	1.42	1.51
Parkmill	2.60	5.63	6.89
Nettle	2.82	5.77	4.96
Old Oak	8.61	11.79	11.24
Trackside	9.80	2.55	2.44
Honeywell	3.24	6.81	6.38
Boxwood	25.47	8.53	9.79
Rabbit	8.86	4.62	5.01
Cedar	9.22	6.98	4.65
Larch	18.50	15.88	12.23
Mean	8.56	7.94	7.10
SD	8.57	6.08	5.20

Table 3.10: Mean of Total Error Areas (UOPM) in has.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Mean of Total error	2	24.6	12.31	0.268	0.765
Residuals	66	3026.3	45.85		

Table 3.11: Mean Total Error Areas (UOPM)

3.3.2 Unconditional Outlier Prediction Model: Sampling Approach

The use of the probabilities of the latrine being an extraterritorial excursion to shrink its distance to the sett is not the best way to use them. The probabilities obtained can be applied in a sampling approach. The idea is to sample sets of latrine labelling (as an outlier or not) from the probability distribution fitted by the UOPM, and use the reconstructions based on them to demonstrate the uncertainty in the model.

If the fitted probability that latrine i is an outlier is p_i, then we randomly sample the "status" of $latrine_i$ as outlier with probability p_i, not an outlier with probability $1 - p_i$. Repeating this for each latrine gives a single reconstruction of the territory as explained below. A large sample was generated of reconstructions to explore the uncertainty in the fitted model.

Note that the states of latrines are sampled independently; this is an inherent limitation of this model, and is addressed in the next chapter. We are also ignoring the uncertainty in the parameters of the model, so this is not a true predictive distribution; there are various ways in which the extra uncertainty could be included, but it is expected to be small compared with the randomness from latrines status.

In practice, the sample of the reconstructions is summarized in two ways. Firstly, a plot was constructed showing, on a grey scale, the probability that any location forms part of the territory. Secondly, since the number of possible reconstructions is limited as there are a limited number of ways the latrines can be classified, and so the most likely reconstruction can be identified as being those most frequently occurring in the sample.

The procedure to do this was as follows:

- 100,000 samples were obtained using the method explained above for each territory.
- A 100% MCP was applied to each one of these samples.
- Plotting of the most likely reconstructions, one at a time.
- Plotting all reconstructions in one single plot, following these steps:
    1 Create a square grid of points, based on their cartesian coordinates, over a rectangular region containing all possible latrines associated with the territory.
    2 For each sample of latrine statuses, determine which points of the grid are inside the reconstruction obtained from this sample.
    3 For each point of the grid, count how many of the reconstructed territories contained that point, convert the count to a probability, and produce a grey-scale plot indicating each point's probability of being in the territory.

For the 'most likely' reconstructions, the procedure is as follows:

For each sampled labelling of the states of the latrines, we construct the 100% MCP as described previously. Then, the number of times each distinct obtained MCP is counted and each area is represented by darker colours the more is counting it has. Different labellings can lead to the same MCP, since whether an 'internal' latrine is identified as an outlier does not change the MCP.

Results

All the territories were reconstructed and are presented in A.2. However, for thesis length limitations, four territories have been chosen as examples to present and discuss the results of the UOPM Sampling Approach: Beech, Junction, Parkmill and Woodrush. They were chosen after all the models proposed in this thesis for reconstruction of badgers' territories were developed, to make easier the discussion and conclusion of this work. Beech was picked randomly since the beginning of the research and all the models and scripts were tested first in this territory; Junction was chosen since it is on the edge of Woodchester Park and it has several shared latrines; Parkmill is on the edge of the park, however it does not have too many shared latrines and Woodrush was selected because it was one of the territories that the UOM (introduced in chapter 5) was able to present valid configurations.

a) ***Beech*** The results using all the reconstructions in one single plot are shown in Fig. 3.5. In this figure darker areas represent a higher probability of belonging to the territory; paler areas represent a smaller probability. For this territory, the biggest uncertainty is present to its south and southwest. From this figure the two most eastern latrines are outliers since none reconstruction include them as part of the territory. The biggest advantage of this presentation is that the whole picture in one single figure, however the main disadvantage is that, since all the samples are used to produce this single reconstruction, it is not possible to identify which reconstruction has the bigger probability.

   Another way to display this results is plotting the most probable reconstructions as in Fig. 3.6 in which the 4 most frequent reconstructions are shown, totalizing for this example for the territory Beech 75%. To produce this figure, the different reconstructions obtained sampling from the UOPM posterior distribution were grouped and their frequency and probability calculated sorting them by this characteristic to plot the desired number of different reconstructions. This presentation is more explicative than the previous one in the sense that each reconstruction can be evaluated individually. In this example, two reconstructions, that only differ in one latrine, account for 61% of the probability, this can not observed if the results are presented only as in Fig. 3.5. Reconstruction with higher probability differs from the other three only in one latrine.

b) ***Junction***

   Fig. 3.7 presents the results using all the reconstructions in one single plot. In this territory, the only boundary with no predicted uncertainty is the south. The bigger uncertainty is predicted by the UOPM to be in the north and east boundaries. When the most frequent reconstruction is presented, as in Fig. 3.8 the 4 most frequent reconstructions have a total probability of 0.84. The two most frequent reconstructions are different in only one latrine, totalizing 0.68. Only the two most frequent reconstructions are needed, for this territory, to account for a probability bigger than 0.50.

c) ***Parkmill***

Territory Parkmill reconstruction is shown in Fig. 3.9 and 3.10. In the first figure the uncertainty predicted by the UOPM is on the mainly in the west side of the territory with some dark grey areas in its north boundary. A probability of 0.84 is explained by the four most frequent territories reconstructions. The first two reconstructions with the highest frequency have a total probability of 0.64.

d) ***Woodrush***

Fig. 3.11 shows uncertainty in all its boundaries but southeast. Bigger uncertainty is present to the west of the territory, predicting, with a low probability, a latrine that seems an outlier. The four most probable reconstructions were needed to sum a total probability bigger than 0.50 (0.56 for this case) as can be observed in Fig. 3.12. The differences of these four reconstruction is on the latrines in the west side of the territory.

3.4 Discussion and Conclusions

From a *Point Estimation Approach*, four comparisons (*Total Area, Under-estimated area, Over-Estimated Area* and *Total Error Areas*) of three methods (*MCP, EOPM* and *UOPM*) were made with the opinion of FERA experts as a reference (Fig. 3.1) using a one-way ANOVA and Tukey's test when statistical differences were found.

1 *Total Area:* As shown in Table 3.4 the mean of the territories in the experts opinion is 24.45, the MCP tends to produce bigger territories (mean=28.89) than the actual ones with more variability (SD=11.97); on the other hand, OPMs produce smaller territories (EOPM mean=19.74 and UOPM mean=20.57) than the Experts territories, however their SD is smaller (6.82 and 7.59 respectively). The model with smaller difference from the Experts territory was the UOPM.

2 *Under-Estimated areas:* The MCP is the method with smaller mean (mean=2.06 Ha) under-estimated area, this is mainly because it produced 11 territories with non-under-estimated areas, followed by the the the UOPM (mean=5.49) and the EOPM (mean=6.32) (Table 3.6).

3 *Over-Estimated:* The OPMs had a smaller mean for this comparison, both mean=1.61), but SD=1.94 and 1.71 for EOPM and UOPM, respectively, with only 4 territories without over-estimating areas. The MCP had and over-estimated mean of 6.49 and bigger variability (SD=6.99) (Table 3.7).

4 *Total Error:* The bigger total error mean and was produced by the MCP, follow by the EOPM and finally by the UOPM and the variability presented the same order (Table 3.10). However the differences were not statistically significant (Table 3.11).

Summarizing, in two of the four comparisons (Total area and over-estimated area) the OPMs had a better prediction than the MCP, however this method produced a better prediction when the under-estimated areas comparison was made. The MCP is well known as a method liable to over-estimation (Moorcroft and Lewis, 2006), as seen in this analysis. Moreover, the model obtained using the UOPM can be utilized in sampling methods as it was made in section 3.3.2. Even when the UOPM presented better predictions and smaller variability than the EOPM in all the comparisons, this differences were not statistically significative (Table 3.4). However, improvement on the AIC value has been done as shown on Table 3.1.

The use of the UOPM with a sampling approach is a better way of using the probabilities for each latrine of being an *'Outlier'* since its uncertainty can be displayed in the plots with the reconstruction of the territories, especially in its boundaries. However, one of the problems of the UOPM that it considers that the state *'Outlier' or 'Part of the territory'* of each latrine is independent of the states of the latrines in the same direction. In the next chapter this problem is addressed.

A simpler way to improve on the '95% MCP' approach would be to use the current set of expert judgments and latrine locations data to estimate a better percentile to use in an 'α% MCP' approach. However, the methods developed here and in later chapters attempt to use more of the information, and to formalize ideas about territory shape that are biologically plausible, to improve performance and generality.

The MCP estimates can be biased. Moreover, the bias increases with sample size. This limitation of the MCP has been shown in several works as in Burgman (Burgman and Fox, 2003). This limitation, usually, is addressed by eliminating some of the observations, commonly 5 %, based only on distance. For the purpose of this thesis, more information of covariates was available to make better estimations of the territories based on models to predict if a given latrine is an 'Outlier' or not. The model was fitted based on a latrine-level data, selecting the covariates using the AIC capturing some of the uncertainty involved, which it is an added benefit compared with the MCP.

Figure 3.5: Probability of inclusion in the Beech territory using the sampling method of the Unconditional Outlier Prediction Model

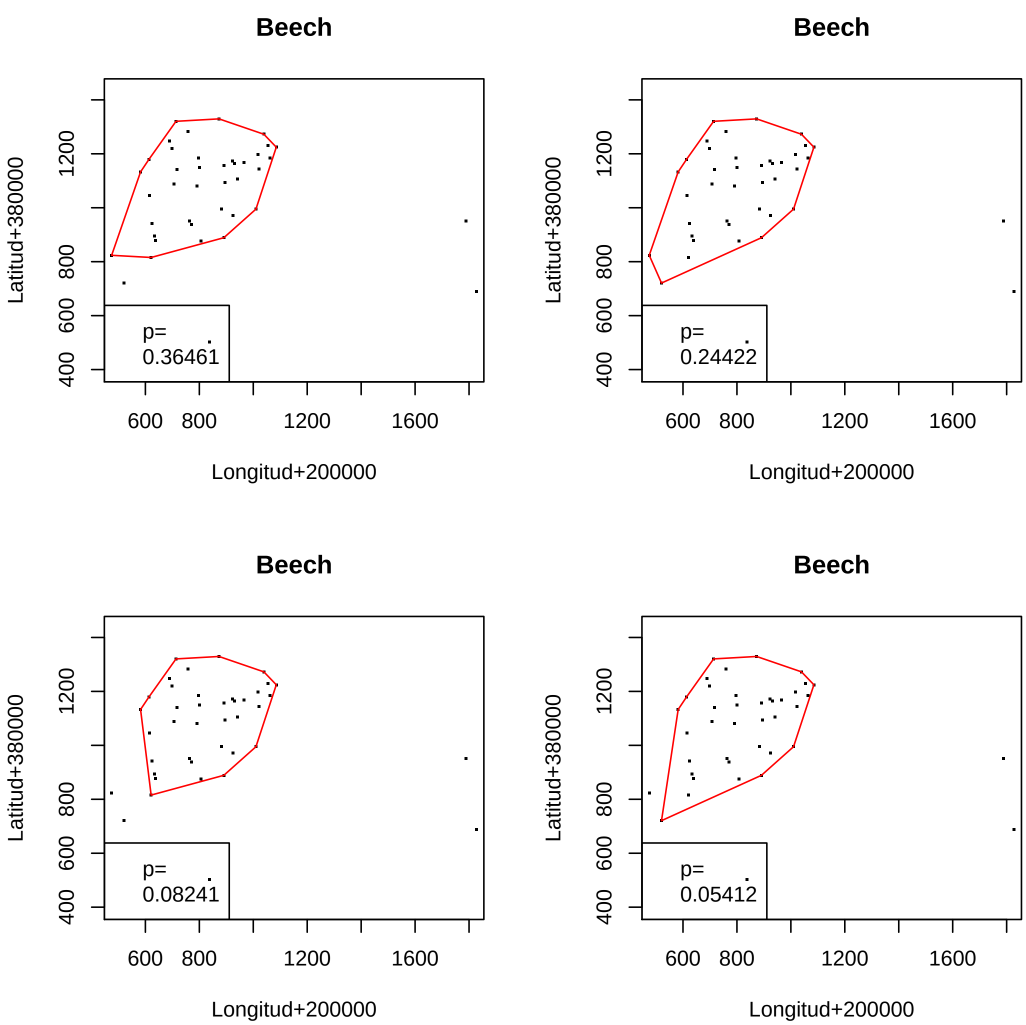


Figure 3.6: Most frequent reconstructions of the Beech territory using the sampling method of the Unconditional Outlier Prediction Model

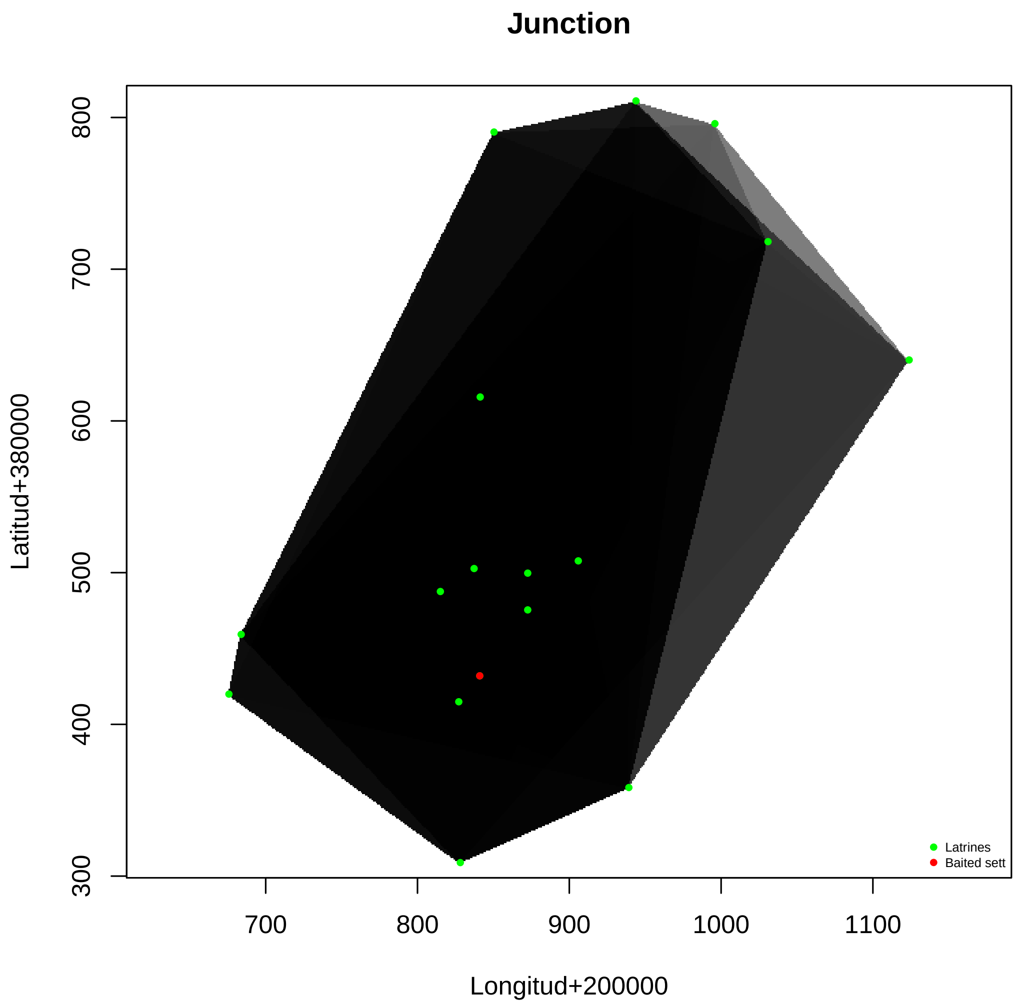


Figure 3.7: Probability of inclusion in the Junction territory using the sampling method of the Unconditional Outlier Prediction Model

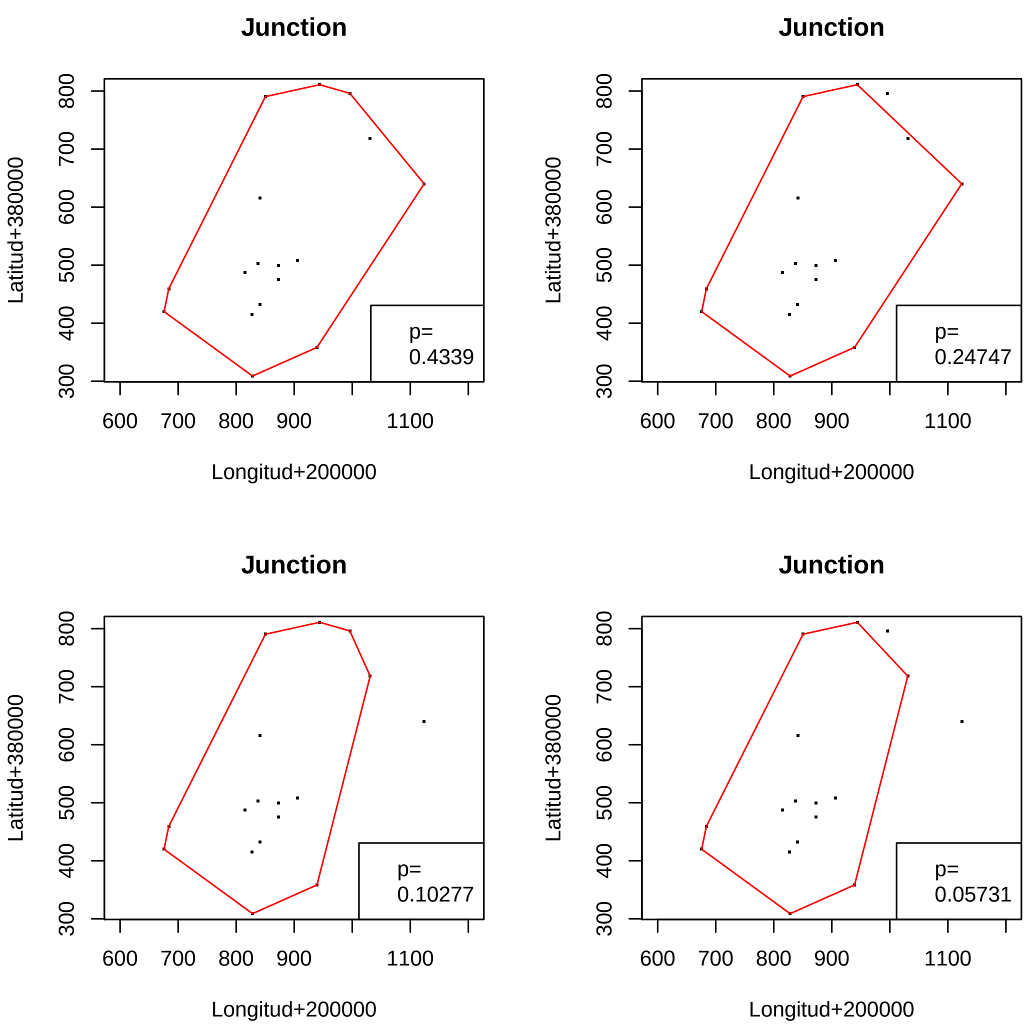


Figure 3.8: Most frequent reconstructions of the Junction territory using the sampling method of the Unconditional Outlier Prediction Model

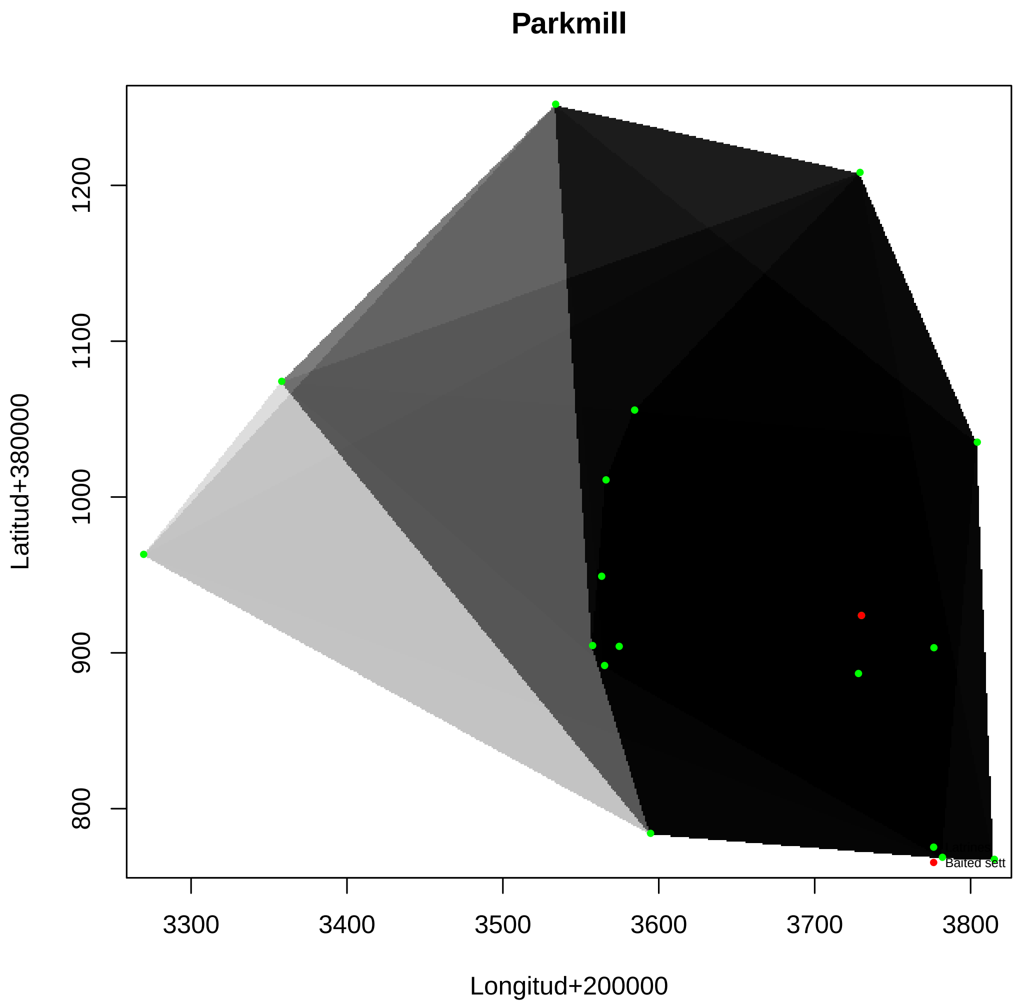


Figure 3.9: Probability of inclusion in the Parkmill territory using the sampling method of the Unconditional Outlier Prediction Model

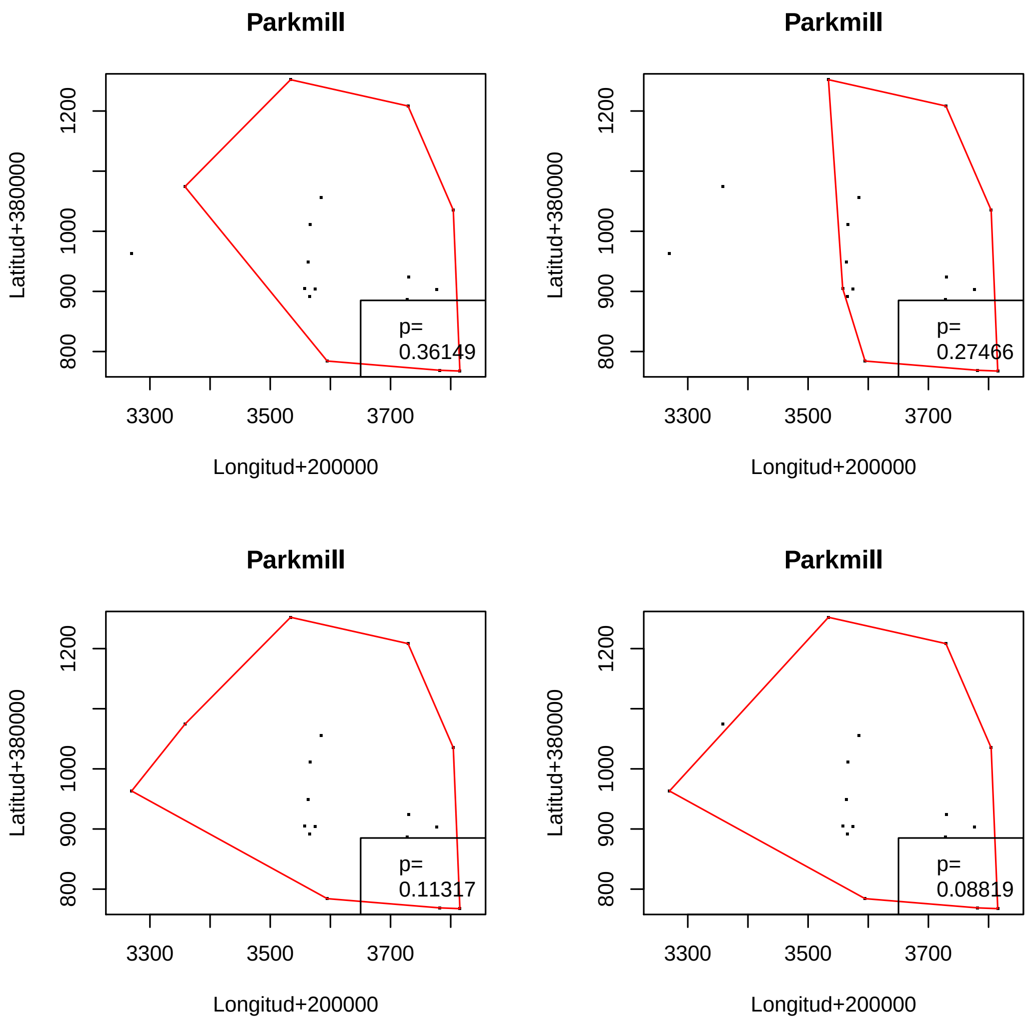


Figure 3.10: Most frequent reconstructions of the Parkmill territory using the sampling method of the Unconditional Outlier Prediction Model

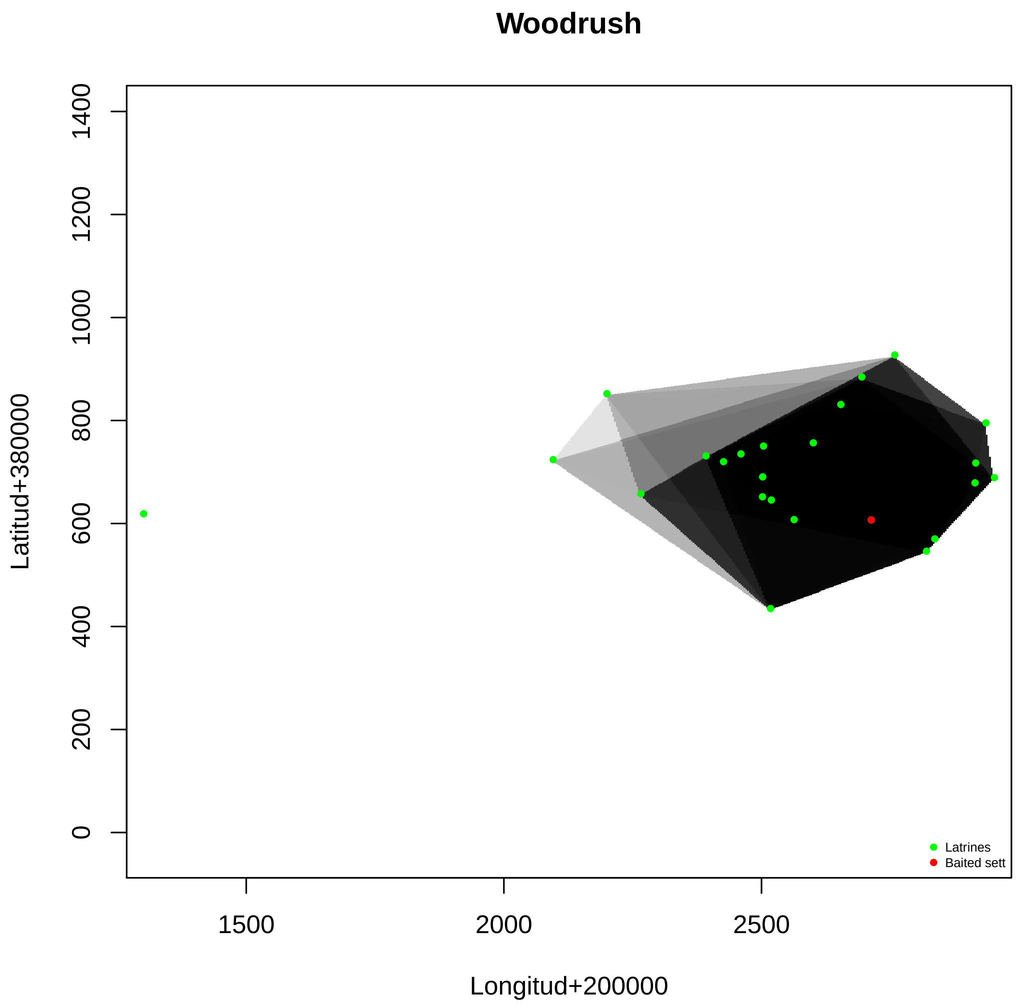


Figure 3.11: Probability of inclusion in the Woodrush territory using the sampling method of the Unconditional Outlier Prediction Model

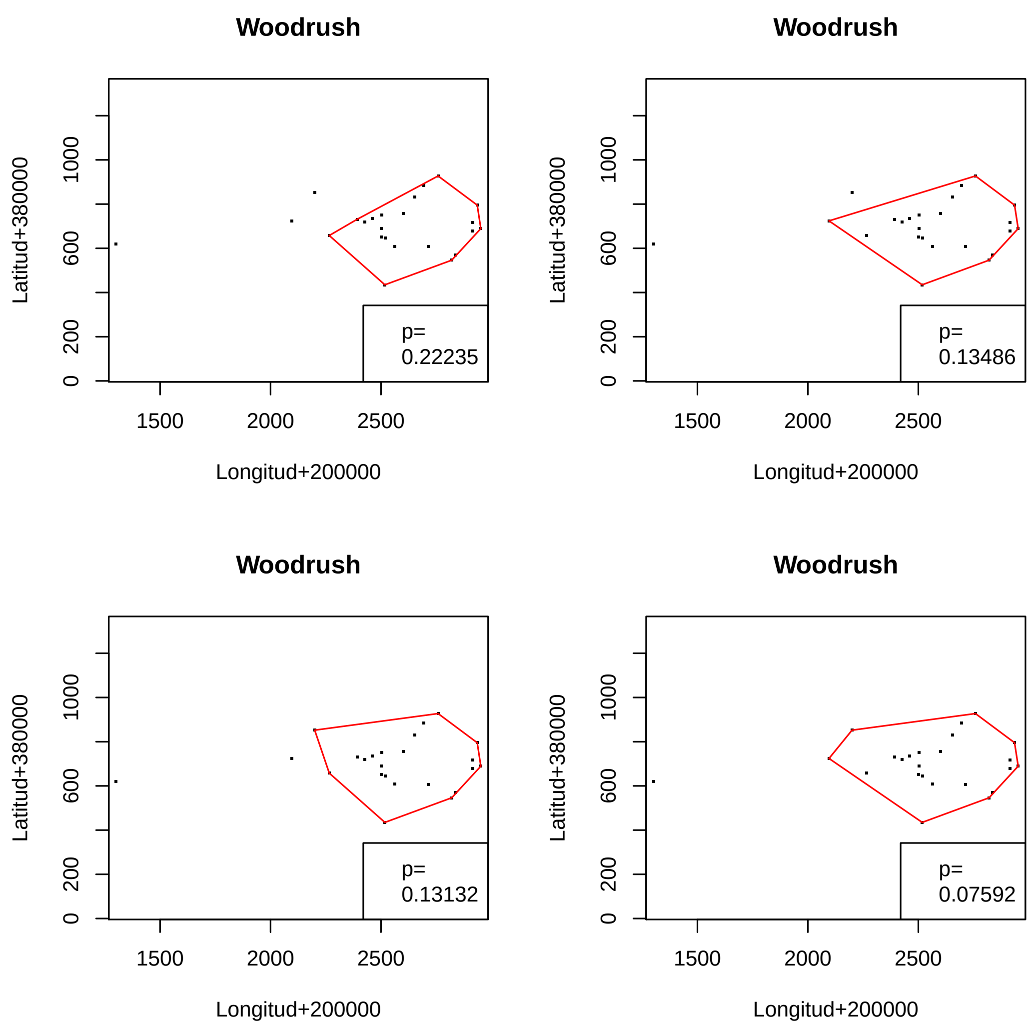


Figure 3.12: Most frequent reconstructions of the Woodrush territory using the sampling method of the Unconditional Outlier Prediction Model

Chapter 4

Conditional Outlier Prediction Model

4.1 Introduction

In this chapter an extension to the Outlier Prediction Models (OPMs), called the Conditional Outlier Prediction Model (COPM) is developed. This extension is 'conditional' in the sense that we now attempt to model whether a latrine is an outlier given knowledge of whether the latrines in the same direction and associated with the same sett are in fact outliers. In isolation this is less useful than the original OPM, because in practice we do not have this knowledge when considering a new territory. However, we can use this conditional model for each latrine as the full conditional distribution for its status within a Gibbs sampling approach, thus fitting a joint model that allows for the dependence between latrines.

4.2 Model and Variables

This model predicts whether a latrine used by the baited clan is part of the territory or an extraterritorial excursion based on covariates, similar to the previous chapter Elstub Outlier Prediction Model (EOPM) and Unconditional Outlier Prediction Models (UOPMs). However, the difference in the COPM is that this model consists of two terms (α and β) (4.1) where, in the α_i term, only non-conditional variables are included (4.2) and in β_{ij} only conditional variables for neighbouring latrines were used (4.3), as explained in detail below. The status of the jth latrine is denoted by

$$I_j = \begin{cases} 0 & = & `Non-Outlier' \\ 1 & = & `Outlier' \end{cases}$$

and the probability of being an outlier is modelled as

$$Logit(P(I_i|I_j, j \neq i)) = \alpha_i + \sum_{\substack{i \sim j, \\ I_j = 0}} \beta_{ij} \tag{4.1}$$

where:

$$\alpha_i = c + ad_i + br_i + eS_i + f\frac{d_i}{\frac{1}{n_i}\sum\limits_{i\sim j} d_j} + g\frac{1}{n_i}\sum_{i\sim j} d_j \tag{4.2}$$

$$\beta_{ij} = h(d_i + d_j) + k(d_i - d_j)^2 \tag{4.3}$$

$i \sim j$ indicates that latrines i and j are 'neighbours' (defined more precisely below) and n_i is the number of neighbours of $latrine_i$. The same definition of neighbours is used in both equations (4.1) and (4.2), but only is the conditional term does the status of neighbours have an effect.

1 In the non-conditional part (α_i) all the variables are calculated whether they are part of the territory or outliers. Variables included are: distance from the sett to the latrines (d), number of returns (r) and shared (s). In addition two calculated variables are part of the α part of the model that only use the distance of the latrines in the same direction of the latrine which value is being calculated angular fit ($AngFit$) (explained in detail in the literature review on page 13) and Mean of Distances of Neighbouring Latrines (MDnl). A graphical example for ($AngFit$) is shown on figure 4.1b and in figure 4.1a the latrines used to calculated its mean is presented.

2 The conditional part of the model (β_{ij}) involves variables with values that are conditional on the state of the neighbouring latrines; these variables are symmetrical as explained later. Terms involving $(d_i + d_j)$ and $(d_i - d_j)^2$ respectively in (equation 4.3). An example of one latrine can be seen in Figure 4.1c.

Only variable interactions within of each part of the models were tested.

where:

Neighbouring latrines ($i \sim j$) Latrines in a similar direction of $latrine_i$; latrines included in a 2θ angle swath centred on $latrine_i$ as shown on figure 3.4, joined to the baited sett by a doted blue line and circled in blue, and the $latrine_j$ are represented in red, bounded by doted-dark red lines. For the $AngFit_i$ the minimum distance in this example was 90m; a black dotted circle surrounding the sett represents it.

Conditional variables To calculate the value of these variables, only those the latrines in the same direction (neighbouring latrines) that are not 'outliers' were used. To ensure consistent joint distribution, only 'symmetrical' variables can be used. Two variables with this characteristic were used: the sum of all the $latrine_j$ with respect to the

$latrine_i$ and their squared differences (4.3). An example is given in figure 4.1c where the $latrine_j$ are in red, the $latrine_i$ in green and circled in blue.

Use of transformed variables Two variable transformation were tested: the log of $Distance$, and the squared root of $NumberofReturns$. (Box and Cox, 1964).

The optimal model was estimated using the original values and the log transformation of $Distance$. The variable Number of Returns was used as factor (as explained in the previous chapter), as quantity and the squared root of Number of Returns when this was measured as quantity, selecting the best model based on the Akaike Information Criterion (AIC) as shown in table 4.1.

Model	No. of Returns as	Distance			Log(Distance)		
		MinDist	Angle	AIC	MinDist	Angle	AIC
OPM	Quantity	90	46	267.98	90	46	269.10
	Factor	90	46	265.39	90	46	267.10
COPM	Quantity	90	42	269.18	90	55	263.64
	Factor	184	85	268.07	90	55	262.04
Elstub	Factor	40	45	286.01	-	-	-

Table 4.1: AIC values for the different models

4.3 Joint Distribution

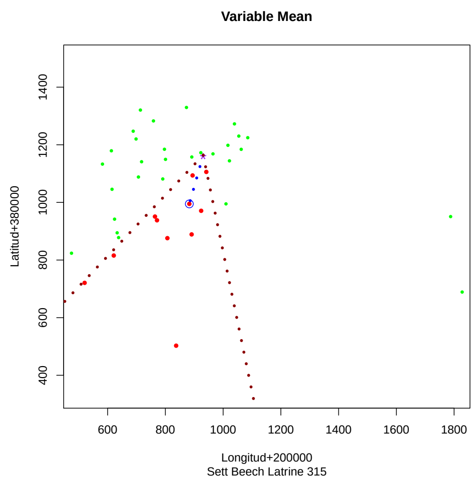
4.3.1 Gibbs Sampling

The COPM models the probability of each latrine being an outlier given all others, that is the 'full conditional' distribution of each I_j. Given these distributions, a Gibbs sampling can be used to sample from the joint distribution of the I_j's.

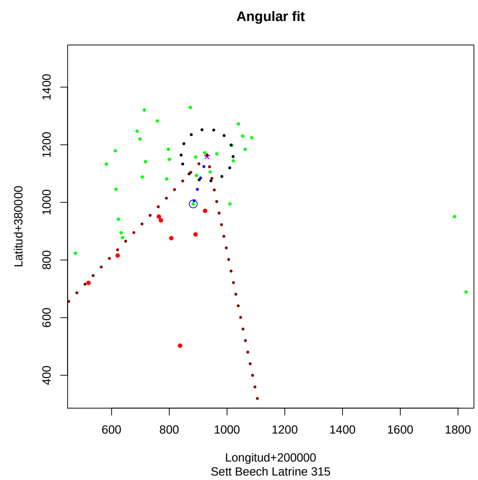
R (R Core Team, 2014) was used to carry out these simulations. 200,000 iterations were obtained for each territory, 100,000 using as initial state of the latrines (non-outlier, outlier) based on the reconstruction of the territories by the Food and Environmental Research Agency (FERA) presented in figure 3.1, cycling through all their latrines. For the other 100,000 iterations, the initial values were the opposite ones.

Convergence

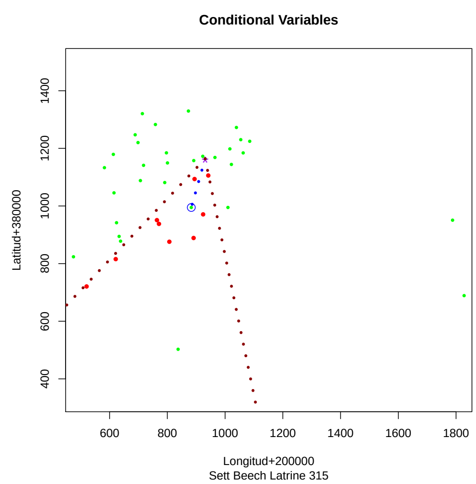
Territory Beech was selected to illustrate some examples about the convergence of the Markov chain Monte Carlo (MCMC). Four latrines were picked (347, outlier; 340, boundary; and 407 and 312, hinterlands) to present graphically using all the interactions. The Brooks, Gelman, & Rubin convergence diagnostics were run on all the latrines of Beech territory using a 50% burn-in. The library boa (Smith, 2007) was used to for convergence diagnostics.



(a) Latrines used to estimate the Mean



(b) Latrines used to estimate Angular Fit



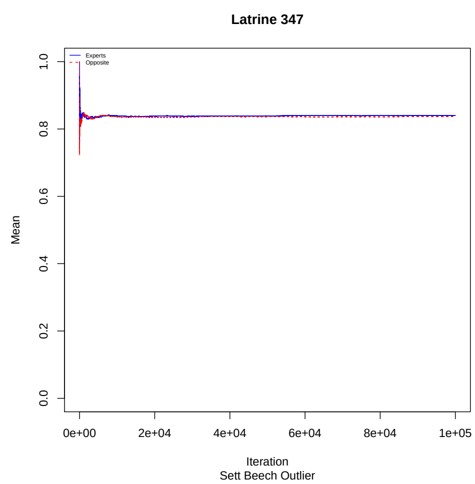
(c) Latrines used to estimate the Conditional Variables

Figure 4.1: Example of Latrines of the same direction used in the estimation of the COPM

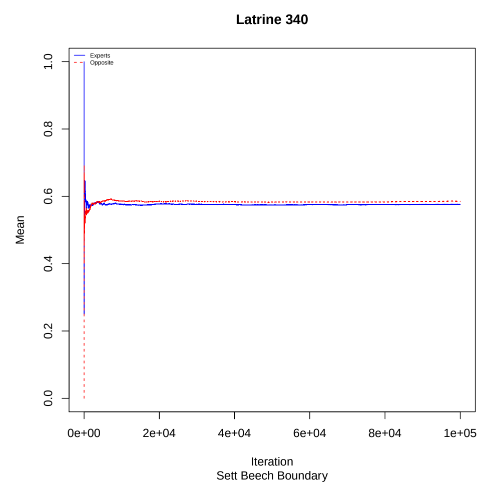
.

Graphically, in figure 4.2 four examples of the simulations are presented to show that the chains converge for both initial values. Figure 4.2a is and example of a latrine classified as an 'outlier', figure 4.2b a boundary latrine and figures 4.2c and 4.2d hinterland latrines. Blue lines are the results using the reconstruction of the territory in Figure 3.1 and the red line the opposite one. These graphs represent the estimated:

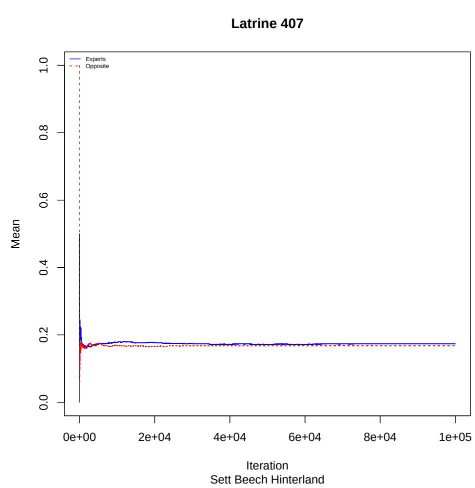
$$P(latrine_j \text{ is an outlier}) = \frac{(\text{number of iterations in which j is an outlier})}{(\text{total number of iterations})} \tag{4.4}$$



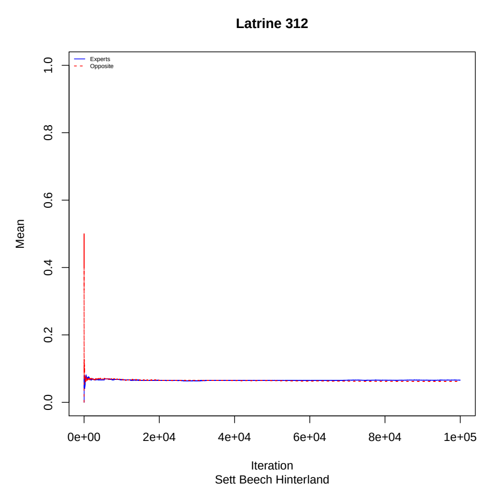
(a) Outlier



(b) Boundary



(c) Hinterland



(d) Hinterland

Figure 4.2: Examples of Latrines in the same direction used in the estimation of the COPM

Table 4.2 shows the results of the Brooks, Gelman, & Rubin convergence diagnostics. The Potential Scale Reduction Factor (PSRF) values for all the latrines and the territory are very close to 1, indicating convergence of the chains, this is corroborated when the Multivariate Potential Scale Reduction Factor (MPSRF) is used since all the values of the 0.975 quantile are smaller than 1.20 (Gelman and Rubin, 1992).

4.3.2 Reconstruction of the territories

To reconstruct the territories, the chain with the data from Woodchester Park was used, after 50% burn-in (for convergence) and a 10 lag thinning to avoid autocorrelation within each iteration. Two graphical reconstructions were generated with this last sample, as in chapter 3; as a reminder, the methods are outlined here.

1 **Probability of inclusion** After the burn-in and thinning of the sample, the frequency of each possible configuration was calculated and a 100% Minimum Convex Polygon (MCP) applied to each one. An 100% MCP is used instead of the widely 95% MCP since the sampling has already been used to take in account possible 'outliers'. A grid was created using the maximum and minimum coordinates of the territory to count the number of times each point of the grid belongs to the reconstructed territory. Once divided by the total number of the sample, a ratio was obtained to represent the uncertainty of that area and to give it a grey scale colour based on this probability.

2 **Most frequent reconstructions** A second approach is to plot the most frequent territories reconstructions individually, using the previous sample. An example is shown in figure 4.4.

Par	Latrine	PSRF	Corrected Scale Reduction Factors	
			Estimate	0.975
1	264	1.000	1.002	1.00
2	265	1.000	1.000	1.00
3	266	1.000	1.000	1.00
4	288	1.000	1.001	1.00
5	291	1.000	1.001	1.00
6	292	1.000	1.000	1.00
7	293	1.000	1.000	1.00
8	294	1.000	1.001	1.00
9	296	1.000	1.000	1.00
10	310	1.000	1.000	1.00
11	311	1.001	1.000	1.00
12	312	1.000	1.000	1.00
13	313	1.000	1.000	1.00
14	314	1.000	1.000	1.00
15	315	1.000	1.000	1.00
16	316	1.000	1.000	1.00
17	318	1.000	1.002	1.00
18	319	1.000	1.000	1.00
19	320	1.000	1.001	1.00
20	338	1.000	1.003	1.01
21	340	1.000	1.001	1.00
22	342	1.000	1.000	1.00
23	347	1.000	1.000	1.00
24	405	1.000	1.000	1.00
25	406	1.000	1.023	1.02
26	407	1.000	1.006	1.01
27	452	1.000	1.000	1.00
28	454	1.000	1.081	1.08
29	455	1.000	1.090	1.09
30	456	1.000	1.000	1.00
31	457	1.001	1.022	1.02
32	458	1.000	1.002	1.00
33	459	1.000	1.000	1.00
34	468	1.000	1.013	1.01
35	469	1.000	1.000	1.00
36	470	1.000	1.021	1.02
37	471	1.000	1.000	1.00
38	472	1.000	1.003	1.00
39	473	1.000	1.000	1.00
40	474	1.000	1.019	1.02
			1.004	

Table 4.2: Brooks, Gelman, & Rubin convergence diagnostics using the COPM to reconstruct the territory Beech.

Graphical Results

Looking for an easier comparison with the UOPM and the Adjusted Ordinal Model (AOM) introduce later, the reconstructions of the badger territories using the COPM, Beech, Junction, Parkmill and Woodrush were chosen to illustrate the COPM.

a) ***Beech***

When using the probability of inclusion approach, presented in Fig. 4.3, a darker area surrounding the main sett, can be observed, representing a higher probability that that zone is part of the territory. Uncertainty can be observed to the south and southwest of the territory with a mostly clear zone due to the most south latrine. This figure shows two latrines to the east of the territory are predicted as outliers by the model. Fig. 4.4 is the reconstruction of the most frequent configurations. The total probability of these four configurations is 0.61 with 0.54 for the first three. When the most frequent territory is compared with the other three, the difference is only one latrine, which are located on the southwest side of the territory.

b) ***Junction***

Reconstruction of Junction territory is presented on Fig. 4.5 and 4.6. In the first one uncertainty is predicted by the COPM in all the boundaries but the south one, with a smaller probability predicted in the northeast and east borders. When the probability of each reconstruction is estimated, the four more frequent territories sum to a probability of 0.59. The sum of the probabilities of the three most frequent configurations is 0.52. If the most frequent reconstruction is compared with the other three, 1 latrine is the difference with the configurations 2 and 4, and 2 with the third most probable reconstruction.

c) ***Parkmill***

The uncertainty observed in Fig. 4.7 when the territory Parkmill is made, is present in the west boundary of the territory. This can also be observed in Fig. 4.8 where, moreover, the probability of the four of the four most frequent reconstructions is presented. They total 0.89 and if only the two configurations with higher probability are considered, the total is 0.79. When the most frequent reconstruction is compared with the other three, the difference is only one latrine.

d) ***Woodrush***

Fig. 4.9 shows the reconstruction of Woodrush territory using the COPM when all the iteration (after burning and thinning) are used. Uncertainty can be observed in all the boundaries of this territory but the southeast one. It is bigger in the west side of the territory. In Fig. 4.10, the probabilities of the four most frequent reconstructions sum 0.50. From all the chosen examples, this is the only one the needed four possible reconstruction to totalize at least a probability of 0.50. The differences between this reconstructions is only 1 or 2 latrines located on the west side of the territory. The most western latrine of the territory is predicted with a low probability by Fig. 4.9 as part of the territory, however is not included in the reconstructions of Fig. 4.10.

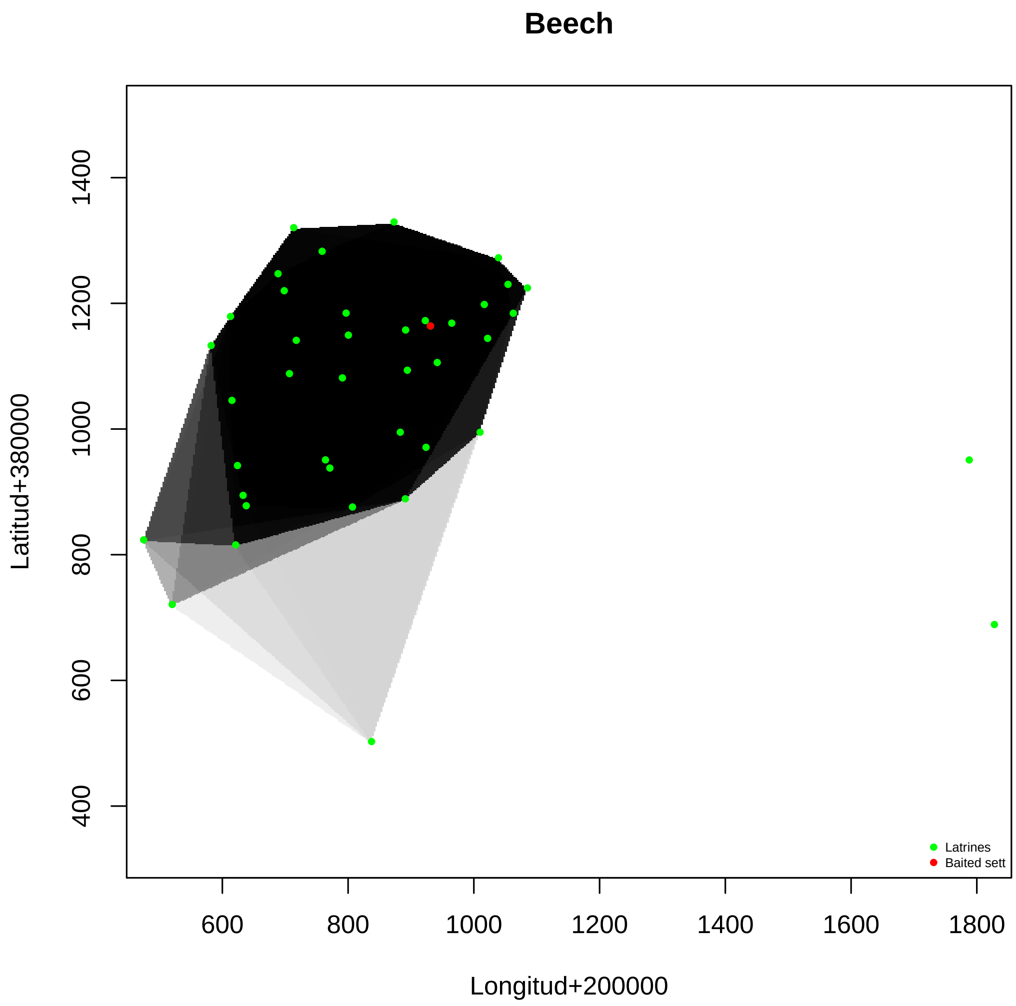


Figure 4.3: Probability of inclusion of the Beech territory using the Conditional Outlier Prediction Model

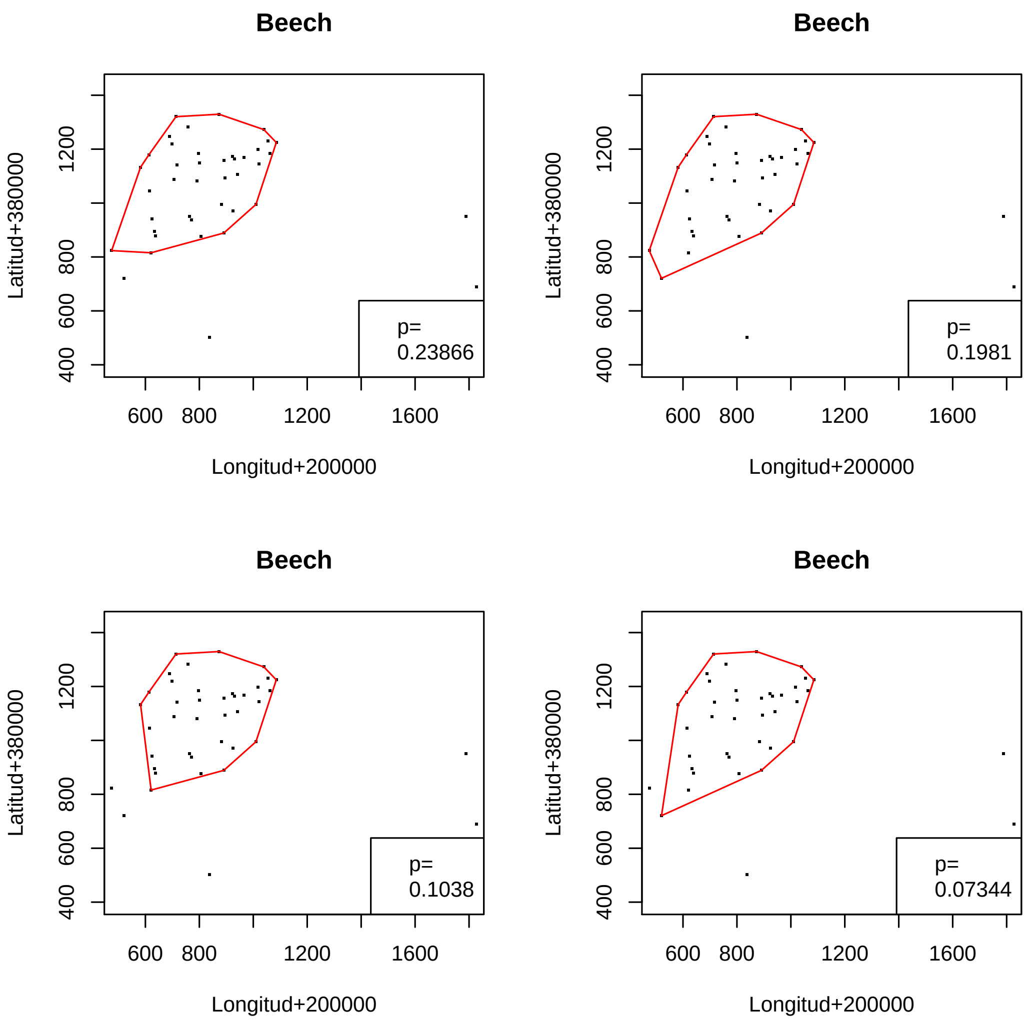


Figure 4.4: Most frequent reconstructions of the Beech territory using the Conditional Outlier Prediction Model

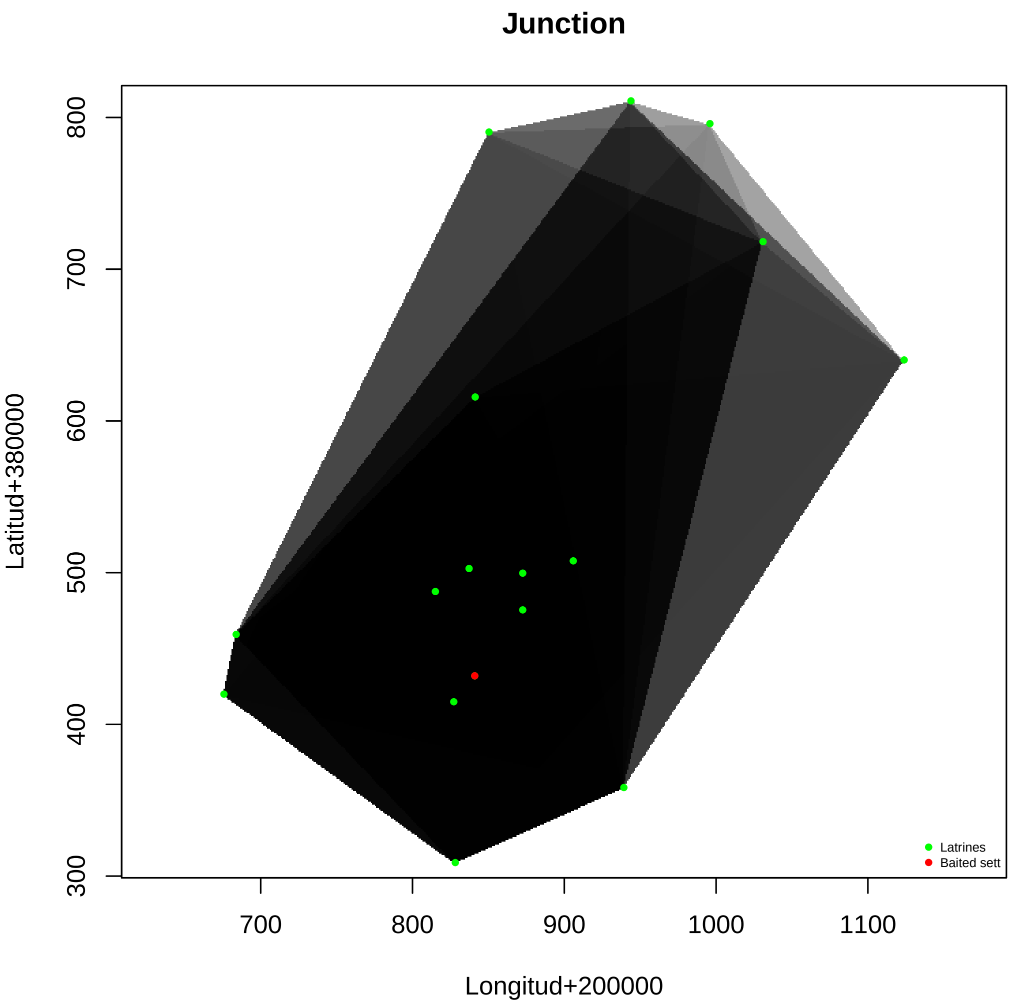


Figure 4.5: Probability of inclusion of the Junction territory using the Conditional Outlier Prediction Model

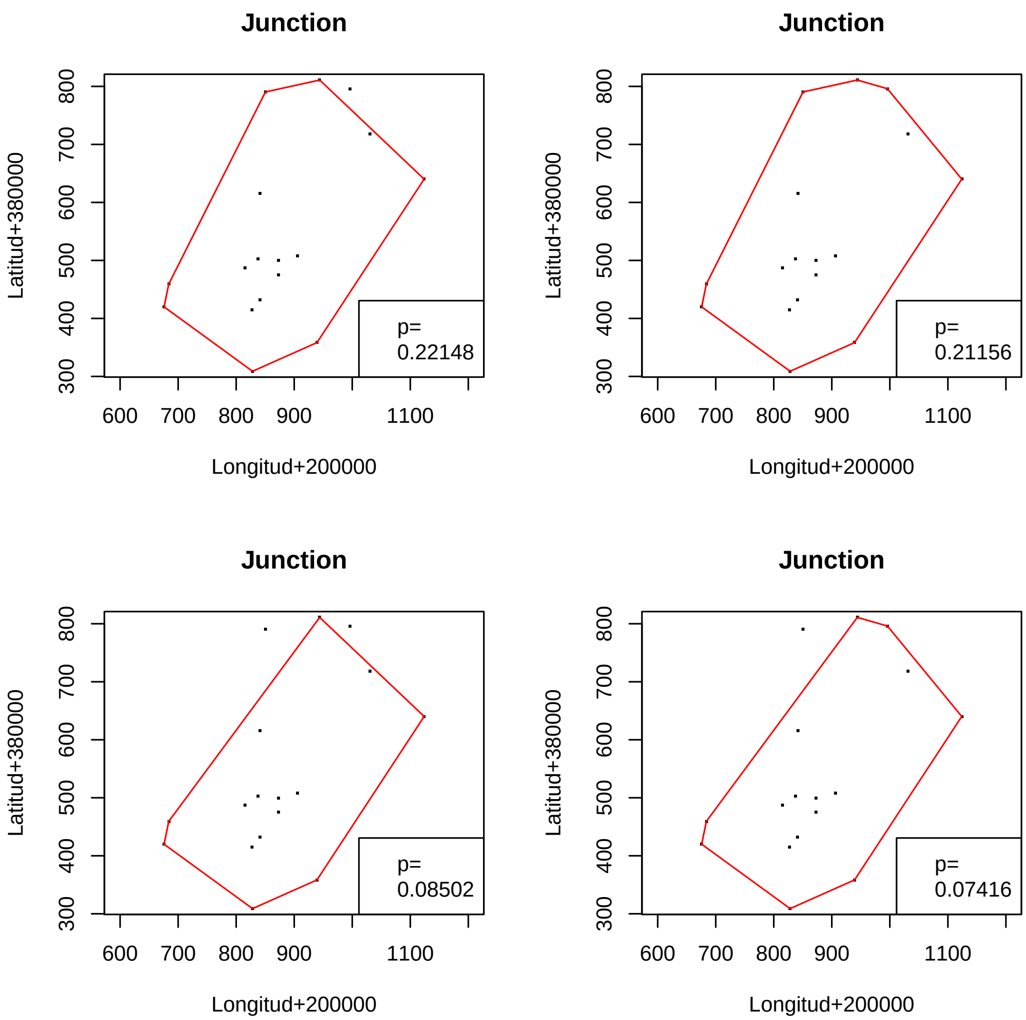


Figure 4.6: Most frequent reconstructions of the Junction territory using the Conditional Outlier Prediction Model

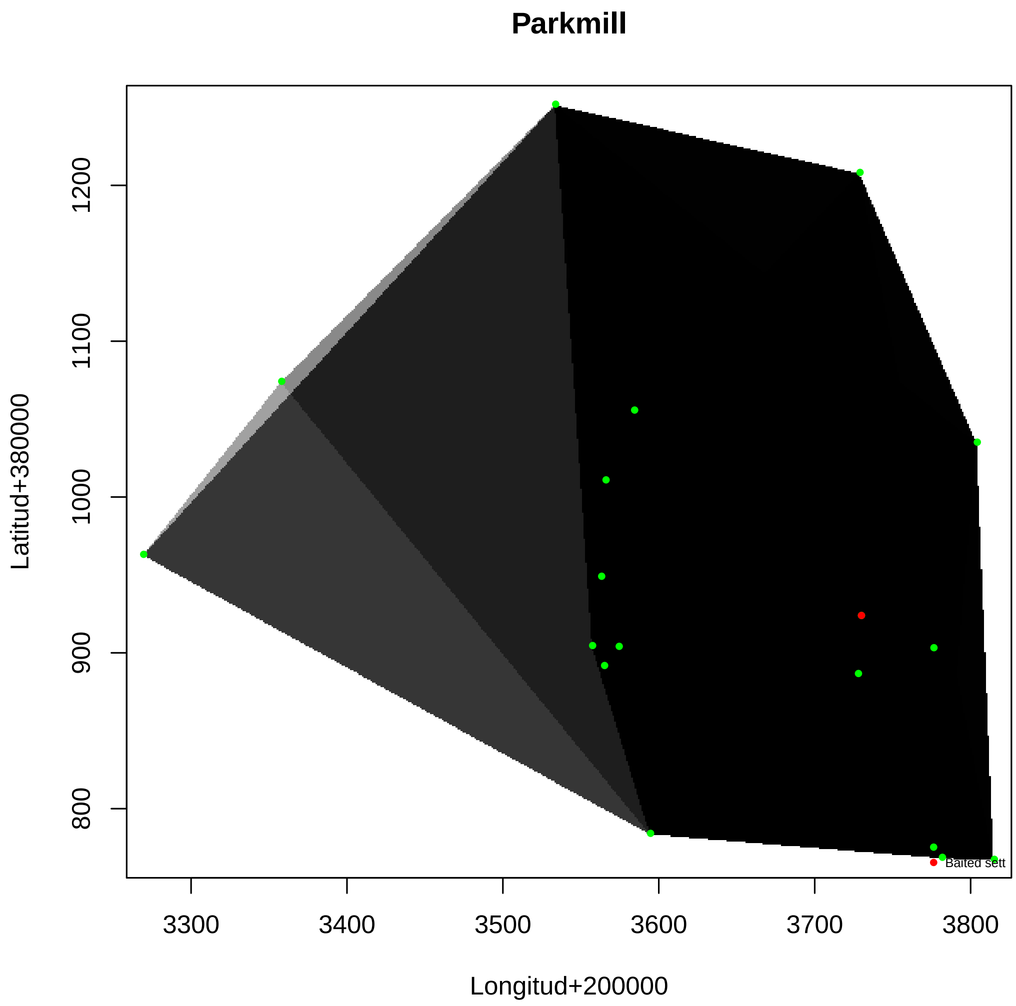


Figure 4.7: Probability of inclusion of the Parkmill territory using the Conditional Outlier Prediction Model

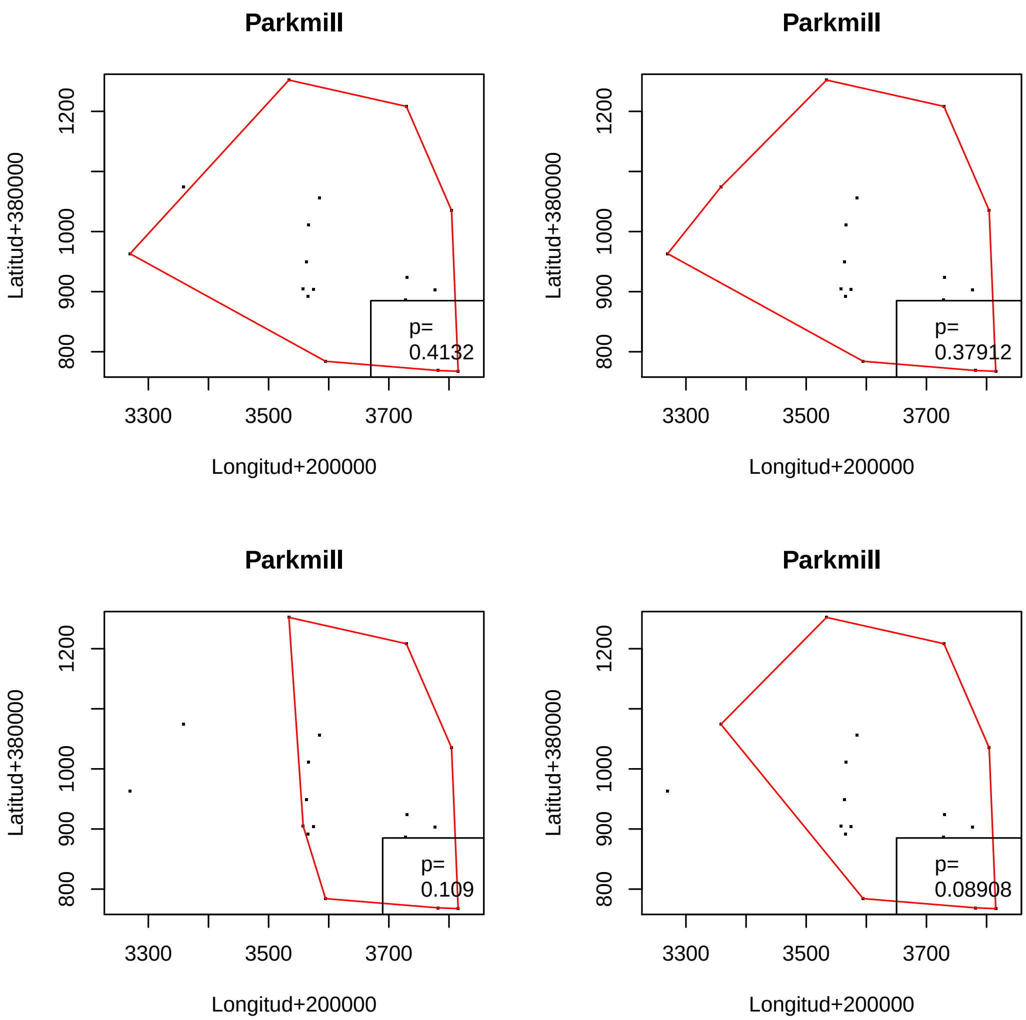


Figure 4.8: Most frequent reconstructions of the Parkmill territory using the Conditional Outlier Prediction Model

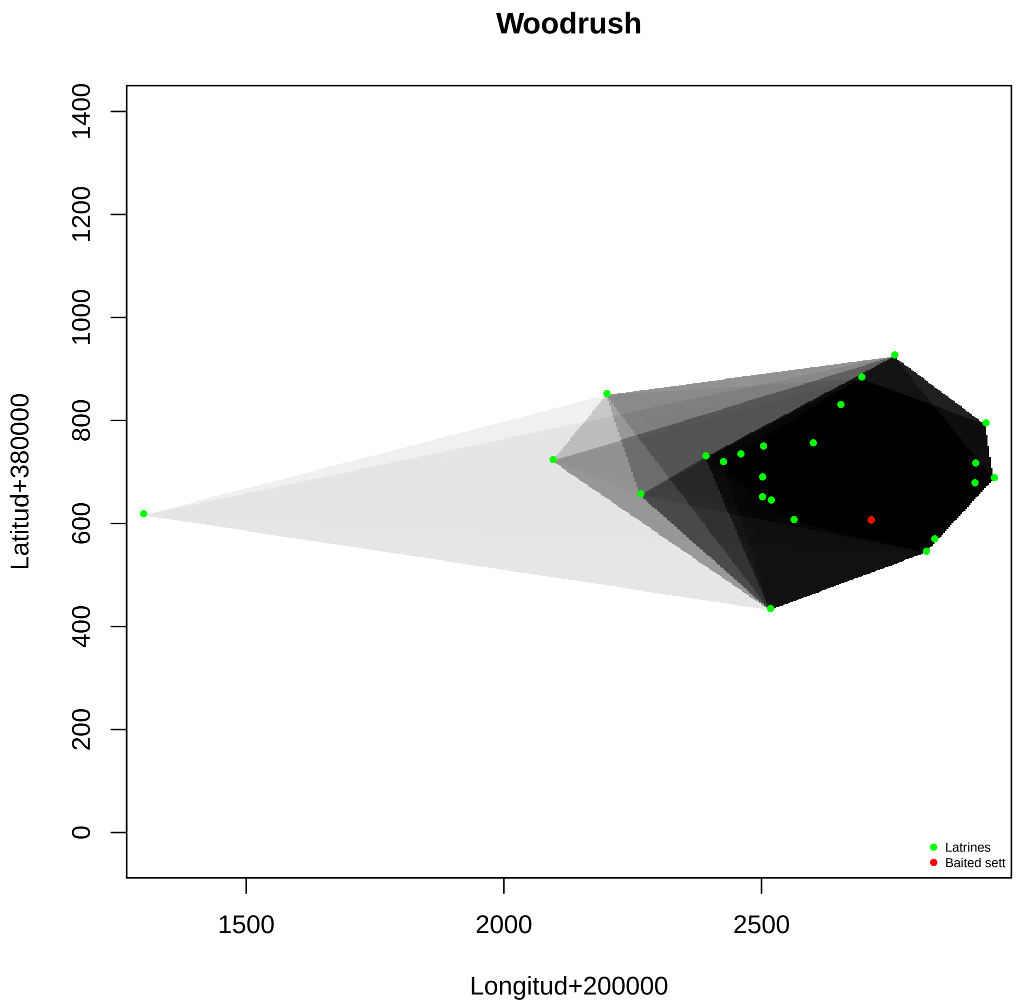


Figure 4.9: Probability of inclusion of the Woodrush territory using the Conditional Outlier Prediction Model

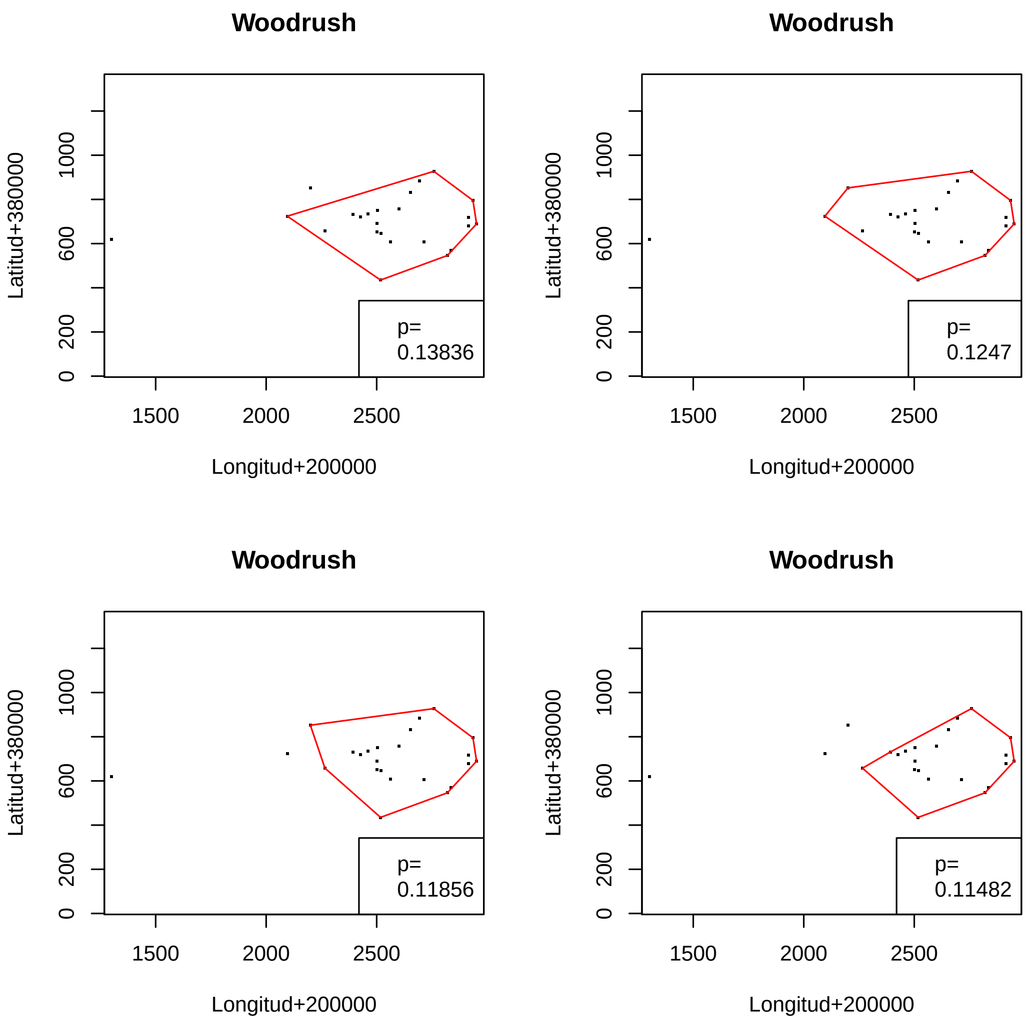


Figure 4.10: Most frequent reconstructions of the Woodrush territory using the Conditional Outlier Prediction Model

4.4 Discussion and Conclusions

When the most frequent reconstruction for each of the examples of the territories is carried out, for *Beech*, the reconstructions are the same and they are in the same order. However, the probabilities are different with 0.36, 0.24, 0.08 and 0.05 (Fig. 3.6) respectively for the UOPM and 0.24, 0.20, 0.10 and 0.07 for the COPM (Fig. 4.4). When the reconstructions using the UOPM and the COPM of the territory *Junction* are compared, only the two more frequent reconstructions are the same (Fig. 3.8 and Fig. 4.6). Moreover, the order and probabilities are different since the most frequent reconstruction for the UOPM has a probability of 0.43 whilst for the COPM is the second most frequent reconstruction with a probability of 0.21. The second most frequent reconstruction of the UOPM has a probability of 0.24, while the same reconstruction is the most frequent for the COPM with a probability of 0.22. The other two reconstructions for both methods are different. The 4 most frequent reconstructions of *Parkmill* territory of the UOPM and the COPM are the same with different order and probabilities: reconstruction 1 (p=0.36) of the UOPM is the same as the 4 (p=0.09) of the COPM; 2 (p=0.27) is the 3 (p=0.11); 3 (p=0.11) is the 1 (p=0.41); and 4 (p=0.09) is the 2 (p=0.38) respectively, as can be observed in Fig. 3.10 and Fig. 4.8. When *Woodrush* territory is compared, the 4 most frequent reconstructions are the same with different order or probability. Reconstruction 1 (p=0.22) of the UOPM is reconstruction 4 (p=0.11) of the COPM; 2 (p=0.13) is 1 (p=0.14); 3 (p=0.13) is 3 (p=0.12); and 4 (p=0.08) is 2 (p=0.12) respectively. When the sampling approach is used, the COPM needs more configurations to account for a probability of at least 0.50. This means that the COPM permits a wider range of reconstructions, so the best have smaller probabilities and more are needed to reach this 0.5 cumulative probability.

The COPM is a very different model from the EOPM and the UOPM in the way that the relationship between latrines is modelled. However, even when the models are completely different internally, they are statistically comparable using their AIC's (Kass and Raftery, 1995). This comparison shows that the COPM is a much better fit than the other two models. The conditional variables and Gibbs Sampling approach to reconstructing the territories allow the COPM to incorporate dependance between latrines. Since this is the main difference between the three models (minimum distance is the same and the Angular fit is very similar), the improvement on the AIC's value, is due to this dependence giving information that, when a latrine is used, the odds of neighbouring latrines to be used, are increased. On the other hand, one disadvantage is that the COPM is a very time consuming method to apply, compared with these other two methods.

All the models so far rely on MCP in the reconstruction, if a concave area is present, it will not be possible to be taken in consideration: all reconstructions are necessarily convex. For this reason, in the next chapter, a different latrine classification is proposed rather than outlier and non-outlier. Moreover, the COPM has more biological implications since the probability of using a neighbouring latrine is bigger than using other one; this is because, as cited in the literature review, even when badgers defend, as clan, a territory, they forage alone going to specific foraging areas, using latrines in the same direction.

Chapter 5

Unadjusted Ordinal Model

5.1 Introduction

The Unconditional Outlier Prediction Model (UOPM) in Chapter 3 and the Conditional Outlier Prediction Model (COPM) from Chapter 4 use information obtained from a logistic regression to estimate a probability, given some covariates, that a latrine is an extraterritorial excursion or 'outlier'. Converting the information about the latrines into a map of the territory has been done by applying the idea of a Minimum Convex Polygon (MCP), producing, for this reason, always convex territories. However some territories are known for presenting concave areas that these methods are unable to predict. Since data provided by 2010 Food and Environmental Research Agency (FERA)'s Woodchester Park baitmarking results (Figure 3.1) gives a natural ordered classification for the latrines (hinterland, boundary and outliers) a model for ordinal response was explored to tackle the problem of reconstructing territories with concave zones.

5.2 Ordinal Models

There are two types of categorical variables: nominal and ordinal. In the previous chapters the latrines were classified on a two-level scale, which can therefore be seen as nominal; we could label either class as 'success' *(Territorial and Outliers latrines)*. This classification did not allow us to model concave areas if they should be present in the reconstruction of the territories since it gives no indication of whether latrines lie on the boundary, if they are not outliers. This means that some other method is needed to determine the boundary; we have use the 100% MCP preventing the generation of concave polygons. An example of this can be viewed on Fig. 3.6 UOPM and on Fig. 4.4 COPM of the Beech territory. To tackle this problem, a slightly different approach was followed, classifying the latrines using ordinal variables. From Fig. 3.1 where stars represent setts, red dots latrines, colour lines joining a sett with a latrine represents a baited sett an the latrine where the plastic pellets were recovered (i.e. latrine used by that clan), black lines are boundaries of the territory, so latrines joined by a black lines are *boundary latrines*, latrines used by a clan inside this boundary are *hinterland latrines* and those outside the territory are *outlier latrines*. From now on, we use an ordinal response with these three levels: hinterland, boundary and outlier.

For binary response variables, in most fields logistic regression has become the standard model for analyzing the effects of explanatory variables. An extension for these models is forming logits for an ordinal scale which applies to response cumulative probabilities (explained below) (Agresti, 2010).

The most popular class of ordinal regression models are *cumulative link models* (explained below) that treat the response variable as categorical, providing the regression framework from linear models.

5.2.1 Cumulative logits

One of the approaches used to analyze ordered categorical response variables is using only the ordering information about the categories. One example of this approach is the model for cumulative response probabilities which is possibly the most popular model for ordinal data. Cumulative probabilities are used up to a threshold, making binary the ordinal categories at that threshold. By subtracting adjacent cumulative probabilities, the individual category probabilities can be calculated.

Cumulative logit models simultaneously fit all $J-1$ logit models for the J categories of the response. The individual category probabilities are found by subtracting adjacent cumulative probabilities.

When response categories are ordered, the logits can utilize the ordering. This results in models that have simple interpretations. A cumulative link model is a model for an ordinal response variable, Y that can fall in $j = 1, \ldots, J$ categories, where $J \geq 2$. If $J = 2$ binomial models also apply, and in fact the cumulative link model is in this situation identical to a generalized linear model for a binomial response. A *cumulative probability* for Y is the probability that Y falls at or below a particular point. For outcome category j, the cumulative probability is

$$P(Y \leq j) = \pi_1 + \cdots + \pi_j, j = 1, \ldots, j$$

The ordering is reflected by the cumulative probabilities, with $P(Y \leq 1) \leq P(Y \leq 2) \leq \cdots \leq P(Y \leq J) = 1$. Since the final cumulative probability , $P(Y \leq J)$, always necessarily equals 1, this model does not use it. The logits of the cumulative probabilities are

$$\begin{aligned} logit[P(Y \leq j)] &= \log\left[\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right] \\ &= \log\left[\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_j}\right], \\ j &= 1, \ldots, j-1 \end{aligned} \tag{5.1}$$

These are called *cumulative logits*. For $J = 3$, for example, models use both $logit[P(Y \leq 1)] = \log[\pi_1/(\pi_2 + \pi_3)]$ and $logit[P(Y \leq 2)] = \log[(\pi_1 + \pi_2)/\pi_3]$. All the response categories are utilized for each cumulative logit.

Cumulative Logit Models with Proportional Odds Property

A model for the cumulative logits that assumes the same covariate effects for each logit but different intercepts is called a *Cumulative Logit model with proportional odds.* When covariates or explanatory variables are incorporated, the following model can be presented. For subject i, let y_i denote the outcome category for the response variable, and let $\mathbf{x}_i$ denote a column vector of the values of the explanatory variables. The model simultaneously uses all $J-1$ cumulative logits having the form

$$logit[P(Y_i \leq j)] = \alpha_j + \boldsymbol{\beta}'\mathbf{x}_i = \alpha_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \tag{5.2}$$

for $J = 1, \ldots, J-1$. The effects of the explanatory variables are described by a column vector $\boldsymbol{\beta}$ of parameters. Each cumulative logit has it own intercept. The reason why α_j are increasing in j is that $P(Y \leq j|\mathbf{x})$ is increased in j for fixed x and the logit is an increasing function of $P(Y \leq j \mid x)$. The same effects $\boldsymbol{\beta}$ are assumed in this model for each logit. The cumulative logit model satisfies 5.2

$$\begin{aligned} & logit[P(Y \leq j|\mathbf{x}_1)] - logit[P(Y \leq j|\mathbf{x}_2)] \\ & = log\left[\tfrac{P(Y \leq j|\mathbf{x}_1)/P(Y > j|\mathbf{x}_1)}{P(Y \leq j|\mathbf{x}_2)/P(Y > j|\mathbf{x}_2)}\right] \\ & = \boldsymbol{\beta}'(x_1 - x_2) \end{aligned} \tag{5.3}$$

An odds ratio of cumulative probabilities is called a *cumulative odds ratio*. The odds of making response making response of j or below at $\mathbf{x} = x_1$ are $\exp[\boldsymbol{\beta}'(x_1 - x_2)]$ times the odds at $\mathbf{x} = x_2$. The log cumulative odds ratio is proportional to the distance between x_1 and x_2. The same proportionality constant applies to each logit. Because of this property, model 5.2 is often called a *proportional odds model* (McCullagh, 1980).

We used the *Cumulative Logit model* to classify the latrines, as mention above, as *hinterland, boundary and outlier*, without considering the information of neighbouring territories, only the information of the clan (sett, latrines used by the sett and its covariates mentioned below) calling it ***Unadjusted Ordinal Model***.

5.3 Model Fitting and Selection

Based on the classification given by FERA 2010 baitmarking Woodchester Park survey shown in Fig. 3.1, three types of latrines can be found: *hinterland*, *boundary* and *outlier*. The idea of using the type of latrine as an ordinal variable was because the territory is made by hinterland and boundary latrine and the former are always inside the territory and the latter in its edge. The classification of the latrine depends of covariates mentioned below and explained in detail in previous chapters.

A Cumulative Logit Model with the proportional odds property (equation 5.3) explained above was fitted to estimate the probabilities of a latrine of being hinterland, boundary and outlier, based on its distance to the main sett, angular fit, mean of latrines in the same direction, number of returns and if the latrine is shared with other clans or not.

The model was fitted using the library VGAM (Vector Generalized Linear and Additive model) which can be used to estimate classical categorical regression models such as the multinomial logit and proportional odd models (Yee, 2010). The Akaike Information Criterion (AIC) (Akaike, 1978) was used to select the best model.

As in the Outlier Prediction Models (OPMs) of the last two chapters, the *minimum distance* to calculate the variable *Angular fit* was varied from 1m to 500m. In the same way the *angle* to consider what latrines are considered *neighbouring latrines* to calculate the variables angular fit and mean of latrines in the same direction, was varied from $10°$ to $360°$. The best AIC value (777.78) was obtained with a *minimum distance* of 85m and a $2\theta = 22°$.

This approach is exactly the same that the UOPM (chapter 3) except that the response variable is ordinal rather that binomial.

$$I_j = \begin{cases} 1 & = & `Hinterland' \\ 2 & = & `Boundary' \\ 3 & = & `Outlier' \end{cases}$$

The idea of this classification is due to a latrine that is part of the territory, can be a hinterland or a boundary latrine and this could depend on the explanatory variables $\mathbf{x}_i$ (explained in detail on sections 2.3.3 and 4.2) in addition to the outlier classification:

- Distance from the baited sett to the latrine (d).
- Angular fit (AF).
- Mean of distances of latrines in the same direction (Md).
- Number of returns: as numeric (transformed and un-transformed) an as indicator (r).
- Shared (S).

Table 5.2 presents the estimates of the coefficients, where the five covariates were selected and its interactions (5 level interaction) which gave the better AIC value when the model selection was carried out. This table presents the estimation of two intercept coefficients. The *Intercept 1* (1.90) is used to calculate the probability of the latrine I_j is a *hinterland latrine*, p(H), and the *Intercept 2* (2.35) is used to calculate the probability the same latrine I_j of being a *boundary latrine*, p(B). Since the total probability is 1, the probability of *outlier latrine*, p(O), for the same latrine I_j, is: $p(O) = 1 - [p(H) + p(B)]$. Due to the model fitted is a cumulative logit with proportional odds property, the coefficients are the same for each intercept.

5.4 Reconstruction of the Territories

The estimated model was used to fit the probabilities, for each latrine, to be a *Hinterland*, *Boundary* or *Outlier*. These probabilities were used in a sampling approach, and the uncertainty quantified. The idea is to sample sets of latrine labellings (hinterland, boundary or outlier) from the probability distribution fitted by the Unadjusted Ordinal Model (UOM), and use the reconstructions based on them to demonstrate the uncertainty of the model.

Each sample represents a territory reconstruction. This reconstruction was made joining the latrines labelled as *boundary*, after sorting them by their angle, clockwise, using the main sett as reference. 100,000 samples were obtained for each territory. Clearly there are other ways in which a polygonal territory could be constructed with the selected boundary latrines. However, with large numbers of boundary points, such polygons in general may have rather complicated shapes, and this method of construction addresses that problem. It is essentially the requirement that the territory be star-shaped with respect to the main sett; that is, every boundary latrine can be reached directly from the main sett, in a straight line, without leaving the territory. (Of course, this property does not apply if the main sett is not inside the territory constructed – see below – but in that case, no boundary based on the selected latrines will include the main sett.) It also implies that the boundary will be 'simple' in that it will not intersect itself. The over-complicated shapes excluded by this method will often be non-convex (since any convex shape is star-shaped with respect to any point that it contains) but many of those that are constructed in this way are still non-convex, as shown below.

Each sample produces a possible territory that was checked for validity according to the next criteria:

- The main sett must be part of the territory. This means that, after joining the latrines labeled as boundary, a territory was reconstructed. Since biologically the main sett is always part of the territory, all reconstructions in which the main sett was not included as part as the territory, were considered as *'not valid'*.
- The second criteria is that after reconstructing the territory using the latrines labeled as boundary, all the latrines labelled as *hinterland* must be inside the reconstruction, otherwise the territory was considered as *'not valid'*.
- Finally, the third requirement for a *valid reconstruction* is that none of the latrines labelled in the sample as *outlier* must be inside the reconstruction.

The last two criteria were analyzed in one single step in R (R Core Team, 2014). The resulting territories with valid configurations were summarized using the two graphical approaches explained in detail in the previous two chapters:

- Most frequent reconstructions.
- Probability of inclusion.

The idea of an alternative model to the UOPM and the COPM – that were developed on the previous two chapters – to reconstruct badger territories that was able to predict concave areas in territories using and an ordinal model approach was reached partially using the UOM. The model was able to predict completely convex territories versions and reconstruct versions with concave areas of the same territory. However, after running 100,000 iterations to reconstruct the territories, 2 territories (Junction and Wood Farm) did not produce a valid territory using the criteria that the main sett must be part of the territory. To this reconstructions that met the 'main sett criteria', the second and third ones was applied ('All the *hinterland latrines* must be part of the territory and no *outlier latrines* must be part of the territory'), with only three territories being able to meet this criteria: Boxwood with 61 valid reconstructions, Mead with 42 and Woodrush with 15 as shown in table 5.1.

One of the advantages of presenting the most probable reconstructions one at a time, is that each possible territory can be observed and the uncertainty estimated by the model interpreted easily on the graphs of the reconstruction of the territories. Moreover, in this case, where the UOM is being evaluated and compared with the previous proposed models to see if it helps to avoid their weaknesses.

This step of checking the validity of territories needs to be carried out separately from the initial classification of latrines, since the classification is done at the level of the individual latrine. While integrating the two would be desirable, it would require a radically different approach from the current one.

Graphical results

Four territories (Beech, Junction, Parkmill and Woodrush) are used to present the results on the UOPM, the COPM and the Adjusted Ordinal Model (AOM). The exception is the UOM since only Boxwood, Mead and Woodrush were able to present valid reconstruction by the UOM. An example of a invalid territory (Parkmill) can be observed in Fig. 5.2, where the baited sett, represented by a dark red '*', is not part of the territory in the most frequent reconstruction.

a) ***Boxwood***

Figure 5.3 is a good example about the ability of this method to reconstruct convex territories, as in Fig. 5.3b and concave areas of them as in the rest of its sub figures, being this the main objective to tackle from the UOPM and the COPM. The main disadvantage of this method can be observed in this figure as well, even the highest probability of the different reconstructions, is very low, only 0.033 (Fig. 5.3a. The other three reconstructions have the same value. Presumably this is because each of them occurred exactly twice in the 61 valid samples. In the probability of inclusion approach, where all the reconstructions were used to produce Fig. 5.4, higher probabilities (darker grey zones) that an area is part of the territory are around the baited

Sett	Unadjusted	
	MS	HO
Beech	62638	0
Boxwood	29204	61
Box	24721	0
Cedar	47124	0
Collliers	18330	0
Honeywell	37351	0
Jacks	96350	-
Junction	0	-
Kennel	29329	0
Larch	37930	0
Mead	69320	42
Nettle	53222	0
Old Oak	38935	0
Parkmill	99709	-
Rabbit	41248	0
Septic tank	33033	0
Top	24313	0
Trackside	49082	0
West	59285	0
Wood farm	0	-
Woodrush	51303	15
Wych Elm	59285	0
Yew	72505	0

Table 5.1: Valid territories using the *Unadjusted Ordinal Model* to reconstruct badger territories from Woodchester Park after 100,000 samples. Column with "MS" header are the valid configurations when the criteria of *Main Sett* was applied. Column with "HO" are the valid reconstructions when the localization of hinterland and outlier latrines requirements are applied.

sett (red dot), especially because one of the restriction for a valid territory is that the main set is always part of the territory. This reconstruction shows a black area surrounding the main sett with different grey scales in all the boundaries. However, darker greys are present to the east side of the reconstruction with paler greys to its west side except for three latrines that make that the reconstruction shows darker grey zones.

b) ***Mead***

Fig. 5.5 presents the most frequent reconstruction of this territory. All of them have the same probability: 0.024. If a pairwise comparison is made, for example, 5.5a with the other three reconstruction, it is different from 5.5b in 1 latrine, of each one of them. Maybe the most interesting comparison is with 5.5c since the difference implies a concave area for 5.5a that is not present in 5.5c.

When the 'Probability of inclusion' graph approach is made, (Fig. 5.6) shows a very well defined dark area surrounding the baited sett. Three dark grey zones are present to the east a south of the territory and a clearer area is present to the northwest and west of the territory.

c) ***Woodrush***

This territory had only 15 valid reconstructions and all of them are different, four of them are presented in Fig. 5.7. Their probability is 0.067. If one the first reconstruction of the territory 5.7a is compared with the other three, only or two latrines differ from one reconstruction to the other. As in the comparisons of the previous territories, the objective of estimating uncertainty in the boundaries of the territory that let to predict concave and non-concave areas was met as it can be observed when 5.7a and 5.7b south area is compared. When all the reconstructions are presented in the 'Probability of inclusion' plot 5.8, a dark area is presented around the main sett. According with this figure, the uncertainty is present to the northeast and west of the territory. However, the tonalities of grey are more clear to the west of the territory, indicating that the probability of these zone being part of the territory are smaller.

5.5 Conclusions and Discussion

The UOM was able to predict concave areas that the UOPM and the COPM were not able to do. This approach is good at producing valid territories when the restriction "the main sett must be part of the territory" is applied, since the UOM produced valid configurations for 21 territories which varied from 18,330 (Colliers) to 99,709 (Parkmill) in a sample of 100,000, not being able to produce valid territories for two of them (Junction and Wood farm). However, when the second restriction was applied, it only valid configurations for only 3 territories. Moreover, the territory with most valid reconstructions when only the first criterion was used, Parkmill, it is not among of the 3 that met both restrictions. The UOM valid territory reconstructions predict higher probabilities around the main sett and the uncertainty is present on the edge of the territory.

The difficulties in obtaining valid configurations seems to be related to the method used to generate the samples. In each case we generated the status of each latrine independently, with probabilities given by the ordinal model fitted, and then conditioned on the criteria for validity being satisfied. In principle this samples from the correctly conditioned distribution, but clearly can fail in practice when the valid configurations are unlikely under the ordinal model. An alternative would be to generate the status of each latrine completely at random, select the valid configurations, and finally weight them according to the probabilities from the ordinal model. We do not pursue that approach here, since in the next chapter we will modify the probabilities to take into account extra information.

Minimum convex polygons (convex hulls) are an internationally accepted, standard method for estimating species' ranges or they territories, particularly in circumstances in which presence-only data are the only kind of spatially explicit data available. One of their main strengths is their simplicity. They are used to make area statements and to assess trends in occupied habitat, and are an important part of the assessment of the conservation status of species. These estimates are biased. The bias increases with sample size, and is affected by the underlying shape of the species habitat, the magnitude of errors in locations, and the spatial and temporal distribution of sampling effort. The errors affect both area statements and estimates of trends (Burgman and Fox, 2003). The UOPM is a proposal to avoid these disadvantages.

Since one of the main reasons of having a different approach was to develop a new method that was able to detect concave areas, as the UOM does, an extension that adjusts the probabilities of each latrine to be classified as *Hinterland, Boundary* or *Outlier* by the estimated probabilities of the same latrine if it is shared by a neighbouring clan is analyzed in the next chapter of this thesis.

Parameter	Coefficient
Intercept 1	$1.899 \times 10^{+01}$
Intercept 2	$2.346 \times 10^{+01}$
returns	-7.927×10^{-01}
Shared(1)	$-1.379 \times 10^{+01}$
Distance	-3.738×10^{-02}
Angular fit (angfit)	$-1.971 \times 10^{+01}$
Mean of dist(DM)	-5.182×10^{-02}
returns:shared	$1.315 \times 10^{+00}$
returns:Dist	-9.901×10^{-03}
shared:Dist	2.498×10^{-02}
returns:angfit	$1.026 \times 10^{+00}$
shared:angfit	$1.711 \times 10^{+01}$
Dist:angfit	3.484×10^{-02}
returns:DM	1.101×10^{-02}
shared:DM	4.608×10^{-02}
Dist:DM	9.823×10^{-05}
angfit:DM	6.334×10^{-02}
returns:shared:Dist	7.428×10^{-03}
returns:shared:angfit	$-2.074 \times 10^{+00}$
returns:Dist:angfit	8.578×10^{-03}
shared:Dist:angfit	-3.238×10^{-02}
returns:shared:DM	-1.353×10^{-02}
returns:Dist:DM	1.230×10^{-05}
shared:Dist:DM	-7.872×10^{-05}
returns:angfit:DM	-1.243×10^{-02}
shared:angfit1:LDM	-7.068×10^{-02}
Dist:angfit:DM	-1.130×10^{-04}
returns:shared:Dist:angfit	-5.174×10^{-03}
returns:shared:Dist:DM	-3.811×10^{-06}
returns:shared:angfit:DM	1.707×10^{-02}
returns:Dist:angfit:DM	-7.609×10^{-06}
shared:Dist:angfit:DM	1.156×10^{-04}
returns:shared:Dist:angfit:DM	-3.714×10^{-06}
Degrees of Freedom:	1262 Total (i.e. Null); 1229 Residual
Null Deviance:	448
Residual Deviance:	711.78
Log-likelihood:	-355.89
AIC:	777.78

Table 5.2: Coefficients using the Cumulative Logit Models with proportional Odds Property

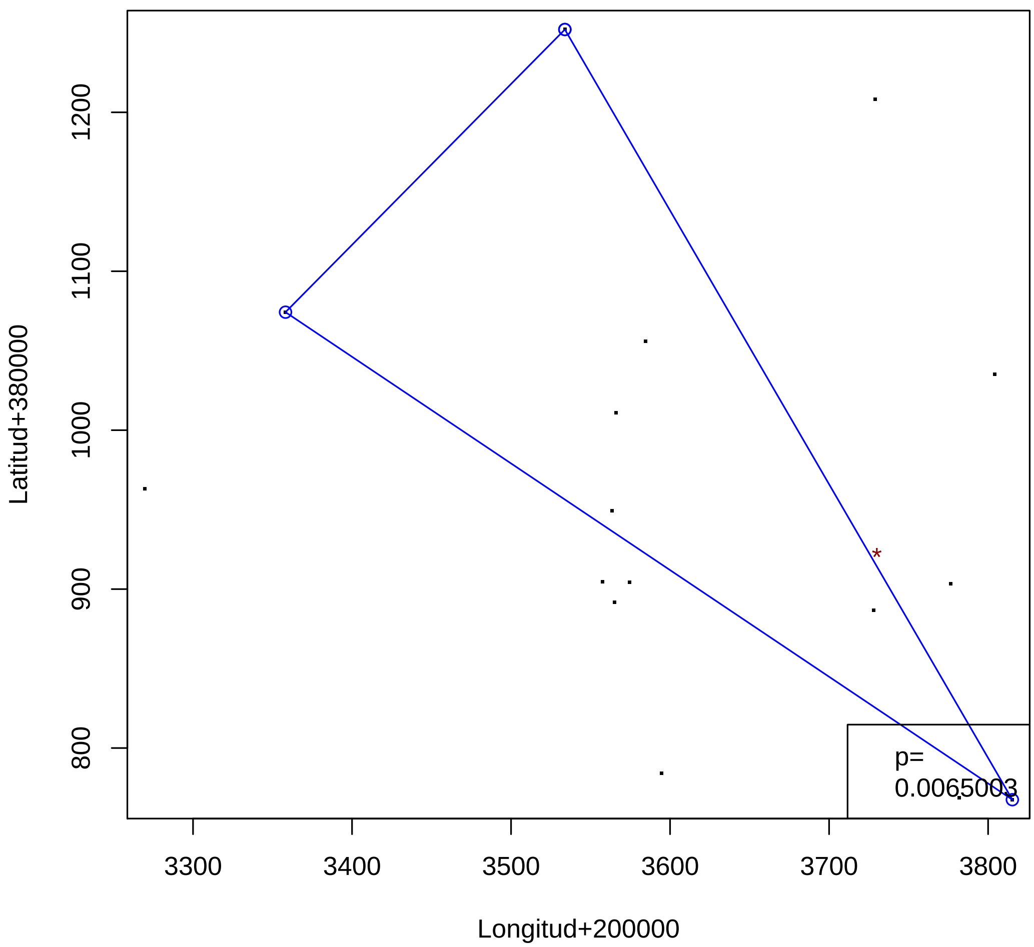
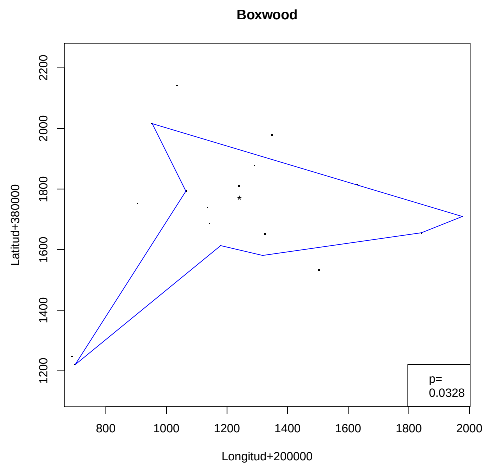


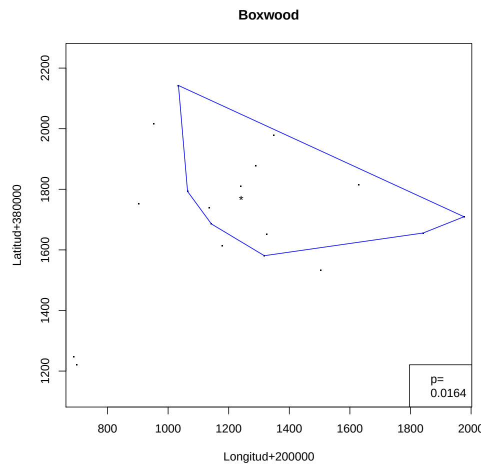
Figure 5.1: Most frequent configuration

Figure 5.2: Most frequent configuration using the Unadjusted Ordinal Model in the reconstruction of Parkmill territory

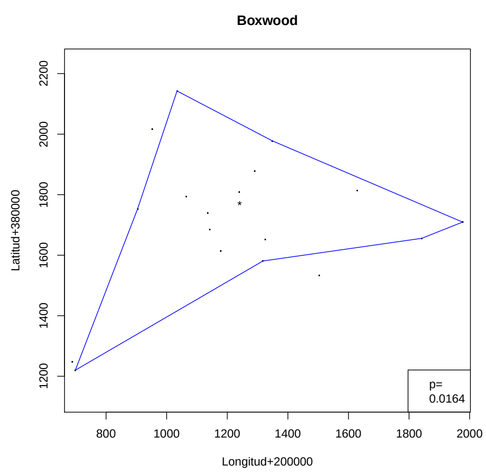
.



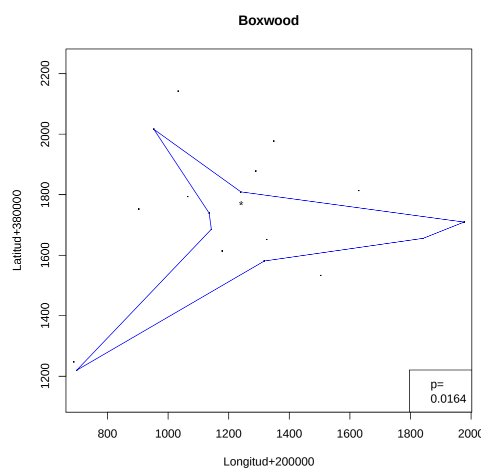
(a) Most frequent configuration



(b) Second most frequent configuration



(c) Third most frequent configuration



(d) Fourth most frequent configuration

Figure 5.3: Most frequent configurations using the Unadjusted Ordinal Model in the reconstruction of Boxwood territory

.

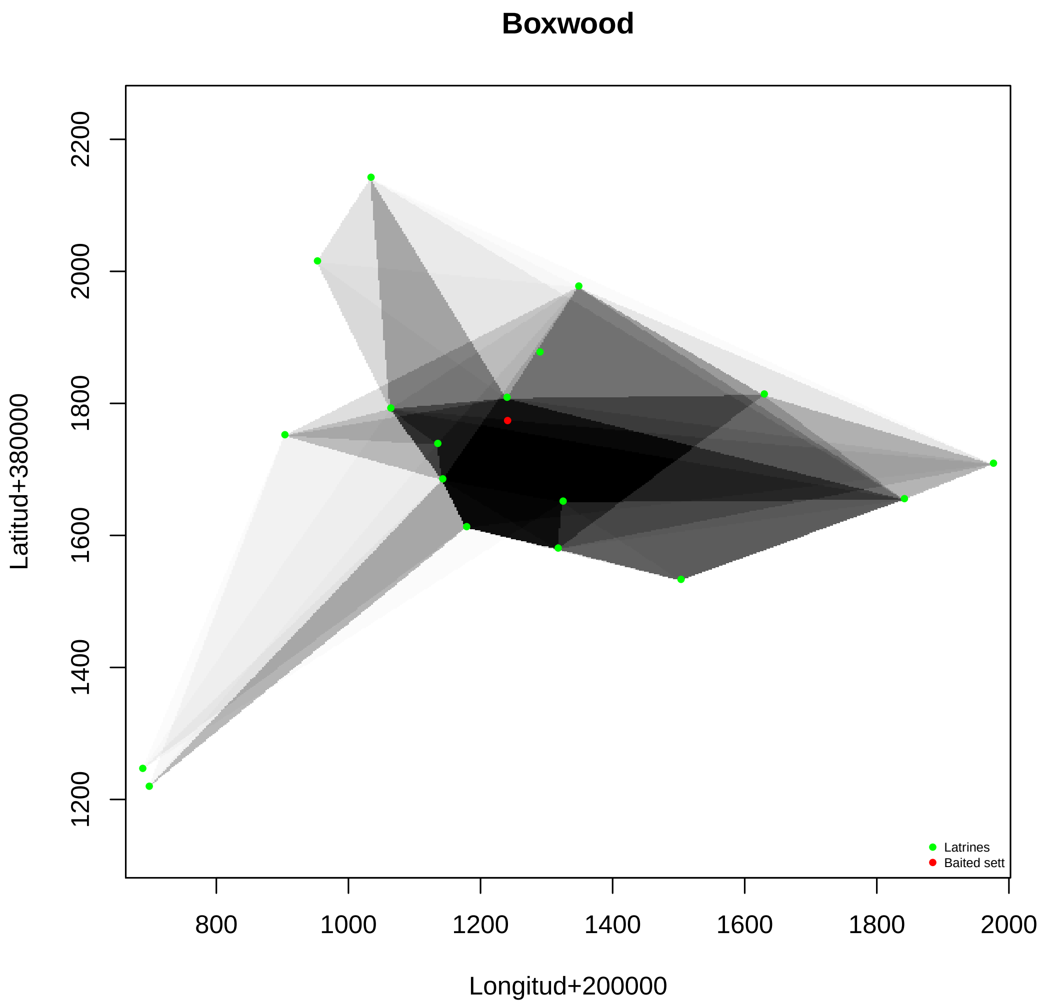
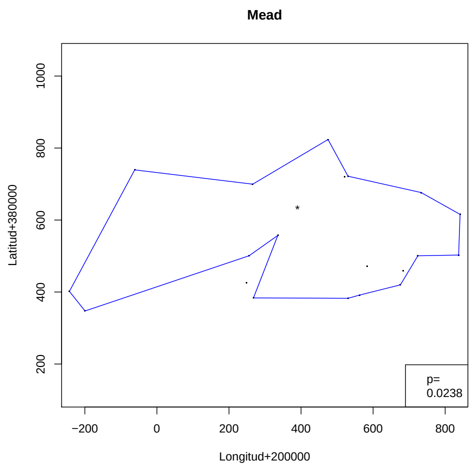
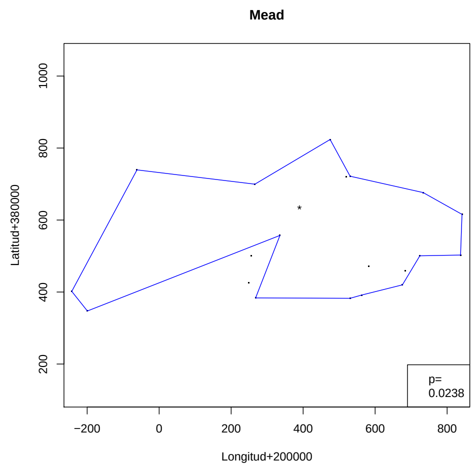


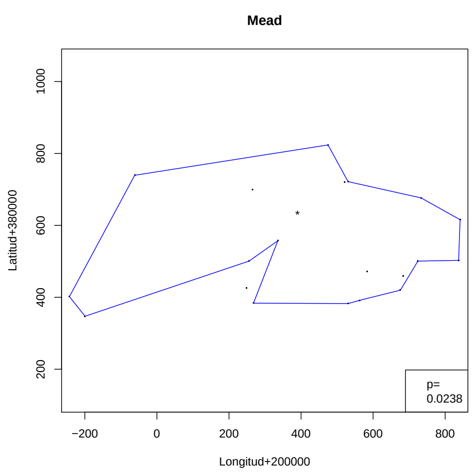
Figure 5.4: Probability of inclusion of the Boxwood territory using the Unadjusted Ordinal Model



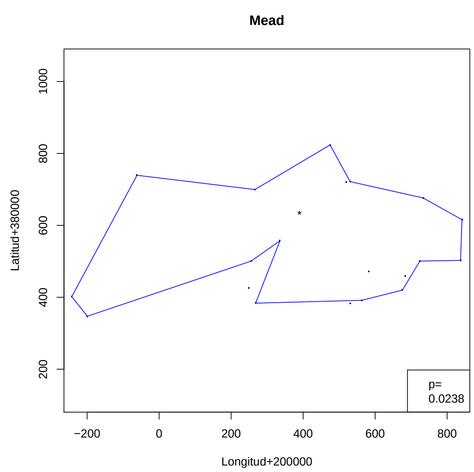
(a) Most frequent configuration



(b) Second most frequent configuration



(c) Third most frequent configuration



(d) Fourth most frequent configuration

Figure 5.5: Most frequent configurations using the Unadjusted Ordinal Model in the reconstruction of Mead territory

.

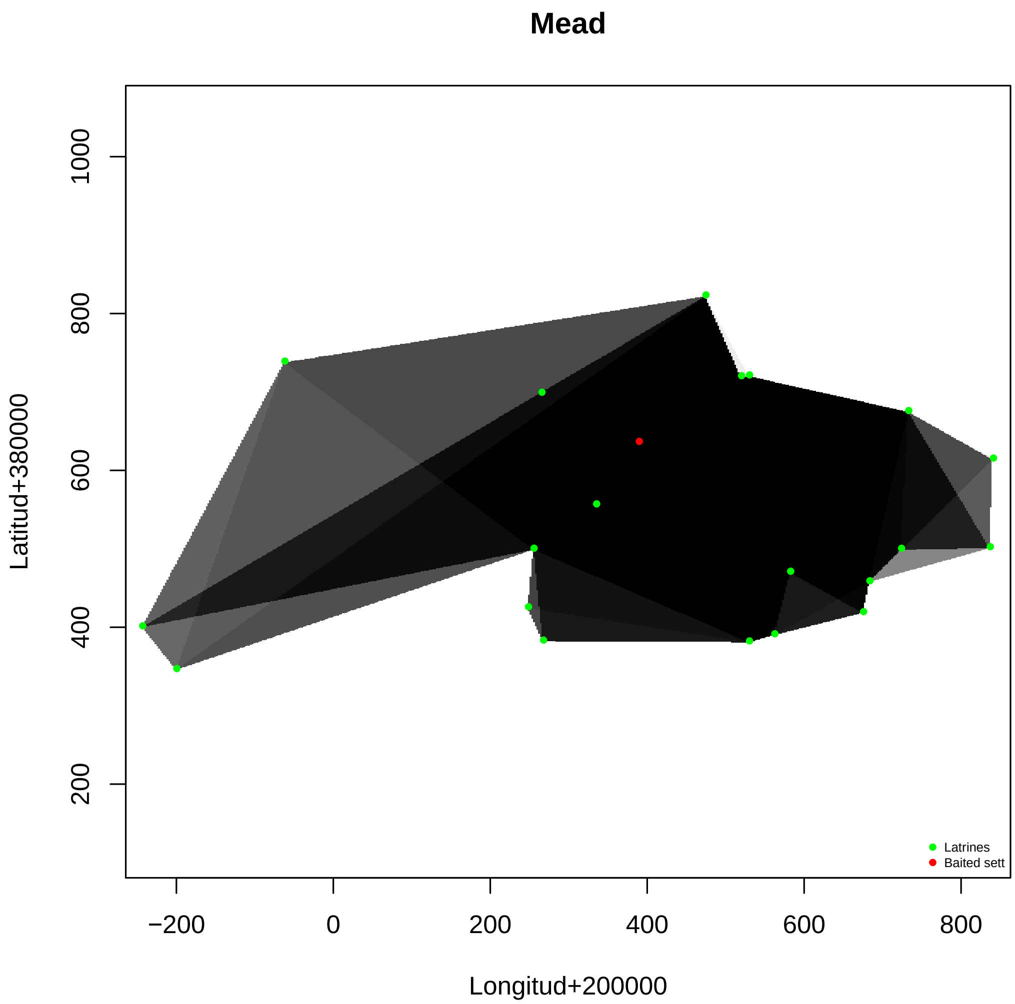
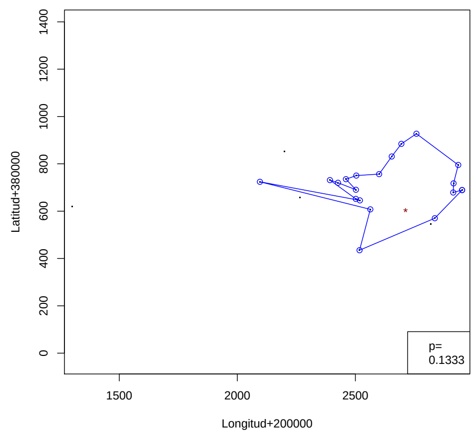
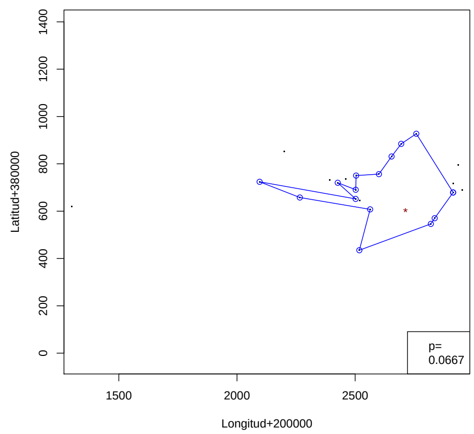


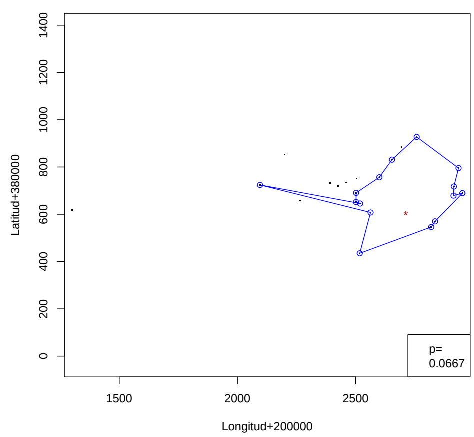
Figure 5.6: Probability of inclusion of the Mead territory using the Unadjusted Ordinal Model



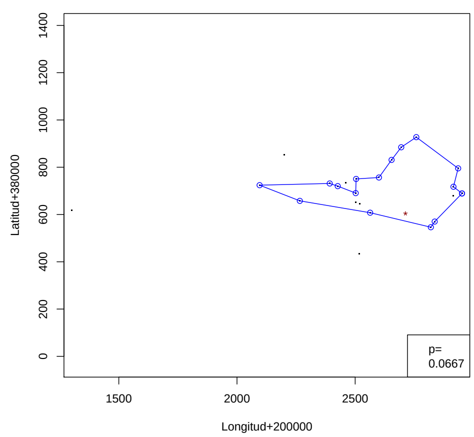
(a) Most frequent configuration



(b) Second most frequent configuration



(c) Third most frequent configuration



(d) Fourth most frequent configuration

Figure 5.7: Most frequent configurations using the Unadjusted Ordinal Model in the reconstruction of Boxwood territory

.

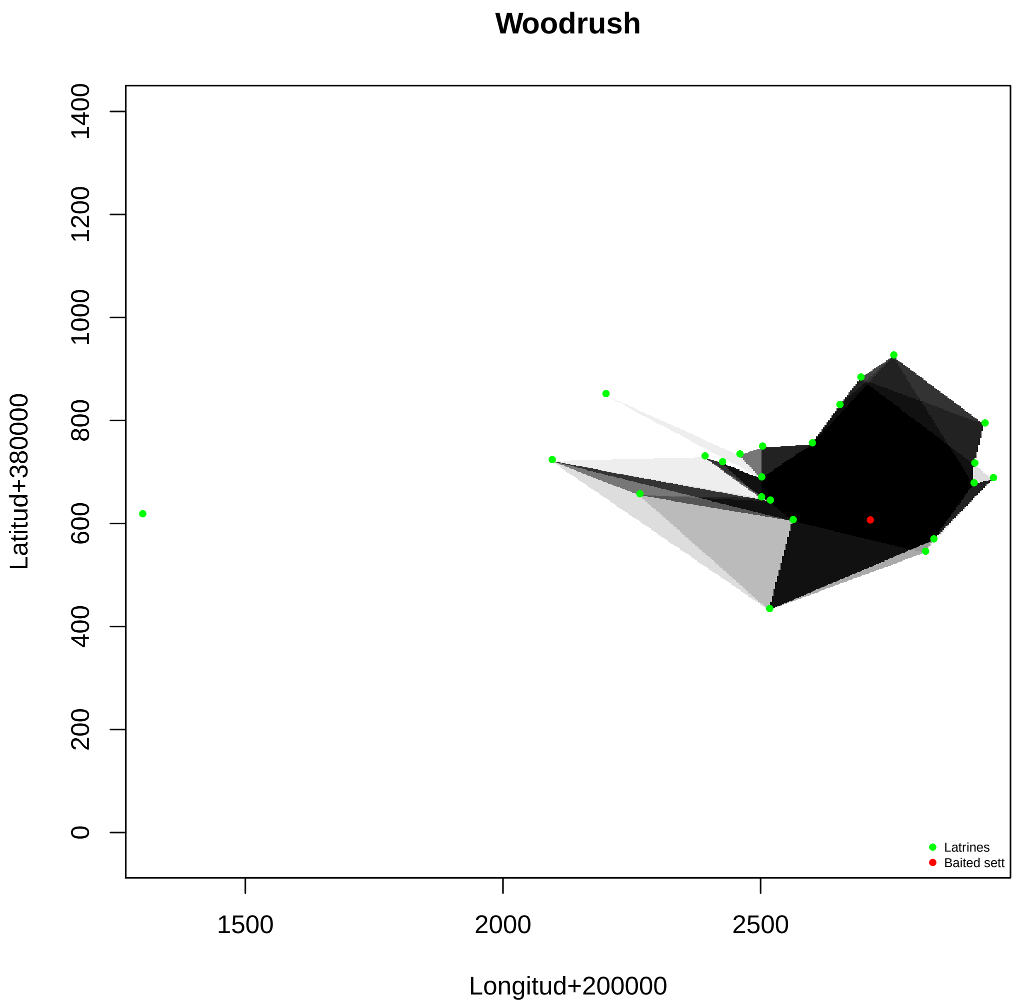


Figure 5.8: Probability of inclusion of the Woodrush territory using the Unadjusted Ordinal Model

Chapter 6

Adjusted Ordinal Model

6.1 Introduction

Classifying badger latrines as *Hinterland, Boundary* and *Outlier* allows the use of ordinal models in its analysis. The development of this method was aiming to solve the problem of undetected concave areas in the reconstruction of badger territories when Outlier Prediction Models (OPMs) were applied. The first attempt was the Unadjusted Ordinal Model (UOM) which was able to produce complete convex territories or versions of the same territory with concave areas on it. Although this method was able to deal with non-prediction of concave territories of the OPMs, requirements of valid territories such as main sett and hinterland latrines being part of the territory and outlier latrines not being part of it, produced a very small number of valid territories. Moreover, these three methods (Unconditional Outlier Prediction Model (UOPM), Conditional Outlier Prediction Model (COPM) and UOM) use probabilities estimated by the model to classify the type of latrine without considering the interaction of the clan which territory is being reconstructed with other clans. The *Adjusted Ordinal Model (AOM)* uses the probabilities estimated by the UOM as a starting point, adjusting them, when they are shared by neighbouring clans.

6.2 Method

The AOM uses *Cumulative Logit Models with Proportional Odds Property* (as in the UOM) as starting point. These models are explained in detail on the previous chapter. The idea was to approach the classification of the latrines as ordinal variables (hinterland, boundary and outliers) obtaining a probability about this classification, for each latrine of the territory analyzed. These probabilities were adjusted when the latrine was shared by another clan and then normalise these quantities, effectively conditioning on the latrine having a set of outcomes allowed by Table 6.1 . Finally, the new set of adjusted probabilities was used to obtain samples from the probability distribution fitted by the AOM to reconstruct the territories plotting them.

6.2.1 Variables and use of the AOM.

The variables used are:

Latrine latrine where plastic markers were recovered.

Sett baited main sett.

Distance from the baited sett to the latrine.

Returns number of returns to the latrine.

Angular fit using a minimum distance of 85m and a $2\theta = 22°$.

Mean of distances mean of distances of latrines in the same direction.

Shared if the latrine is shared by two or more territories.

The UOM (Chapter 5) uses a *Cumulative Logit Model with proportional odds property* (equation 5.3) (McCullagh, 1980) to estimate probabilities to classify $latrine_i$ of $Sett_j$ as $Type_{ij}$ (*Hinterland, Boundary* or *Outlier*) using covariates information for a given territory from data provided by Food and Environmental Research Agency (FERA) Woodchester Park 2010 baitmarking survey as explained in section 2.3.3. Results of this survey were used to generate Figure 3.1. Moreover, this figure shows that the territory configuration for $Sett_j$ is affected by their neighbours since territories do not overlap (Kilshaw et al., 2009), (although home range can). This figure provided extra information used by the AOM to reconstruct the territories such as: classification of the $latrine_i$ (Delahay et al., 2000) according to each $Sett_j$ that used it and if the $latrine_i$ is shared by two or more $Sett_j$.

6.2.2 Adjustment of probabilities.

$Latrine_i$ can be categorized differently, if it is used by more than one clan: it can be a *hinterland latrine* for $Sett_j$ and an *outlier latrine* for $Sett_j$ clan 'B', for example. However, not all the combinations are possible, for example it is not possible that a latrine to be categorized as *hinterland* for two territories. Table 6.1 presents, in green, valid possibilities and, in red, those that are not possible. A latrine can not be *hinterland* for two or more territories nor *hinterland-boundary* since the territories do not overlap. This information is used to adjust the probabilities for a clan 'A' latrine to be categorized as a *hinterland, boundary* or *outlier*, using the probabilities for the same latrine but estimated when it belongs to clan 'B'.

The adjustment of the probabilities is carried out as follows The UOM produces three probabilities for $latrine_i$ of $Sett_j$. α_j corresponds to each type of latrine.

The UOM produces three probabilities for a given latrine, as it relates to $sett_j$ say. For convenience, we write these unadjusted probabilities simply as p(H), p(B), p(O) for hinterland, boundary and outlier respectively. If the latrine is shared, it will have corresponding probabilities based on its role in the n other territories that share it,

		Sett 'B'		
	Type of latrine	Hinterland	Boundary	Outlier
Sett'A'	Hinterland			
	Boundary			
	Outlier			

Table 6.1: Valid possibilities to adjust probabilities for clan 'A' by clan 'B' when a latrine is shared.

denoted here by $p(h_1), p(b_1), p(o_1), p(h_2), \ldots, p(o_n)$; often we have $n = 1$, where a latrine is shared by only one other territory, but this is not always the case. By adjusting the probabilities, we mean taking the unconditional probabilities based on modelling the latrine's status separately within each territory and then conditioning on the fact that its status in the focal territory must be consistent with its status within each of the other territories sharing it.

In particular: if a latrine is a hinterland latrine for the focal territory, it must be an outlier for any other territory; if it is a boundary latrine for the focal territory, it must not be a hinterland latrine for any other territory; if it is an outlier for the focal territory, it can have any status for the other territories.

Writing C for the event that the statuses are consistent in this sense, we then have adjusted probabilities given by:

$$\begin{aligned} p(H|C) &= p(H \cap C)/p(C) \\ &= p(H) \times p(o_1) \times \cdots p(o_n)/p(C), \qquad (6.1) \\ p(B|C) &= p(B \cap C)/p(C) \\ &= p(B) \times (1 - p(h_1)) \times \cdots (1 - p(h_n))/p(C), \qquad (6.2) \\ p(O|C) &= p(O \cap C)/p(C) \\ &= p(O)/p(C). \qquad (6.3) \end{aligned}$$

The adjusted probability as *Hinterland* of $latrine_i$ is its estimated as hinterland (H) multiplied by the probability of outlier (o) of the other territories sharing the $latrine_i$ 6.1. To adjust the probability as *Boundary* of $latrine_i$, it probability as boundary (B) is multiplied by $1 - p(h)$ of all the territories sharing $latrine_i$ 6.2. Finally, there is no need to adjust the probability of $latrine_i$ of being *Outlier* $p(O)$ since it is the same 6.3.

Note that none of the adjusted probabilities is set to zero by this conditioning, as there is always some possible joint classification of a latrine, for each territory sharing it, that that permits any given status for the current territory of interest. The probabilities can be radically changed, however; for example, a latrine is much less likely to be hinterland for territory j if the ordinal model shows that it is likely to be hinterland for some other territory, since that *joint* classification is not permitted.

As used here, the unadjusted probabilities are from an ordinal model that estimates the category of each latrine separately, as in Chapter 5. It would be possible to use a conditional model instead, by analogy with Chapter 4, but that has not been implemented.

Currently we carry out the adjustment separately for each territory, based on its probabilities and those of its neighbours. In principle, the adjustment should be carried out simultaneously for all territories, but in practice the calculation would be prohibitive.

6.3 Sampling

The new set of adjusted probabilities was used to obtain 100,000 samples (300,000 for Beech) from the probability distribution fitted by the AOM to reconstruct the territories using the same idea than the UOM: labelling the latrine as hinterland, boundary and outlier; reconstructing the territory joining the latrines labeled as boundary and evaluation of valid reconstructions using the same criteria explained in detail in 5.4:

- The main sett must be part of the territory.
- Hinterland latrines must be always inside the reconstruction of the territory.
- Outlier latrines are not part of the territory.

The valid reconstructions were presented in the same plots as in the previous three chapters:

- Most frequent reconstructions.
- Probability of inclusion.

6.4 Results

The same three criteria –explained above– to obtain valid reconstructions used in the UOM were applied to the 100,000 samples (300,000 for Beech) obtained from the probability distribution fitted by the AOM. A summary of the valid reconstructions is presented in table 6.2 and compared with the results from the UOM. An improvement was reached to produce valid territories when compared with the UOM for both: Main sett (MS) and location of Hinterland and Outlier latrines (HO). After the adjustment of the probabilities by the neighbour territories, only one territory was not able to produce a valid reconstruction (Parkmill).

Sett	Unadjusted		Adjusted	
	MS	HO	MS	HO
Beech	62638	0	299970	569
Boxwood	29204	61	98211	8878
Box	24721	0	99995	30090
Cedar	47124	0	100000	15110
Collliers	18330	0	99736	4119
Honeywell	37351	0	99872	4283
Jacks	96350	-	86724	51633
Junction	0	-	95057	4721
Kennel	29329	0	99990	4720
Larch	37930	0	100000	31141
Mead	69320	42	99033	2761
Nettle	53222	0	99804	6254
Old Oak	38935	0	99992	22866
Parkmill	99709	-	-	-
Rabbit	41248	0	100000	2208
Septic tank	33033	0	100000	855
Top	24313	0	99979	43070
Trackside	49082	0	99745	10006
West	59285	0	99538	5490
Wood farm	0	-	83257	54349
Woodrush	51303	15	99762	2602
Wych Elm	59285	0	99985	57855
Yew	72505	0	98081	6837

Table 6.2: Comparison of numbers of valid territories using the UOM and the AOM after 100,000 samples (Beech, 300,00 samples). Columns with “MS”headers are the valid configurations when the criteria of *Main Sett* was applied. Columns with “HO” are the valid reconstructions when the localization of hinterland and outlier latrines requirements are applied.

6.4.1 Graphical results

Four territories have been chosen as examples to present and discuss the results of the AOM: Beech, Junction, Parkmill and Woodrush. Beech was picked randomly since the beginning of the research and all the models and scripts were tested firstly in this territory; Junction was chosen since it is in the edge of Woodchester Park and it has several shared latrines; Parkmill is in the edge of the park, however it does not have too many shared latrines and Woodrush was selected because it was one of the territories that the UOM was able to present valid configurations.

a) ***Beech***

Fig. 6.1 is the Probability of inclusion of the reconstruction of Beech territory. It also presents the reconstruction of the most frequent reconstruction of the territories that Beech is sharing at least one latrine with. Beech shares latrines with Boxwood, Box, Cedar, Junction, Larch, Mead, Septic Tank, Top and Yew territories. Red dot is the sett and green dots are the latrines where plastic markers were recovered. The darker the area, the higher the probability of being part of the territory. The reconstruction of this territory using the the AOM shows a darker region surrounding the sett with grey zones on the edge. There is a latrine just southwest of the main sett that it is shared by the territory Larch, predicting that it could be a boundary latrine instead of a hinterland, producing a grey triangular zone. There are two latrines to the east of the territory that are shared with Top territory, that have a high probability of being categorized as outliers. One of them seems to be a hinterland latrine of Top and the other one a boundary latrine.

From the second graphical presentation approach, the initial sample generated for this territory was of size 100,000, as for the other territories. However, the number of valid territories after adjustment was very small, and the frequencies of the most likely were all equal. Since this territory has been looked at closely in earlier chapters, a larger sample of size 300,000 was generated, and the results from this are shown in Fig. 6.2. The most frequent configuration is shown in Fig. 6.2a with a probability of 0.007, it differs from the other three more frequent configurations in only one latrine. The next two configurations have a probability of 0.006 (Fig. 6.2b and 6.2c) and, finally, Fig. 6.2d presents the fourth more frequent configuration.

b) ***Junction***

Fig. 6.3 presents the reconstruction of Junction territory using all the iterations in one single graph. This territory shares at least one latrine with Beech, Jacks, Mead, Septic Tank and Yew territories. Junction territory is on the edge of the park. The reconstruction shows grey areas in all its boundaries, especially on it west side where Mead territory is located, but the east side where there is no surveyed clans. The second graph (Fig. 6.4) presents the four most frequent configurations. The most frequent configuration is shown in Fig. 6.4a with a probability of 0.083, when compared with the other three more frequent configurations, it differs in only one latrine. The second configuration has a probability of 0.082 (Fig. 6.4b. Third reconstruction has a probability of 0.049 6.4c) and, finally, Fig. 6.4d presents the

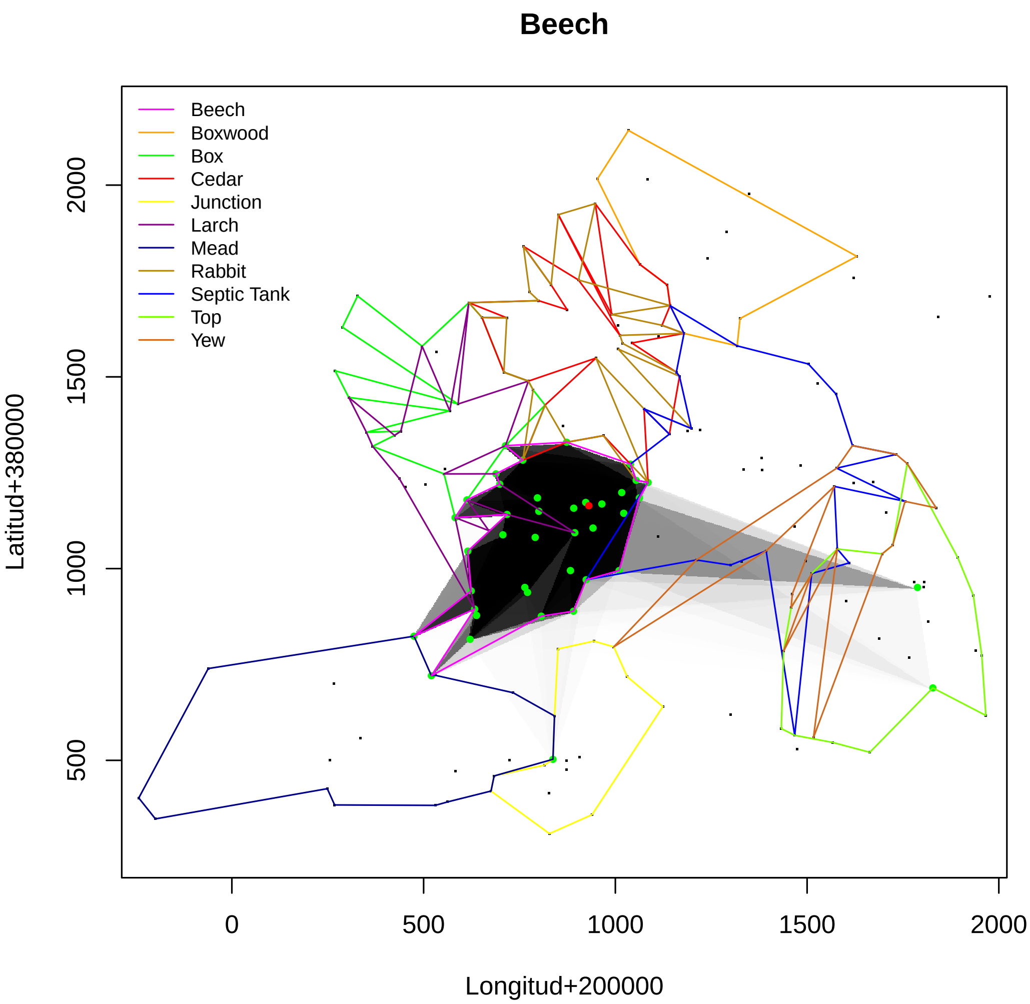
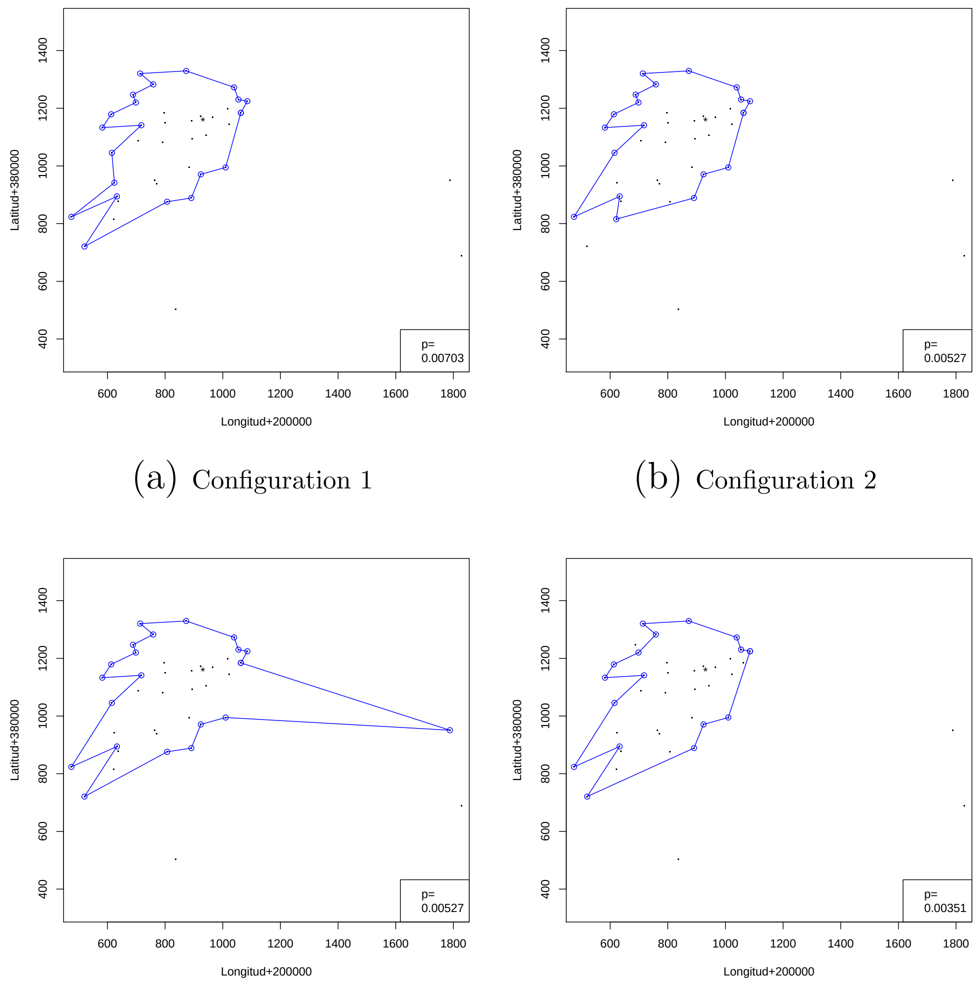


Figure 6.1: Probability of inclusion of the Beech territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.



(a) Configuration 1

(b) Configuration 2

(c) Configuration 3

(d) Configuration 4

Figure 6.2: Reconstructions of the territory Beech with higher frequency using the Adjusted Ordinal Model.

fourth more frequent configuration with a probability of 0.044. The differences of this reconstructions are in the west and north borders of the territory.

c) ***Parkmill***

This territory was selected because it was the only one that the AOM was not able to produce valid reconstructions. Fig. 6.5 displays its reconstruction. This territory is located on the edge of the park and only shares latrines with the Neetle territory. The figure shows no dark areas, only grey ones. In the most probable reconstruction (pink lines) the main sett (red dot) does not belong to the territory. In the second graphical presentation approach, Fig. 6.6, can be seen that the main sett is not inside the reconstructions.

d) ***Woodrush***

Woodrush was selected as an example of the AOM since it was one of the three territories that the UOM was able to generate valid reconstructions. Fig. 6.7 is the reconstruction using the AOM. This territory has shared latrine with Colliers, Kennel, Nettle, Rabbit, Wood Farm and Wych Elm territories. The reconstruction of this territory shows a dark area surrounding the main sett and grey zones to its north, west and southwest. The uncertainty is mainly due to outlier latrines.

From the second graphical presentation approach, Fig. 6.8 shows the fourth most frequent configurations after 100,000 iterations. The most frequent configuration is shown in Fig. 6.8a with a probability of 0.045, it is different from the the next two more frequent configurations in only one latrine and in two from the fourth one. The second configuration has a probability of 0.039 (Fig. 6.8b, the third one 0.031 and 6.8c) and, finally, Fig. 6.8d presents the fourth more frequent configuration with an estimated probability of 0.023.

6.5 Conclusions

The idea of using ordinal models to reconstruct badger territories with concave areas was achieved using the UOM. However, this method was able to predict only 3 valid territories. An extension to this method was proposed, adjusting the estimation of these probabilities, when the latrines were shared by two or more territories, depending on their classification. This adjustment produced an improvement to the reconstruction in all the valid territorial configurations frequencies, as shown on table 6.2 increasing the number of reconstruction of territories that contained the main sett and all the latrines predicted as hinterland.

The AOM presents a multi-territorial approach since it uses information from neighbouring clans. The relationships between territories are illustrated in figures 6.1, 6.3, 6.5 and 6.7 which each shows the probabilities of inclusion for a single territory along with the most probable reconstruction for each territory it shares at least one latrine with.

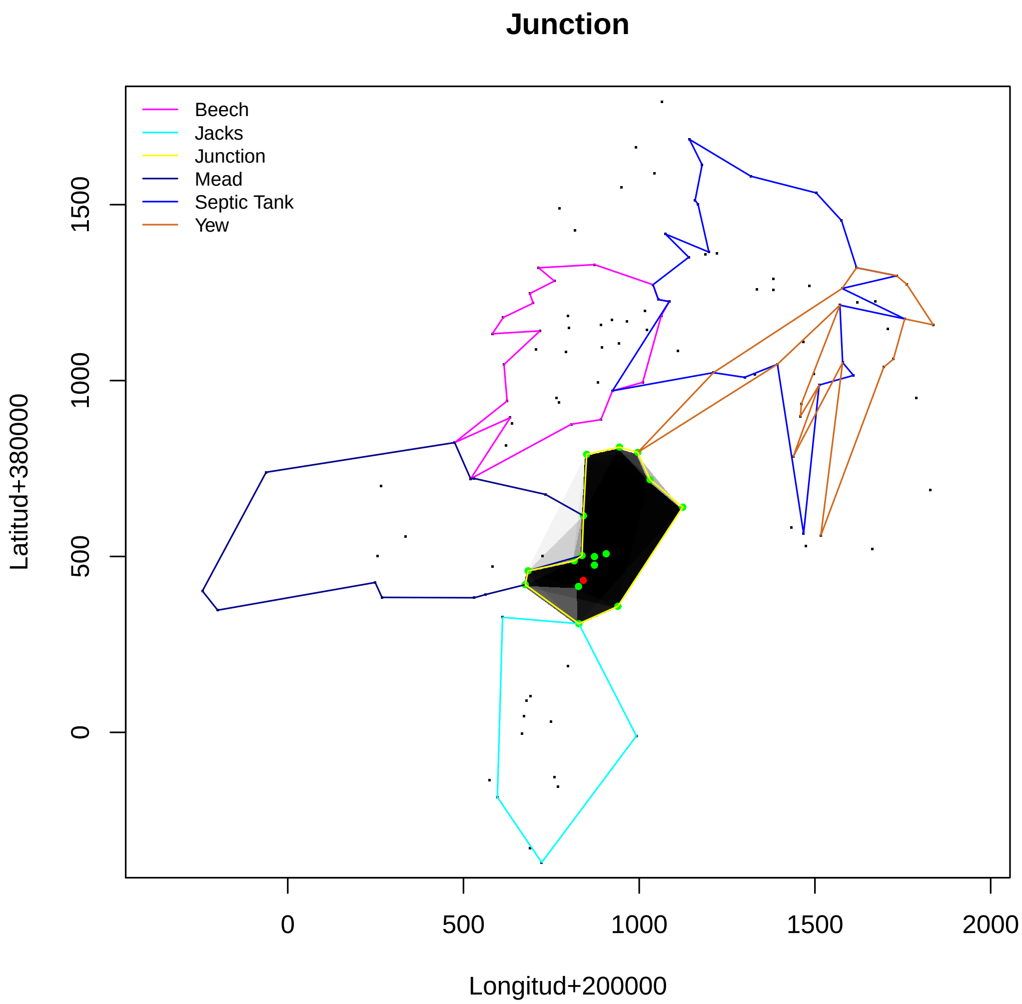
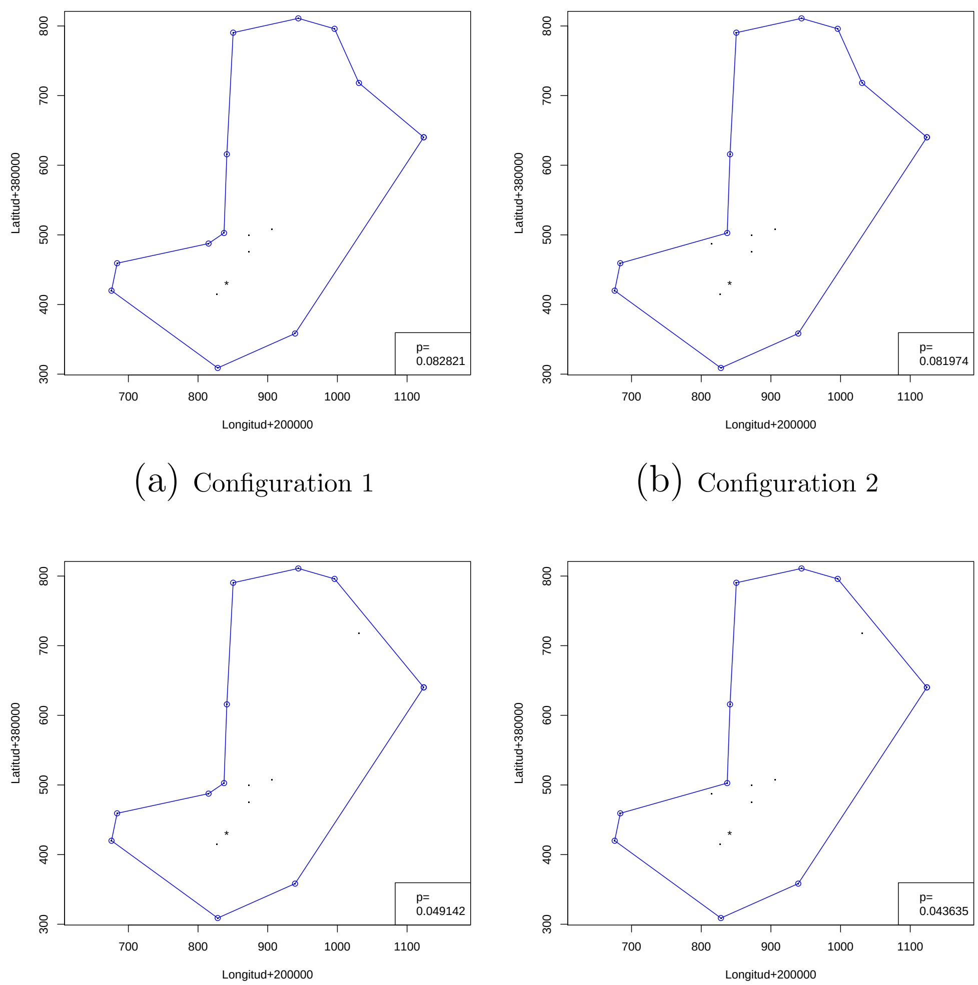


Figure 6.3: Probability of inclusion of the Junction territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.



(a) Configuration 1

(b) Configuration 2

(c) Configuration 3

(d) Configuration 4

Figure 6.4: Reconstructions of the territory Junction with higher frequency using the Adjusted Ordinal Model.

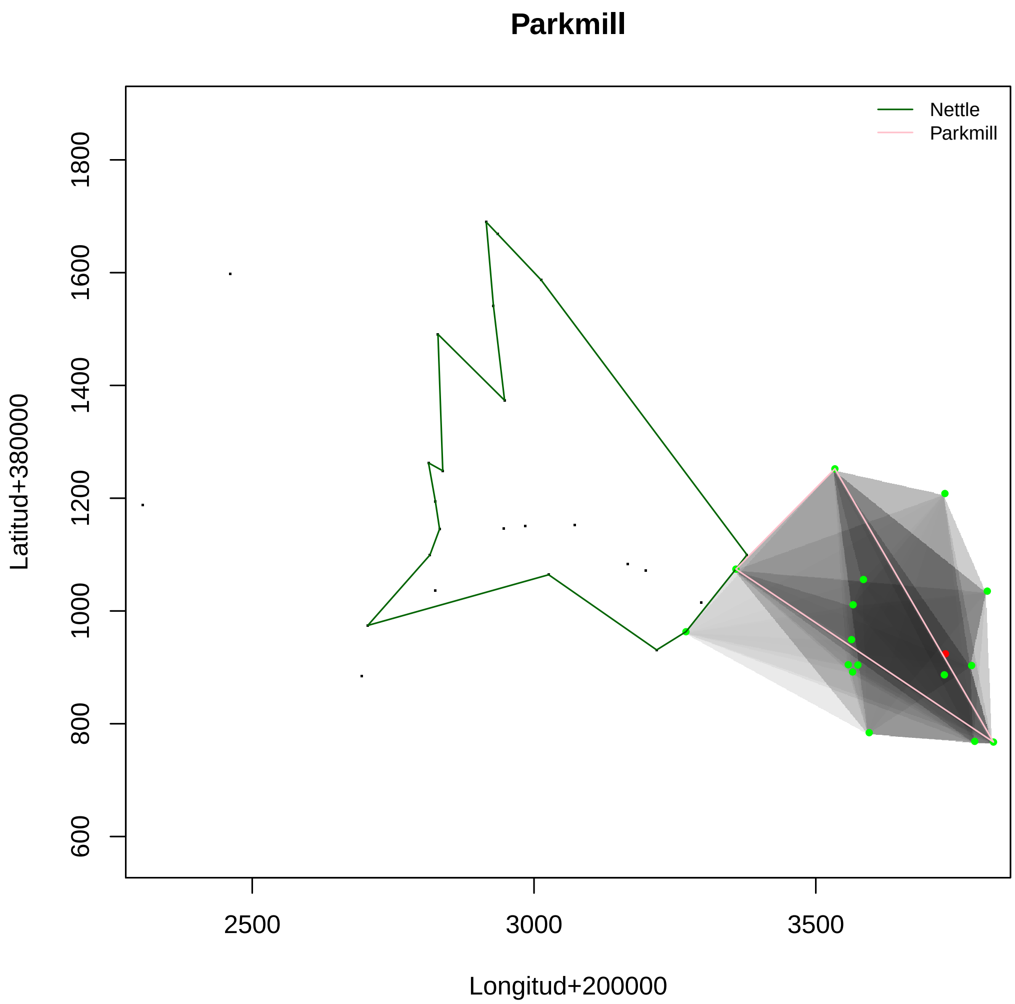
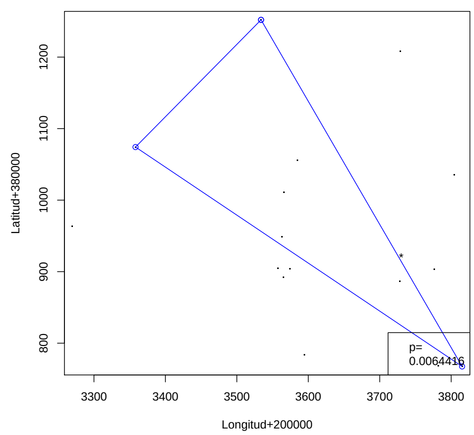
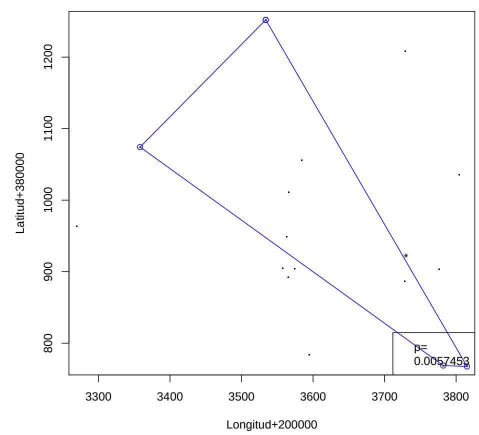


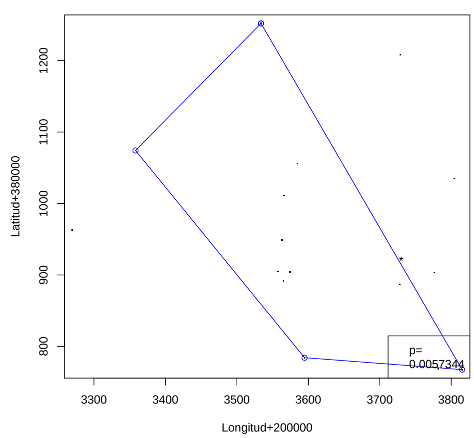
Figure 6.5: Probability of inclusion of the Parkmill territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.



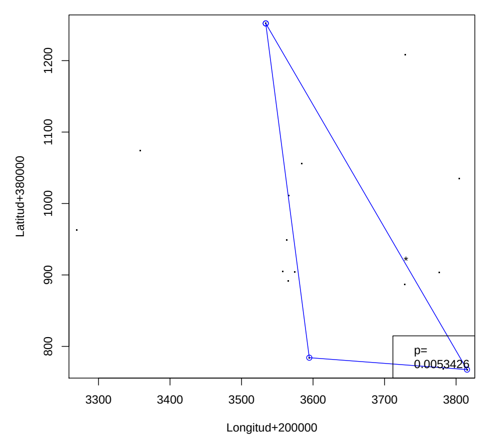
(a) Configuration 1



(b) Configuration 2



(c) Configuration 3



(d) Configuration 4

Figure 6.6: Reconstructions of the territory Parkmill with higher frequency using the Adjusted Ordinal Model.

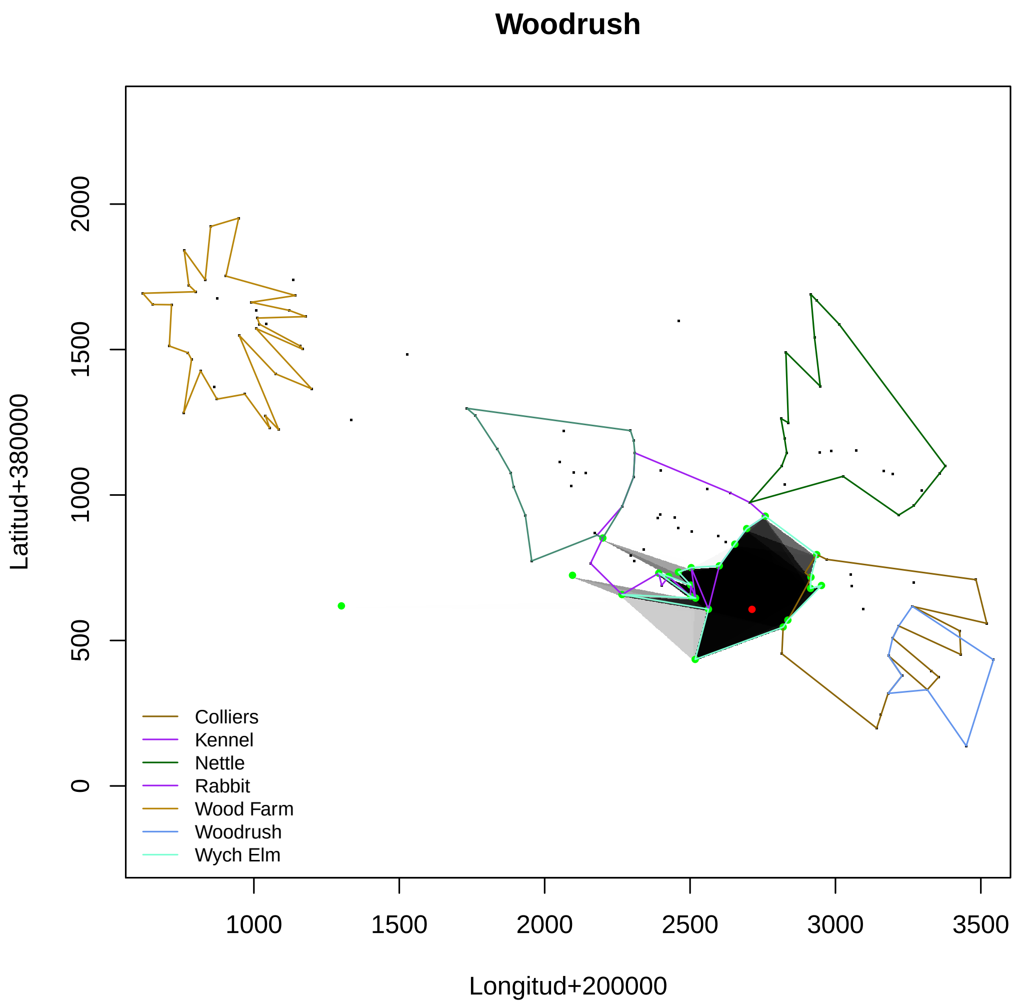
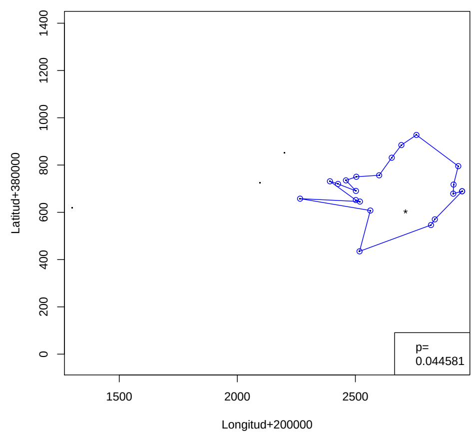
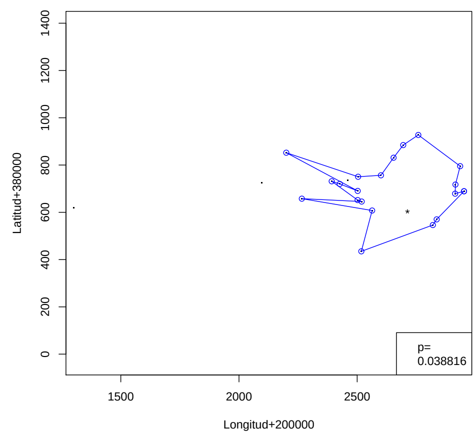


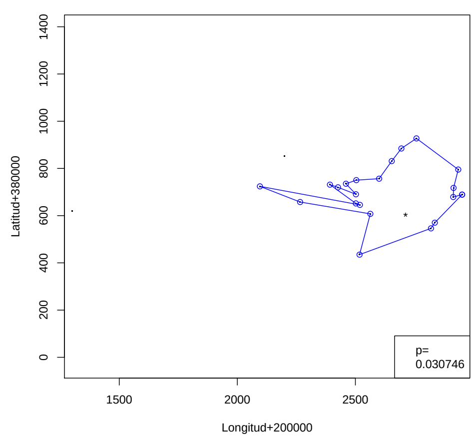
Figure 6.7: Probability of inclusion of the Woodrush territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.



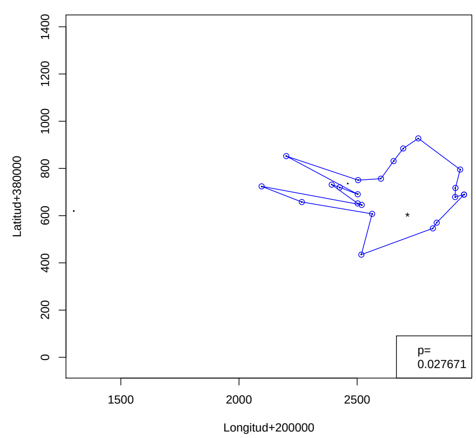
(a) Configuration 1



(b) Configuration 2



(c) Configuration 3



(d) Configuration 4

Figure 6.8: Reconstructions of the territory Woodrush with higher frequency using the Adjusted Ordinal Model.

Chapter 7

General Discussion and Conclusions

Badgers are territorial animals. They defend their territories with direct aggression and with several types of visible signs to mark them such as paths, latrines and tussocks used for 'squat-marking'. All these signs are very important in high density areas. The most important signs of badger territory boundaries are the latrines. These are aggregations of up to 60 defecations, in small pits dug by the badgers themselves. Usually there are signs of scratching of the ground around the latrine. Latrines are used at all times of the year, but especially in spring with a second lower peak in October. Many latrines are used by badgers for several years (Kruuk, 1978b; Kruuk, 1989; Kruuk and Dekock, 1981; Kruuk et al., 1984).

Information from the territories of the badgers in Woodchester Park, was provided by researchers of Food and Environmental Research Agency (FERA). This information consists of a set of points that represent badger setts or latrines and their coordinates. This information was obtained through a baitmarking survey, explained in detail in Chapter 3. After the survey is finished, each latrine position and the origins of its coloured markers are displayed on a map of the study area, connecting each latrine with a line to the main sett at which those markers were fed. Latrines where more than one type of marker is found, known as shared latrines, are shown as linked to both or all the corresponding setts. These points are the 'raw data' for reconstructing the badgers' territories (Delahay et al., 2000).

Some of the disadvantages of this method of reconstruction of the territories are: skilled staff is required for best results, subjective judgment required for interpretation, accurate delineation of a single social group territory also requires feeding at adjacent setts (Delahay et al., 2000). Another, and probably the most commonly used method to reconstruct territories from information like this is the Minimum Convex Polygon (MCP). A different and more robust approach to reconstruct the territories of the badgers is to create covariates such as distance from the sett and the latrines used by the badger clan, direction from the latrines in relationship with the sett, {whetherif the latrine is used by different clans, as explained in detail in Chapter 3, and then use this biological information in statistical models to estimate, for example, the probability that a latrine is part of a badger clan territory.

As mentioned above, the information from Woodchester Park are points representing the setts of badgers and latrines which were obtained from a latrine survey. The Woodchester Park scientist reconstruct the territories of the badgers based on this information. Even when this information can lead to the conclusion that the territories are as represented in this reconstruction, uncertainty exists since the judgment and experience of the researchers of Woodchester Park is involved. However, this uncertainty can not be quantified or modeled from the information provided.

This research started estimating the Unconditional Outlier Prediction Model (UOPM), which is an extension of an unpublished paper, the Elstub Outlier Prediction Model (EOPM), that uses logistic regression to estimate the probability of a latrine being or not part of a badger territory. This method uses this probability to create a new effective location for the latrine and, combined with the 100% MCP, reconstruct the territory. The difference between these two methods (the EOPM and the UOPM) is how to calculate covariates that measure the concept of *Neighbouring latrines*, mainly with the covariate 'Angular fit', which is the ratio that compares the distance of one latrine from the sett against the distance between the sett and the other latrines in a similar direction, excluding latrines within a distance from the sett, because, if a latrine has only latrines that are close to the sett in a similar direction, this would otherwise result in a very high angular fit for the latrine and may lead to incorrectly identifying it as an outlier. This covariate is not reported in the literature before, so the angle and minimum distance used in its calculation needed to be tested for the calculation of the 'Angular fit'. Elstub used a fixed angle and minimum distance of $2\theta = 45°$ and m=40m, respectively. For the UOPM 170,340 different combinations of minimum distance and angle were run to select the best model based on the Akaike Information Criterion (AIC).

Firstly these methods were compared through a *point estimation approach* where the coefficients of the model were used to estimate the probability of a given latrine of not being part of the territory using this value to multiply its distance from the main sett and then calculating the 100 % MCP of the new points. Since the use of the probabilities of the latrine being an extraterritorial excursion to shrink its distance to the sett is not the best way to use them, the model obtained was applied in a sampling approach. The idea was to sample sets of latrines labelling them (as an outlier or not) from the probability distribution fitted by the UOPM, with the 100 % MCP of the sampled setts giving a sample of reconstructions of the territory.

As discussed in Chapter 3, instead of just selecting a 'crude' α% MCP, where the only important covariate was *distance* and α was pre-specified, the model was fitted to the latrine-level data using a range of covariates selected using the AIC, to predict the probability that a given latrine is an outlier. Even if only the covariate *distance* would be selected by the UOPM, this model would be better since it captures some of the uncertainty involved and the probabilities are estimated from the data, which are added benefits over the α% MCP.

Two ways of presenting the results not used before were utilized: *probability of inclusion* and *most frequent reconstructions* for this approach and the next methods investigated. These complementary methods allow both the reconstructed territories and the associated uncertainty to be visualised.

The next stage was to develop a *conditional model* that let us use a Gibbs sampling to allow for dependence between latrines. This approach is conditional in the sense that whether a latrine is an outlier or not can incorporate knowledge about whether the latrines in the same direction and associated with the same sett are in fact outliers. This model has a non-conditional and a conditional part. As cited in the literature review, badgers forage and feed individually. However, they defend a common territory. This behaviour is intended to be accounted for with this model, since not all the badgers of the clan use the whole territory.

The Conditional Outlier Prediction Model (COPM) is a very different model from the EOPM and the UOPM in the way that the relationship between latrines is modelled. However, even when the models are completely different internally, they are statistically comparable using their AIC's (Kass and Raftery, 1995). This comparison shows that the COPM is a much better fit than the other two models. The conditional variables and Gibbs Sampling approach to reconstructing the territories allow the COPM to incorporate dependance between latrines. Since this is the main difference between the three models (minimum distance is the same and the Angular fit is very similar), the improvement on the AIC's value, is due to this dependence giving information that, when a latrine is used, the odds of neighbouring latrines to be used, are increased. On the other hand, one disadvantage is that the COPM is a very time consuming method to apply, compared with these other two methods.

The previous two methods use in the reconstruction a 100% MCP. The main disadvantage is that can only produce convex territories. This is not always true biologically. For example, some natural structures can be present in the boundary of the territory, such as as a lake. It is obvious that this area will not be part of the territory of the badgers. A different approach was taken. An ordinal model, the Unadjusted Ordinal Model (UOM) was used classifying the latrines as hinterland, boundary and outlier that let us delineate territories with concave zones. Even when this method was able to produce territories with this characteristic, it has the problem that most of the reconstructions were not valid from a biological point of view since not all of the hinterland latrines were included the reconstruction, since outlier latrines were inside it. This clearly limits the usefulness of the UOM as it is. An alternative approach to sampling the reconstructions, described in Section 3.3.2, is one possible remedy. Instead, we concentrate on using extra information that we have, the status of the latrines from the perspective of other territories.

The final method, the Adjusted Ordinal Model (AOM), uses the classification probabilities from neighbouring territories relating to shared latrines to refine these for the 'focal territory', making it a method that has an estimation of overall territories.

Even though the AOM is a method which is able to reconstruct badger territories with concave and convex areas, that can measure the uncertainty of the reconstruction and uses information of neighbouring clans, it would be interesting to test this method with information from other areas or years; to evaluate if the territories change from one year to another and by the intervention of external factors e.g. badger culling for bovine tuberculosis control. Investigating a Conditional Adjusted Ordinal Model, in which the probabilities were calculated using a conditional version of the ordinal model, and then adjusted as described in Section 6.2.2, could provide a most robust method in the reconstruction of the territories.

In summary, this thesis started replicating the work of Elstub (Elstub, 2006), for a different data set where she proposed a logistic model classifying the latrines of a given badger territory as an outlier and non-outlier. Then, since no biological justification for the values in the calculation of the covariate 'Angular fit' were found, the values of minimum distance and angle used in its calculation were varied. A better model (based on AIC values) were estimated. When the reconstructions of both methods were compared using the reconstructions of the badger territories by researchers of Woodchester Park as a reference, better approaches were obtained. A disadvantage of those two methods is that, in the reconstruction of the territories, the probability estimated by the model of a given latrine was an outlier, was multiplied by the covariate 'distance' and then the territory was reconstructed. Since there is no biological support to use the probability estimated by the model of the latrine was an outlier in this way, a second approach was taken (a 'sampling approach'). Different reconstructions of the same territory was predicted and its uncertainty was quantified. A second model was proposed using the dependance between 'neighbouring latrines'. The biological reason that support this approach is that the use of a latrine is not independent of the use of latrines in the same direction.

As mentioned above, both methods relied on the MCP in the reconstruction of the territories. This does not allow to represent concave areas of the territories. An ordinal model was used after classifying the latrines as 'hinterland', 'boundary' and 'outlier'. This model was able to construct territories with concave areas. Restrictions of 'Valid territories' was applied in the use of this model: Main sett and all latrines predicted as Hinterland must be inside of the reconstruction, latrines classified as Outliers must be outside of the reconstruction of the territory. This model produced very few valid territories.

Finally, a new approach using the same type of classification of the latrines was used. The previous models ignore the state of neighbouring territories, but it is known that the shape and size of the territories of the badgers in Woodchester Park depends of the neighbouring clans.

The probability of the classification of each latrine was adjusted if the latrine was shared by another badger clan. This adjustment was an improvement since it takes in consideration a very important aspect, the relationship with neighbouring clans. Moreover, it produced more valid reconstructions of the territories.

From the data of Woodchester Park, the reconstructions of the badger territories from the researchers of the Park and the models proposed in this thesis some conclusions can be presented: the European badger badger is a territorial animal that mark its territory with latrines; Latrines can be classified as hinterland, boundary or outlier (extraterritorial excursion); Latrines can be used (shared) by different clans; the shapes of the territories can present convex and/or concave areas; the uncertainty in the reconstructions of the territories can be quantified and represented graphically.

An idea that can be explored, in future work, is to instead of 'Angular fit' (latrines in the same direction) or in addition to it, the concept of neighbouring latrines, measuring the distance between latrines, could be modeled. If more information such as structure (fences, badger paths, lakes, etc.) is provided, better models can be estimated.

Some of these ideas of reconstructing animal territories could be applied to other animal species that use visible and detectable signs to mark their territories such as foxes, coyotes, wolves, etc. Since they quantify the uncertainty of the reconstruction of the territory, hypothesising about changes in it or that some neighbouring clans could merge can lead to future research.

Appendix A

Appendix Unconditional Outlier Prediction Model

A.1 Tables of the Unconditional Outlier Prediction Model

Outlier Prediction Model Using *Number of Returns as Quantitative Variable and Distance Un-transformed*.

Parameter	Coefficient
Intercept	-17.360
angfit1	13.296
Dist	0.004
Shared(1)	-8.731
TR	4.551
angfit1: shared(1)	9.556
angfit1:TR	-5.074
Degrees of Freedom:	630 Total (i.e. Null); 624 Residual
Null Deviance:	448
Residual Deviance:	254
AIC:	267.9775

Table A.1: Outlier Prediction Model Using *Number of Returns as Quantitative Variable and Distance Un-transformed*

Parameter	Coefficient
Intercept	-28.002
angfit1	15.439
LD	1.702
shared(1)	-9.322
TR	6.634
angfit1:shared(1)	10.062
angfit1:TR	-7.034
Degrees of Freedom:	630 Total (i.e. Null); 624 Residual
Null Deviance:	448
Residual Deviance:	254.1
AIC:	268.1

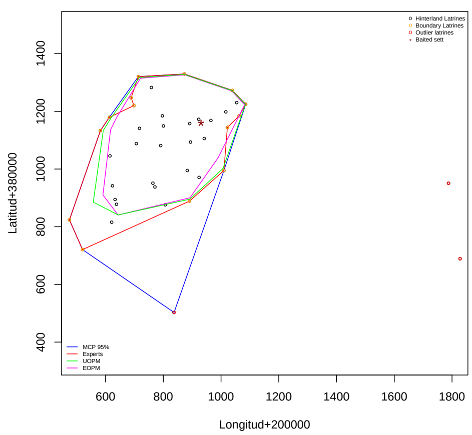
Table A.2: Outlier Prediction Model Using *Number of Returns as Quantitative Variable and using the log of Distance*

Parameter	Coefficient
Intercept	-21.765
angfit1	7.977
LD	1.810
Shared(1)	-10.059
R2(1)	15.414
angfit1:shared(1)	10.840
angfit1:R2(1)	-15.681
Degrees of Freedom:	630 Total (i.e. Null); 624 Residual
Null Deviance:	448
Residual Deviance:	253.2
AIC:	267.2

Table A.3: Outlier Prediction Model Using *Number of Returns as Quantitative Variable and the Log of Distance*

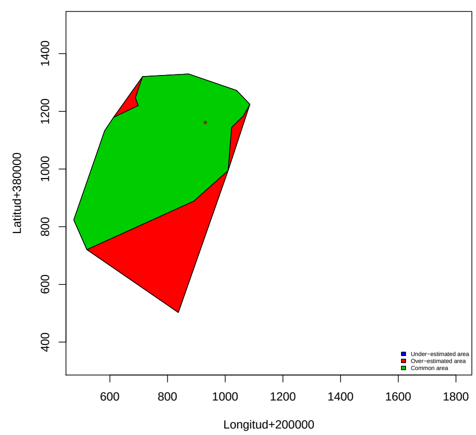
A.2 Figures of the Unconditional Outlier Prediction Model

A.2.1 Areas Comparisons of the Reconstruction of the Badger Territories using the Point Estimation Approach

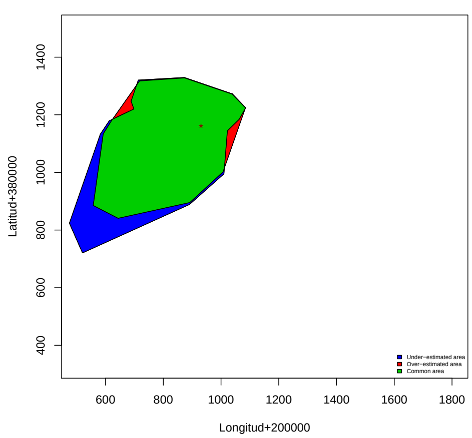


[b]

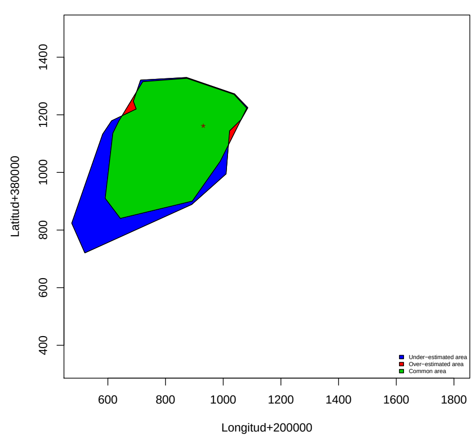
(a) Total Area



(b) 95% MCP-Experts

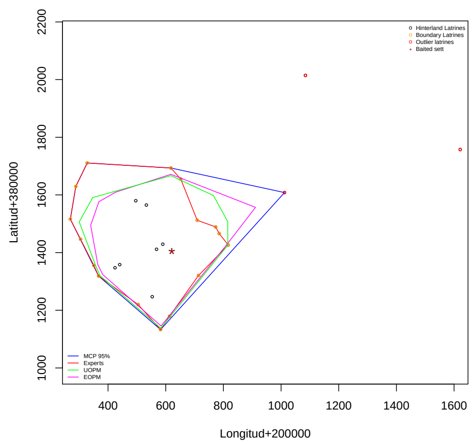


(c) UOPM-Experts

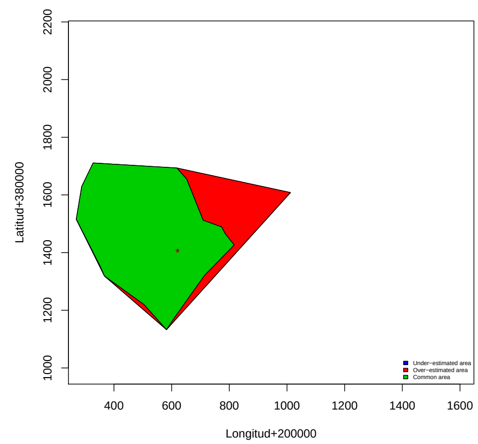


(d) EOPM-Experts

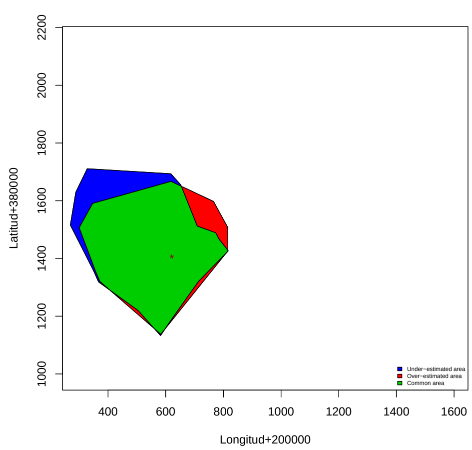
Figure A.1: Comparison of the reconstruction of the territory Beech using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



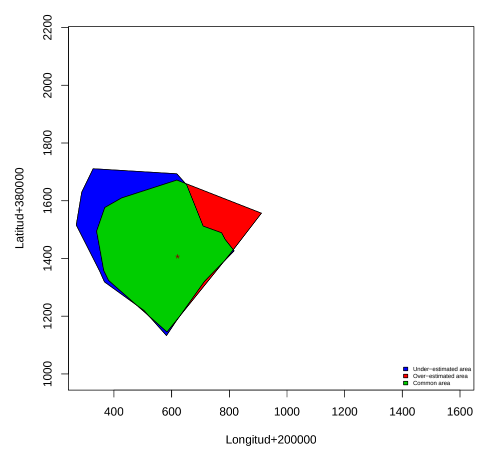
(a) Total Area



(b) 95% MCP-Experts

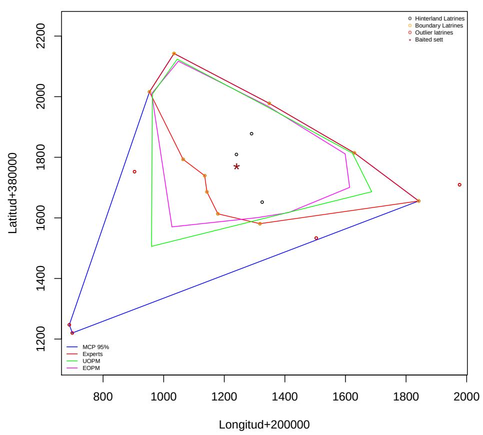


(c) UOPM-Experts

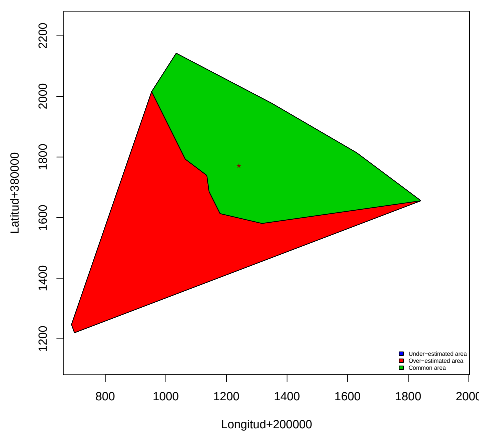


(d) EOPM-Experts

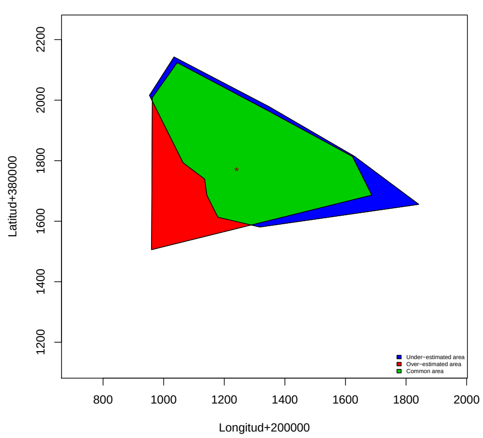
Figure A.2: Comparison of the reconstruction of the territory Box using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



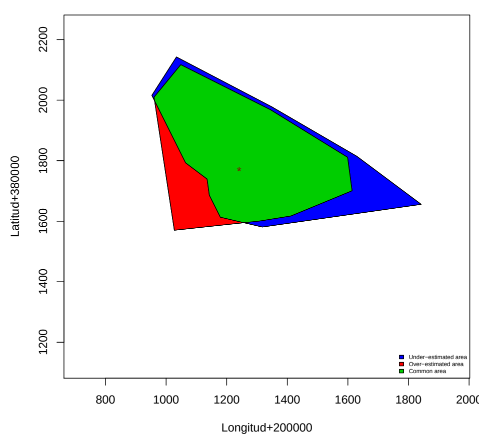
(a) Total Area



(b) 95% MCP-Experts

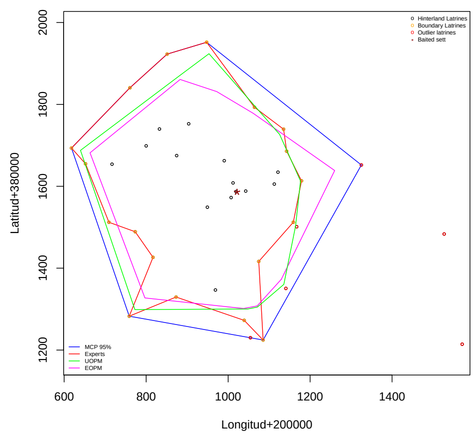


(c) UOPM-Experts

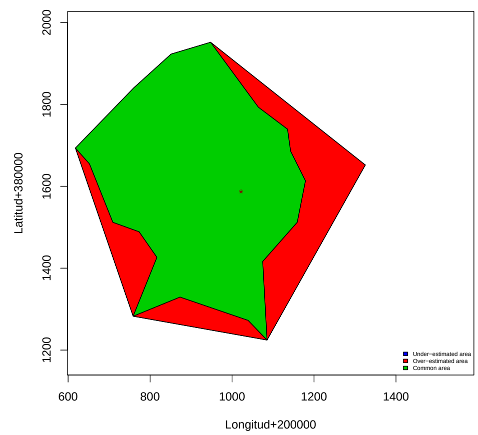


(d) EOPM-Experts

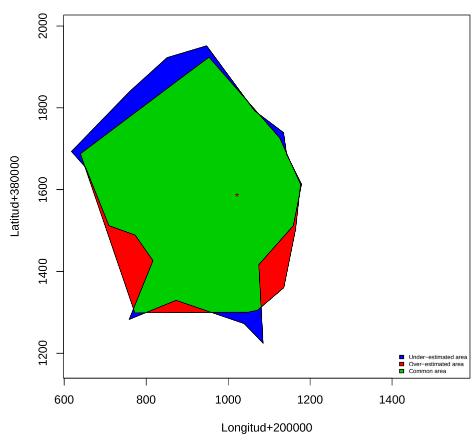
Figure A.3: Comparison of the reconstruction of the territory Boxwood using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



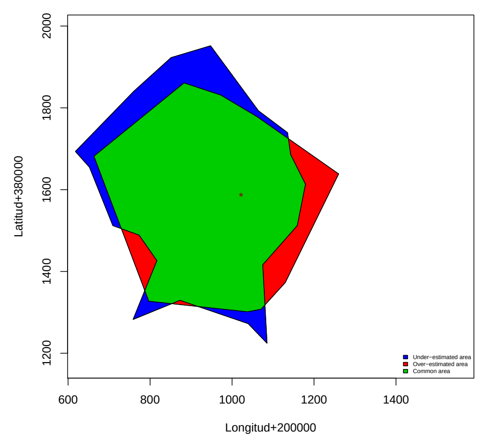
(a) Total Area



(b) 95% MCP-Experts

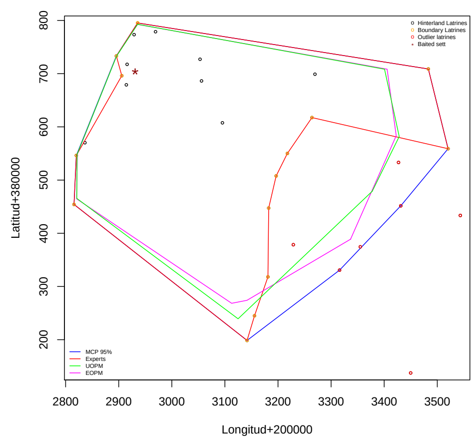


(c) UOPM-Experts

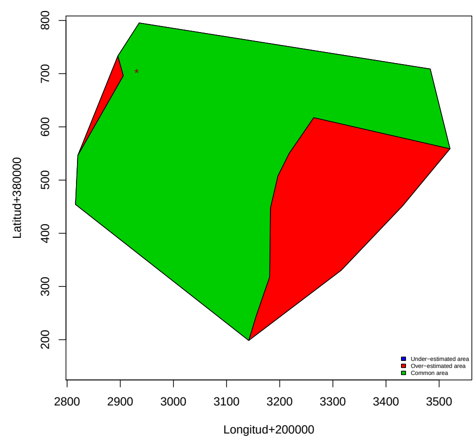


(d) EOPM-Experts

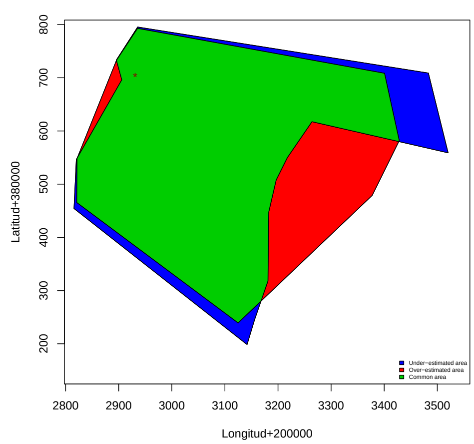
Figure A.4: Comparison of the reconstruction of the territory Cedar using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



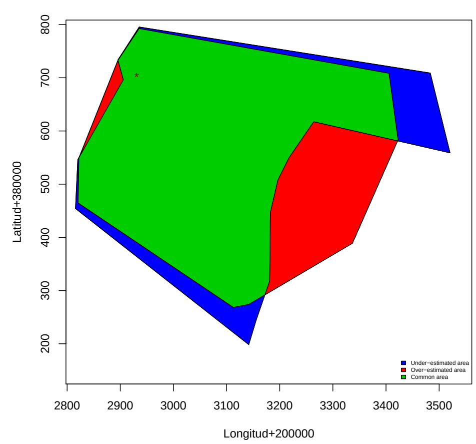
(a) Total Area



(b) 95% MCP-Experts

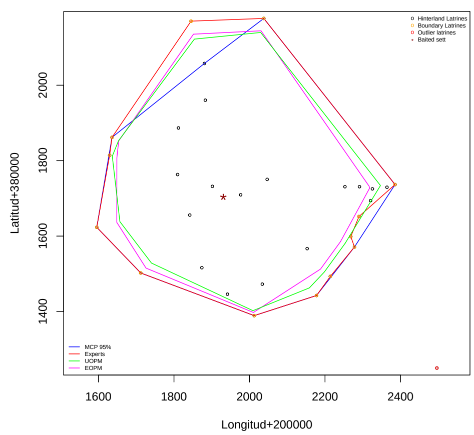


(c) UOPM-Experts

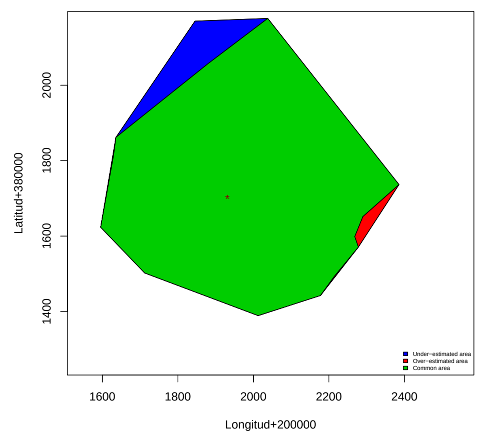


(d) EOPM-Experts

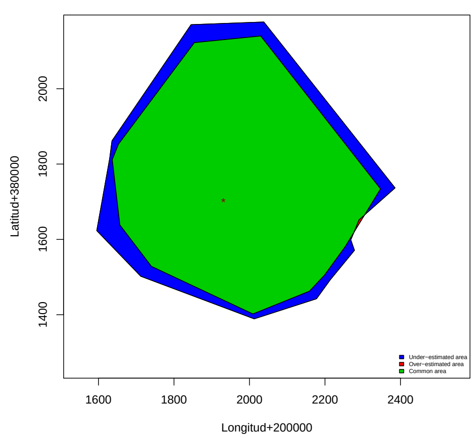
Figure A.5: Comparison of the reconstruction of the territory Colliers using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



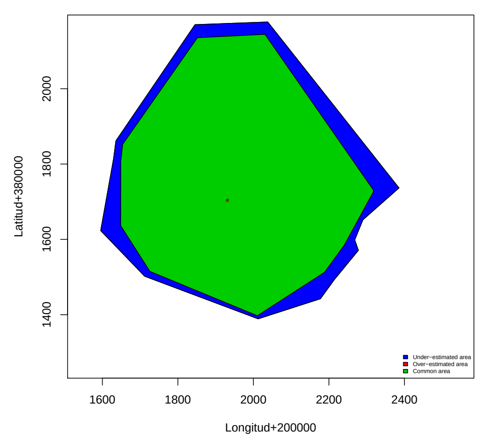
(a) Total Area



(b) 95% MCP-Experts

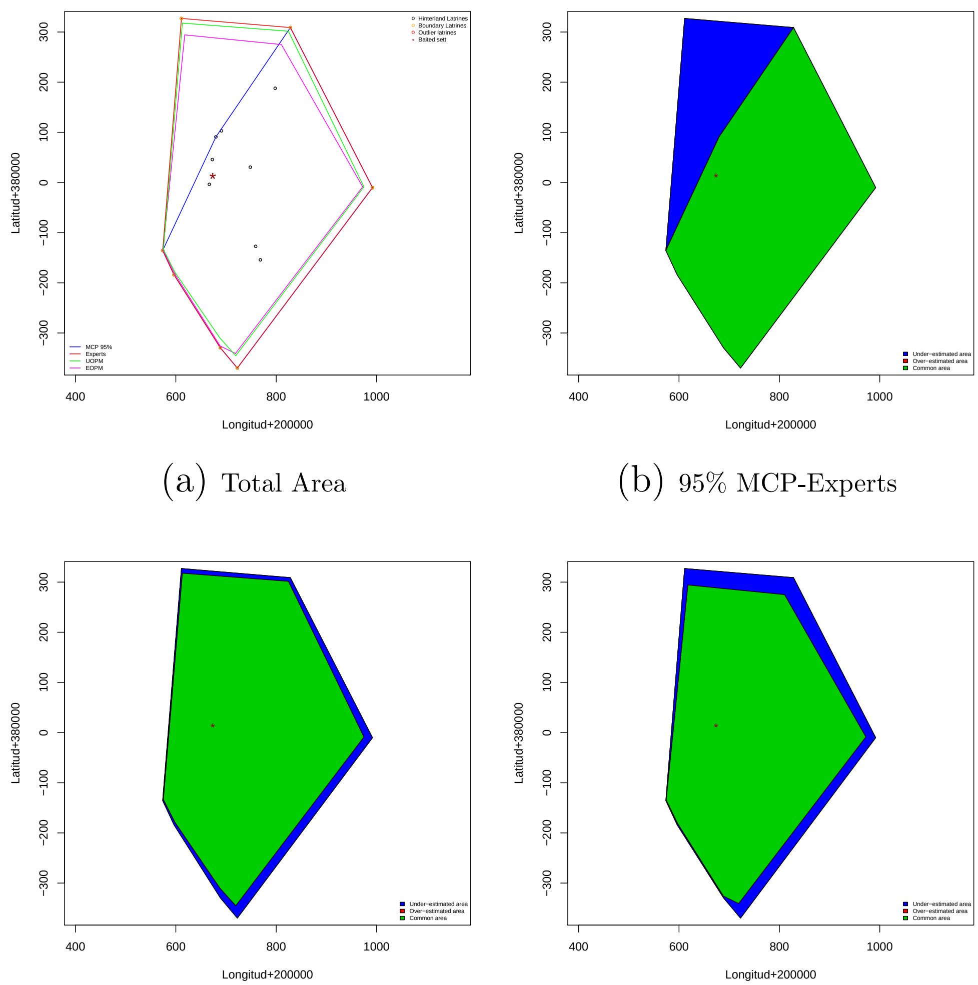


(c) UOPM-Experts



(d) EOPM-Experts

Figure A.6: Comparison of the reconstruction of the territory Honeywell using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



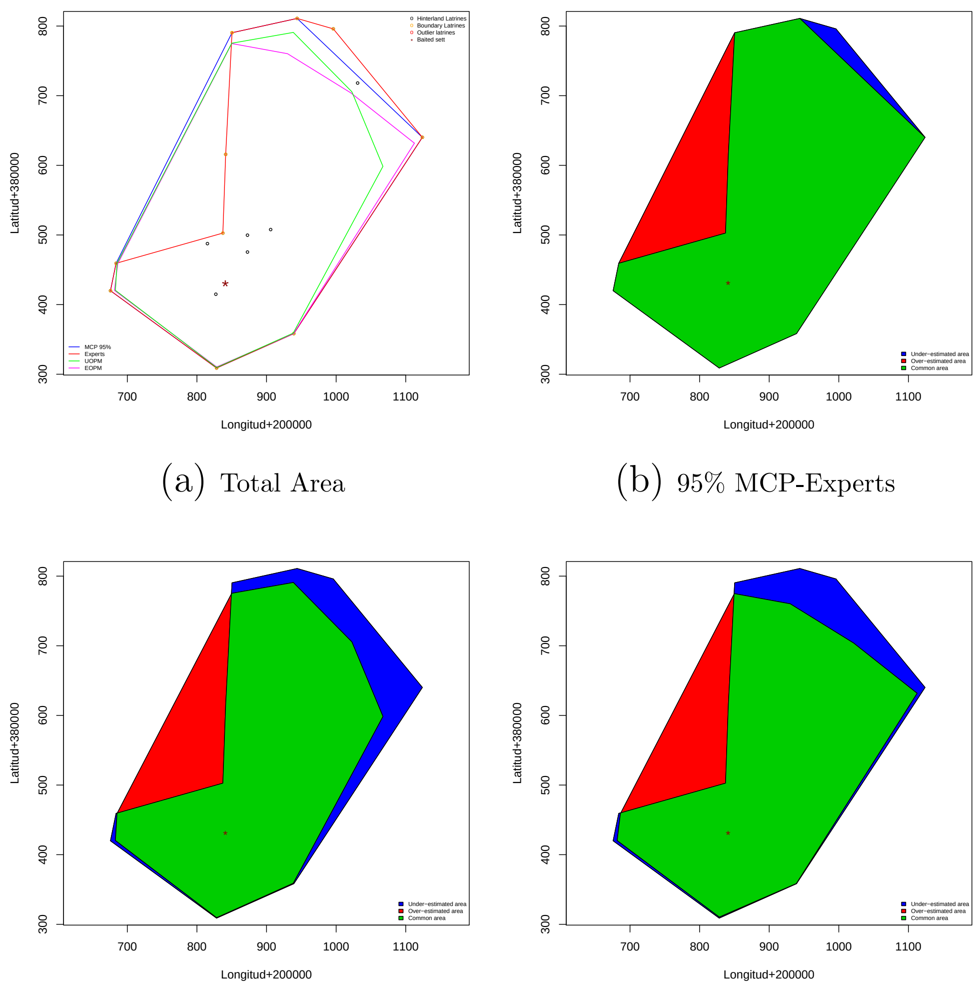
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

Figure A.7: Comparison of the reconstruction of the territory Jacks using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



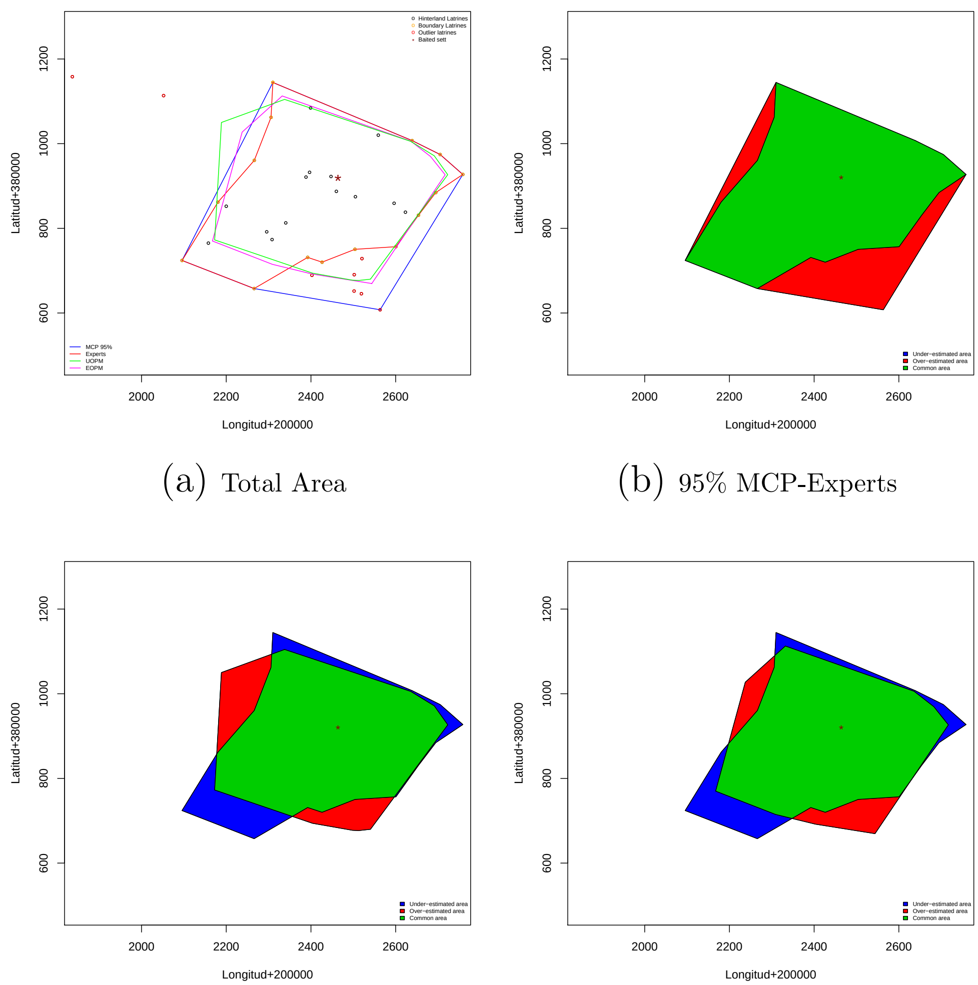
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

Figure A.8: Comparison of the reconstruction of the territory Junction using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



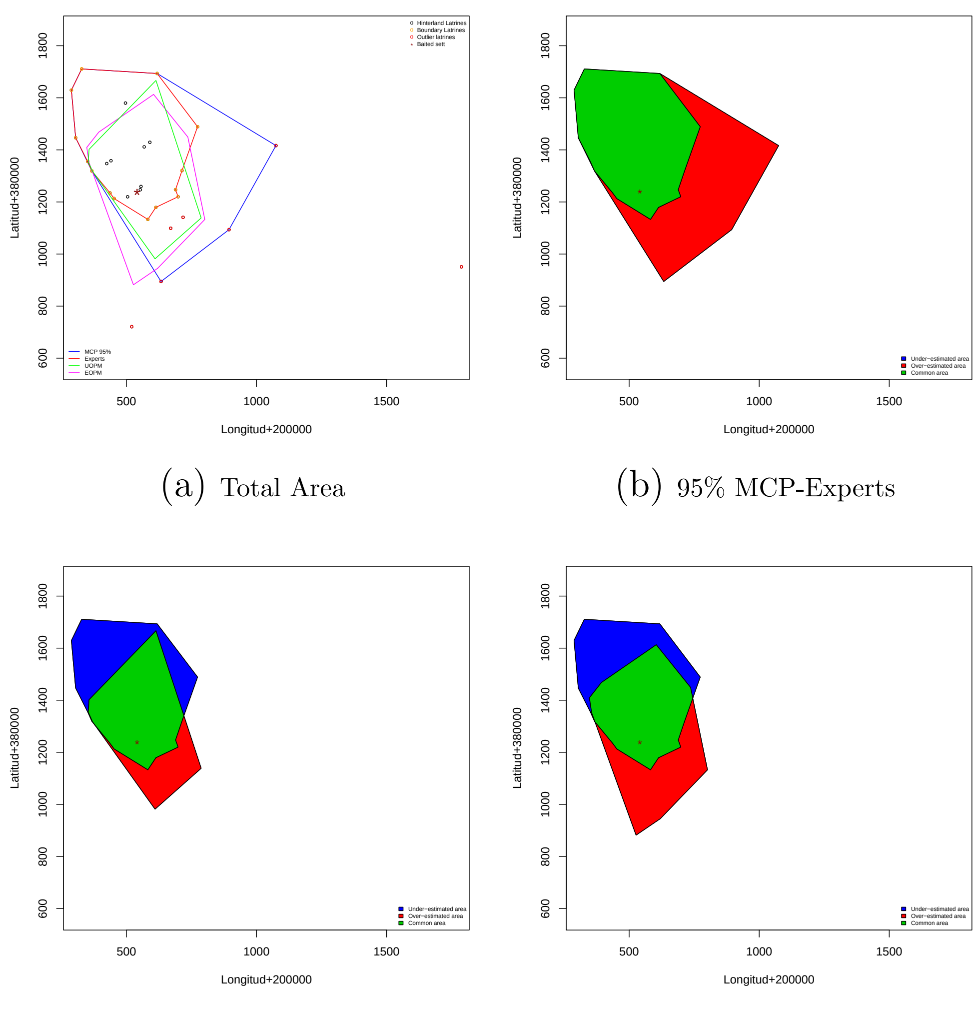
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

Figure A.9: Comparison of the reconstruction of the territory Kennel using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



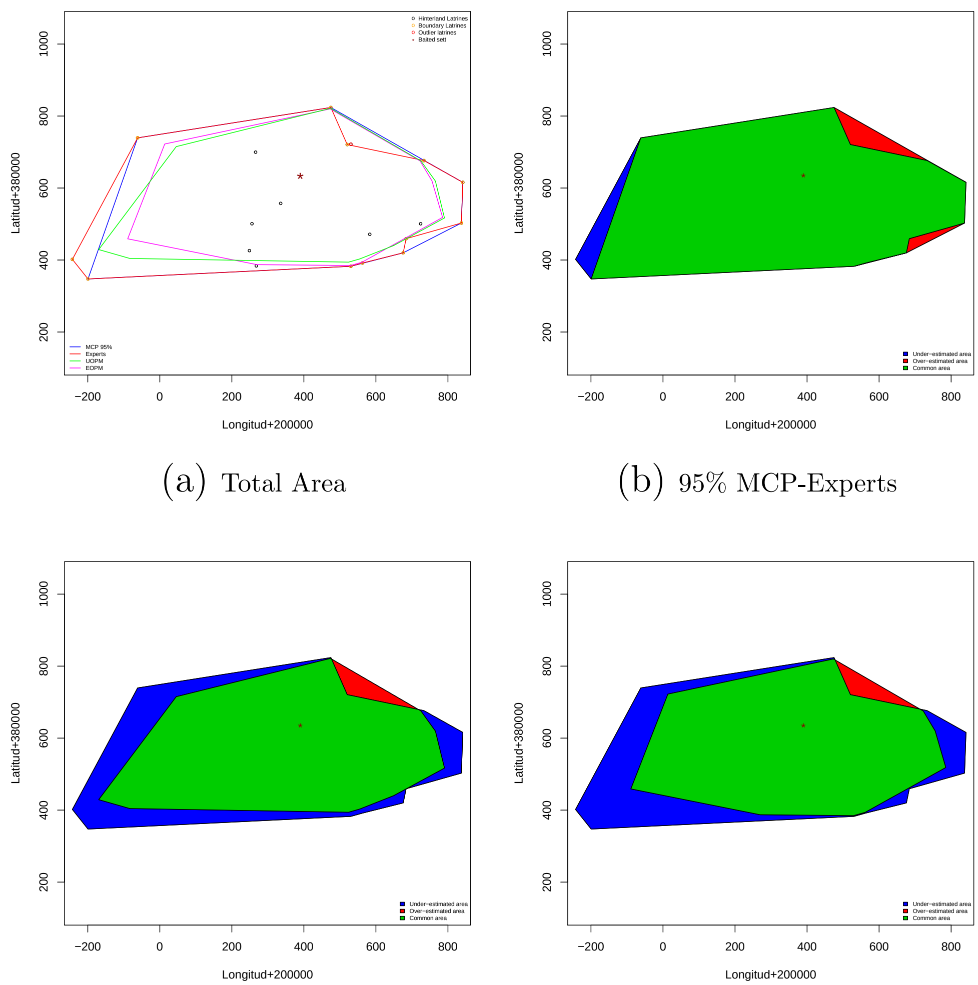
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

Figure A.10: Comparison of the reconstruction of the territory Larch using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



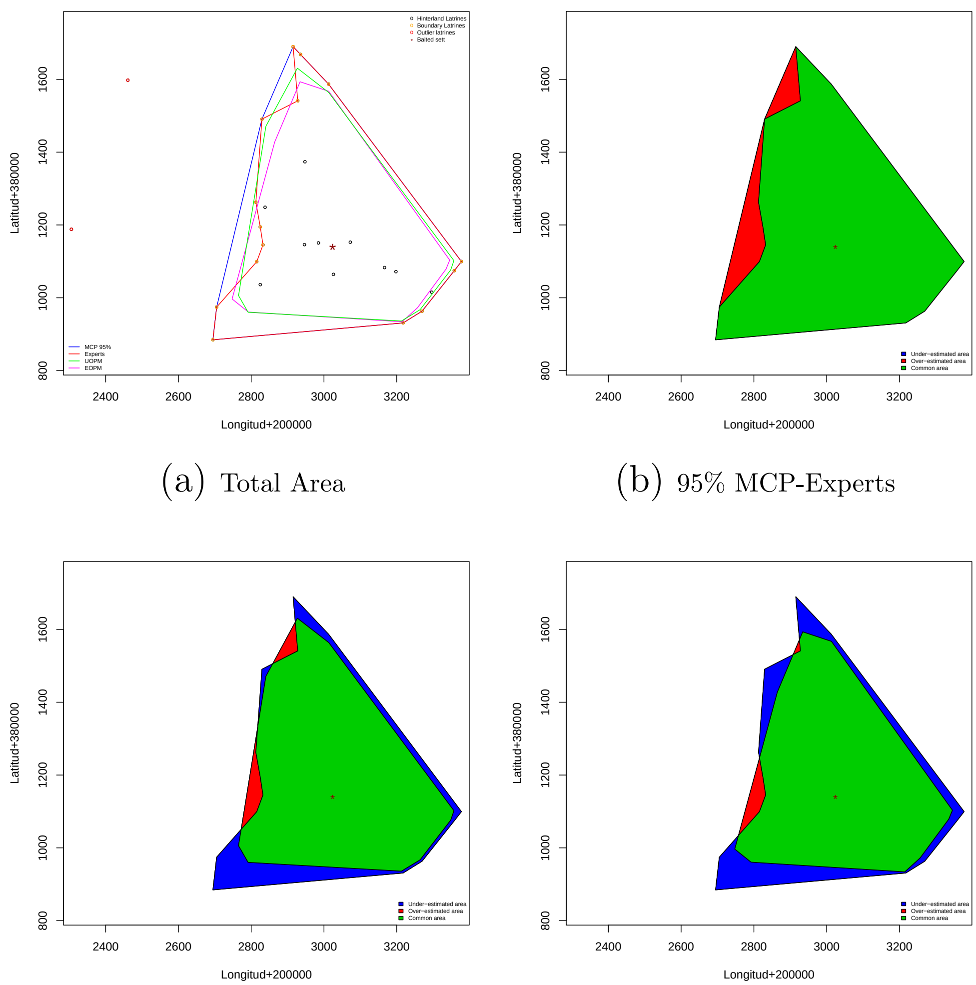
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

Figure A.11: Comparison of the reconstruction of the territory Mead using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



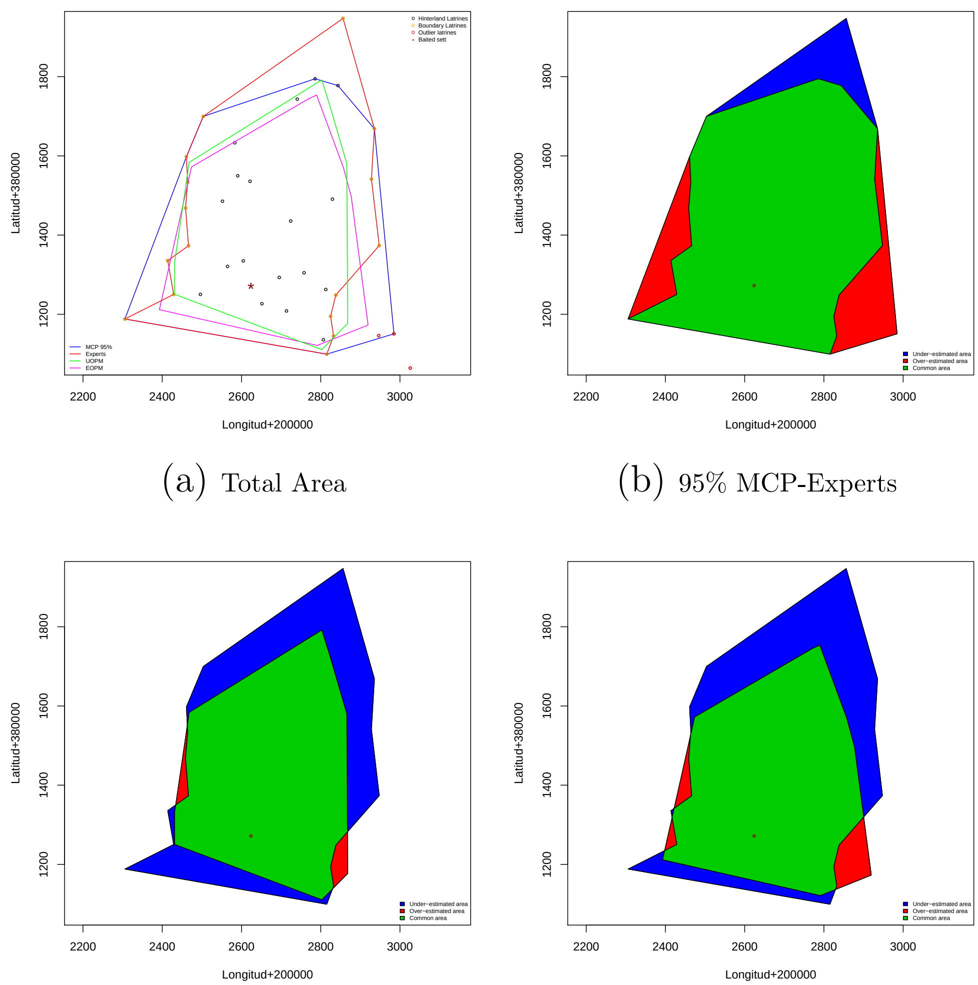
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

Figure A.12: Comparison of the reconstruction of the territory Nettle using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



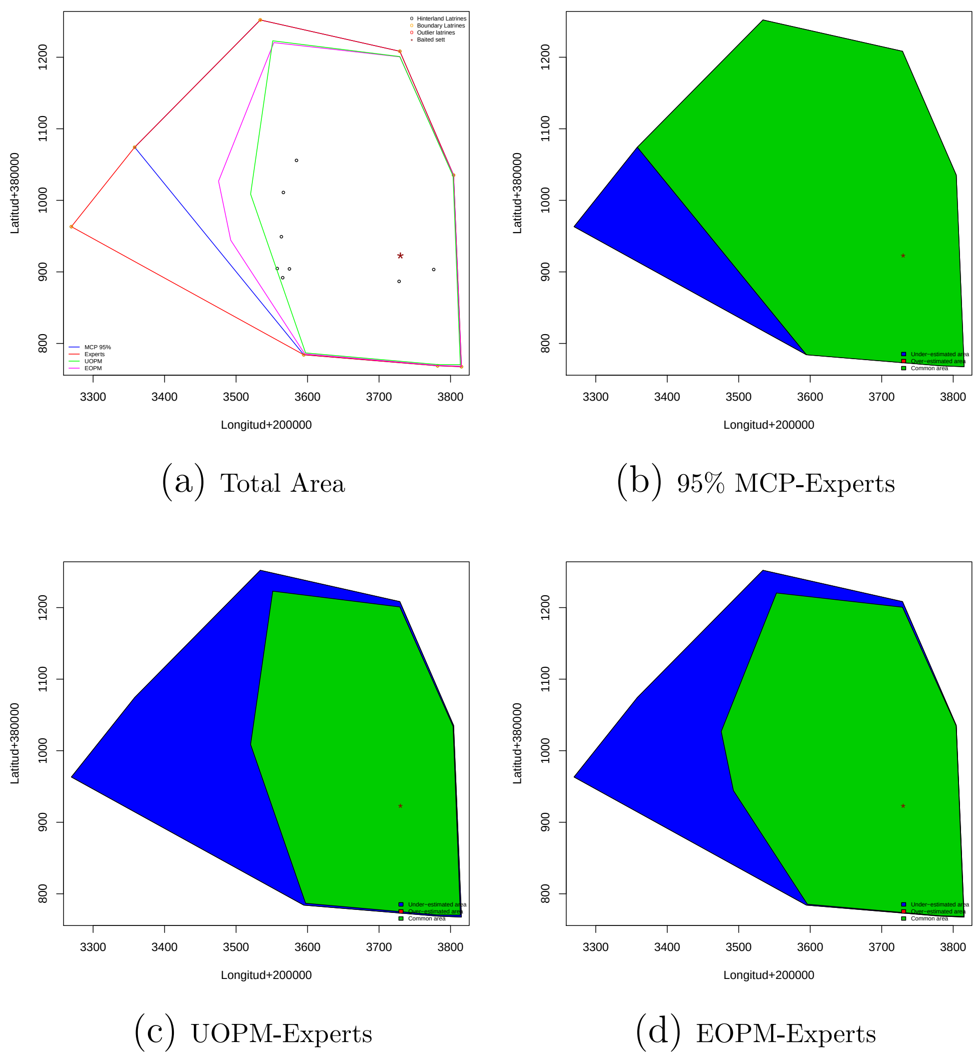
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

Figure A.13: Comparison of the reconstruction of the territory Old Oak using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



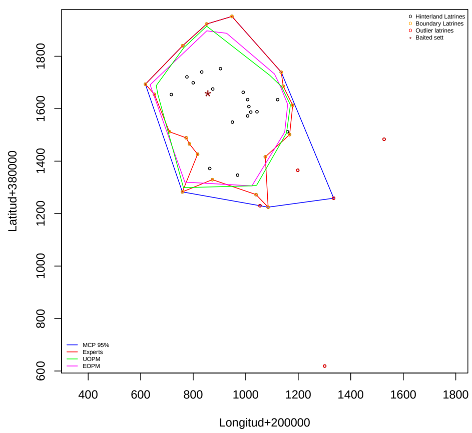
(a) Total Area

(b) 95% MCP-Experts

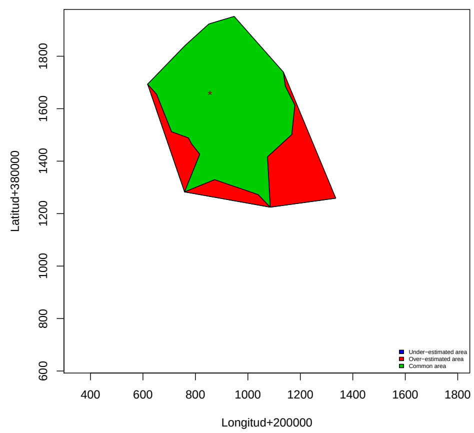
(c) UOPM-Experts

(d) EOPM-Experts

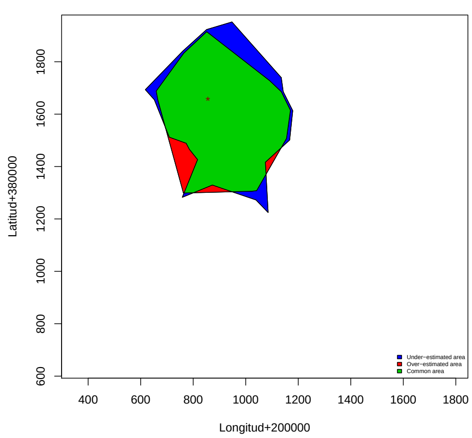
Figure A.14: Comparison of the reconstruction of the territory Parkmill using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



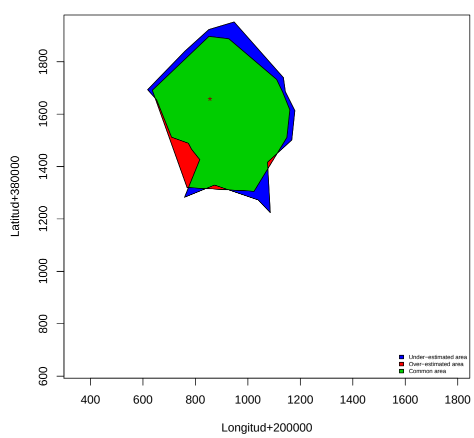
(a) Total Area



(b) 95% MCP-Experts

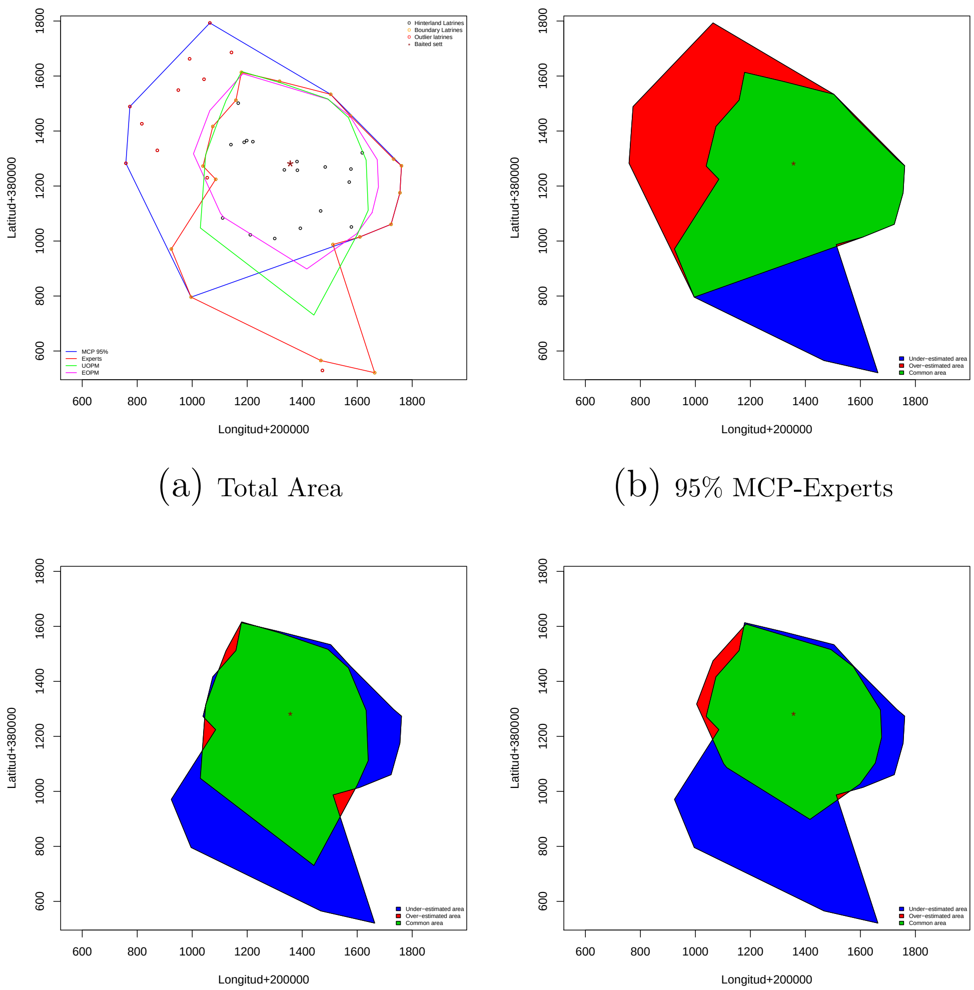


(c) UOPM-Experts



(d) EOPM-Experts

Figure A.15: Comparison of the reconstruction of the territory Rabbit using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



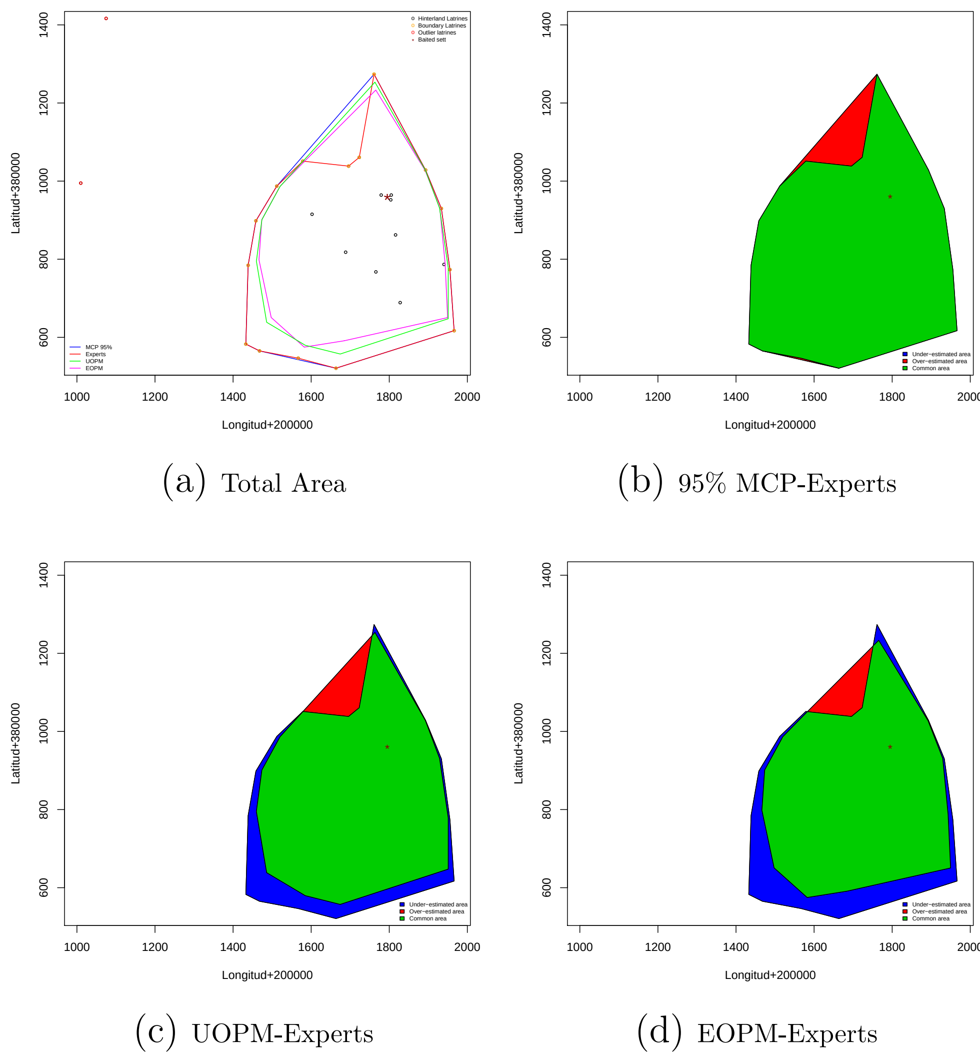
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

Figure A.16: Comparison of the reconstruction of the territory Septic Tank using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



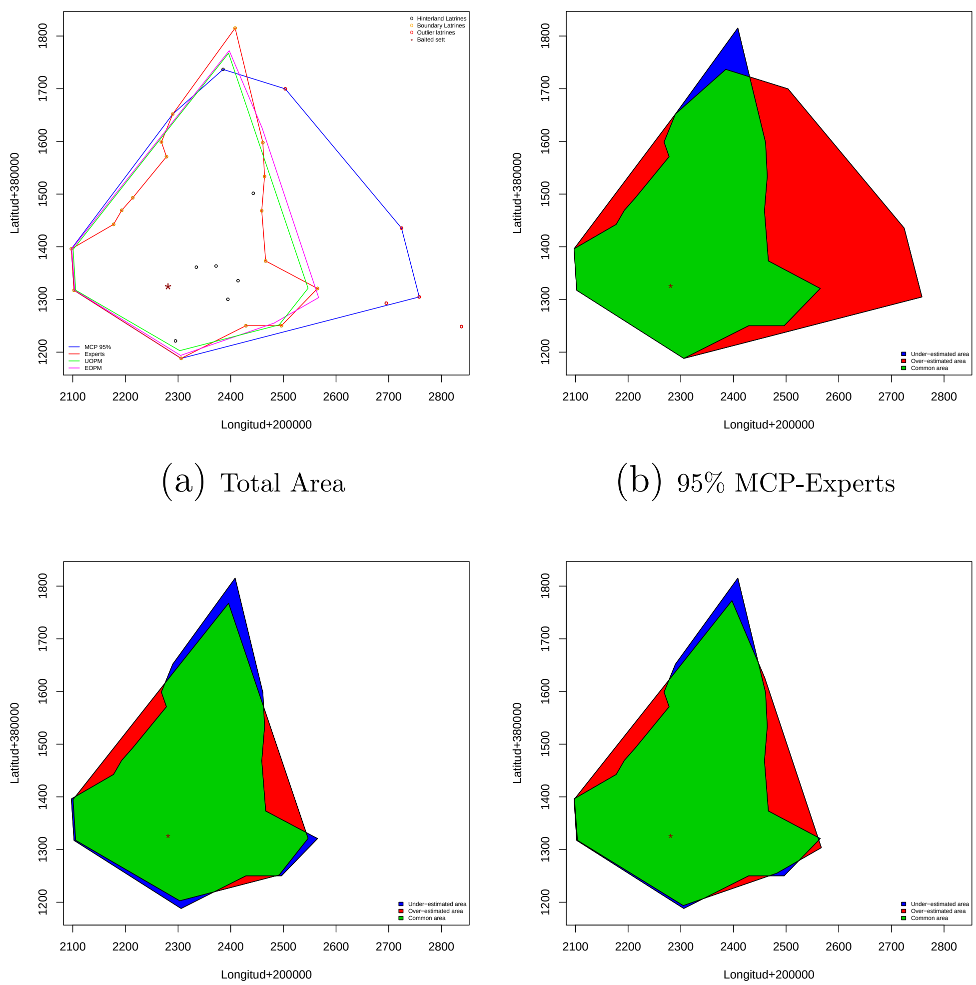
(a) Total Area

(b) 95% MCP-Experts

(c) UOPM-Experts

(d) EOPM-Experts

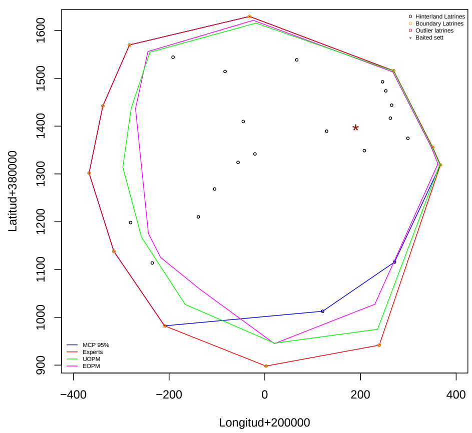
Figure A.17: Comparison of the reconstruction of the territory Top using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



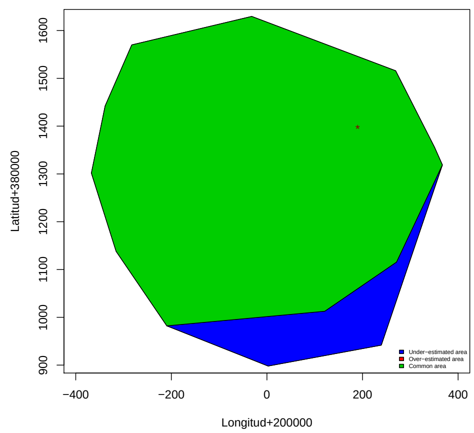
(a) Total Area (b) 95% MCP-Experts

(c) UOPM-Experts (d) EOPM-Experts

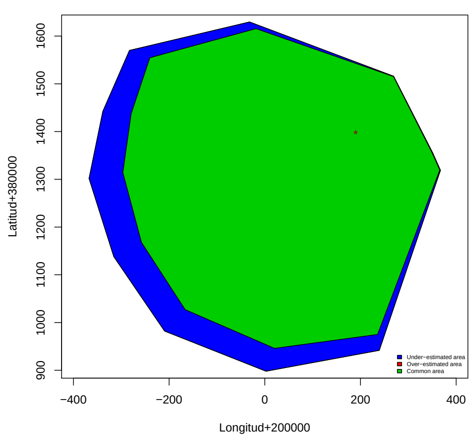
Figure A.18: Comparison of the reconstruction of the territory Trackside using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



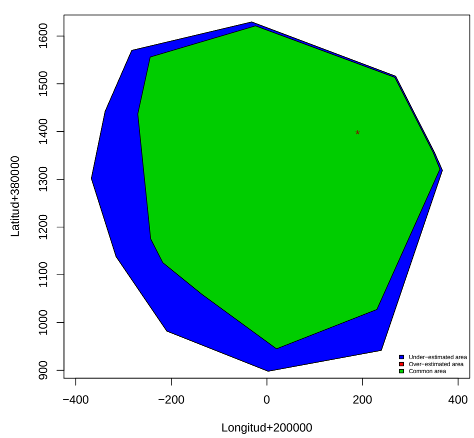
(a) Total Area



(b) 95% MCP-Experts

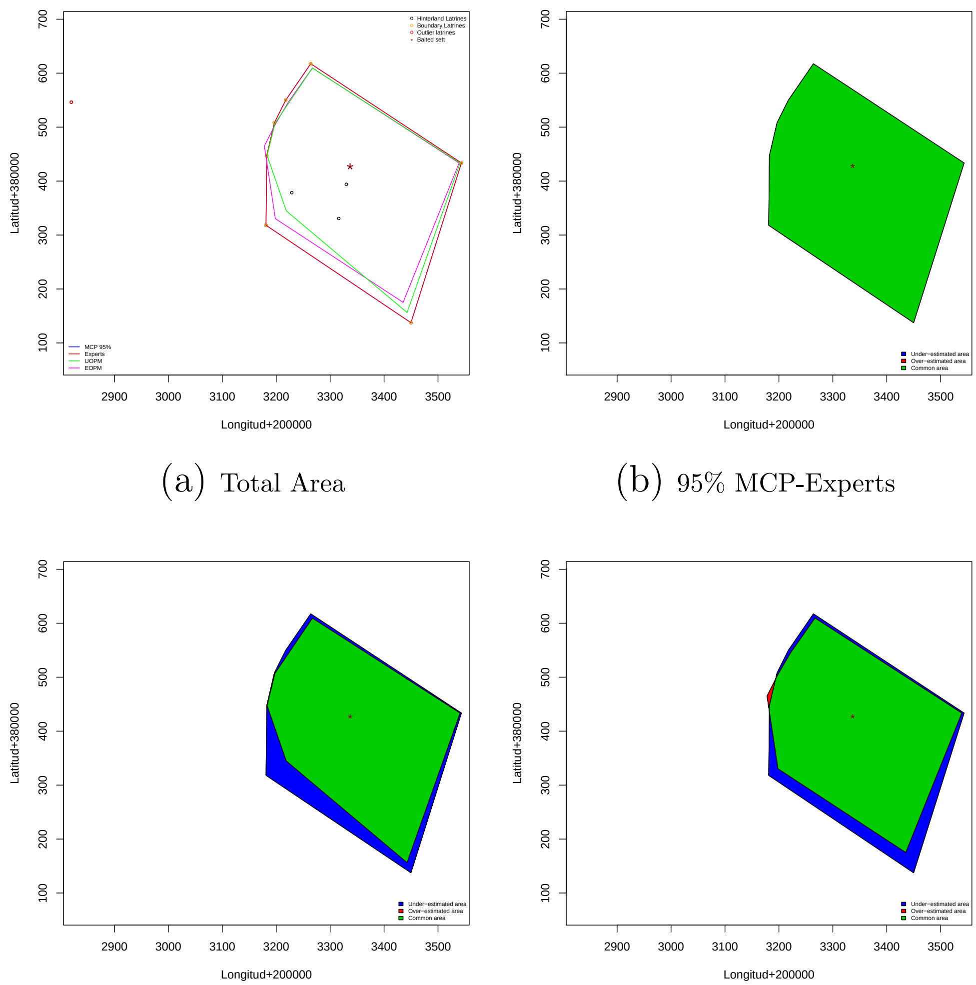


(c) UOPM-Experts



(d) EOPM-Experts

Figure A.19: Comparison of the reconstruction of the territory West using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



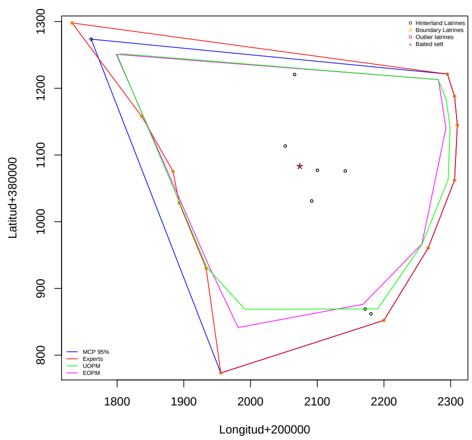
(a) Total Area

(b) 95% MCP-Experts

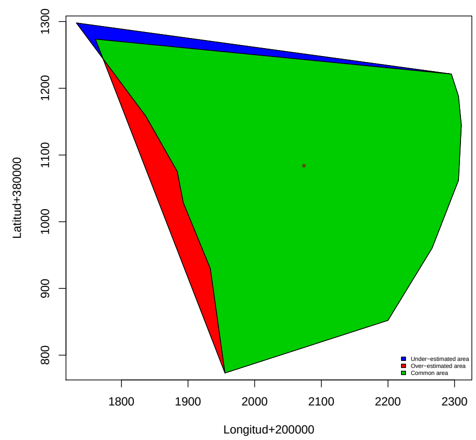
(c) UOPM-Experts

(d) EOPM-Experts

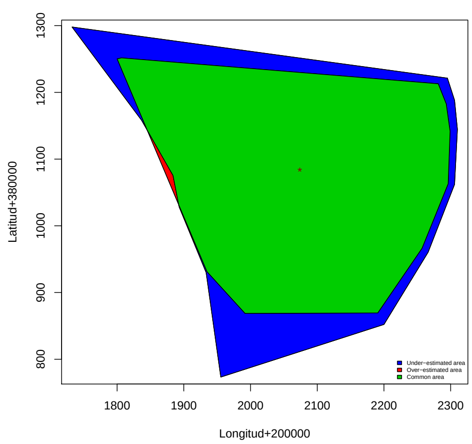
Figure A.20: Comparison of the reconstruction of the territory Wood Farm using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



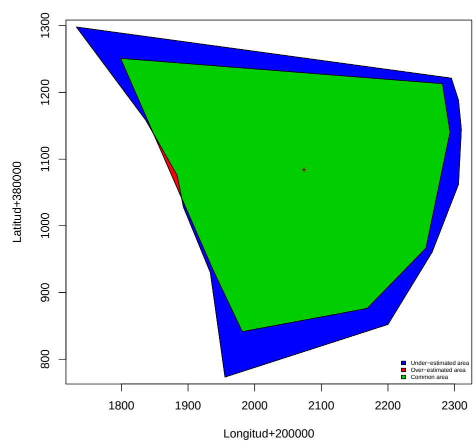
(a) Total Area



(b) 95% MCP-Experts

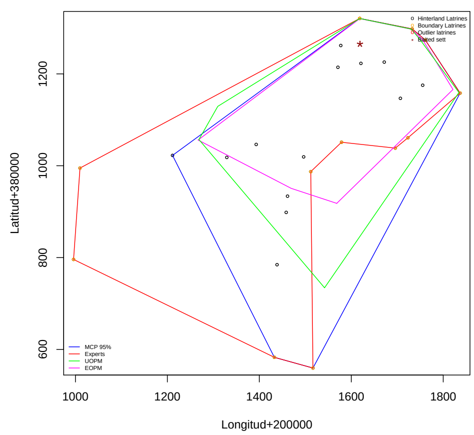


(c) UOPM-Experts

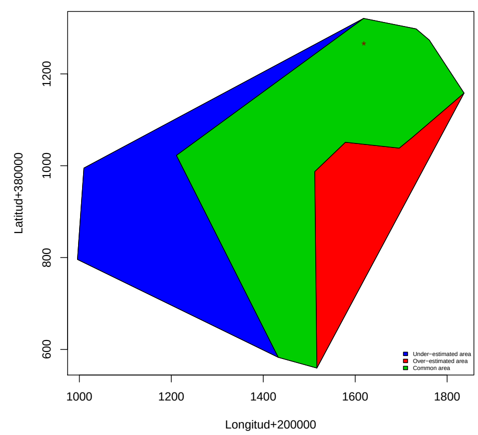


(d) EOPM-Experts

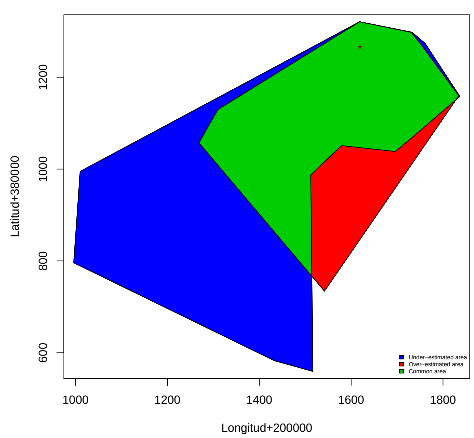
Figure A.21: Comparison of the reconstruction of the territory Wych Elm using the Point Estimation Approach showing Total, Under and Over-Estimated areas.



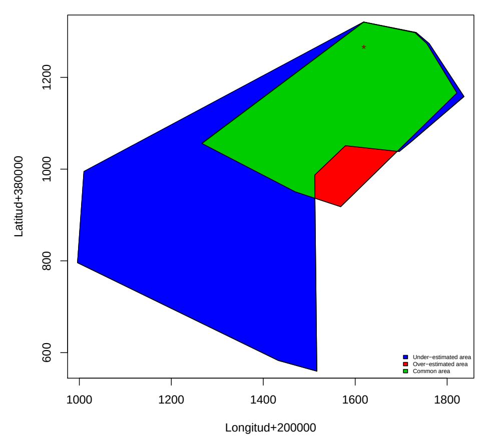
(a) Total Area



(b) 95% MCP-Experts



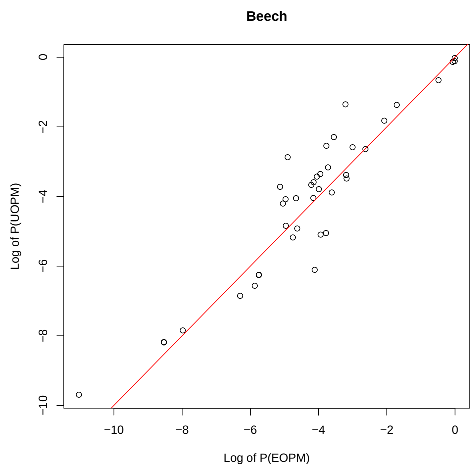
(c) UOPM-Experts



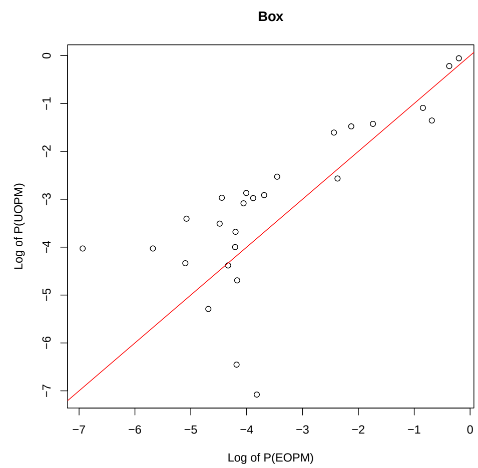
(d) EOPM-Experts

Figure A.22: Comparison of the reconstruction of the territory Yew using the Point Estimation Approach showing Total, Under and Over-Estimated areas.

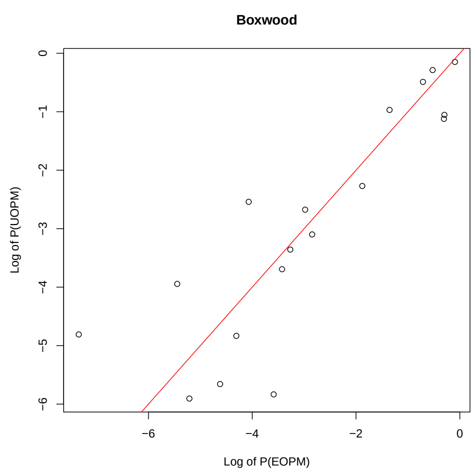
A.2.2 Comparisons of the log p(UOPM) and the log p(EOPM)



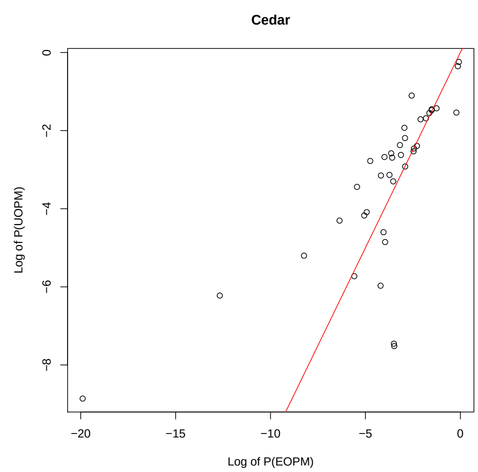
(a) Beech territory



(b) Box territory

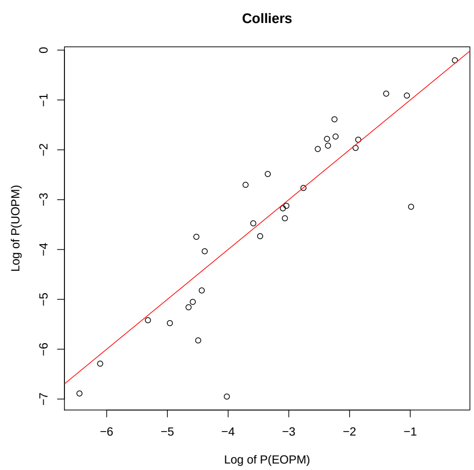


(c) Boxwood territory

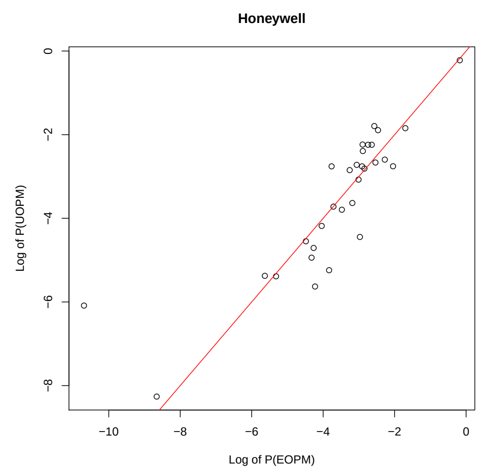


(d) Cedar territory

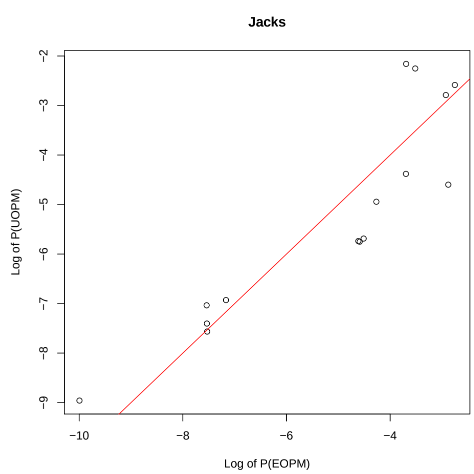
Figure A.23: Comparisons of the log p(UOPM) and the log p(EOPM)



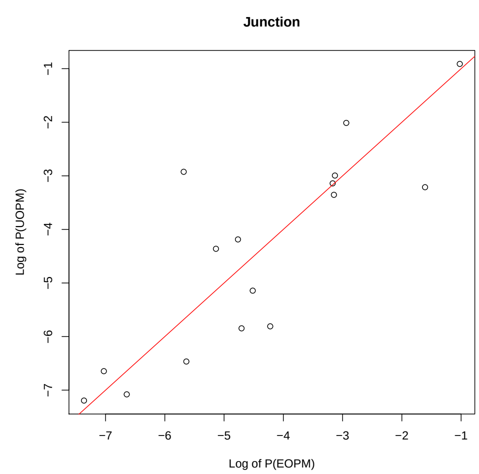
(a) Colliers territory



(b) Honeywell territory

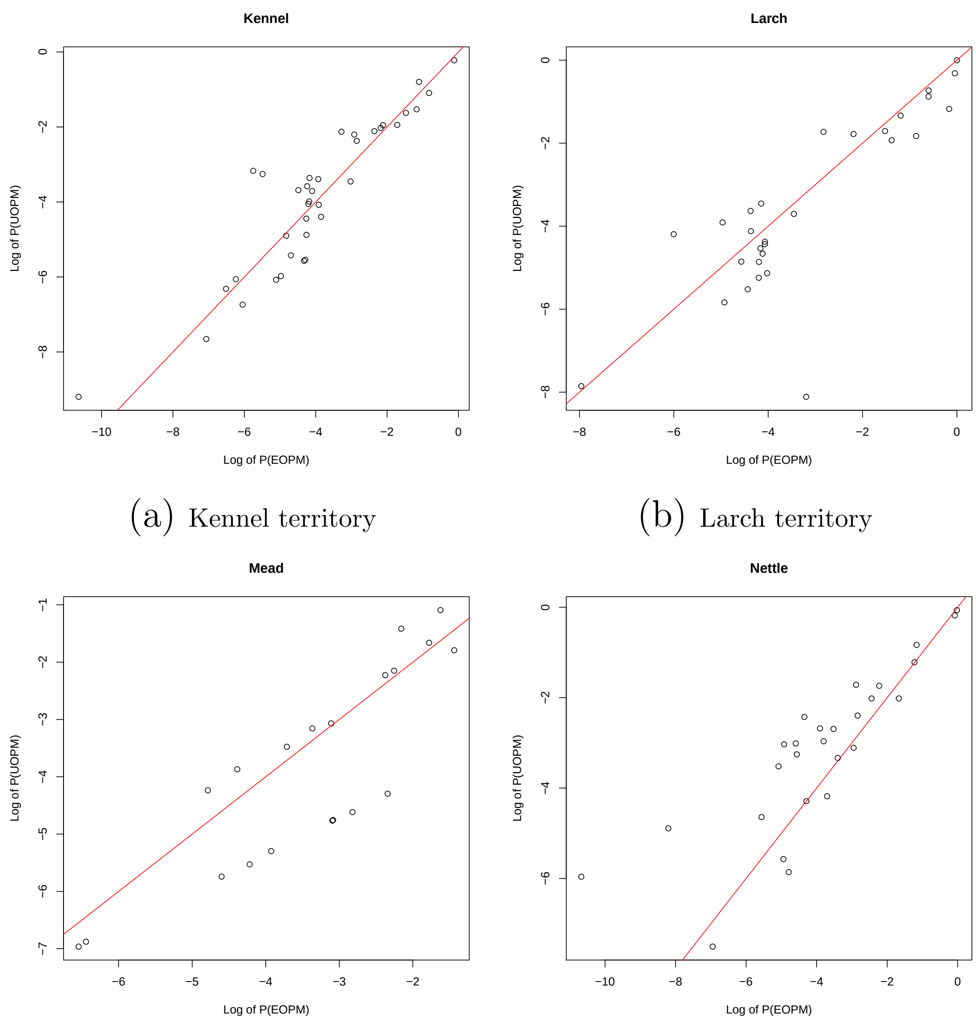


(c) Jacks territory



(d) Junction territory

Figure A.24: Comparisons of the log p(UOPM) and the log p(EOPM)



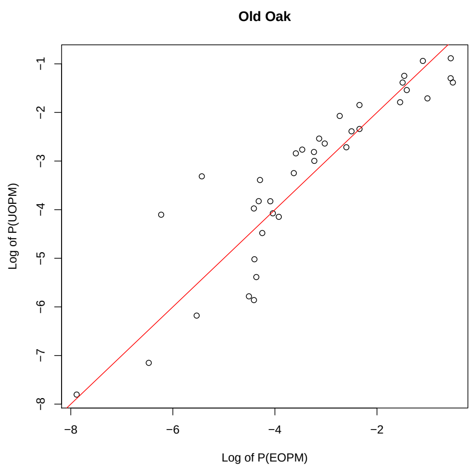
(a) Kennel territory

(b) Larch territory

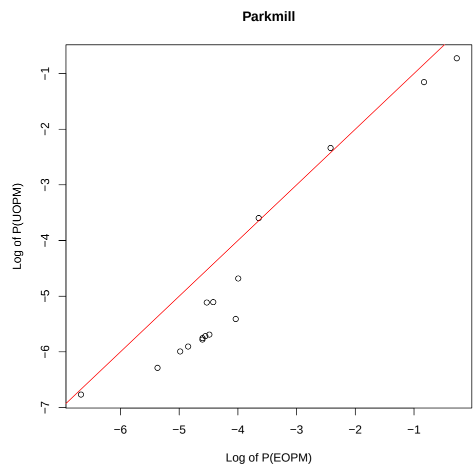
(c) Mead territory

(d) Nettle territory

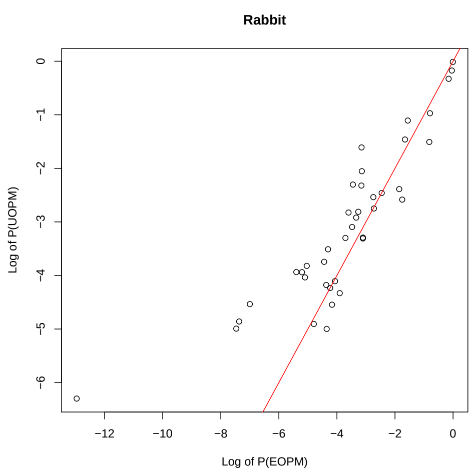
Figure A.25: Comparisons of the log p(UOPM) and the log p(EOPM)



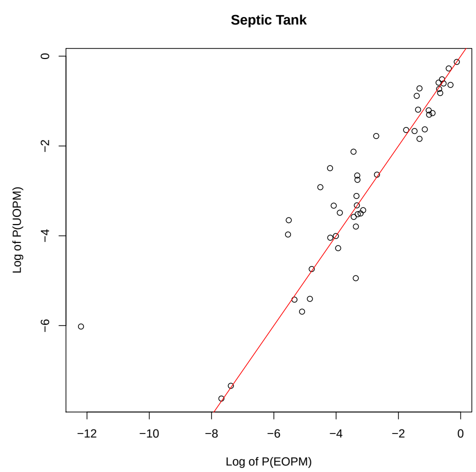
(a) Old Oak territory



(b) Parkmill territory

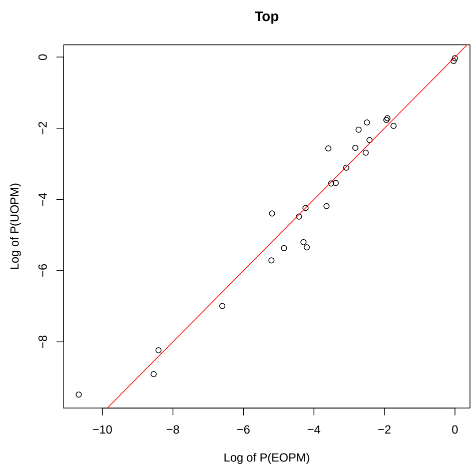


(c) Rabbit territory

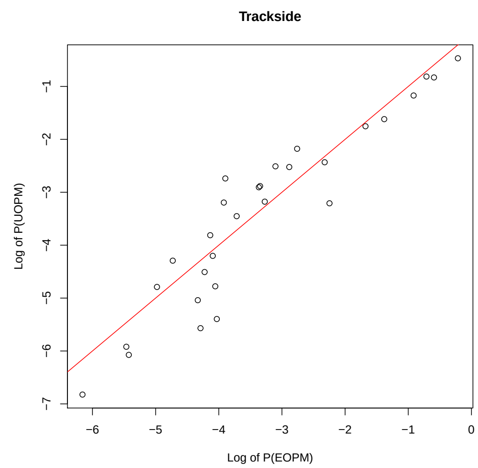


(d) Septic Tank territory

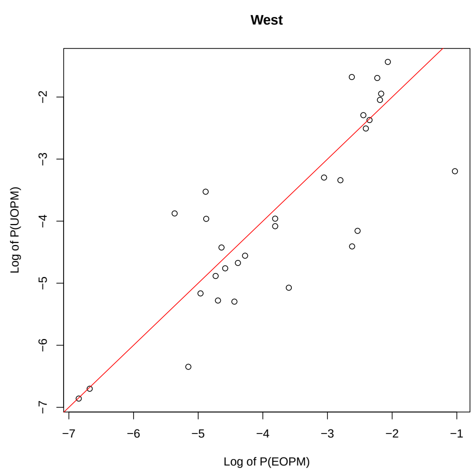
Figure A.26: Comparisons of the log p(UOPM) and the log p(EOPM)



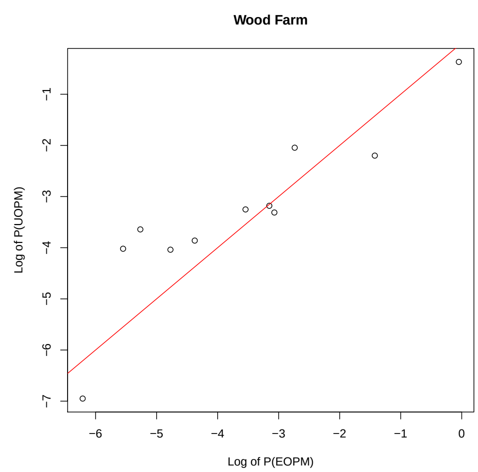
(a) Top territory



(b) Trakside territory

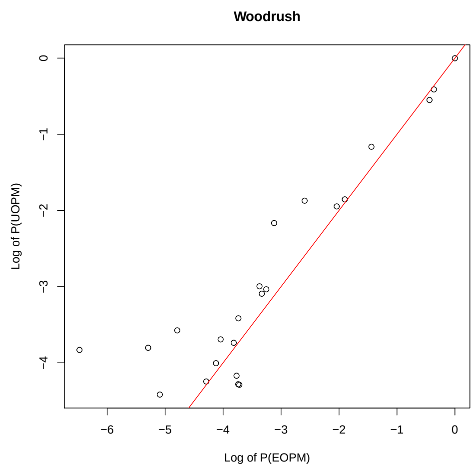


(c) West territory

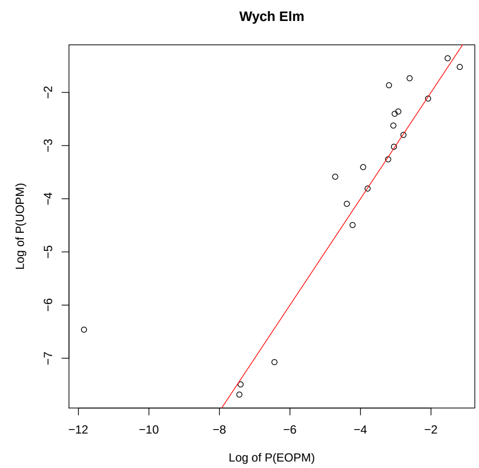


(d) Wood Farm territory

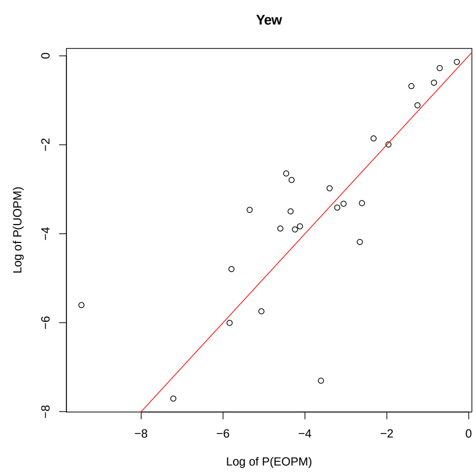
Figure A.27: Comparisons of the log p(UOPM) and the log p(EOPM)



(a) Woodrush territory



(b) Wych Elm territory



(c) Yew territory

Figure A.28: Comparisons of the log p(UOPM) and the log p(EOPM)

A.2.3 Reconstruction of the badger territories using the Sampling approach

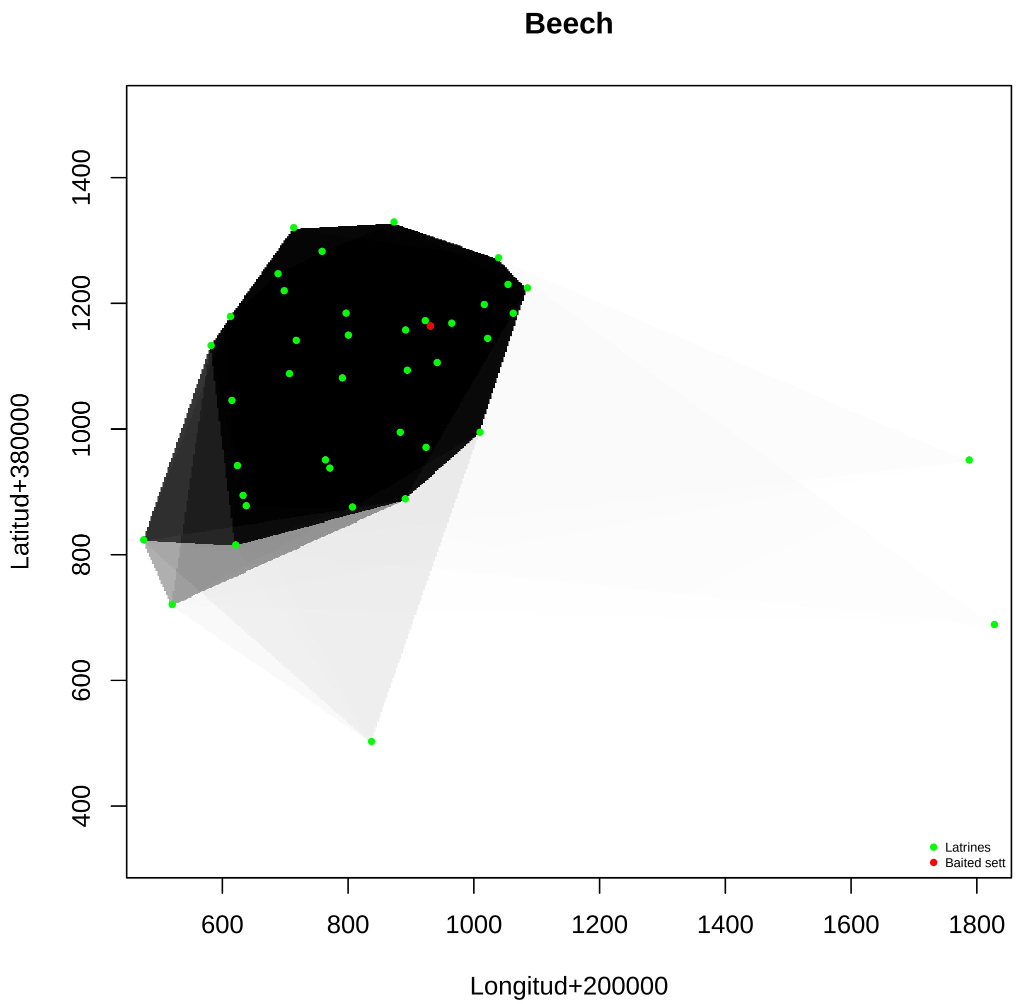


Figure A.29: Probability of inclusion, Beech territory using the Unconditional Outlier Prediction Model sampling method.

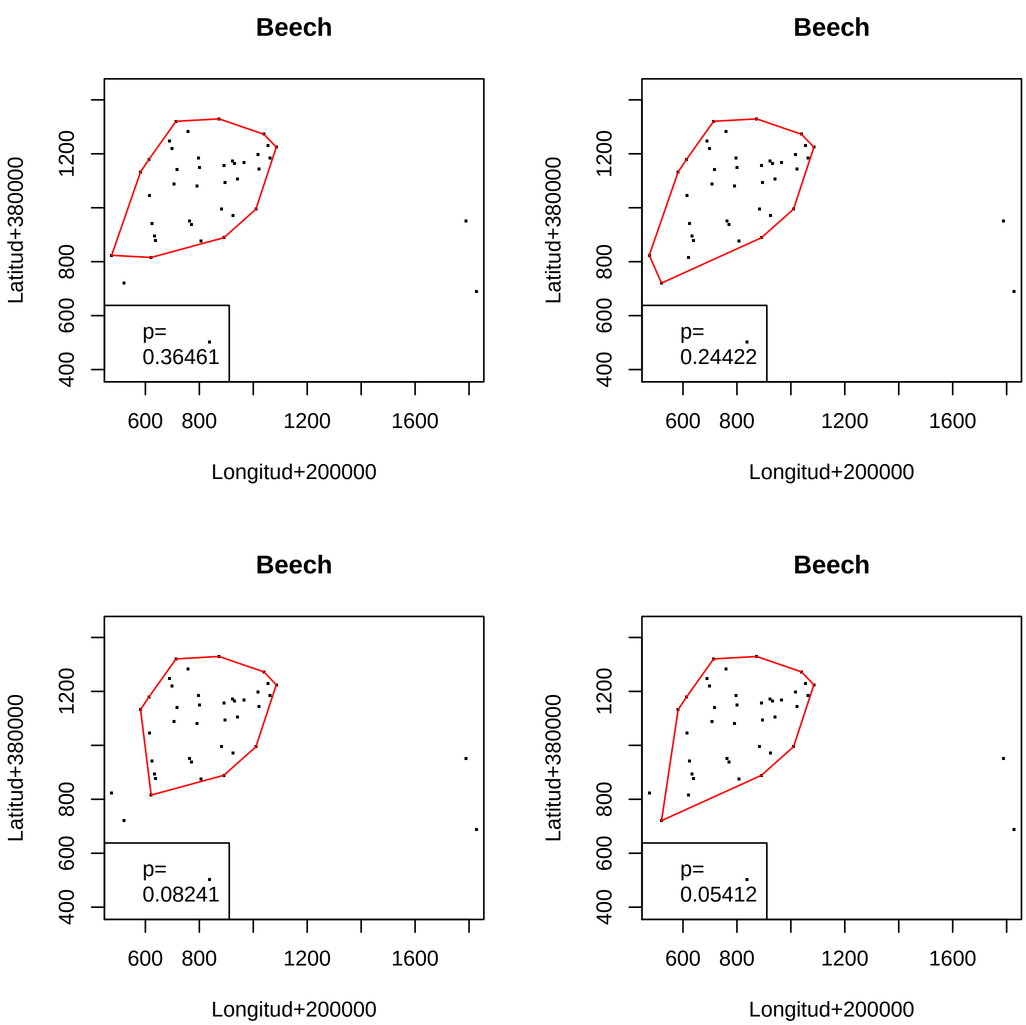


Figure A.30: Most frequent reconstructions of the Beech territory using the sampling method of the Unconditional Outlier Prediction Model

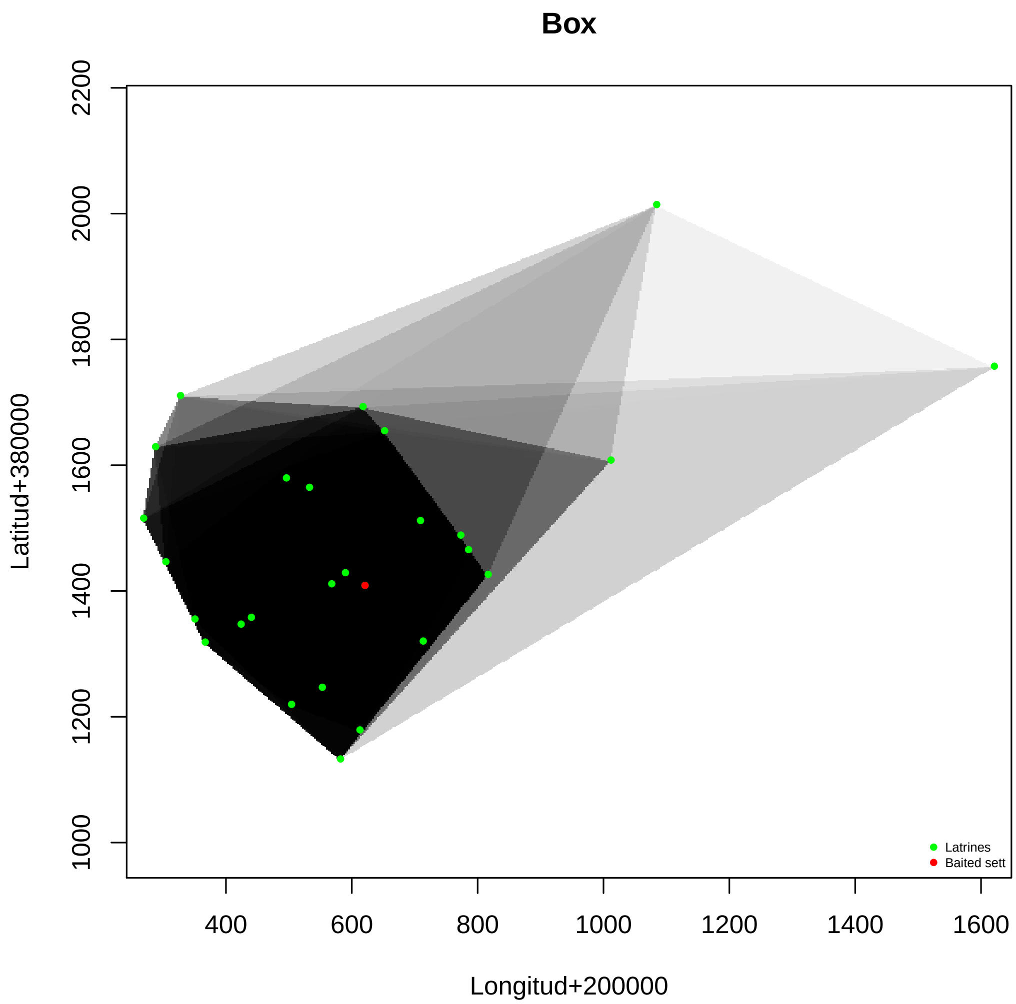


Figure A.31: Probability of inclusion, Box territory using the Unconditional Outlier Prediction Model sampling method.

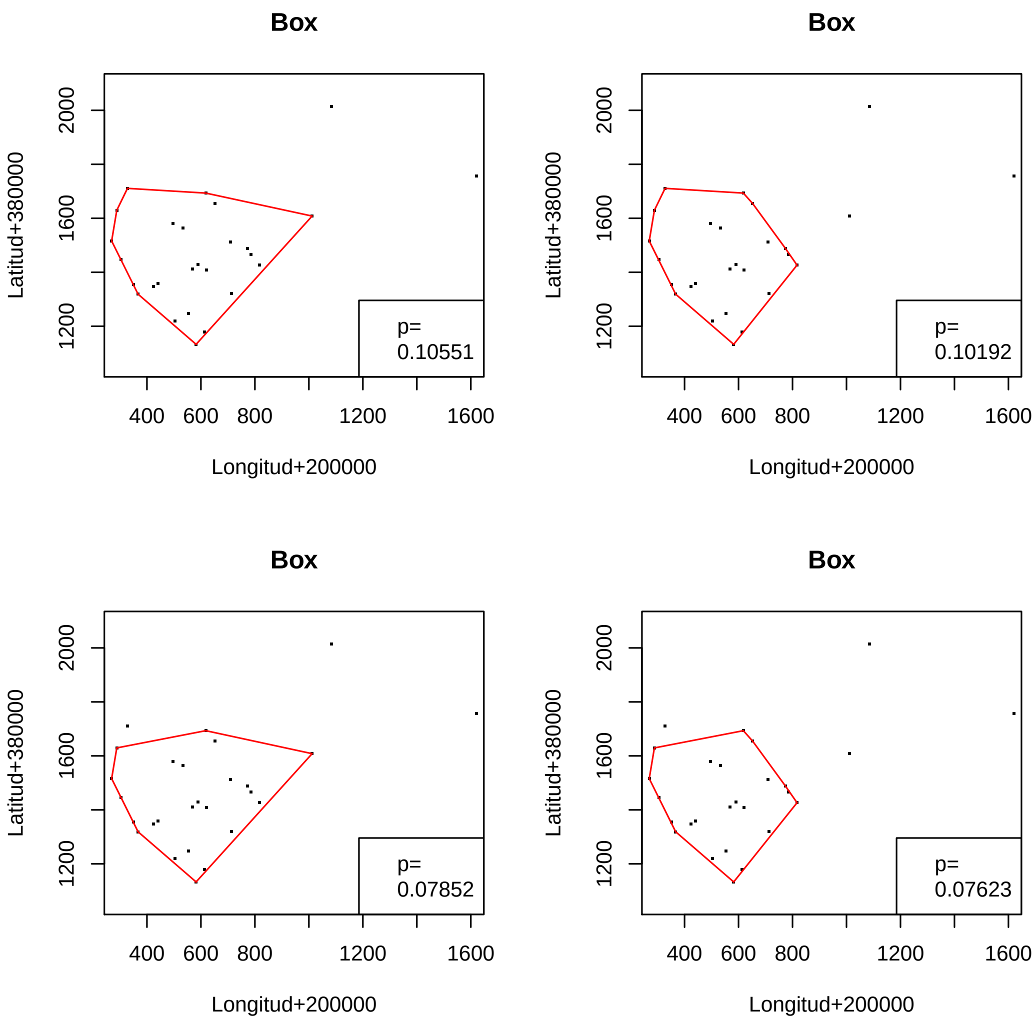


Figure A.32: Most frequent reconstructions of the Box territory using the sampling method of the Unconditional Outlier Prediction Model

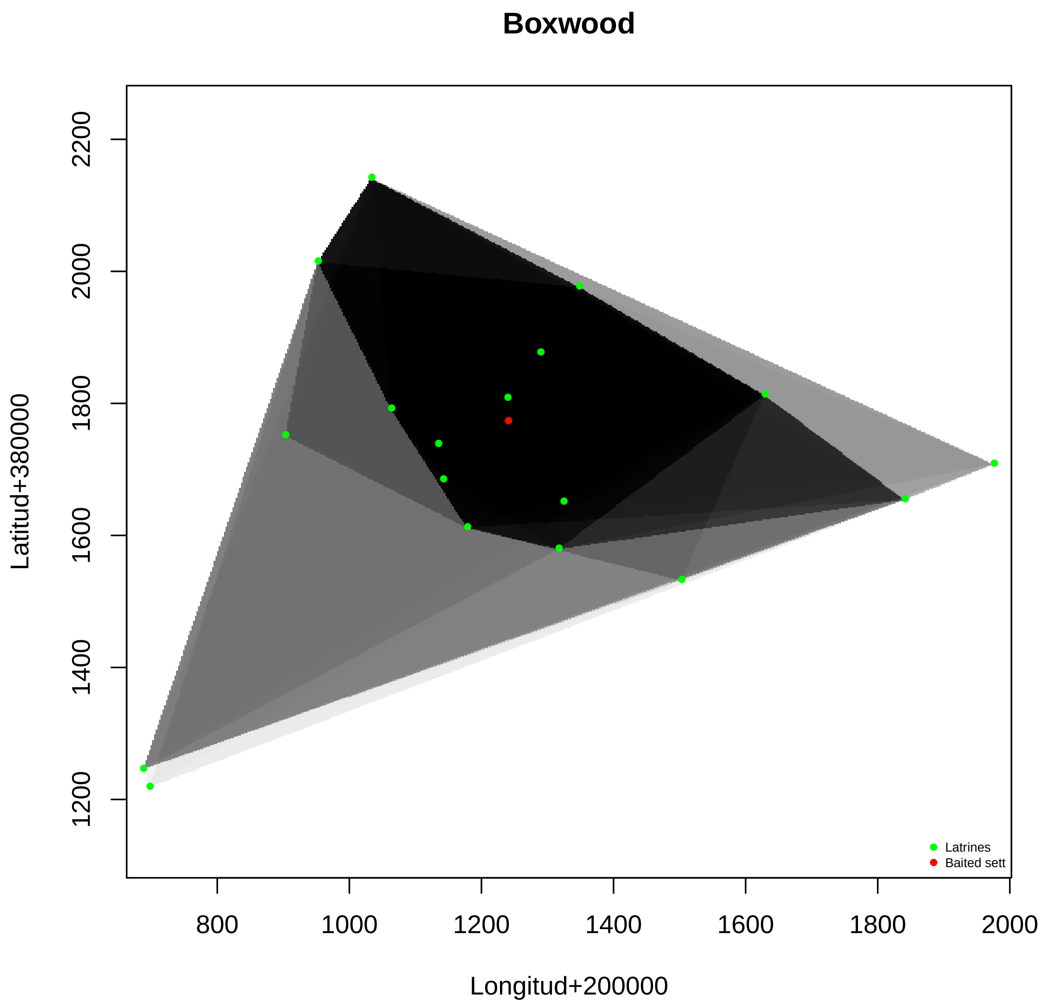


Figure A.33: Probability of inclusion, Boxwood territory using the Unconditional Outlier Prediction Model sampling method.

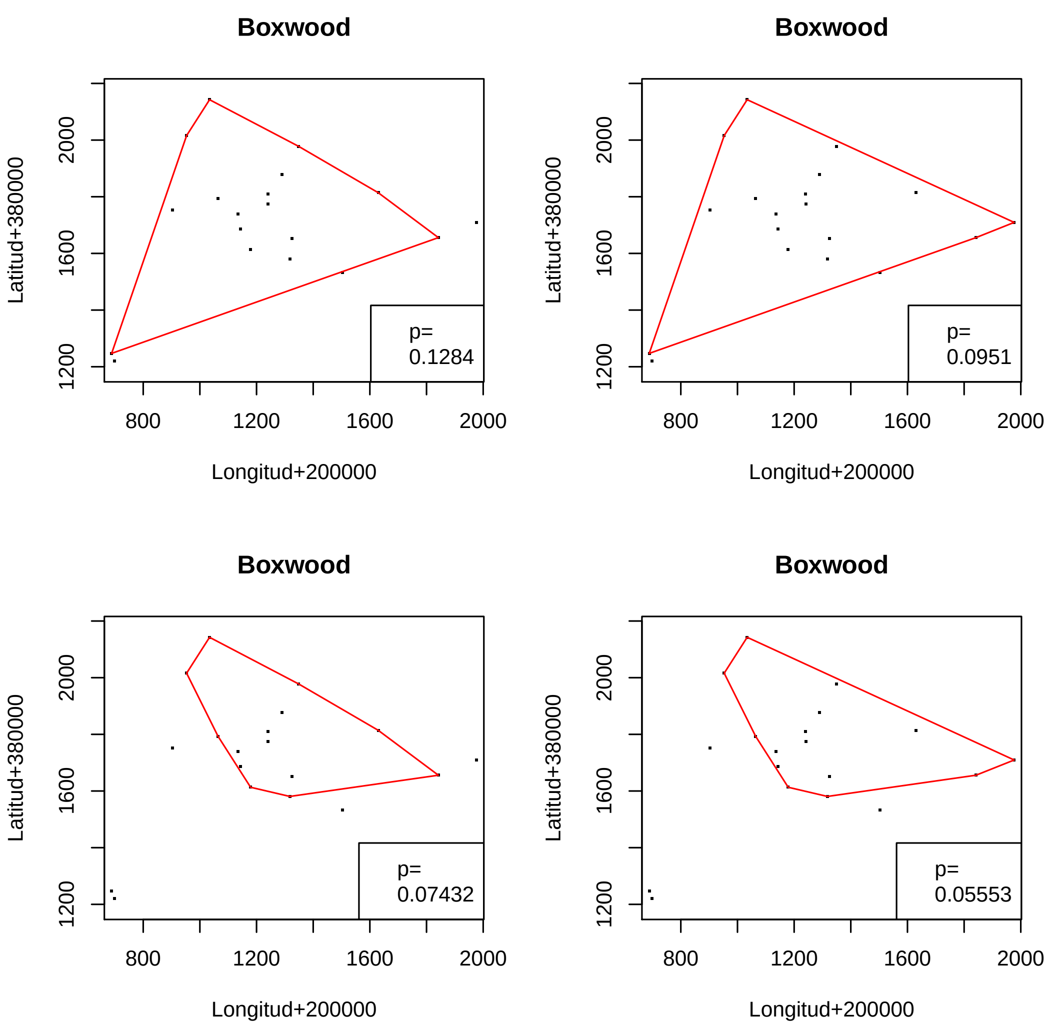


Figure A.34: Most frequent reconstructions of the Boxwood territory using the sampling method of the Unconditional Outlier Prediction Model

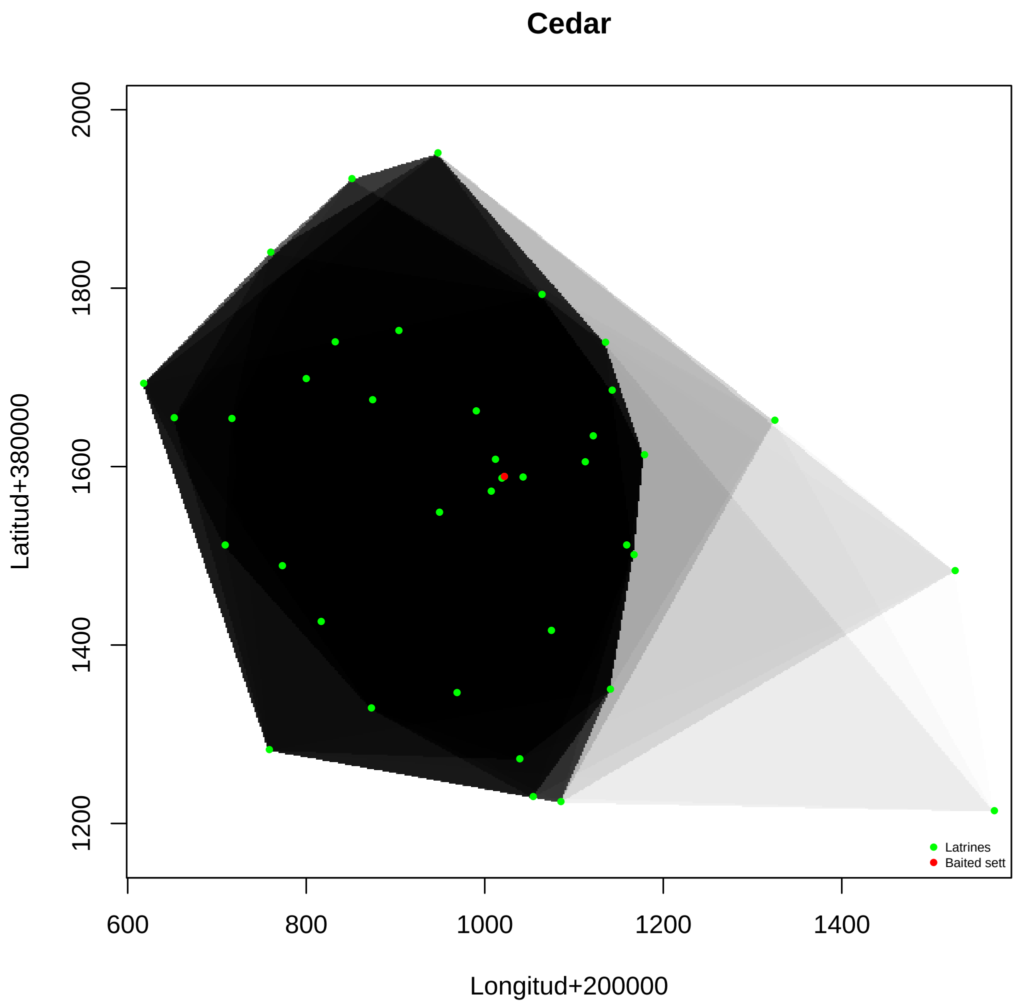


Figure A.35: Probability of inclusion, Cedar territory using the Unconditional Outlier Prediction Model sampling method.

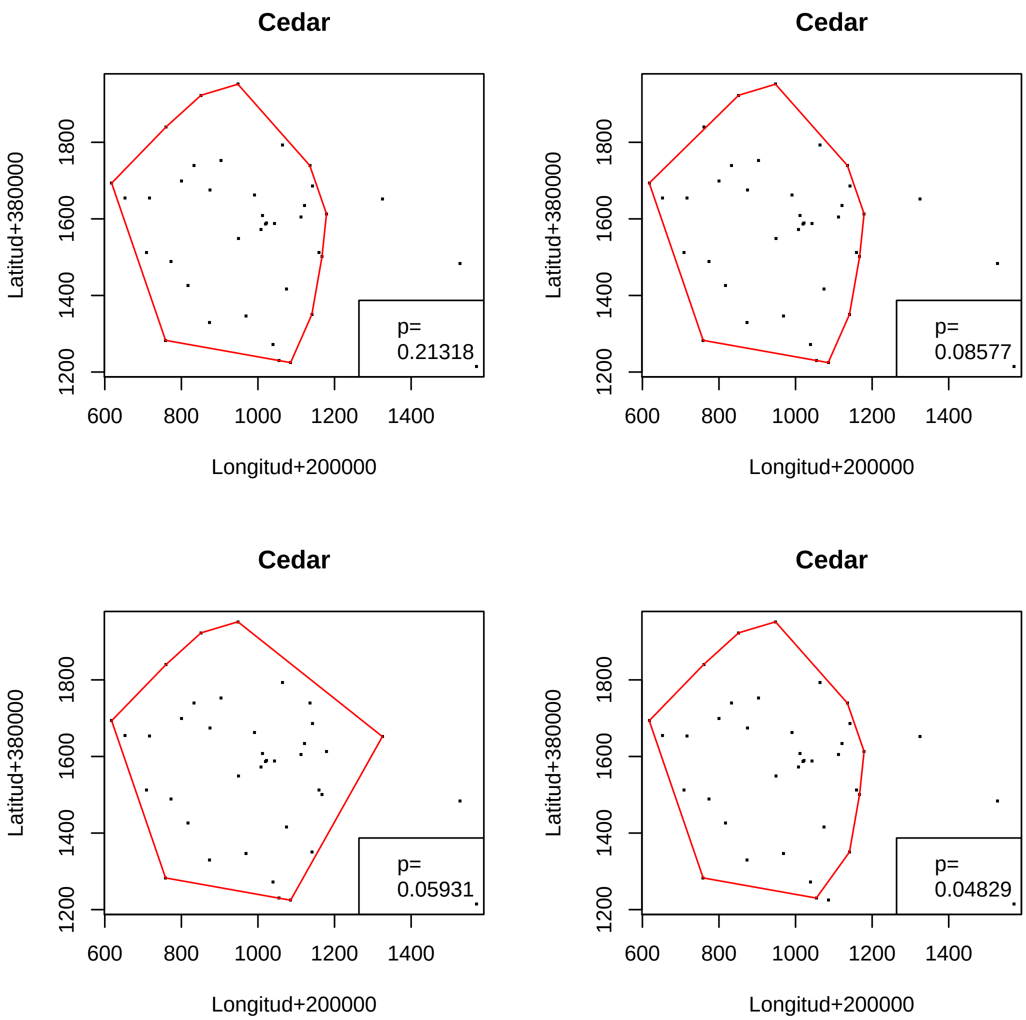


Figure A.36: Most frequent reconstructions of the Cedar territory using the sampling method of the Unconditional Outlier Prediction Model

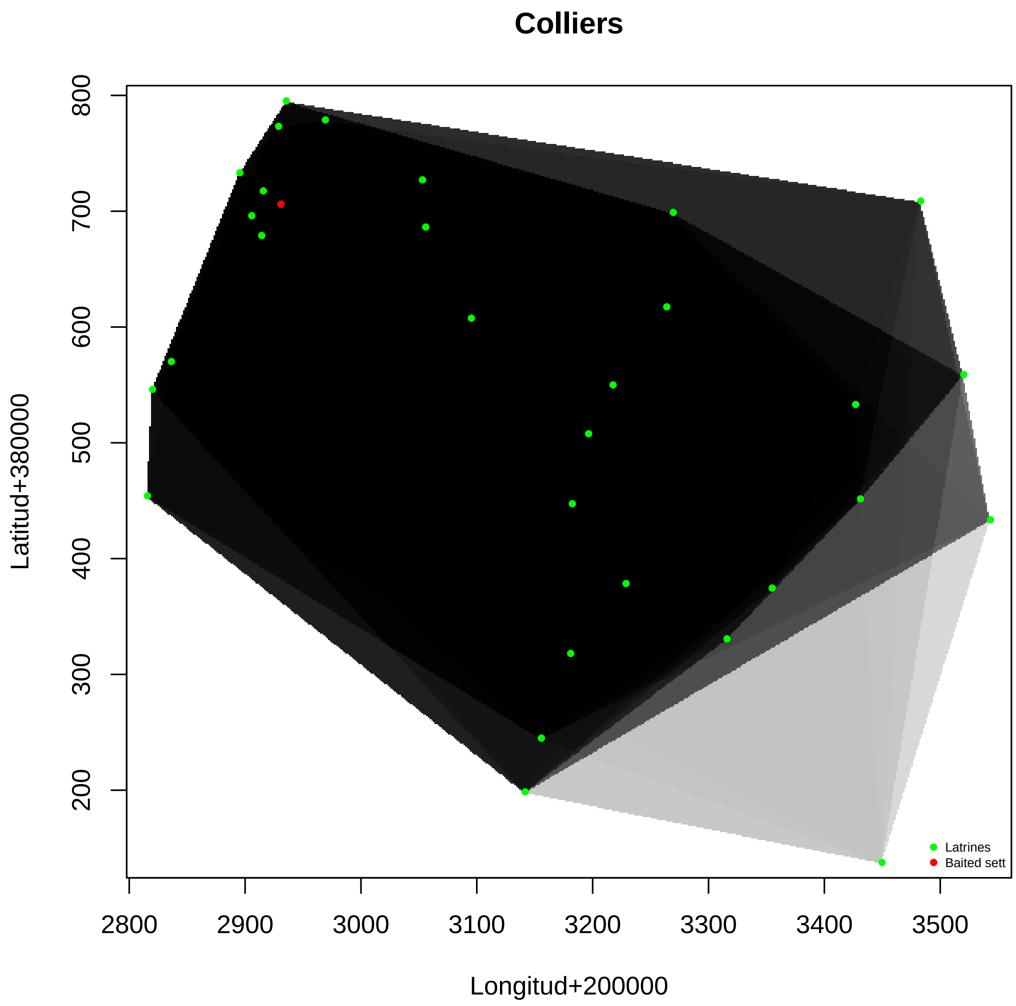


Figure A.37: Probability of inclusion, Colliers territory using the Unconditional Outlier Prediction Model sampling method.

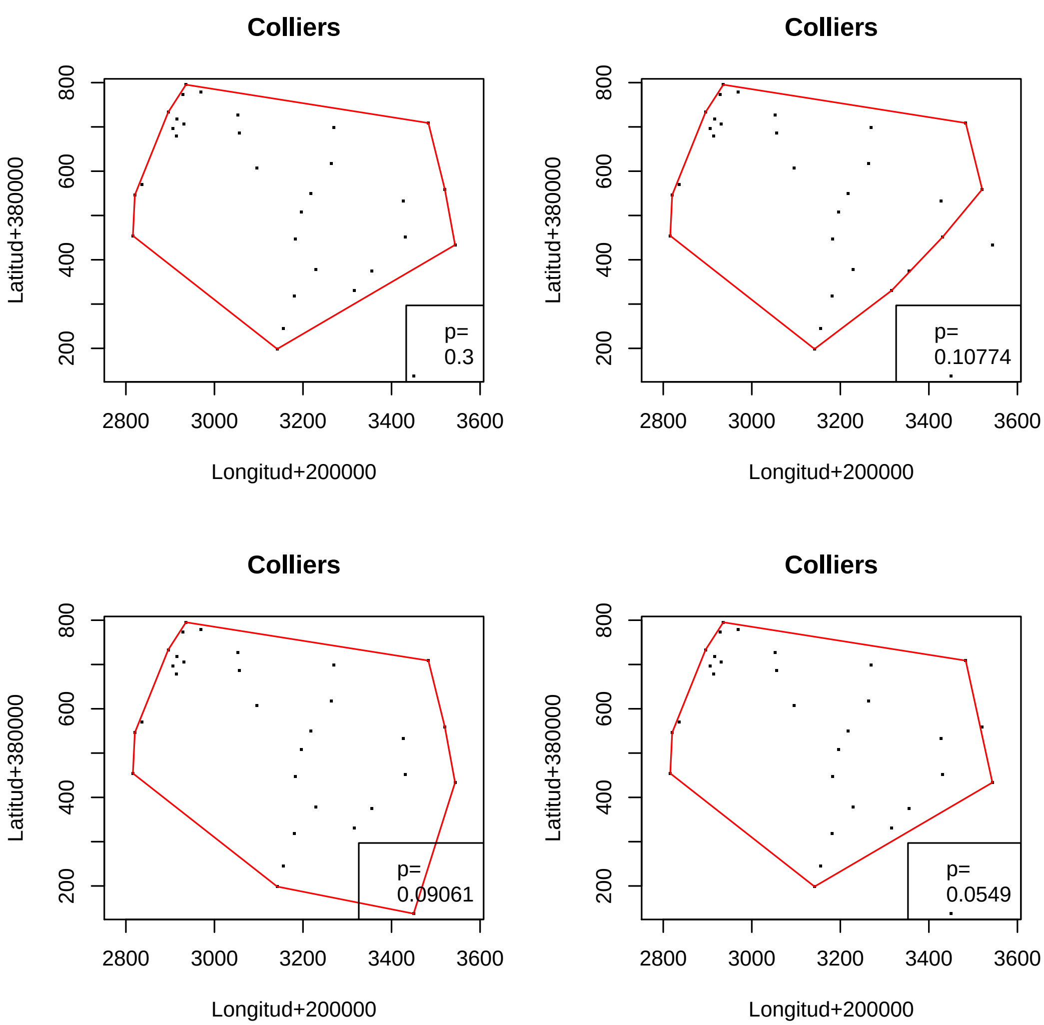


Figure A.38: Most frequent reconstructions of the Colliers territory using the sampling method of the Unconditional Outlier Prediction Model

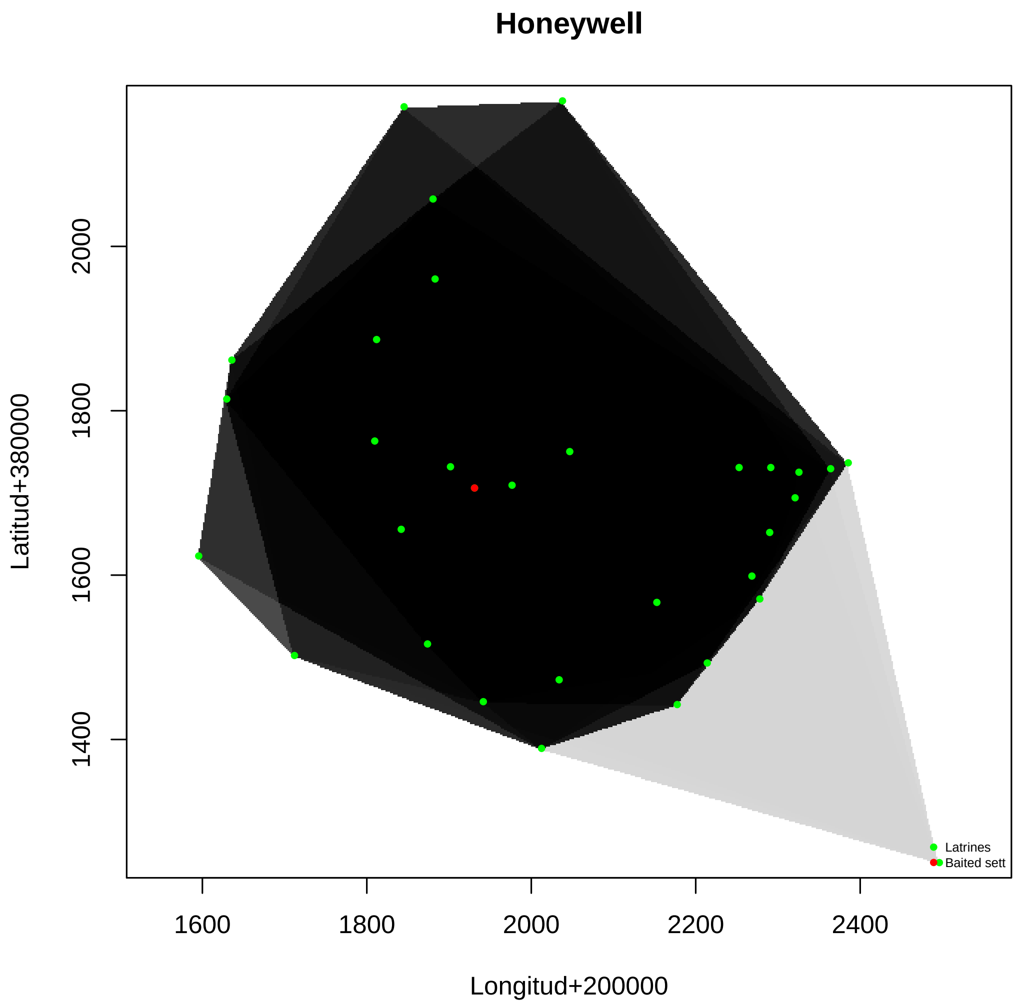


Figure A.39: Probability of inclusion, Honeywell territory using the Unconditional Outlier Prediction Model sampling method.

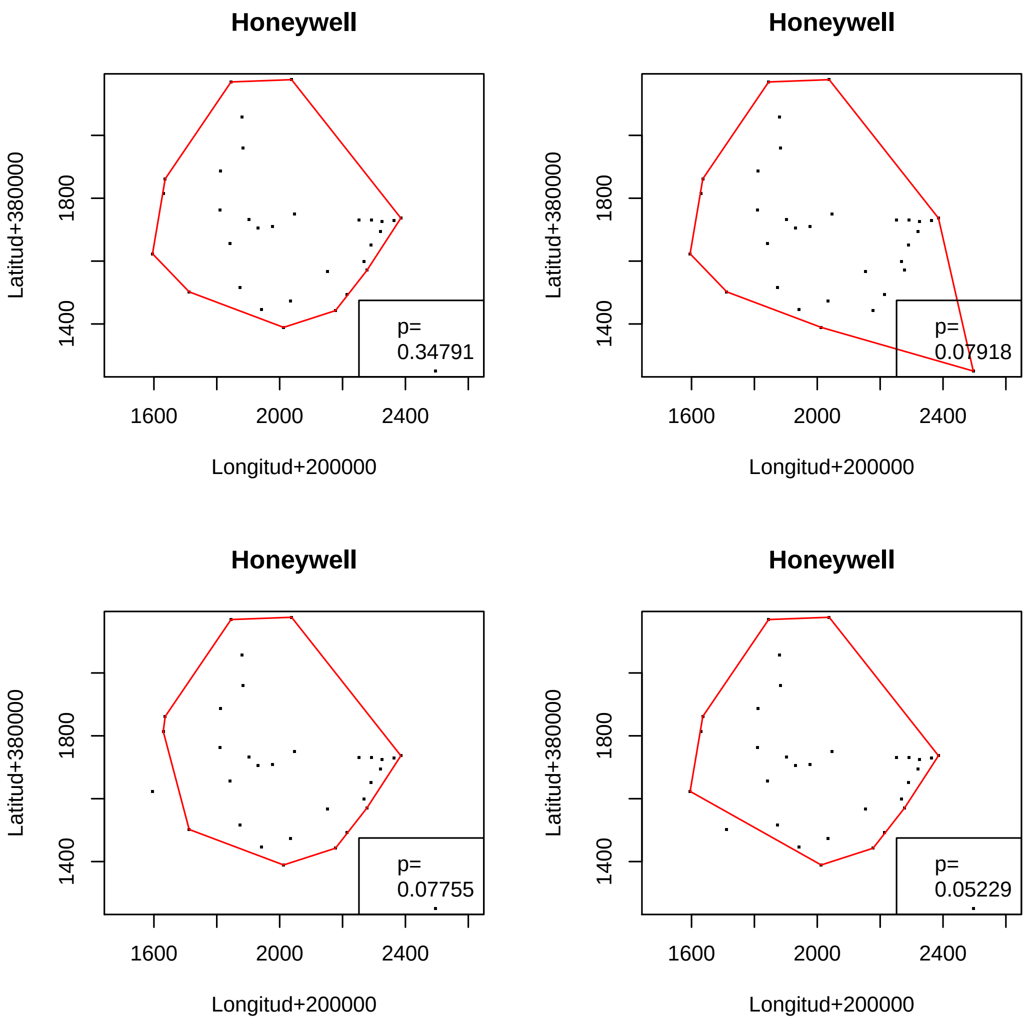


Figure A.40: Most frequent reconstructions of the Honeywell territory using the sampling method of the Unconditional Outlier Prediction Model

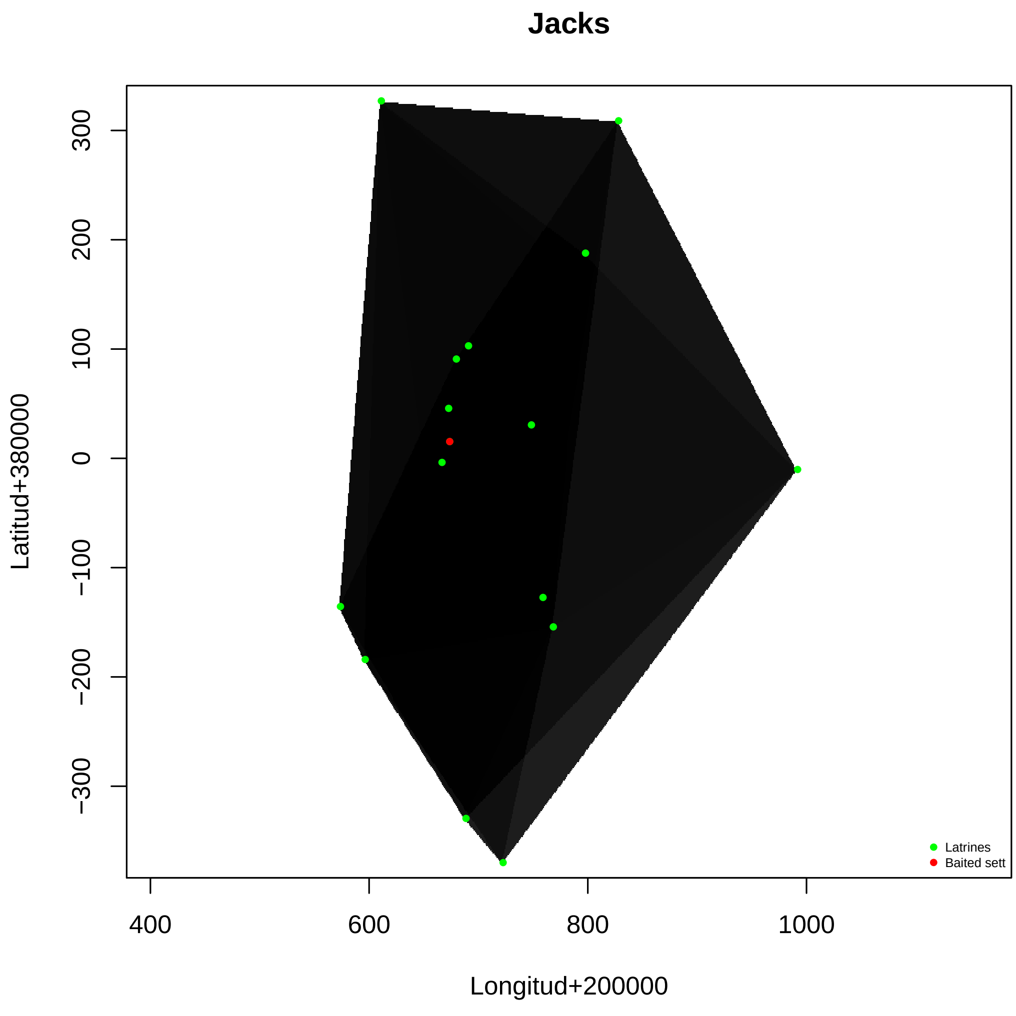


Figure A.41: Probability of inclusion, Jacks territory using the Unconditional Outlier Prediction Model sampling method.

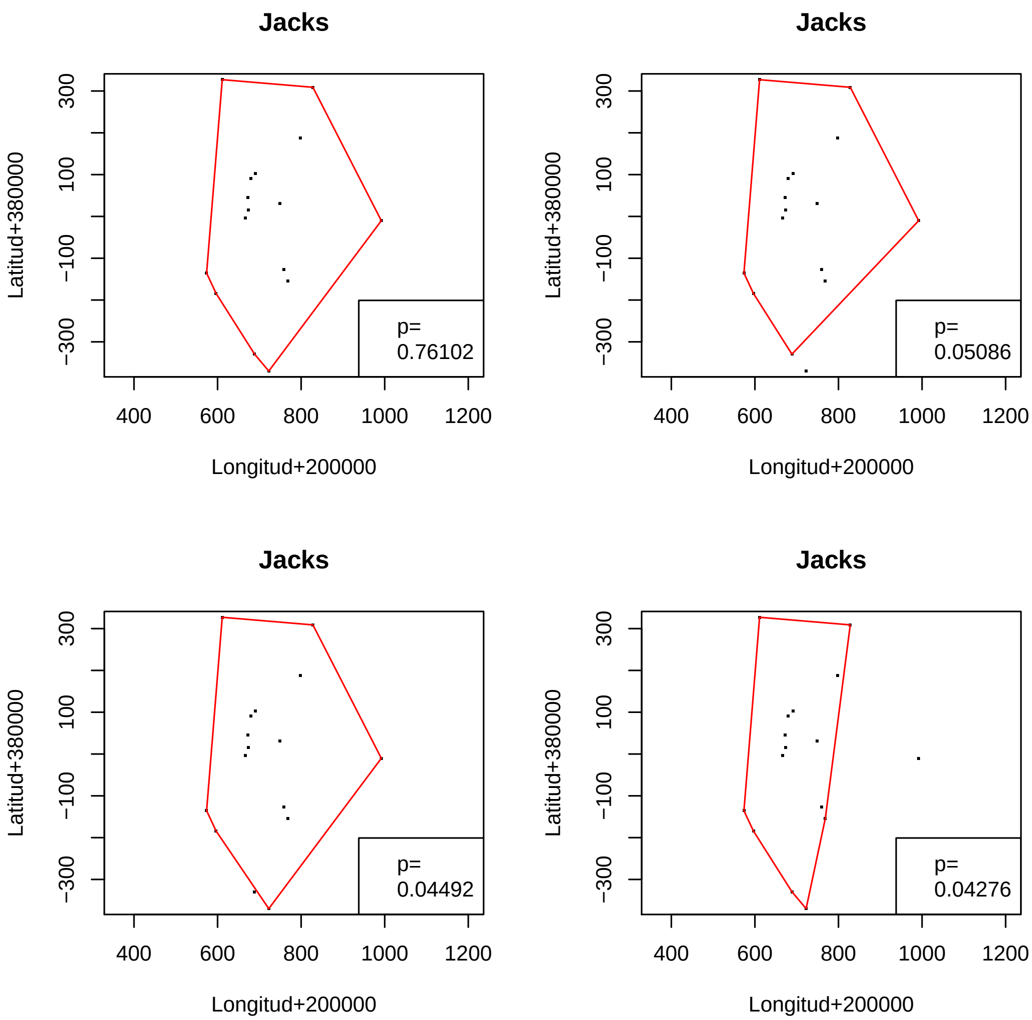


Figure A.42: Most frequent reconstructions of the Jacks territory using the sampling method of the Unconditional Outlier Prediction Model

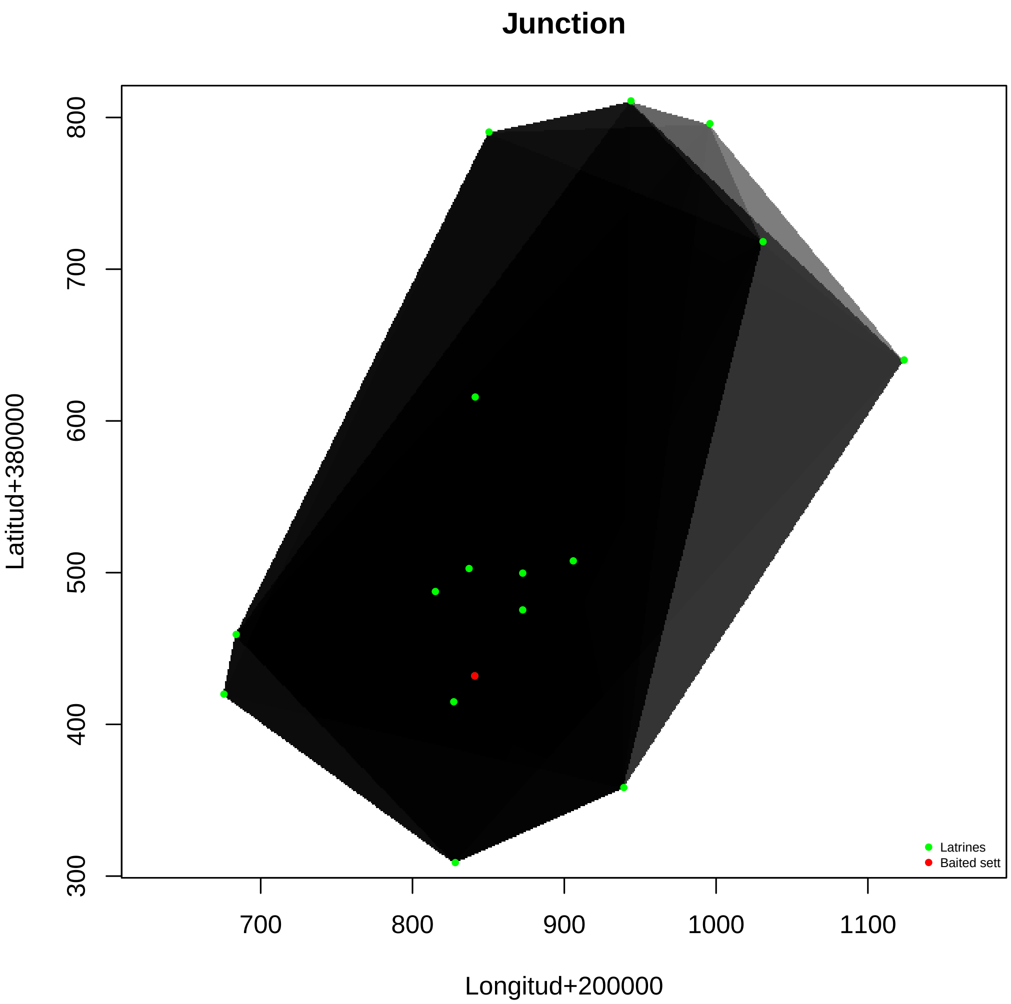


Figure A.43: Probability of inclusion, Junction territory using the Unconditional Outlier Prediction Model sampling method.

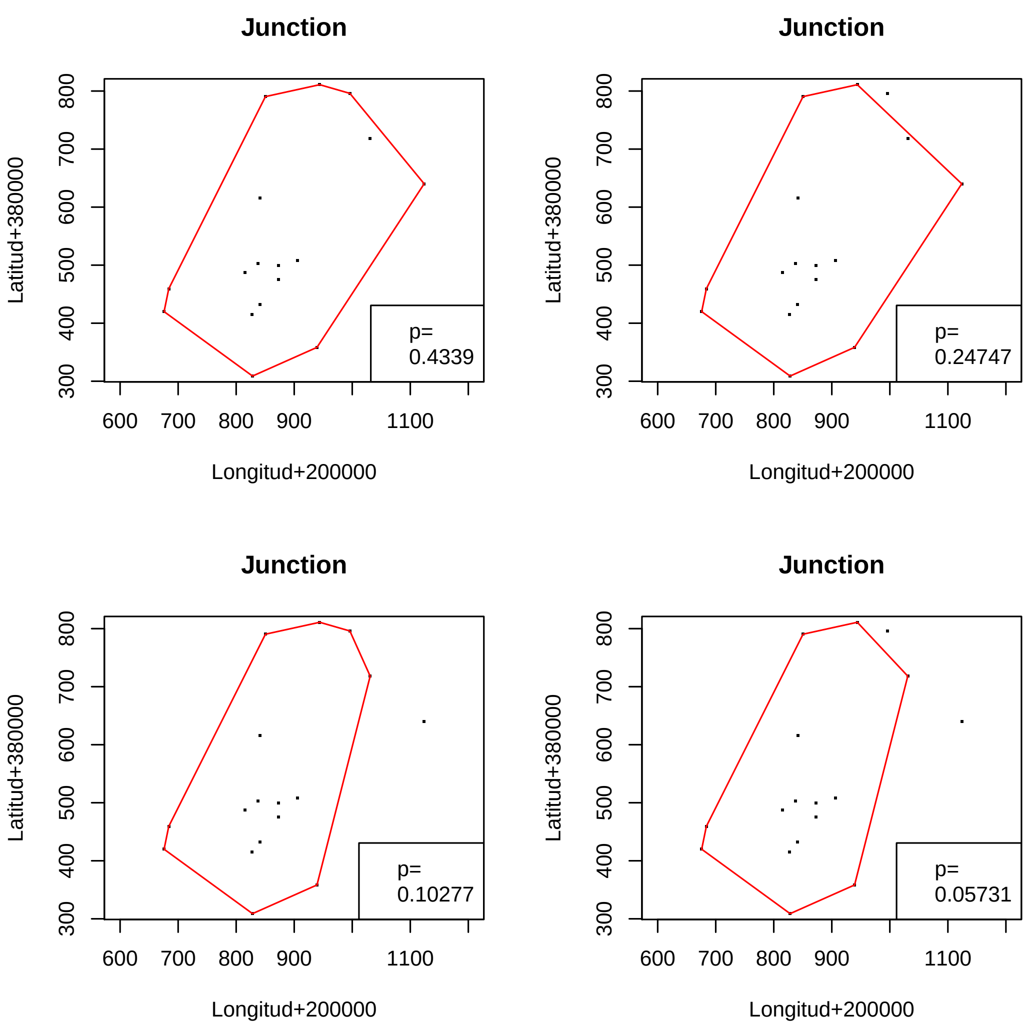


Figure A.44: Most frequent reconstructions of the Junction territory using the sampling method of the Unconditional Outlier Prediction Model

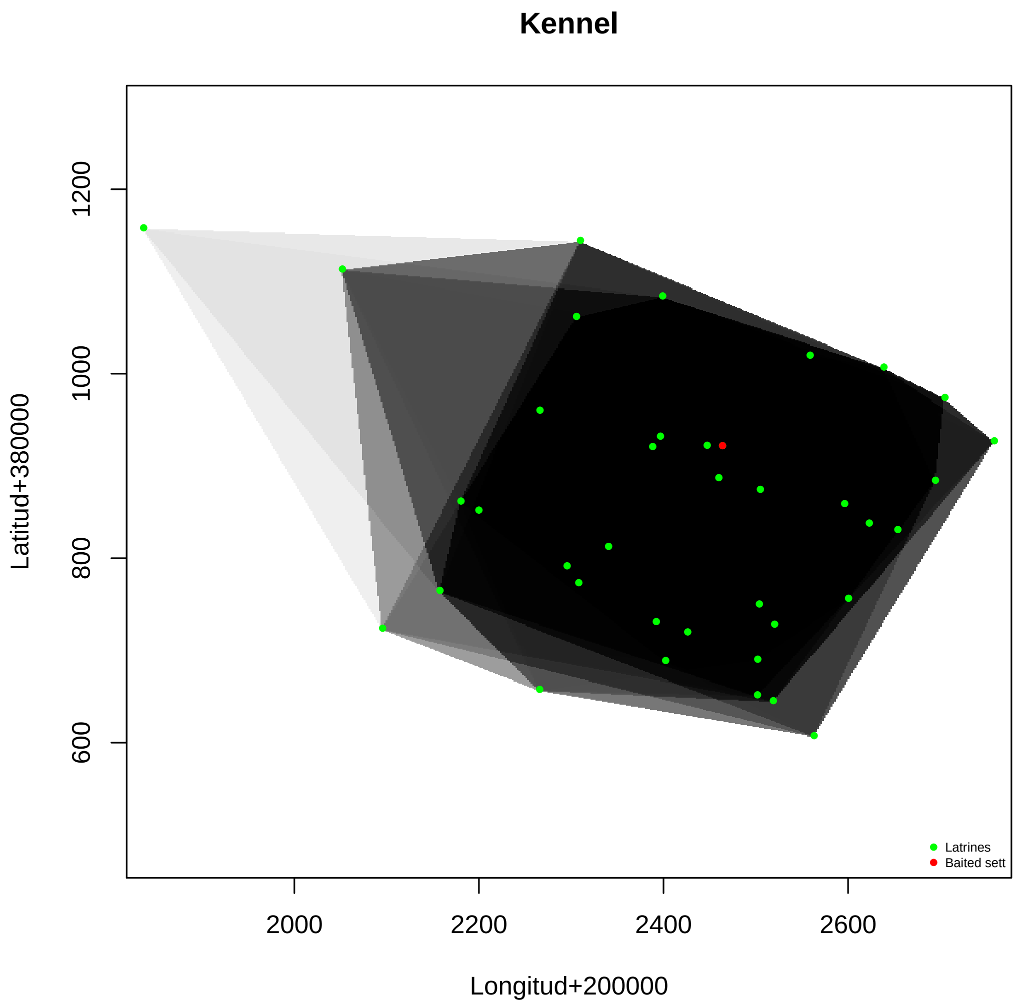


Figure A.45: Probability of inclusion, Kennel territory using the Unconditional Outlier Prediction Model sampling method.

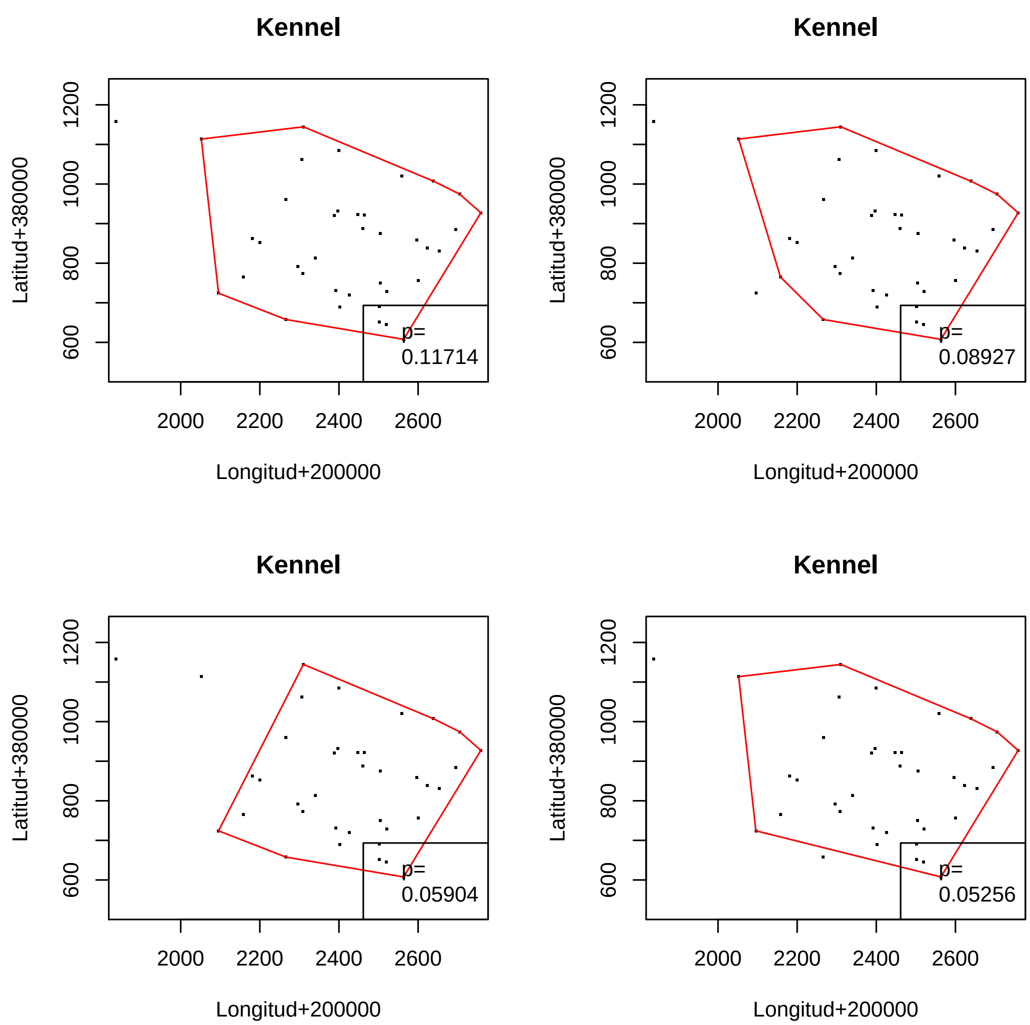


Figure A.46: Most frequent reconstructions of the Kennel territory using the sampling method of the Unconditional Outlier Prediction Model

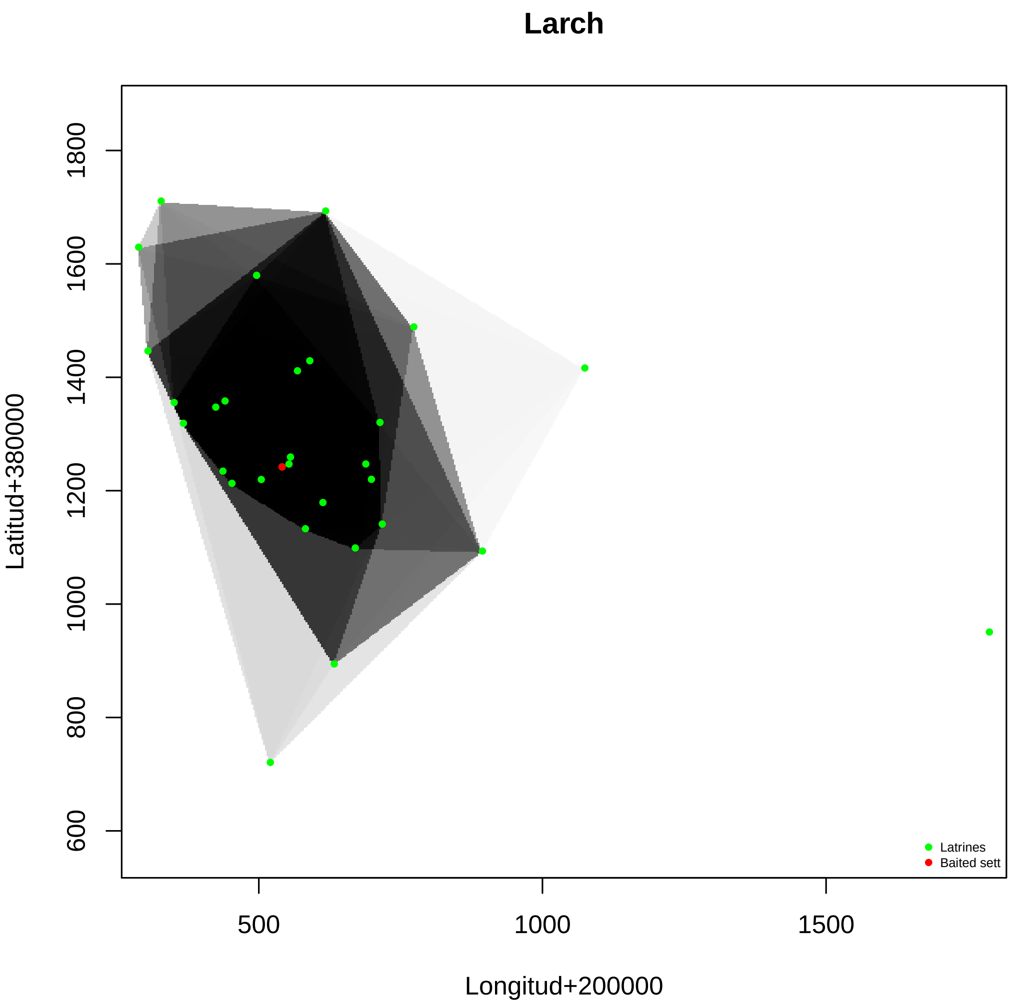


Figure A.47: Probability of inclusion, Larch territory using the Unconditional Outlier Prediction Model sampling method.

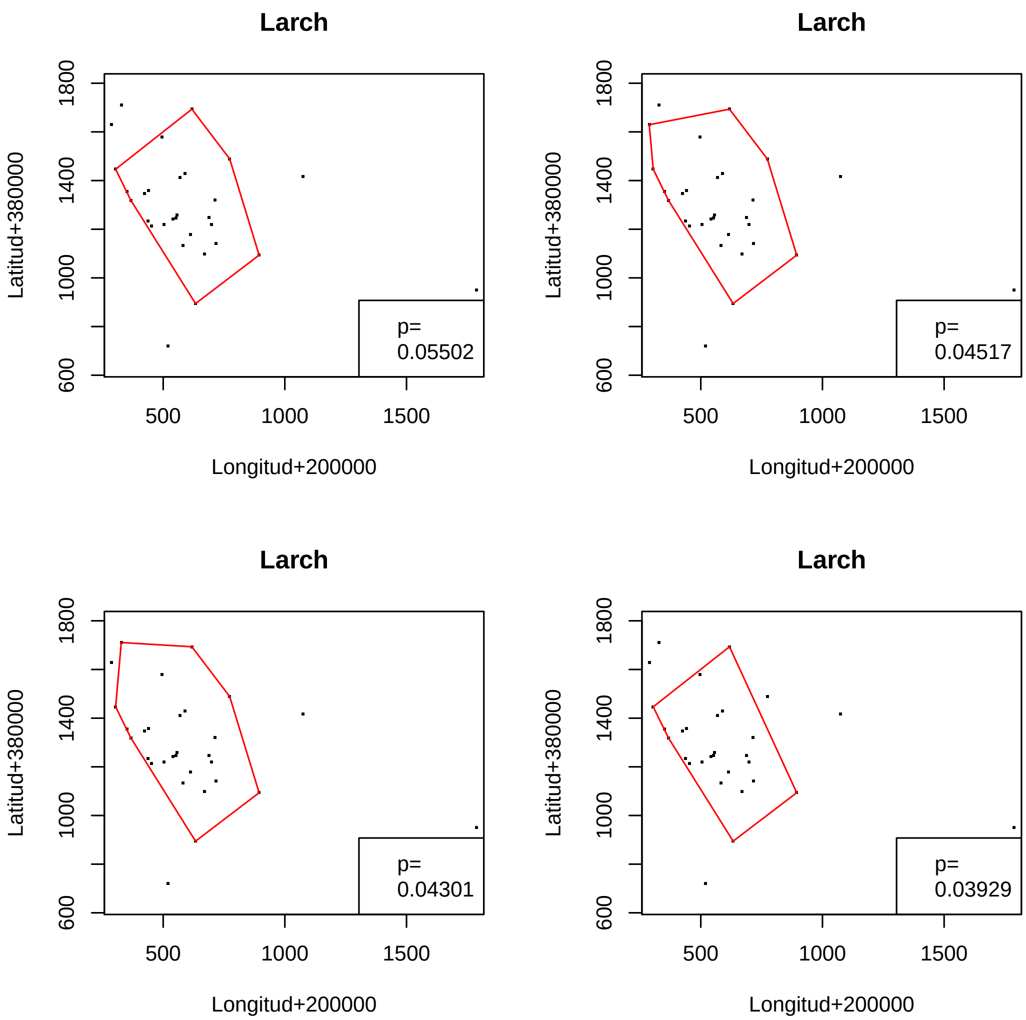


Figure A.48: Most frequent reconstructions of the Larch territory using the sampling method of the Unconditional Outlier Prediction Model

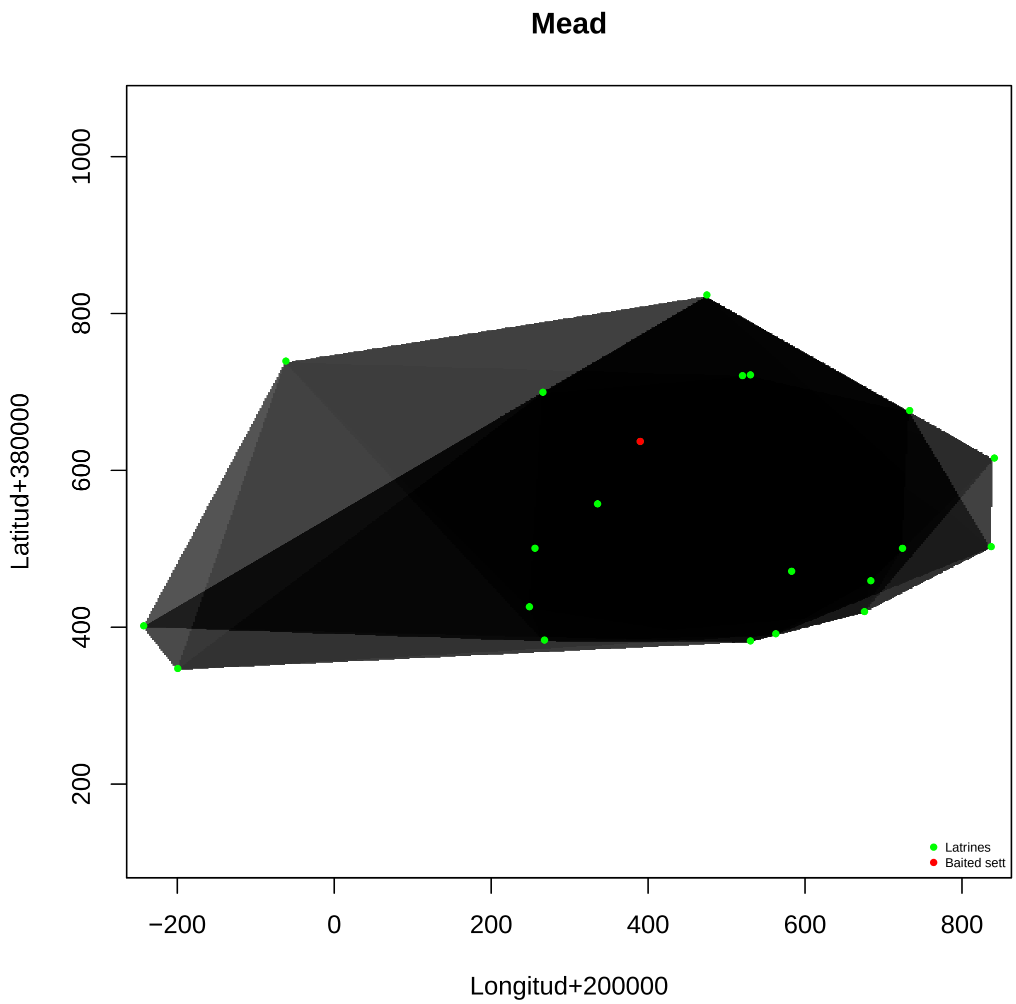


Figure A.49: Probability of inclusion, Mead territory using the Unconditional Outlier Prediction Model sampling method.

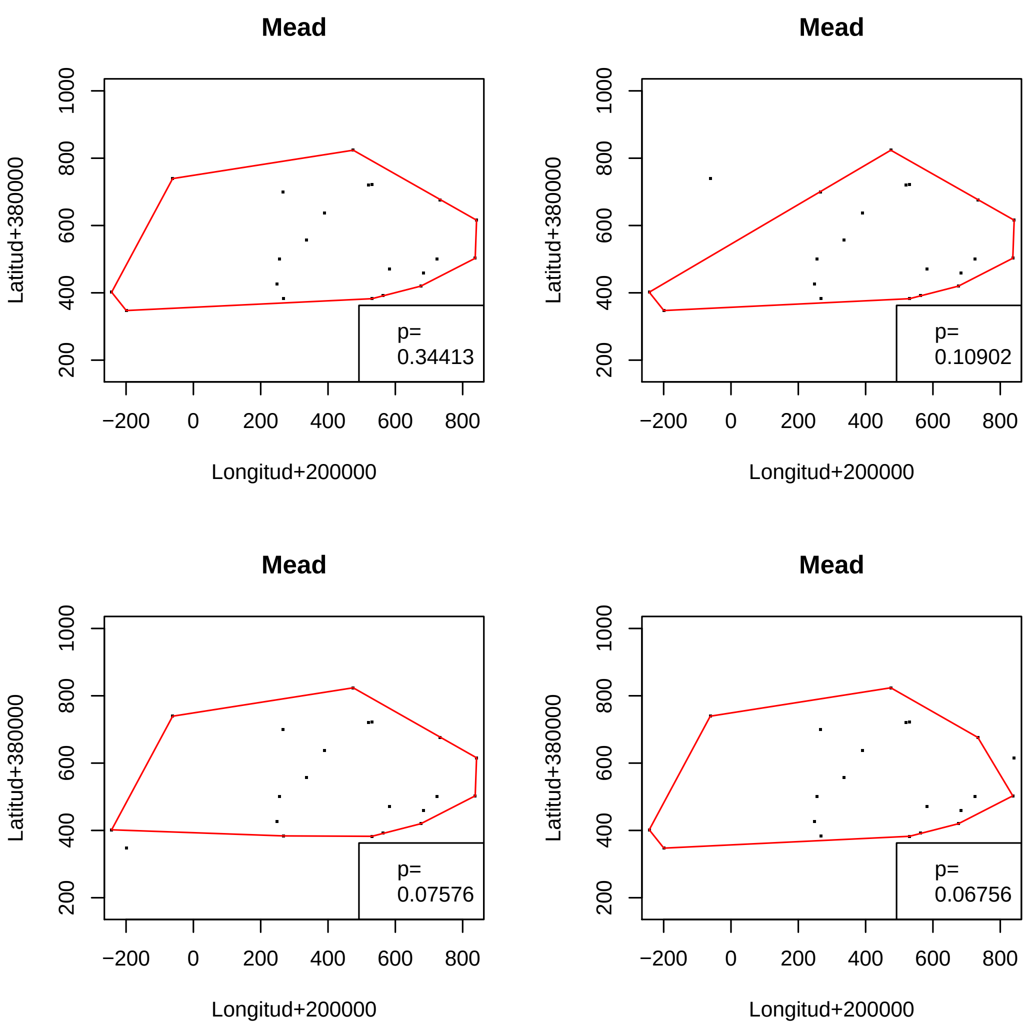


Figure A.50: Most frequent reconstructions of the Mead territory using the sampling method of the Unconditional Outlier Prediction Model

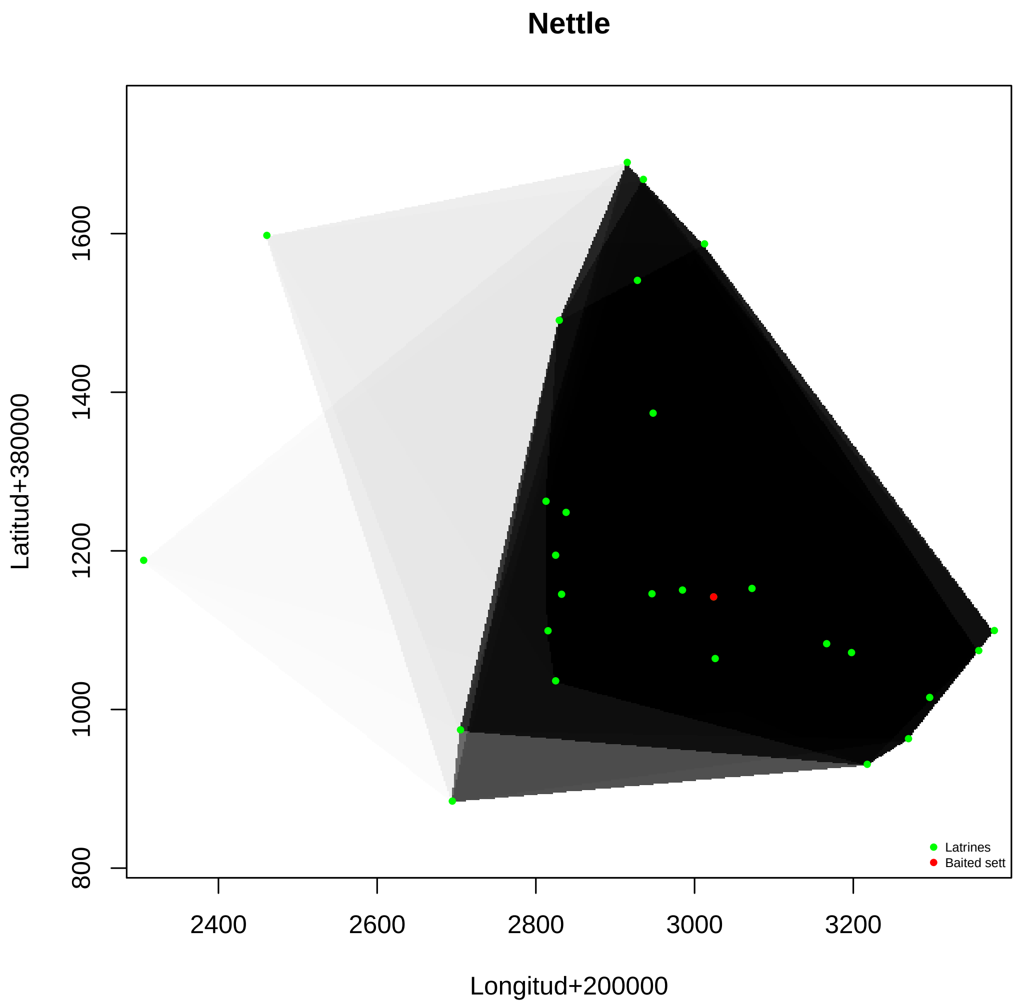


Figure A.51: Probability of inclusion, Nettle territory using the Unconditional Outlier Prediction Model sampling method.

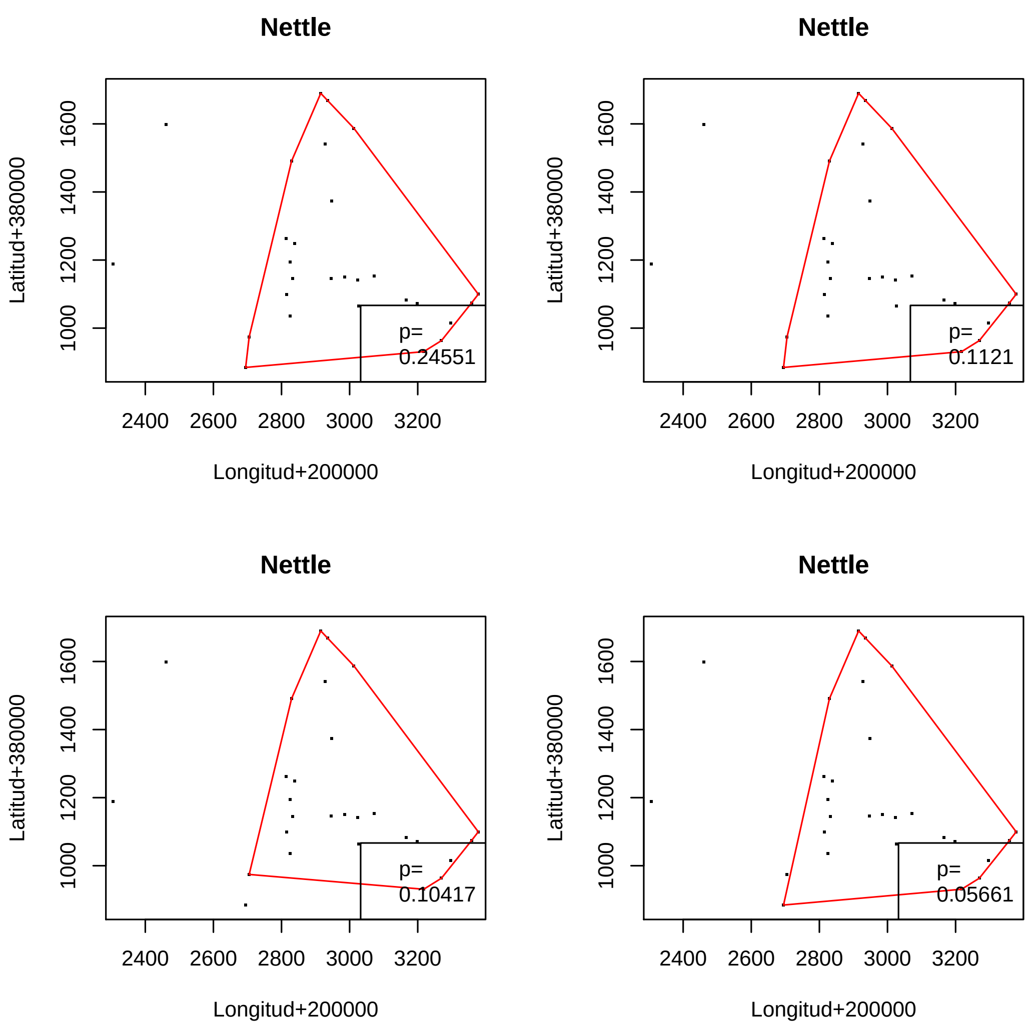


Figure A.52: Most frequent reconstructions of the Nettle territory using the sampling method of the Unconditional Outlier Prediction Model

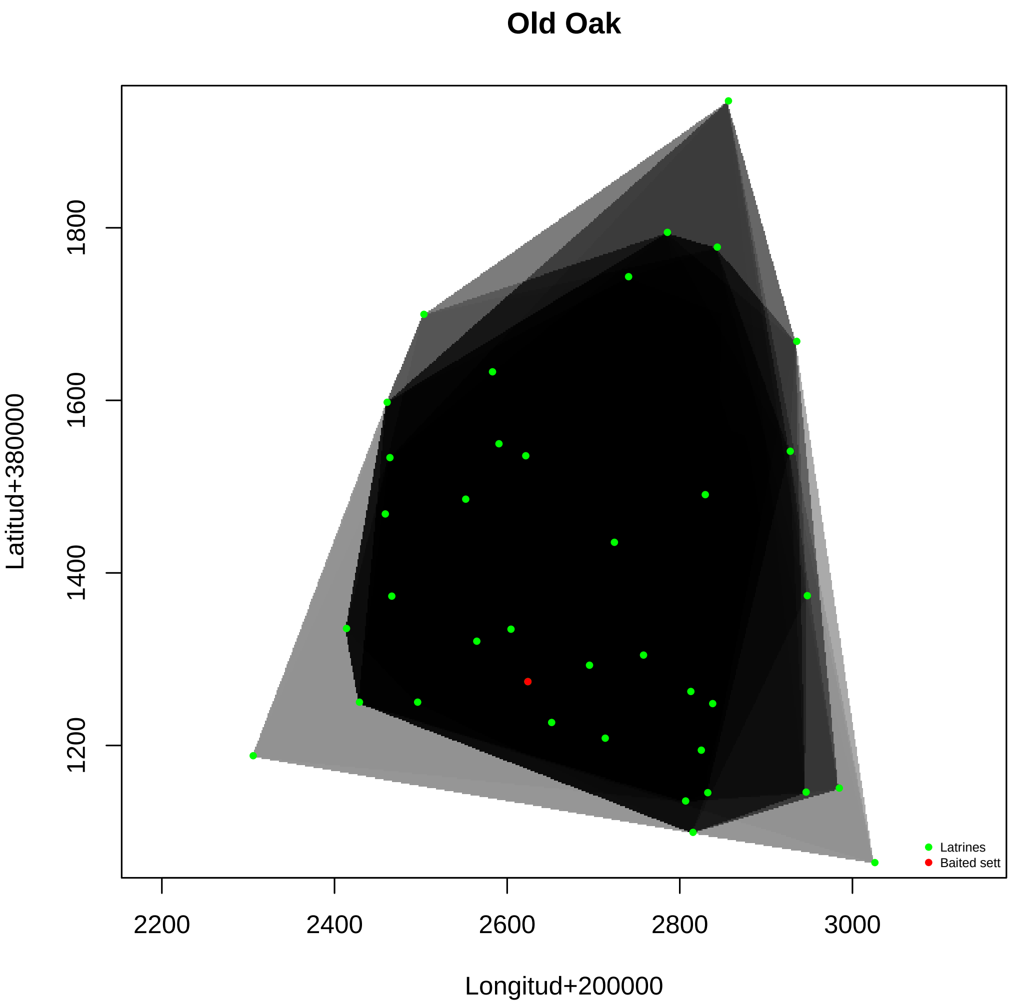


Figure A.53: Probability of inclusion, Old Oak territory using the Unconditional Outlier Prediction Model sampling method.

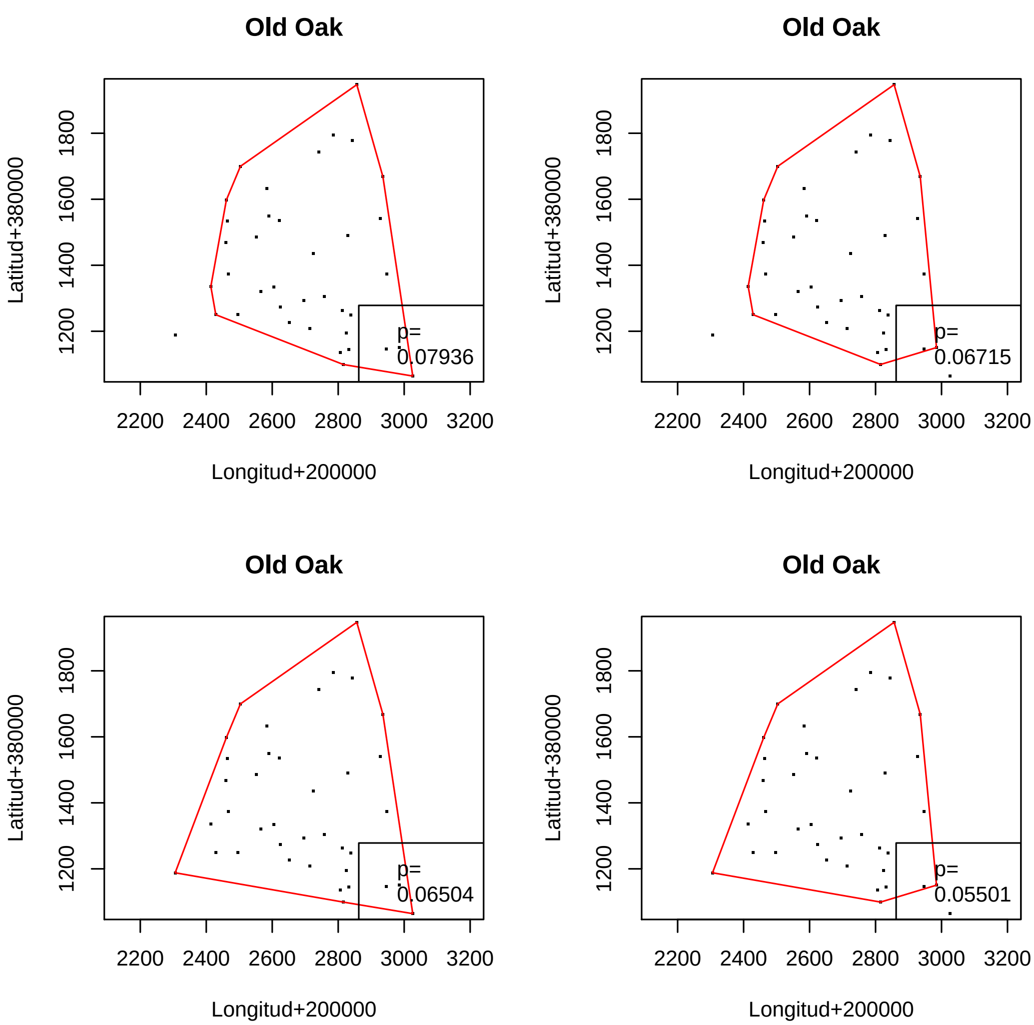


Figure A.54: Most frequent reconstructions of the Old Oak territory using the sampling method of the Unconditional Outlier Prediction Model

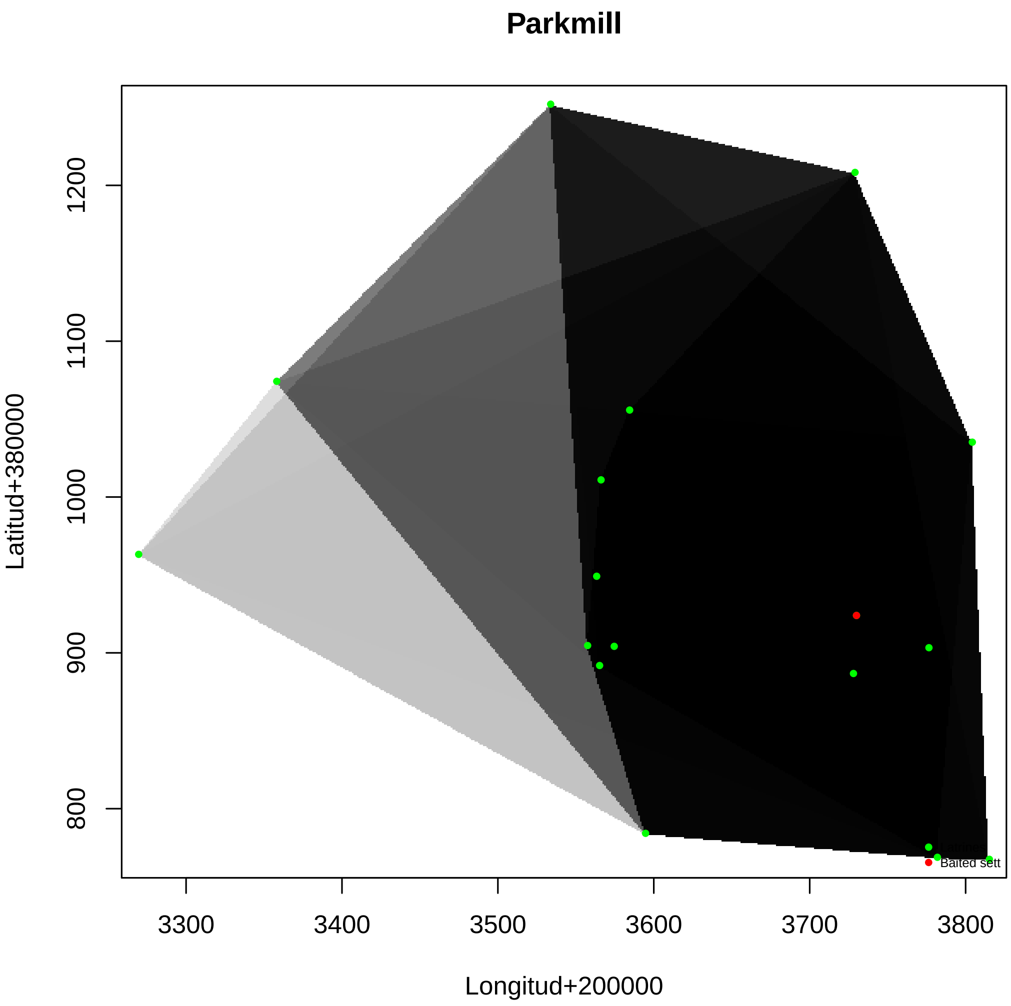


Figure A.55: Probability of inclusion, Parkmill territory using the Unconditional Outlier Prediction Model sampling method.

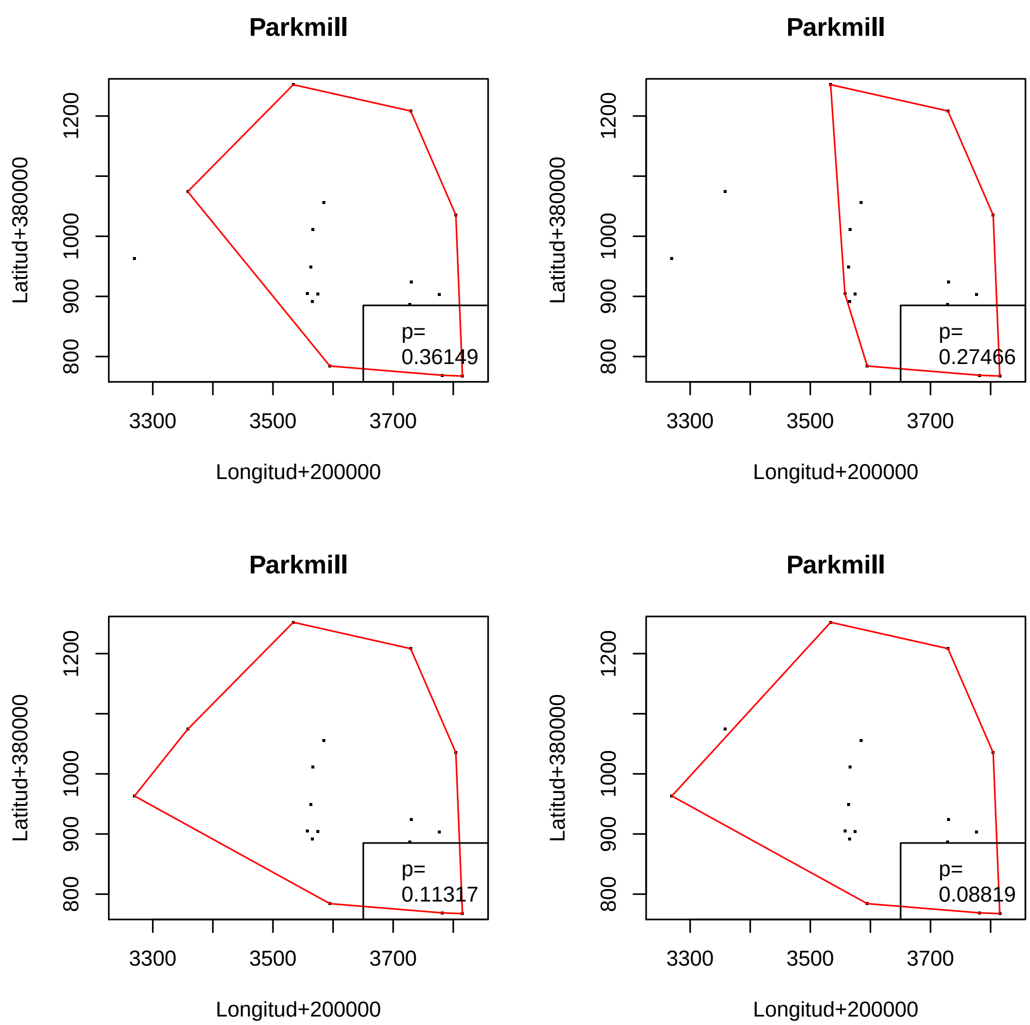


Figure A.56: Most frequent reconstructions of the Parkmill territory using the sampling method of the Unconditional Outlier Prediction Model

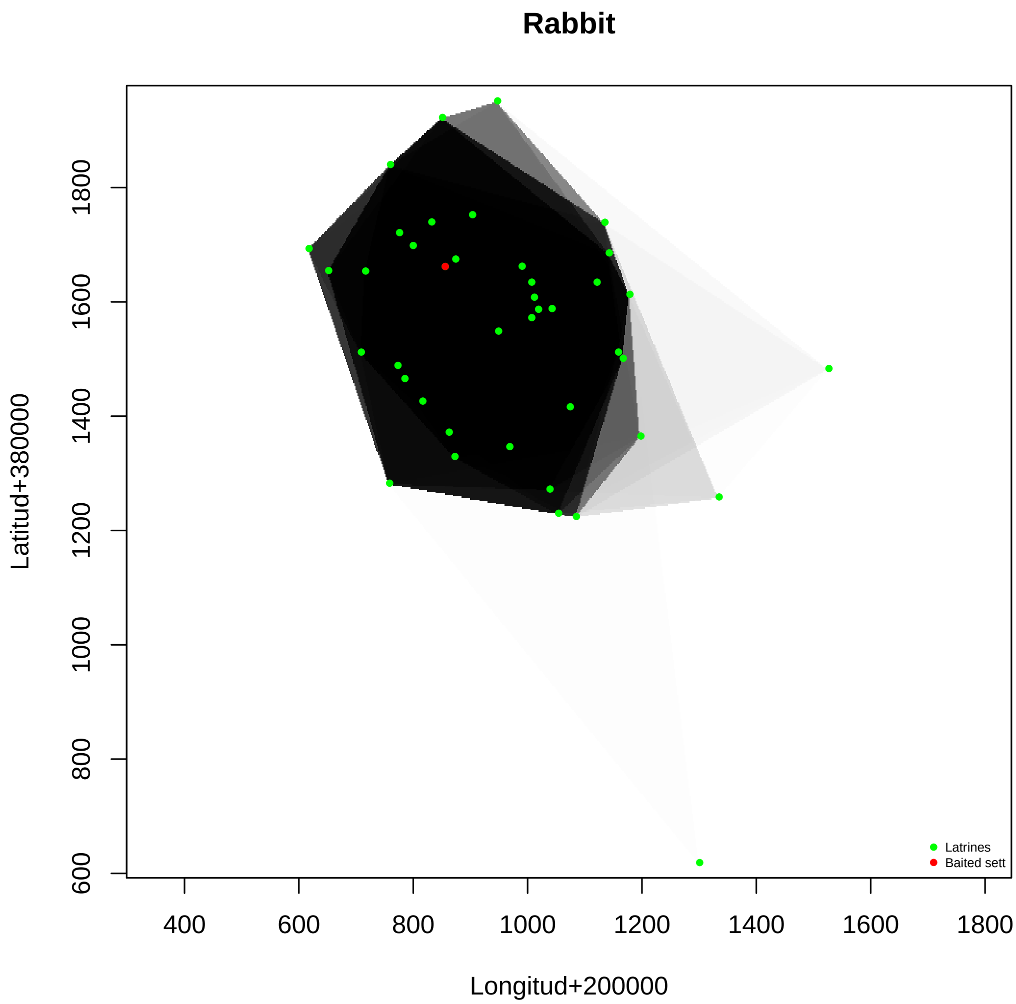


Figure A.57: Probability of inclusion, Rabbit territory using the Unconditional Outlier Prediction Model sampling method.

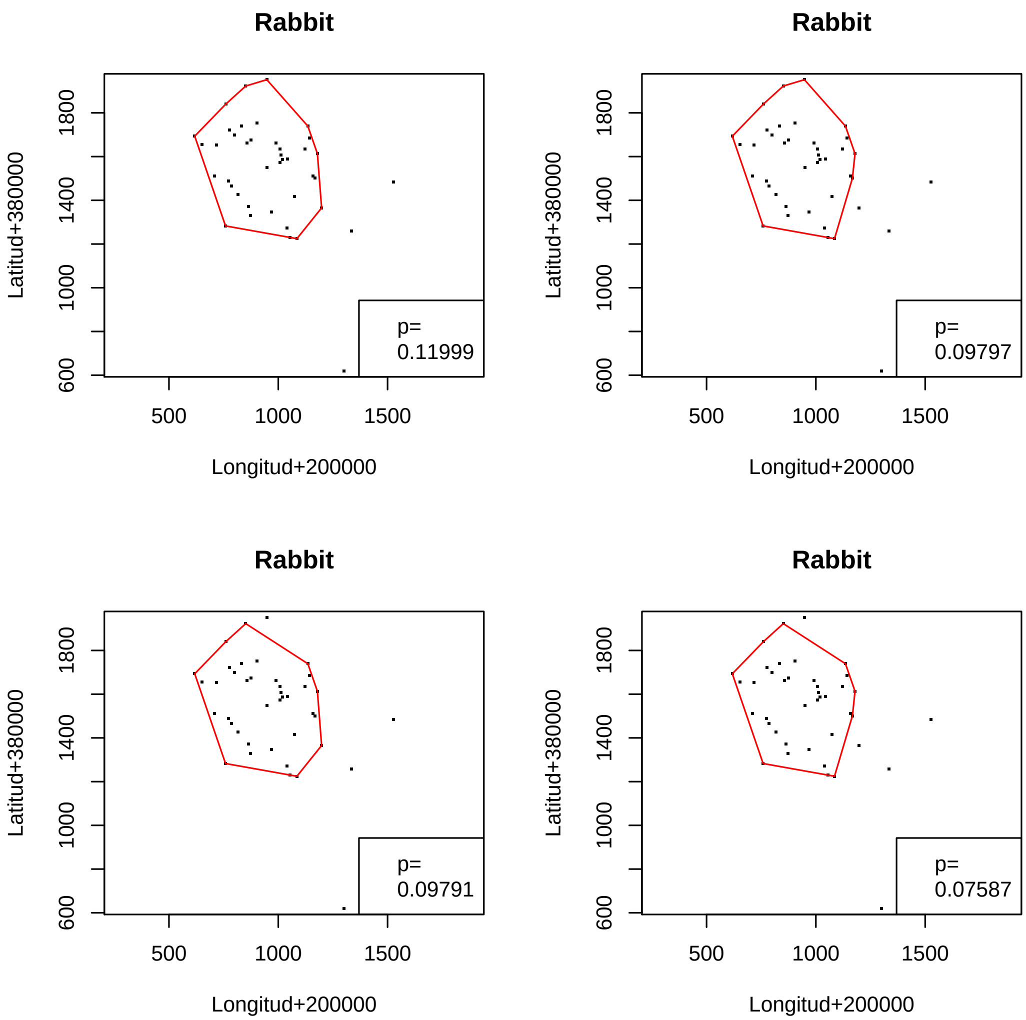


Figure A.58: Most frequent reconstructions of the Rabbit territory using the sampling method of the Unconditional Outlier Prediction Model

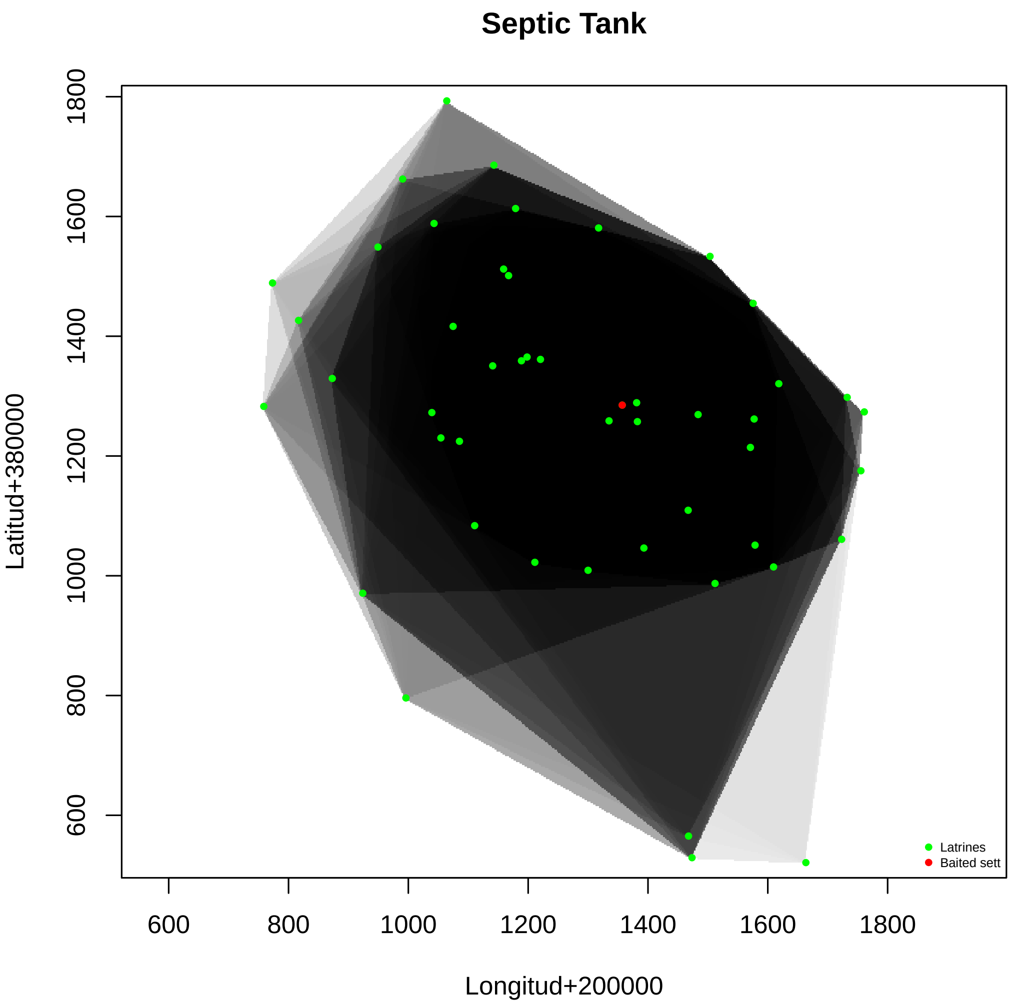


Figure A.59: Probability of inclusion, Septic Tank territory using the Unconditional Outlier Prediction Model sampling method.

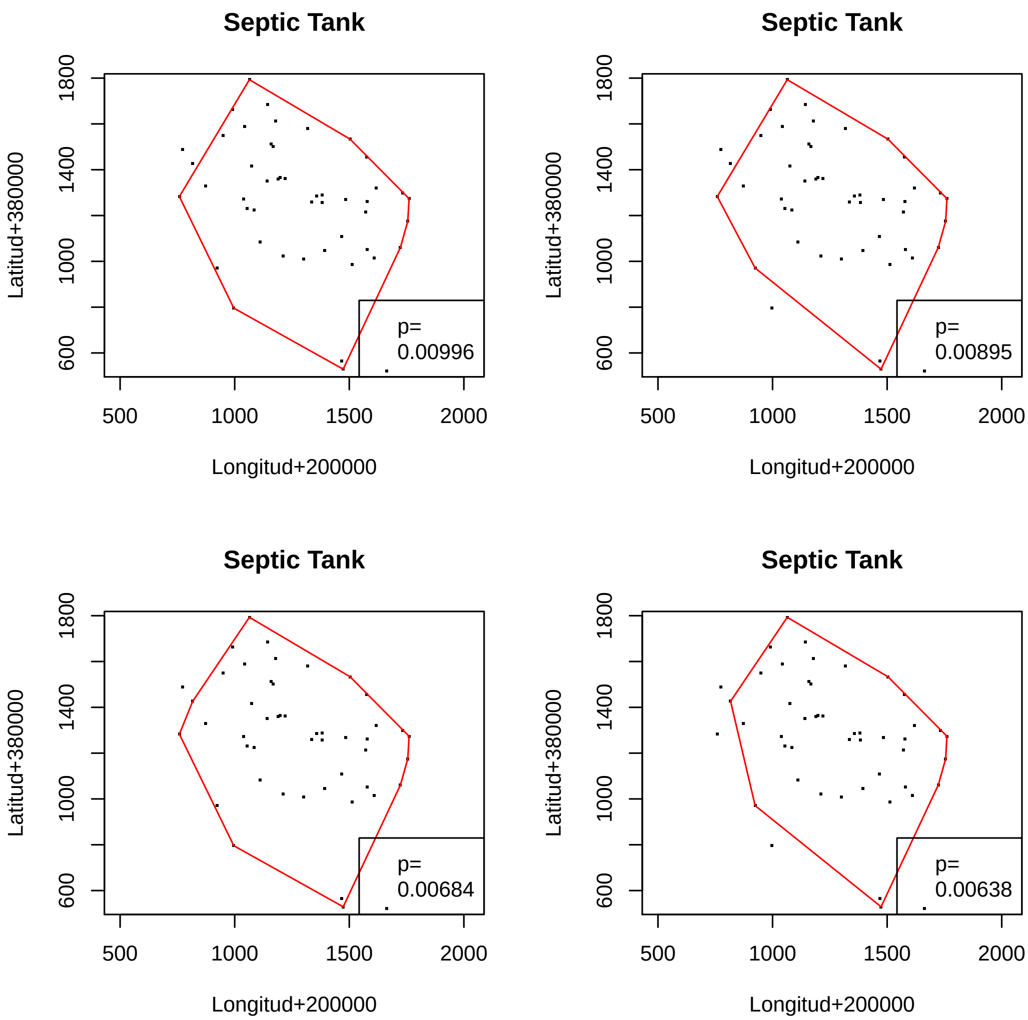


Figure A.60: Most frequent reconstructions of the Septic Tank territory using the sampling method of the Unconditional Outlier Prediction Model

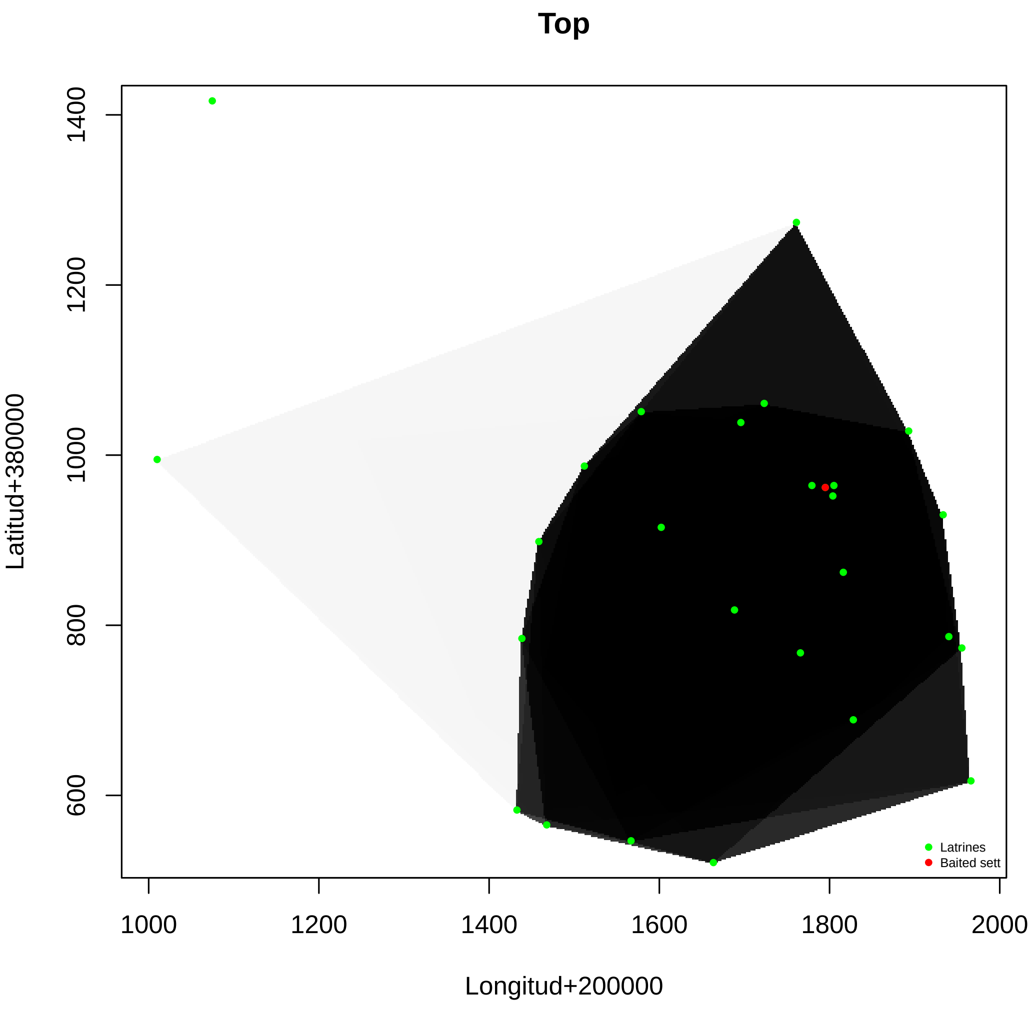


Figure A.61: Probability of inclusion, Top territory using the Unconditional Outlier Prediction Model sampling method.

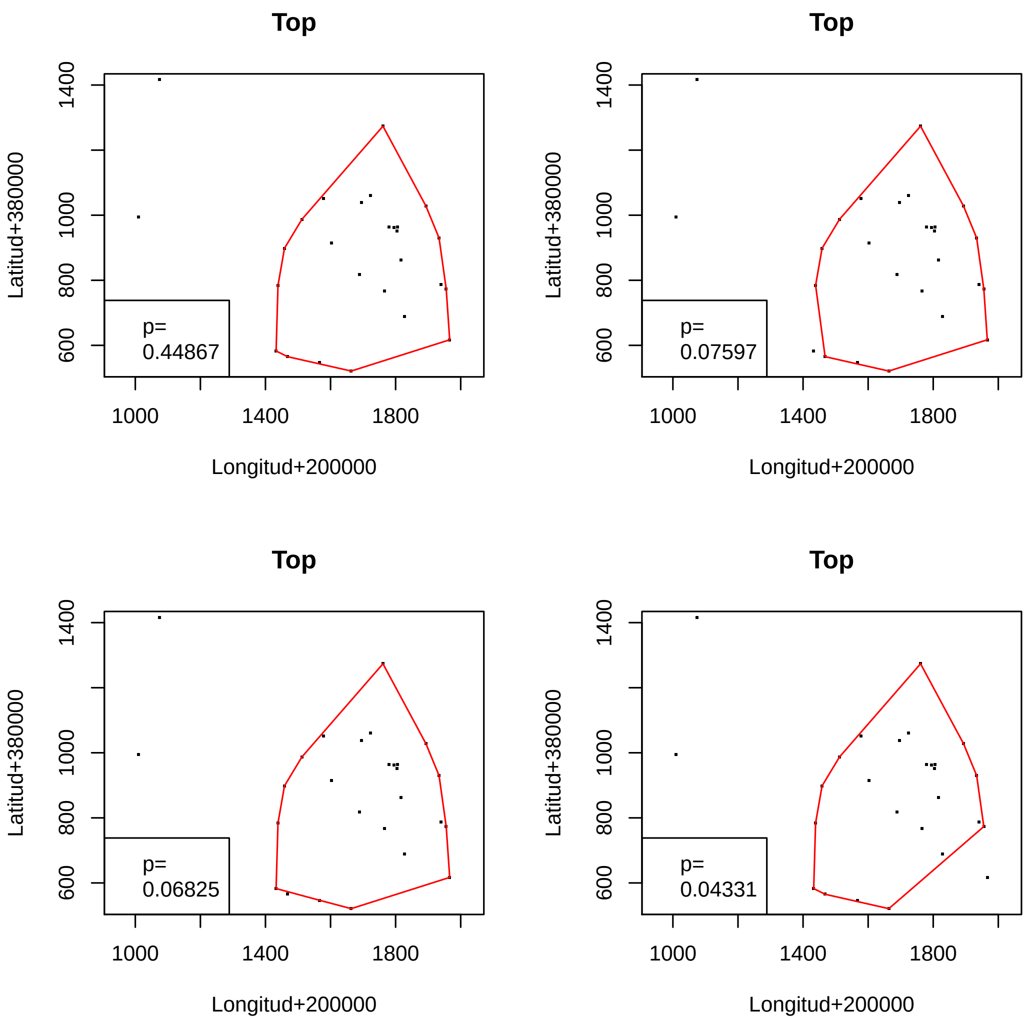


Figure A.62: Most frequent reconstructions of the Top territory using the sampling method of the Unconditional Outlier Prediction Model

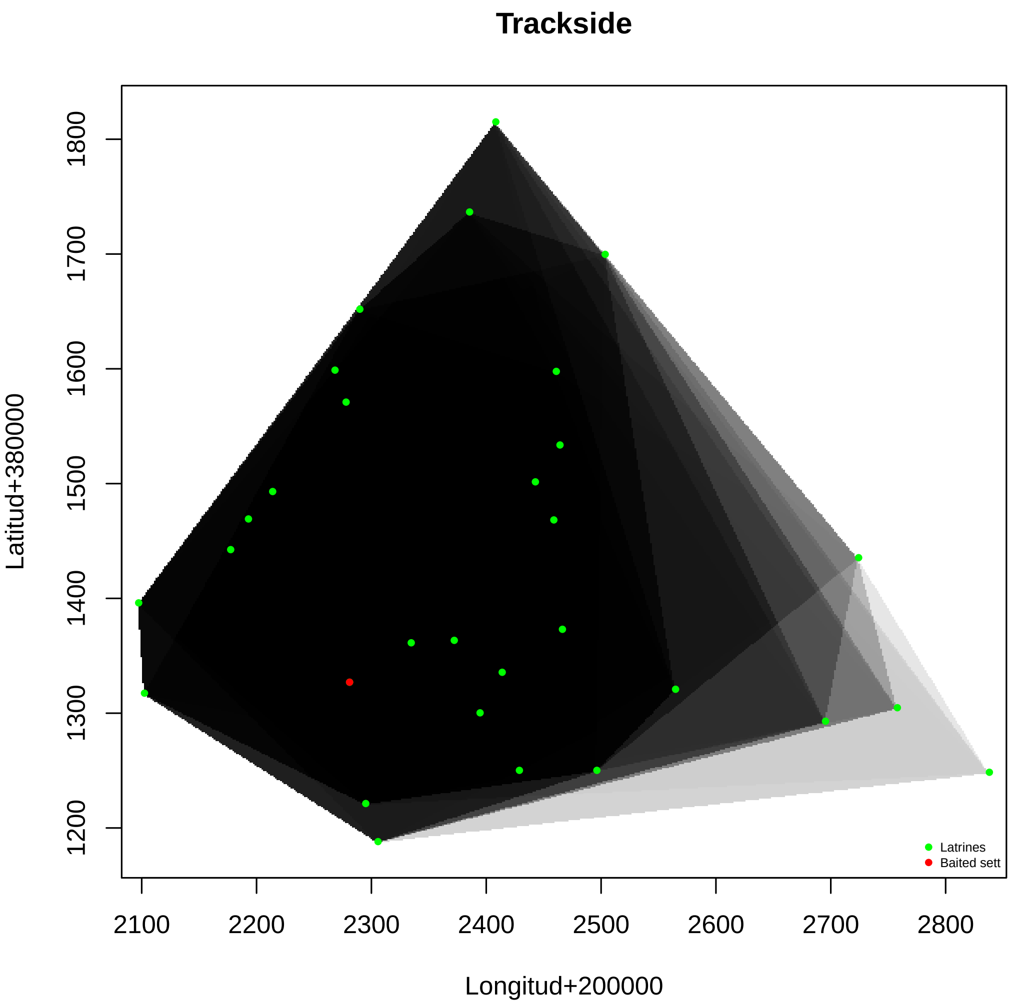


Figure A.63: Probability of inclusion, Trackside territory using the Unconditional Outlier Prediction Model sampling method.

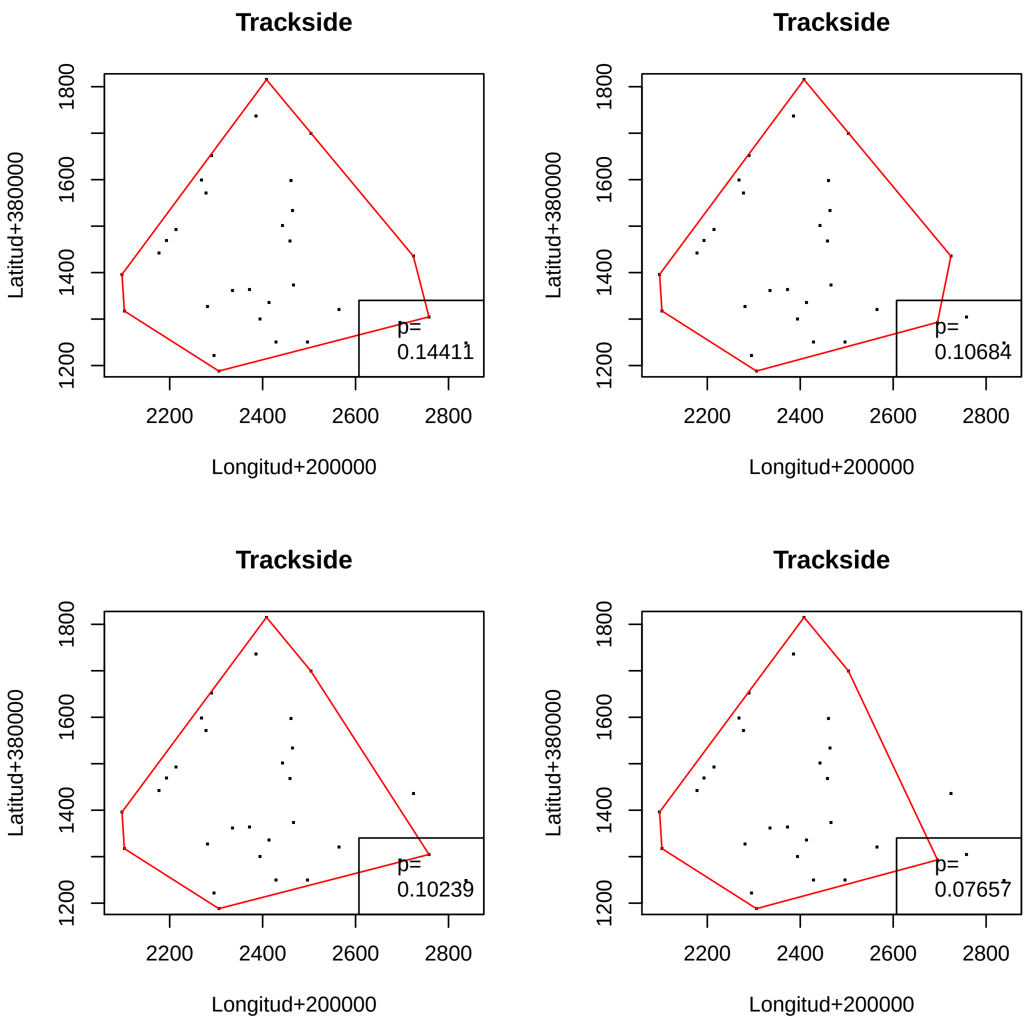


Figure A.64: Most frequent reconstructions of the Trackside territory using the sampling method of the Unconditional Outlier Prediction Model

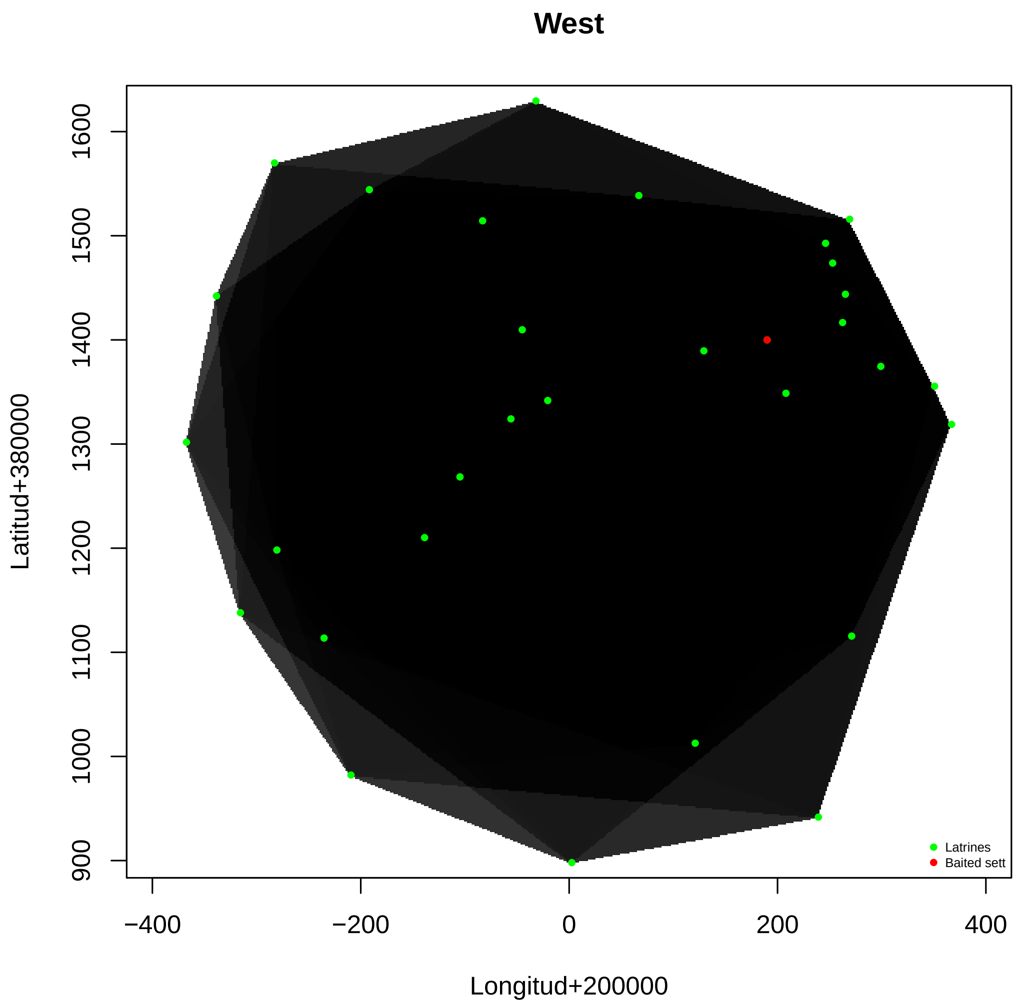


Figure A.65: Probability of inclusion, West territory using the Unconditional Outlier Prediction Model sampling method.

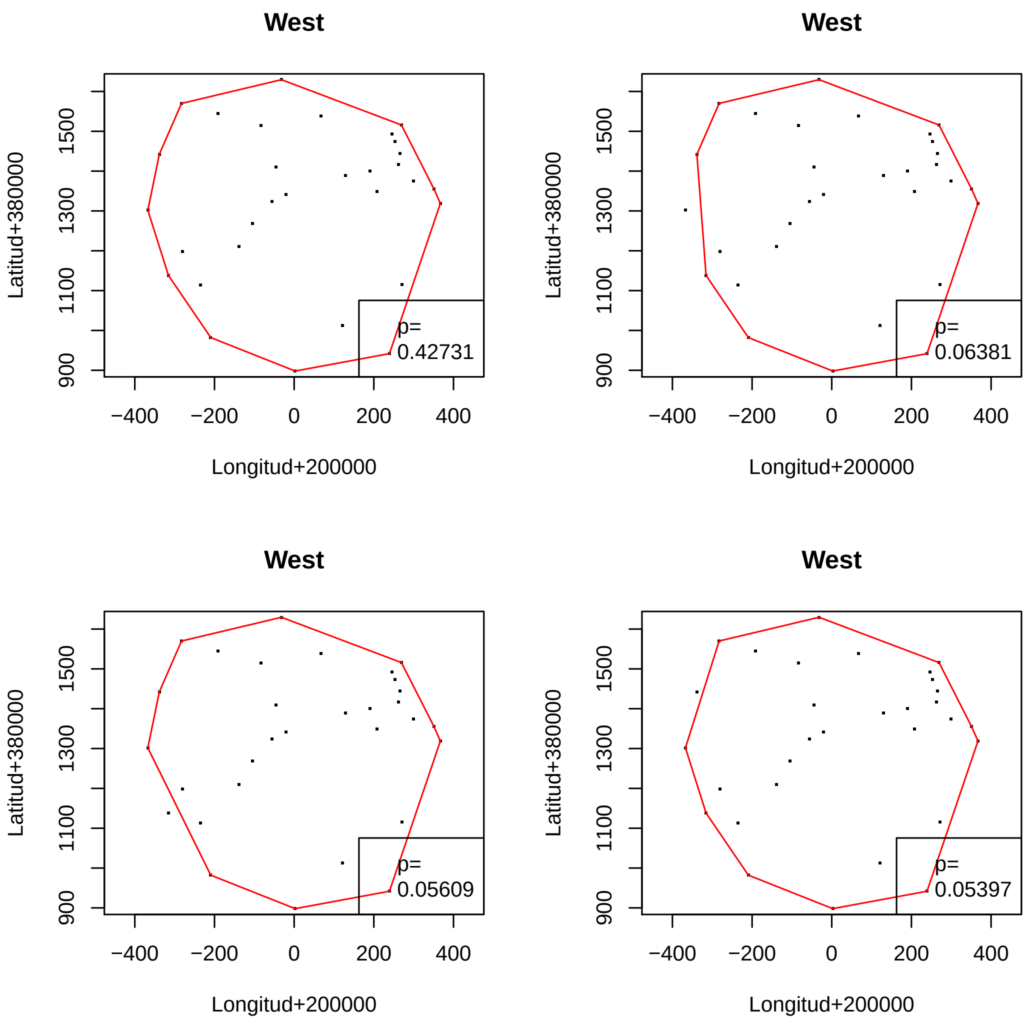


Figure A.66: Most frequent reconstructions of the West territory using the sampling method of the Unconditional Outlier Prediction Model

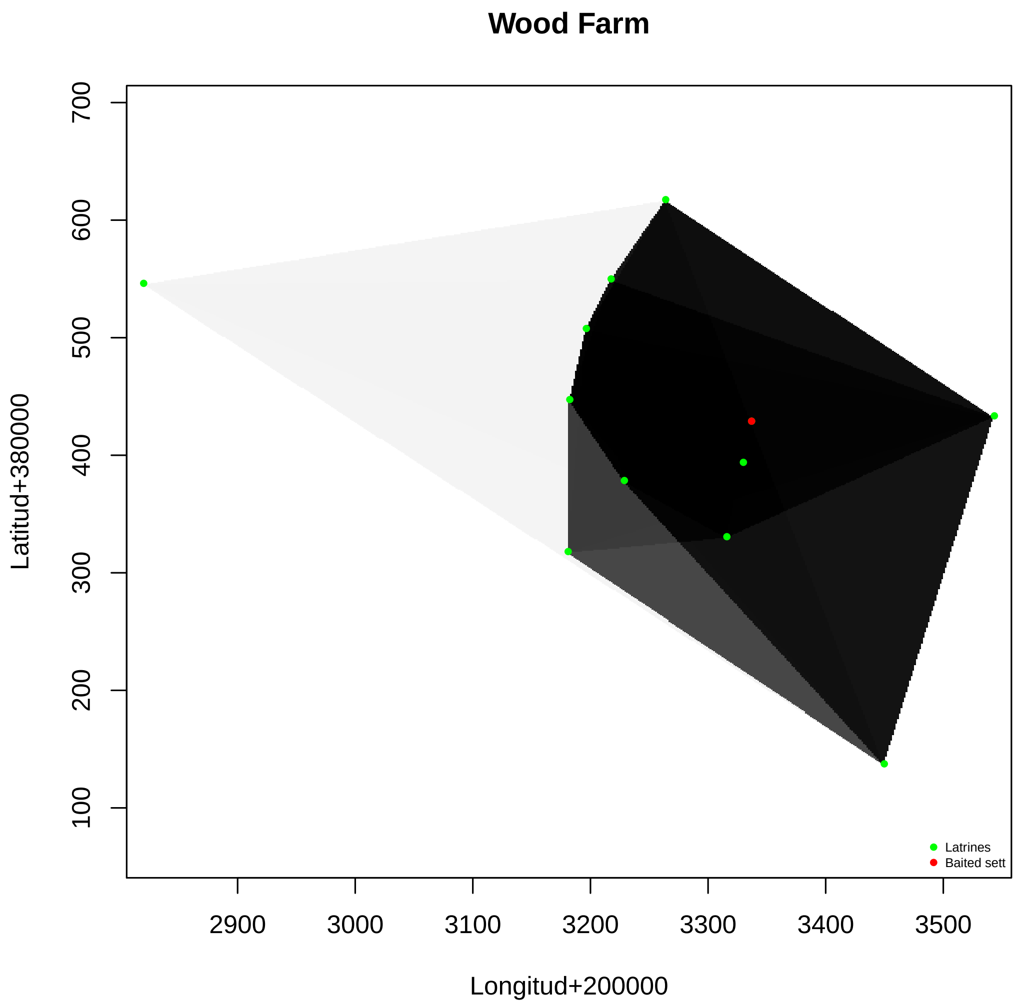


Figure A.67: Probability of inclusion, Wood Farm territory using the Unconditional Outlier Prediction Model sampling method.

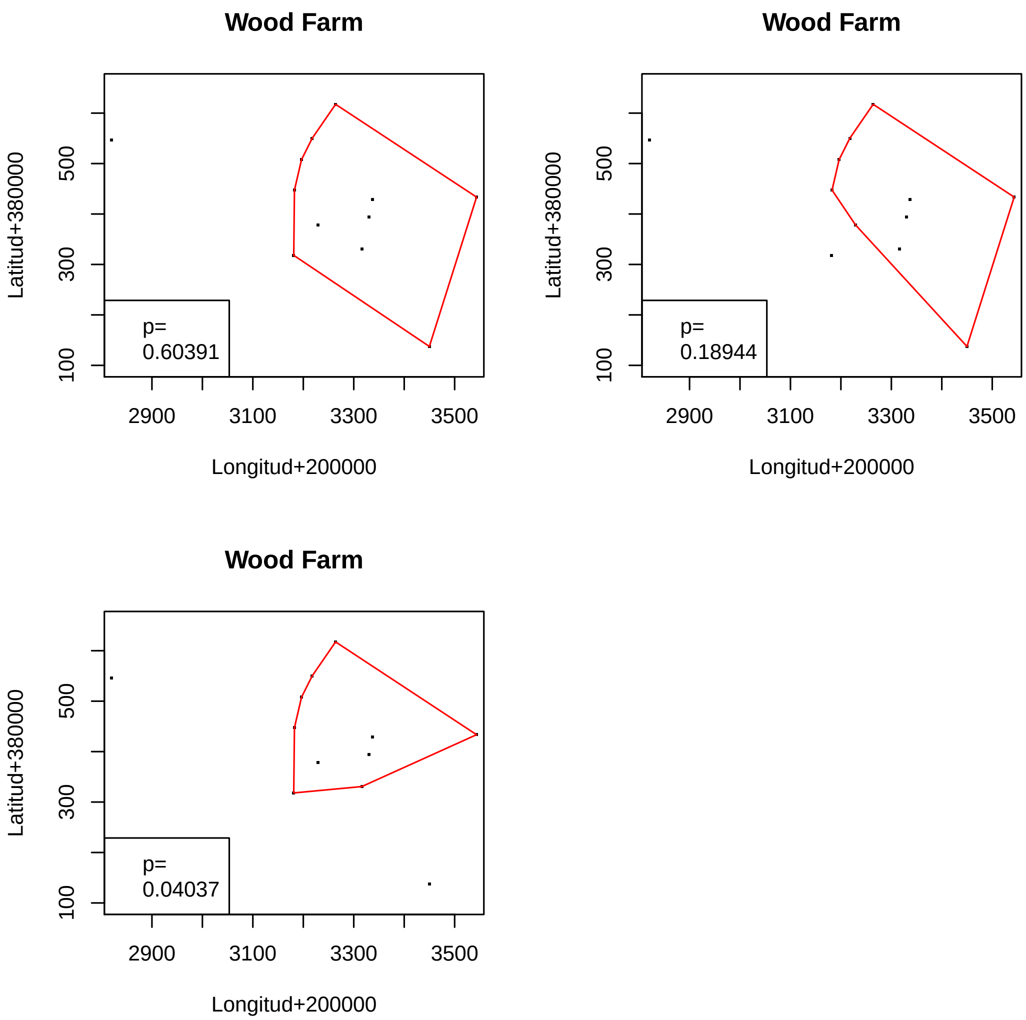


Figure A.68: Most frequent reconstructions of the Wood Farm territory using the sampling method of the Unconditional Outlier Prediction Model

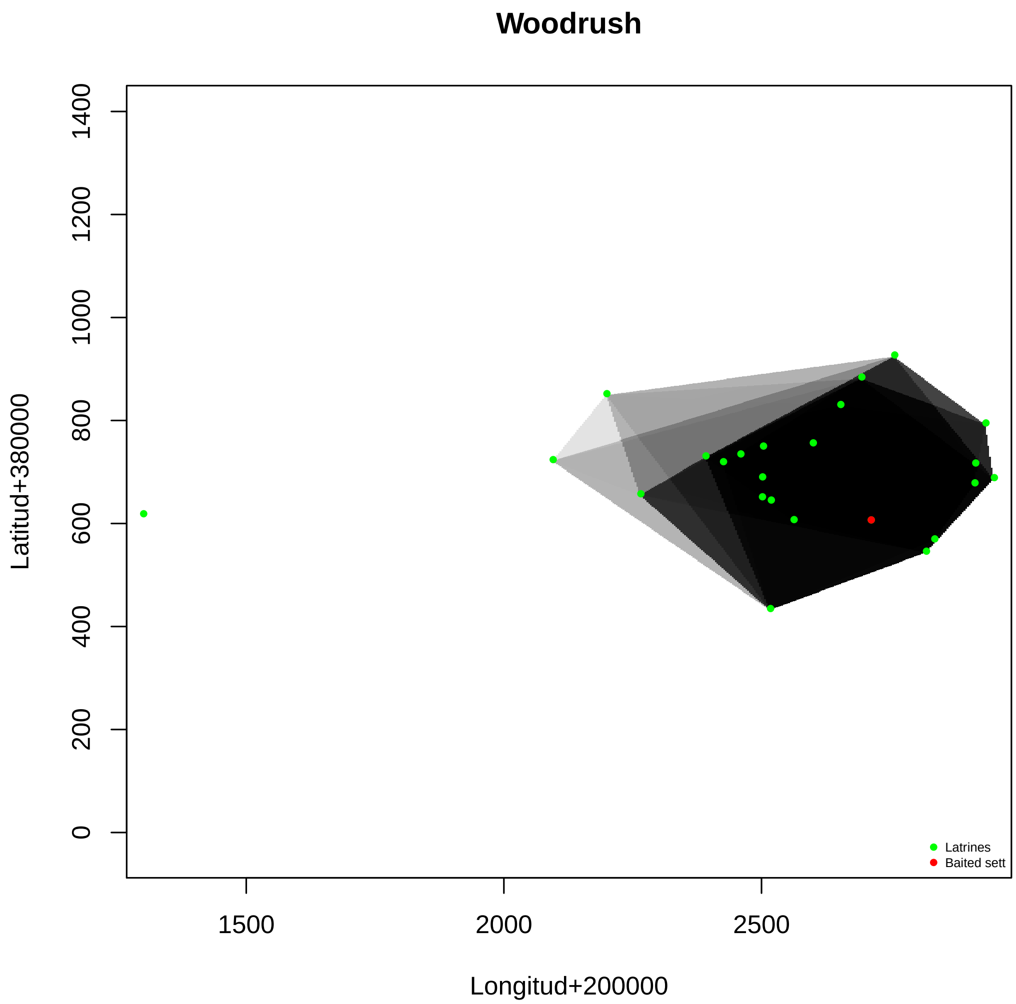


Figure A.69: Probability of inclusion, Woodrush territory using the Unconditional Outlier Prediction Model sampling method.

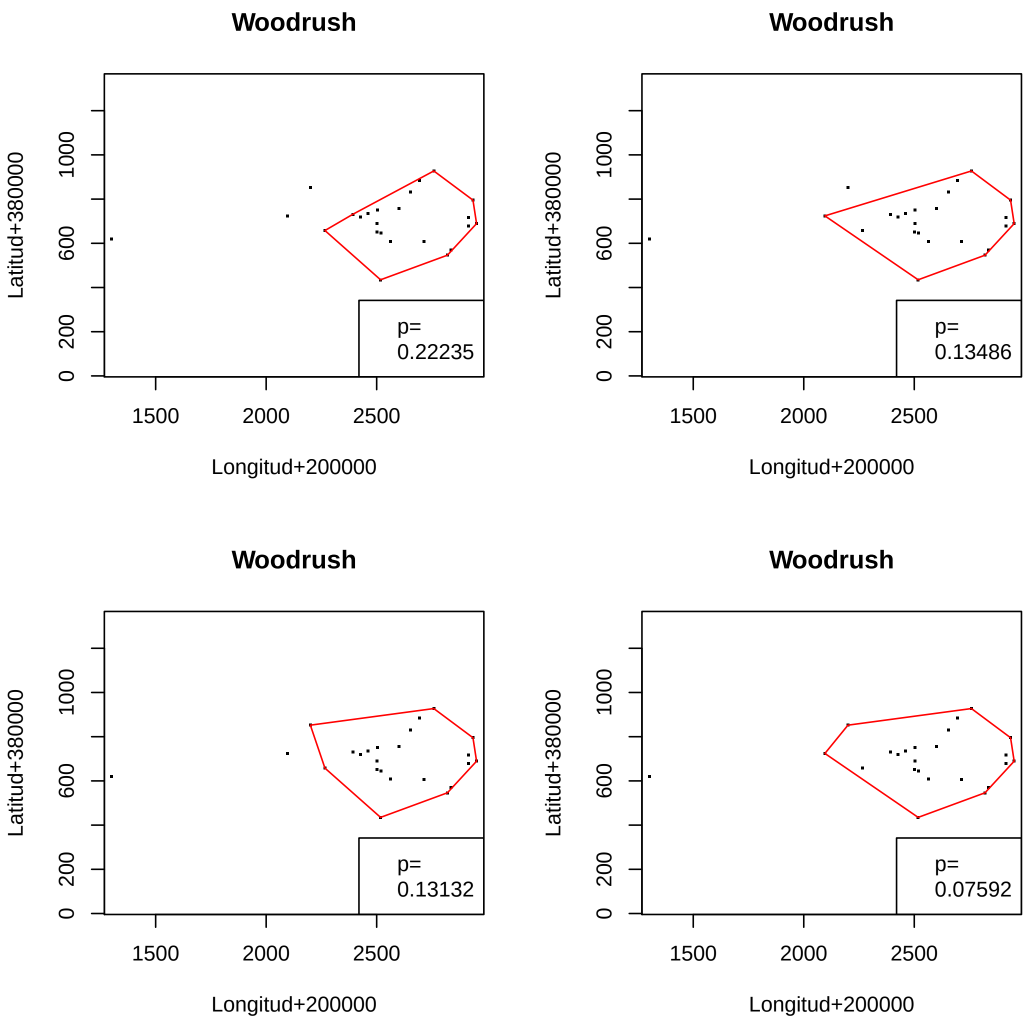


Figure A.70: Most frequent reconstructions of the Woodrush territory using the sampling method of the Unconditional Outlier Prediction Model

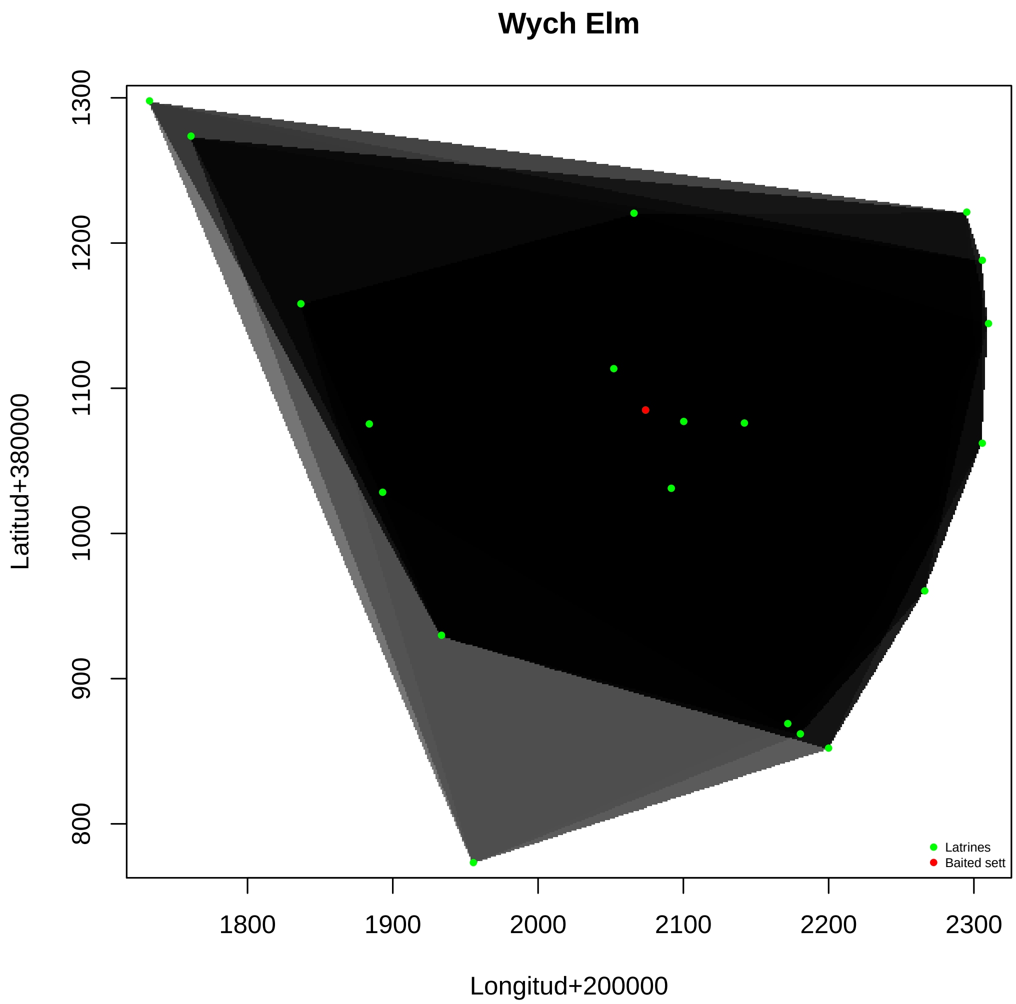


Figure A.71: Probability of inclusion, Wych Elm territory using the Unconditional Outlier Prediction Model sampling method.

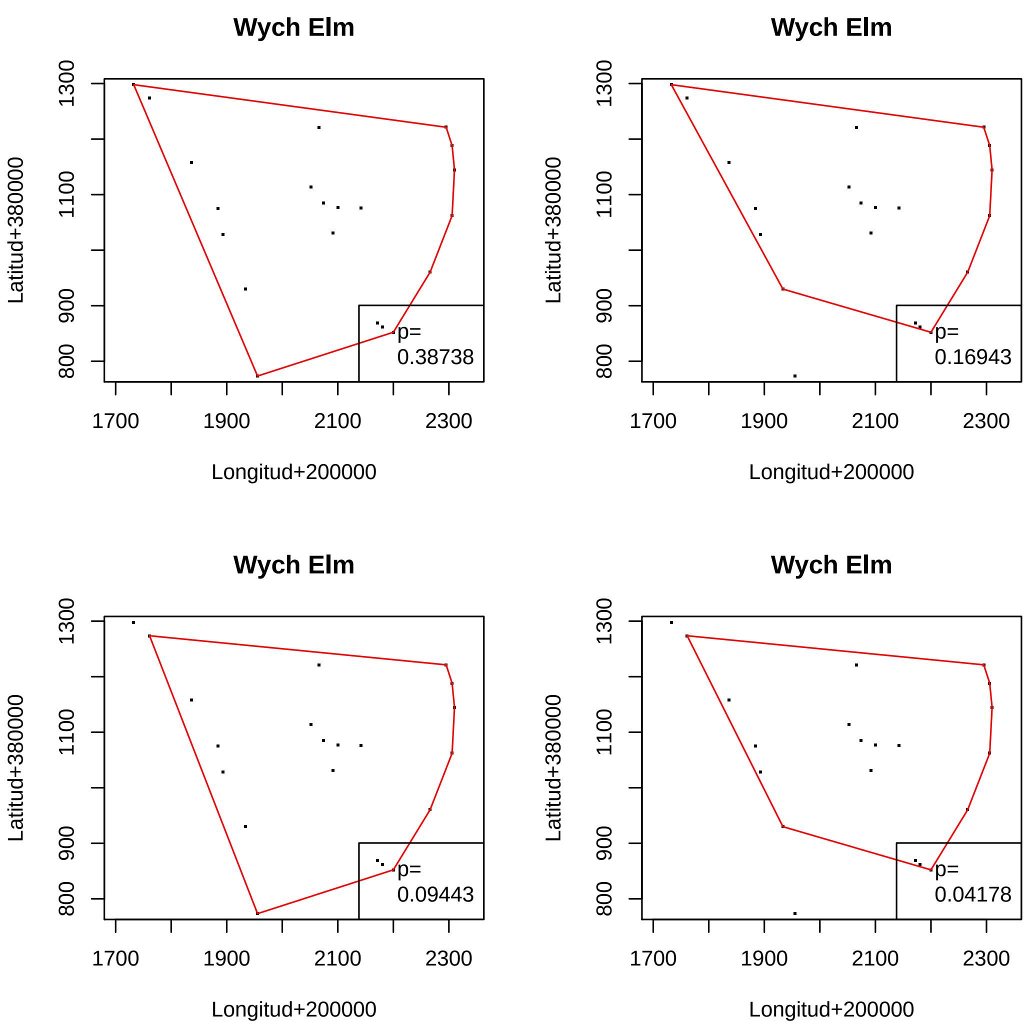


Figure A.72: Most frequent reconstructions of the Wych Elm territory using the sampling method of the Unconditional Outlier Prediction Model

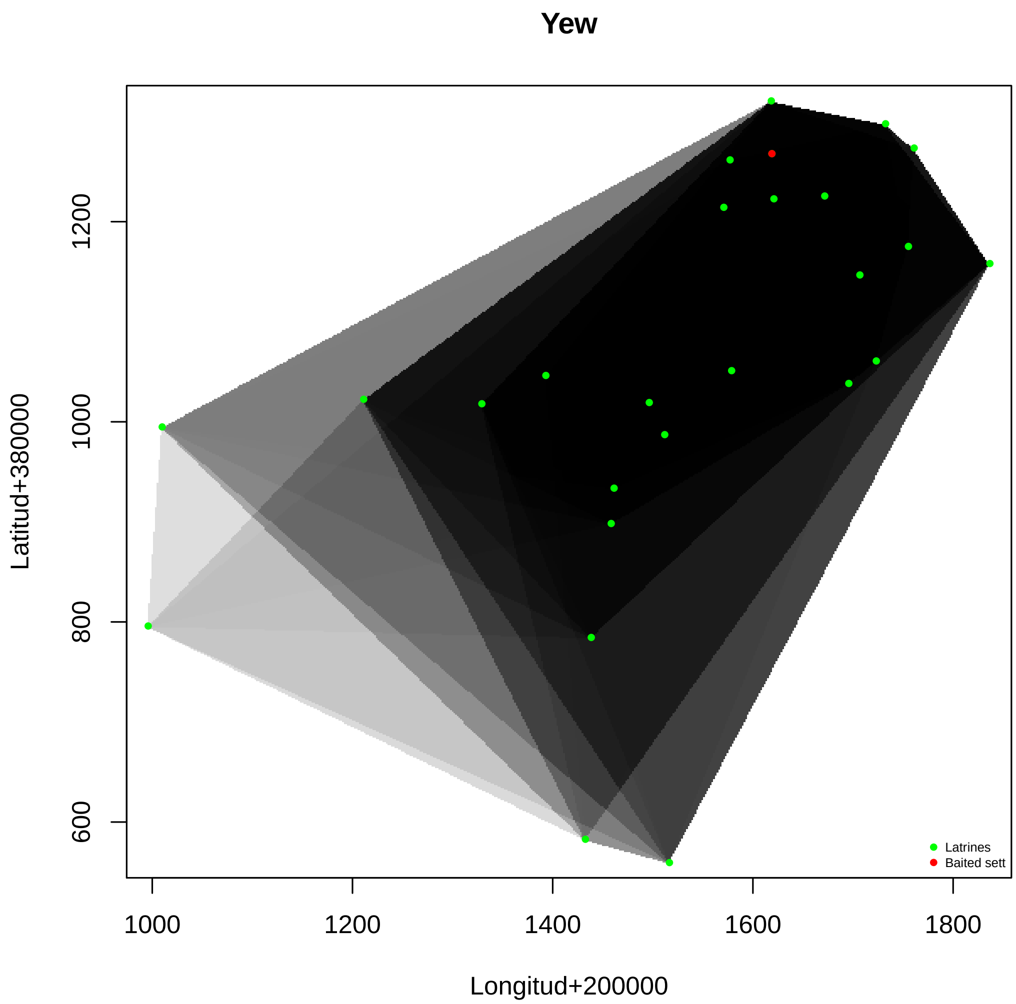


Figure A.73: Probability of inclusion, Yew territory using the Unconditional Outlier Prediction Model sampling method.

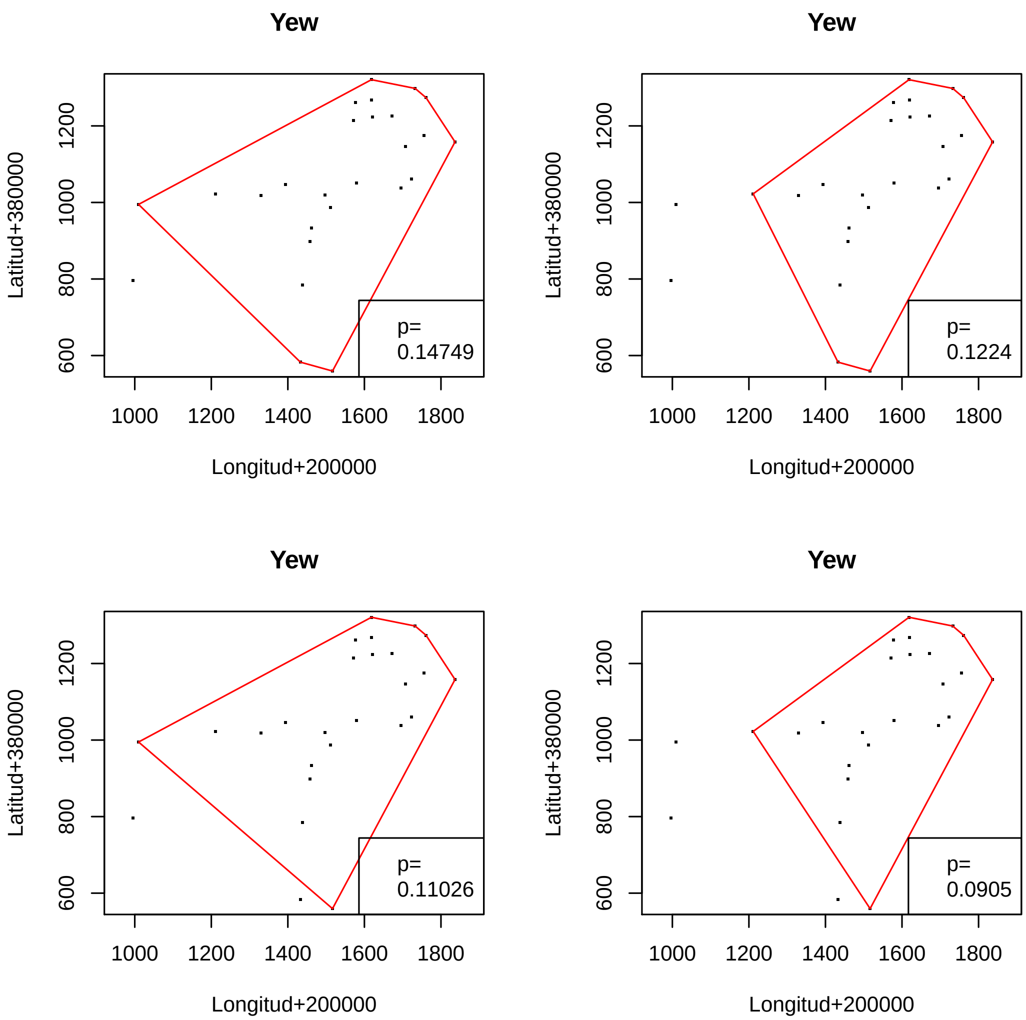


Figure A.74: Most frequent reconstructions of the Yew territory using the sampling method of the Unconditional Outlier Prediction Model

Appendix B

Appendix Conditional Outlier Prediction Model

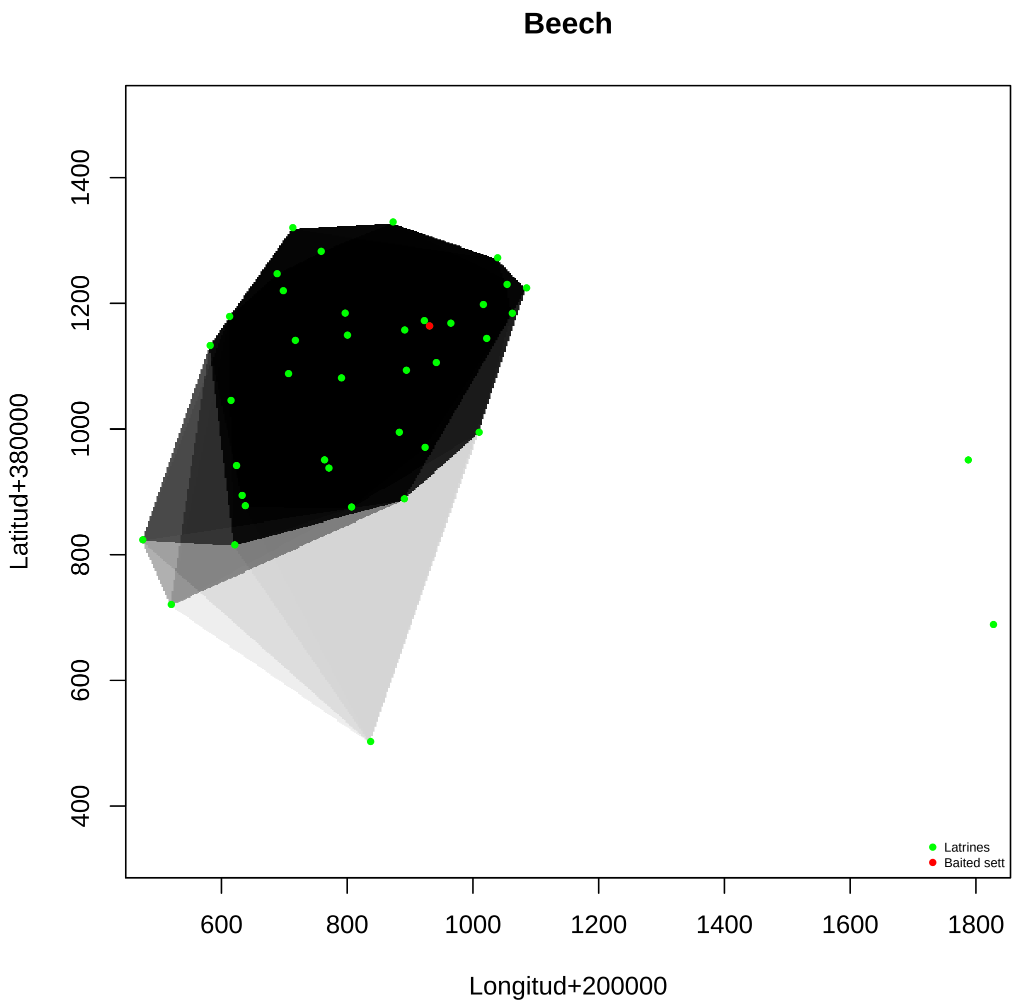


Figure B.1: Probability of inclusion of the Beech territory using the Conditional Outlier Prediction Model

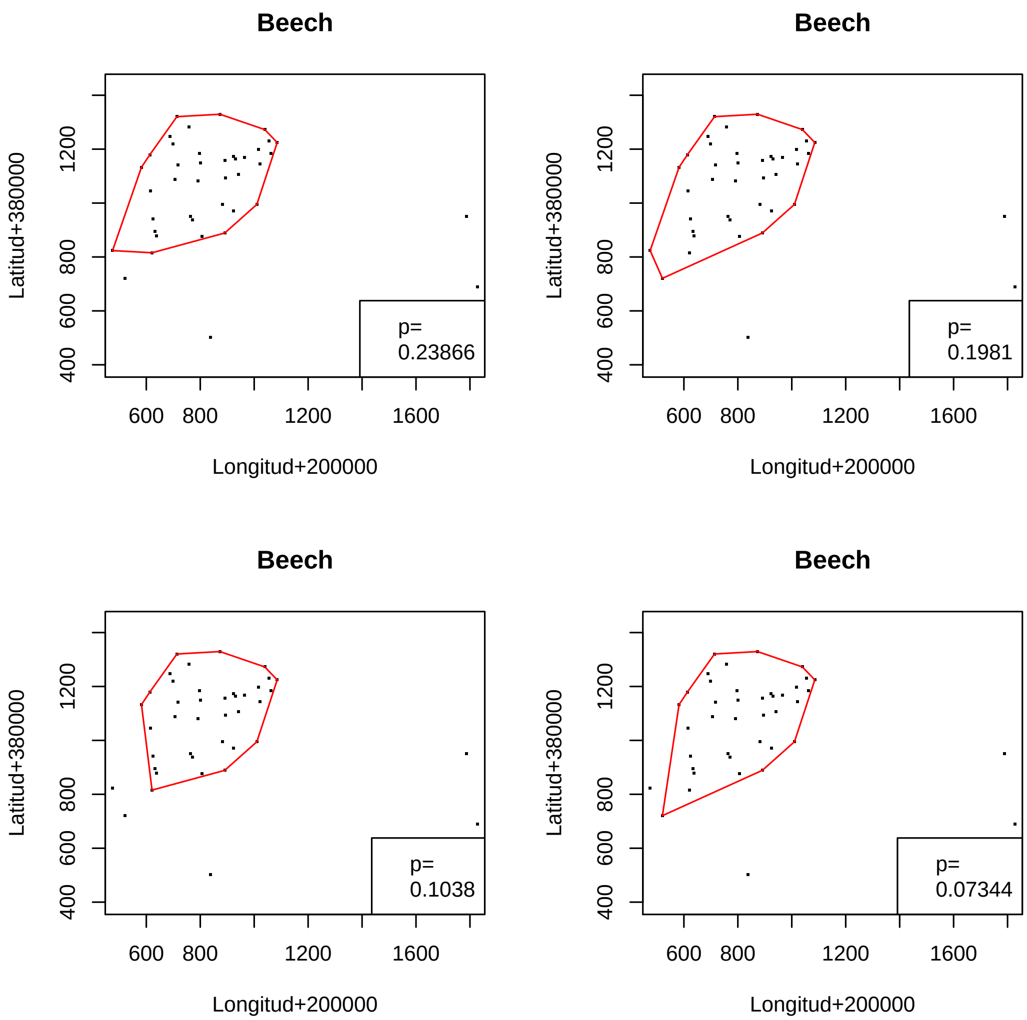


Figure B.2: Most frequent reconstructions of the Beech territory using the Conditional Outlier Prediction Model

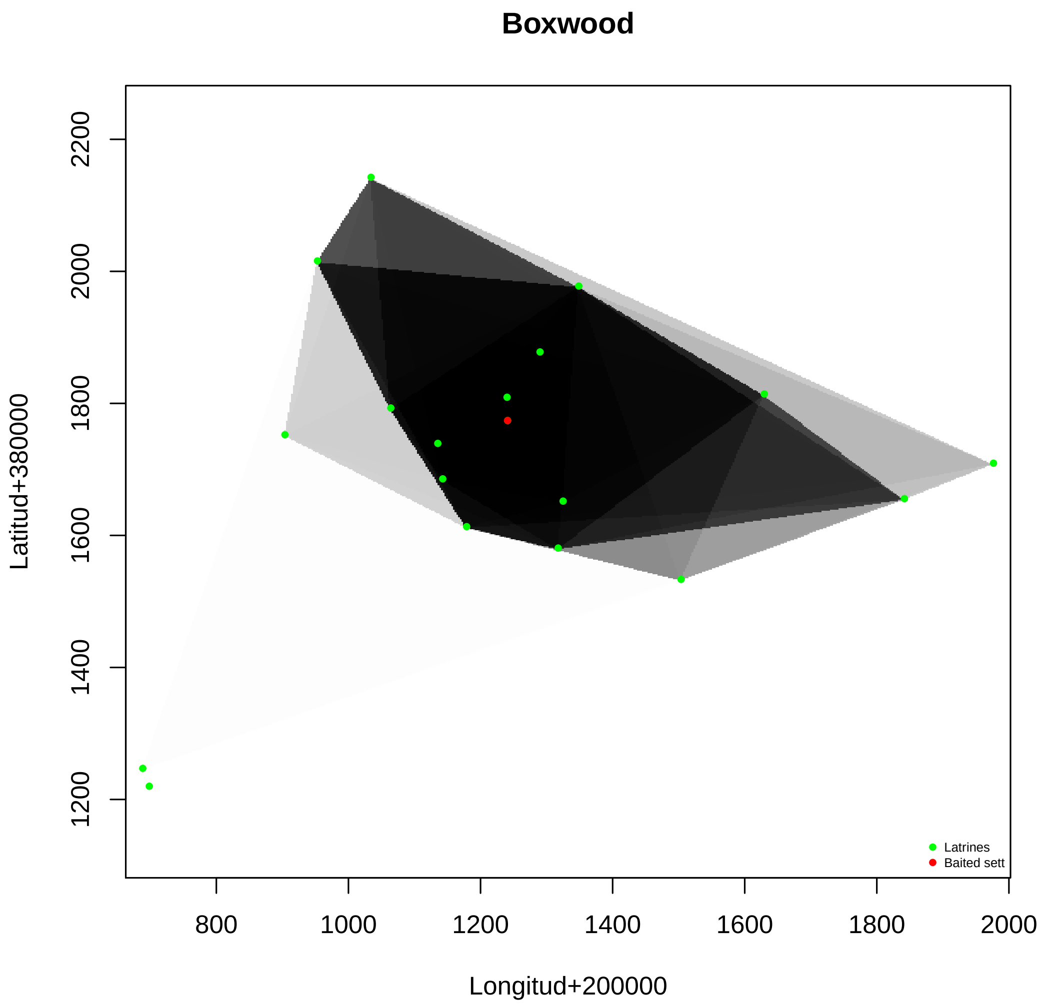


Figure B.3: Probability of inclusion of the Boxwood territory using the Conditional Outlier Prediction Model

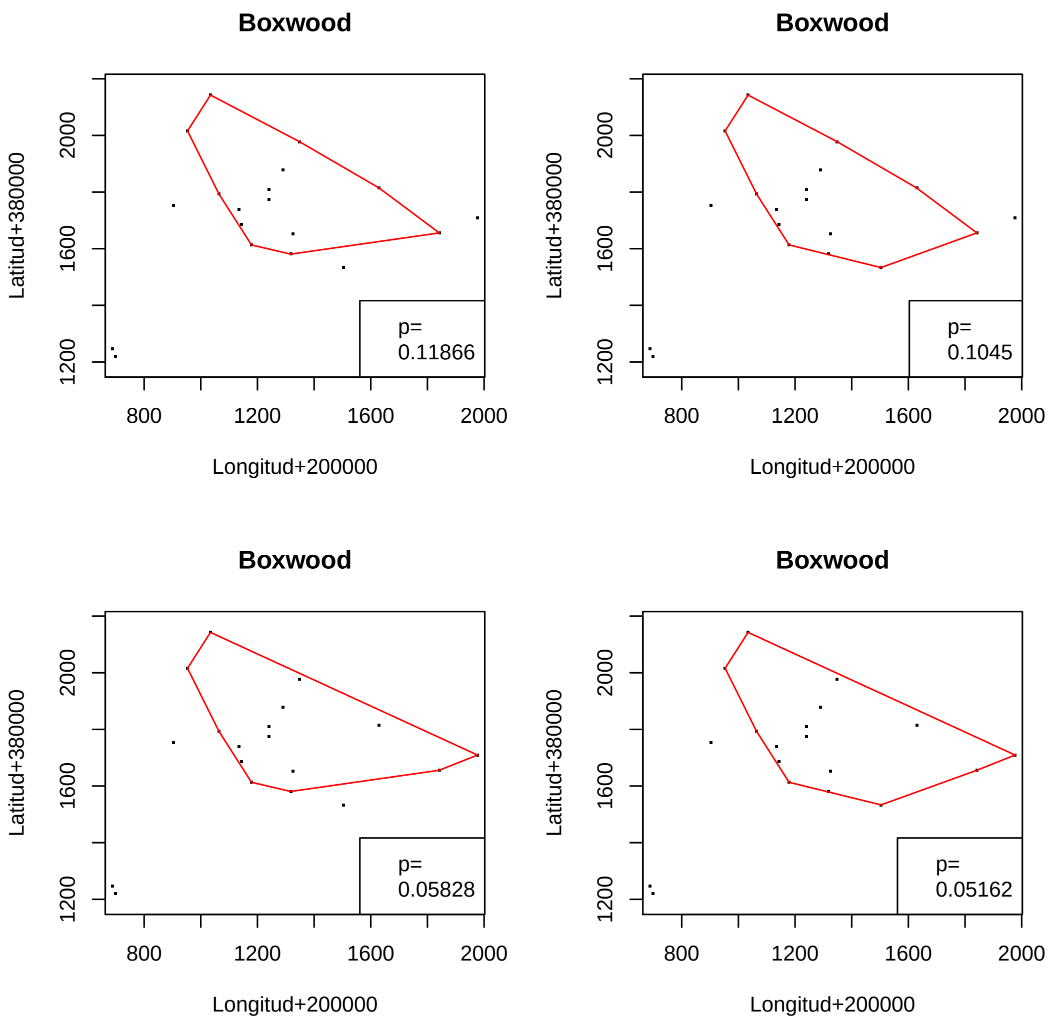


Figure B.4: Most frequent reconstructions of the Boxwood territory using the Conditional Outlier Prediction Model

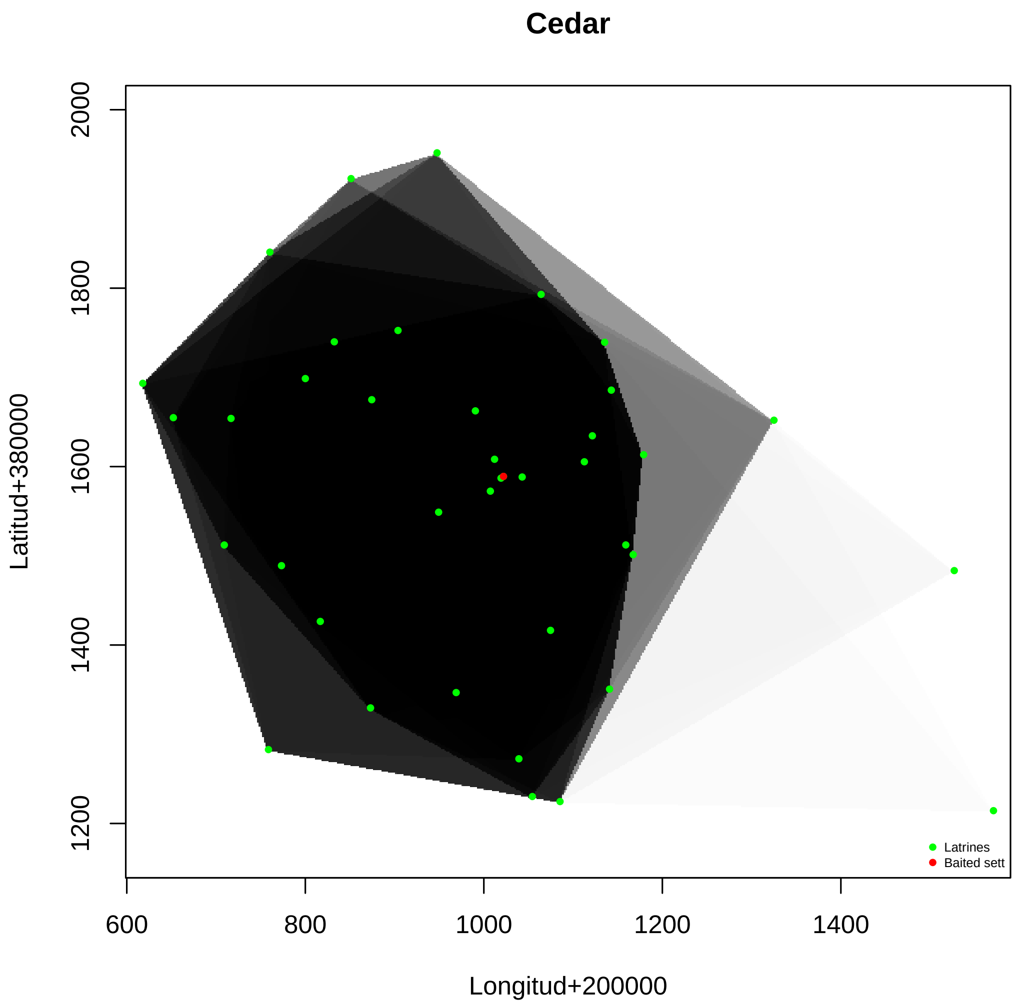


Figure B.5: Probability of inclusion of the Cedar territory using the Conditional Outlier Prediction Model

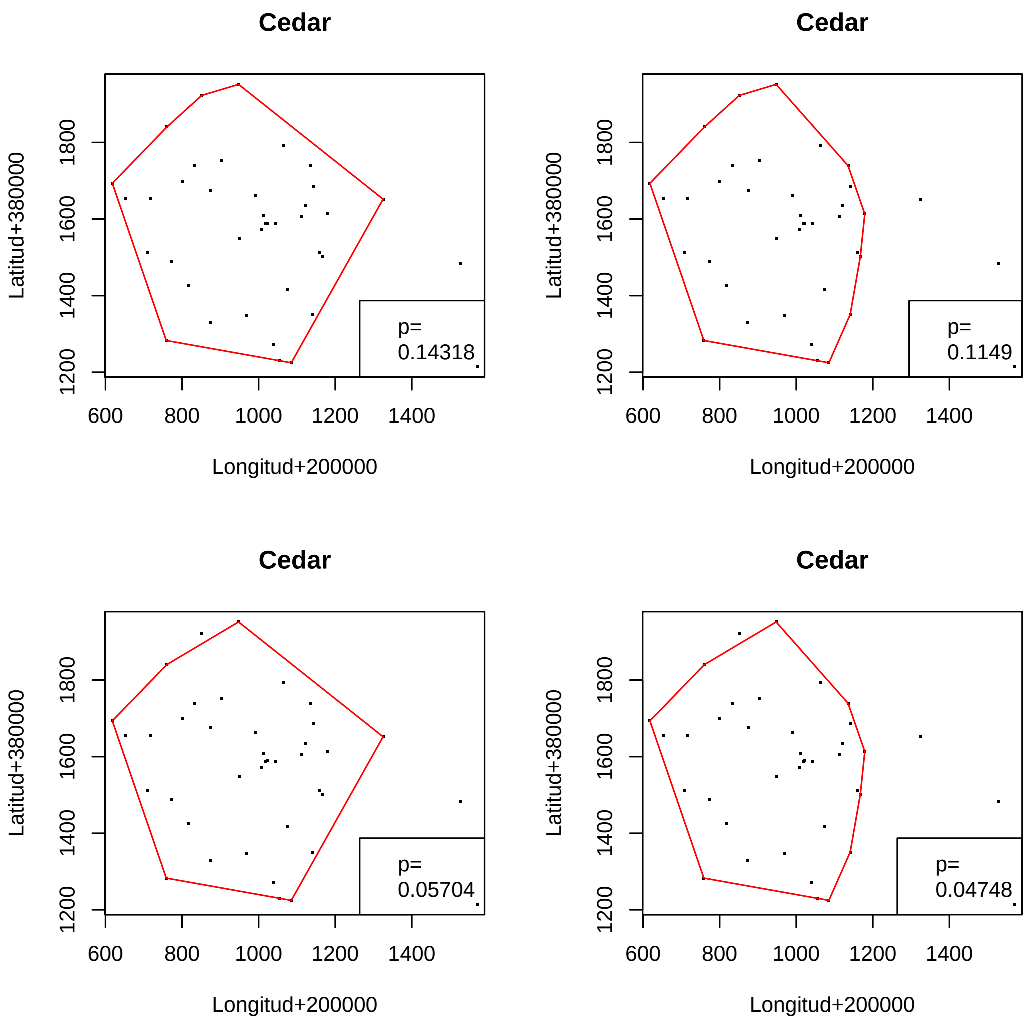


Figure B.6: Most frequent reconstructions of the Cedar territory using the Conditional Outlier Prediction Model

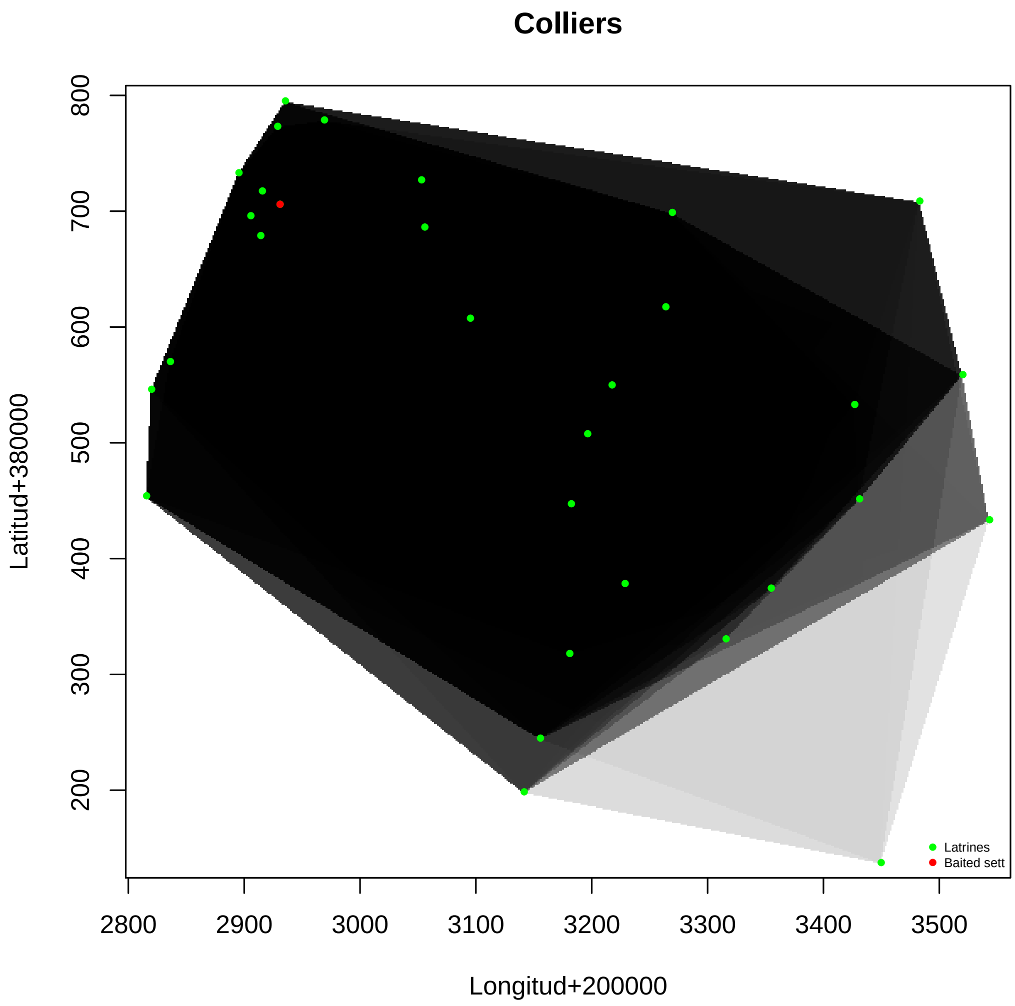


Figure B.7: Probability of inclusion of the Colliers territory using the Conditional Outlier Prediction Model

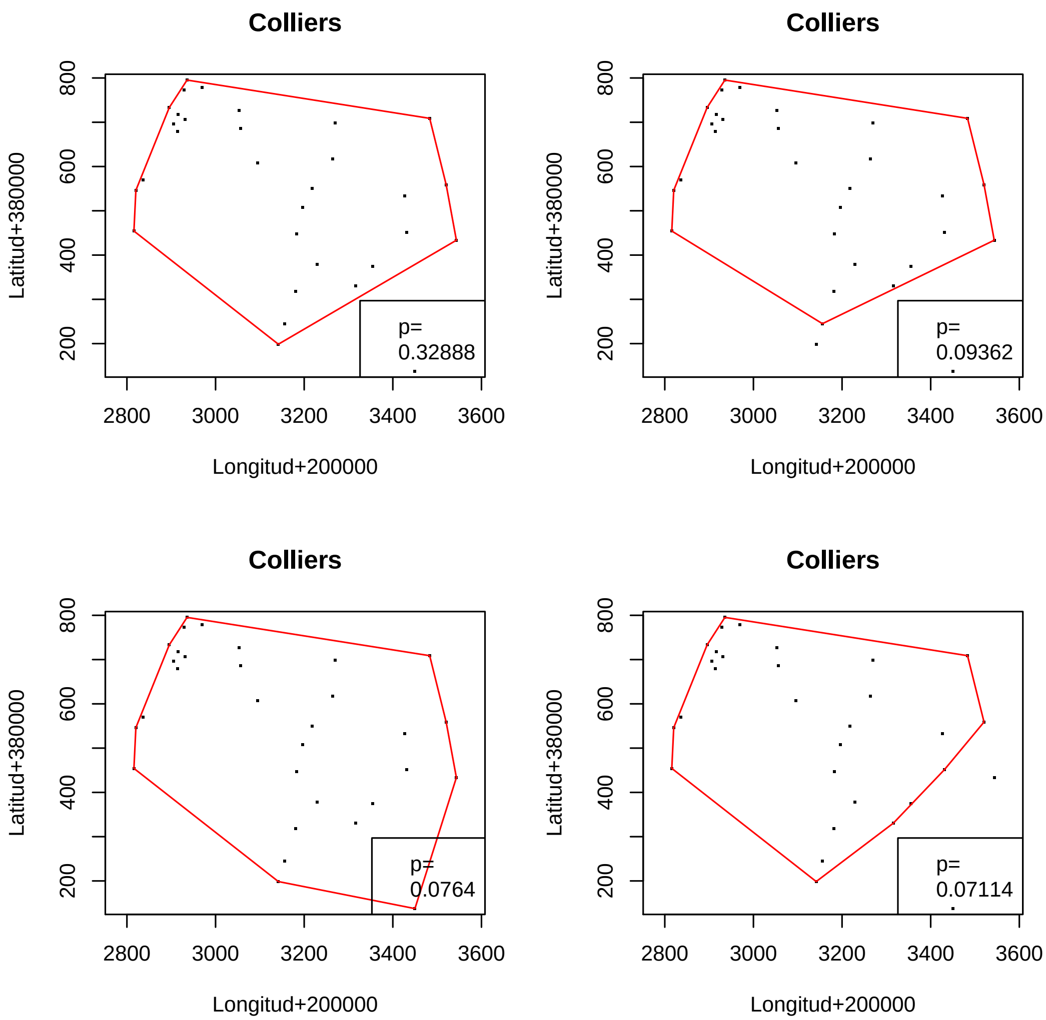


Figure B.8: Most frequent reconstructions of the Colliers territory using the Conditional Outlier Prediction Model

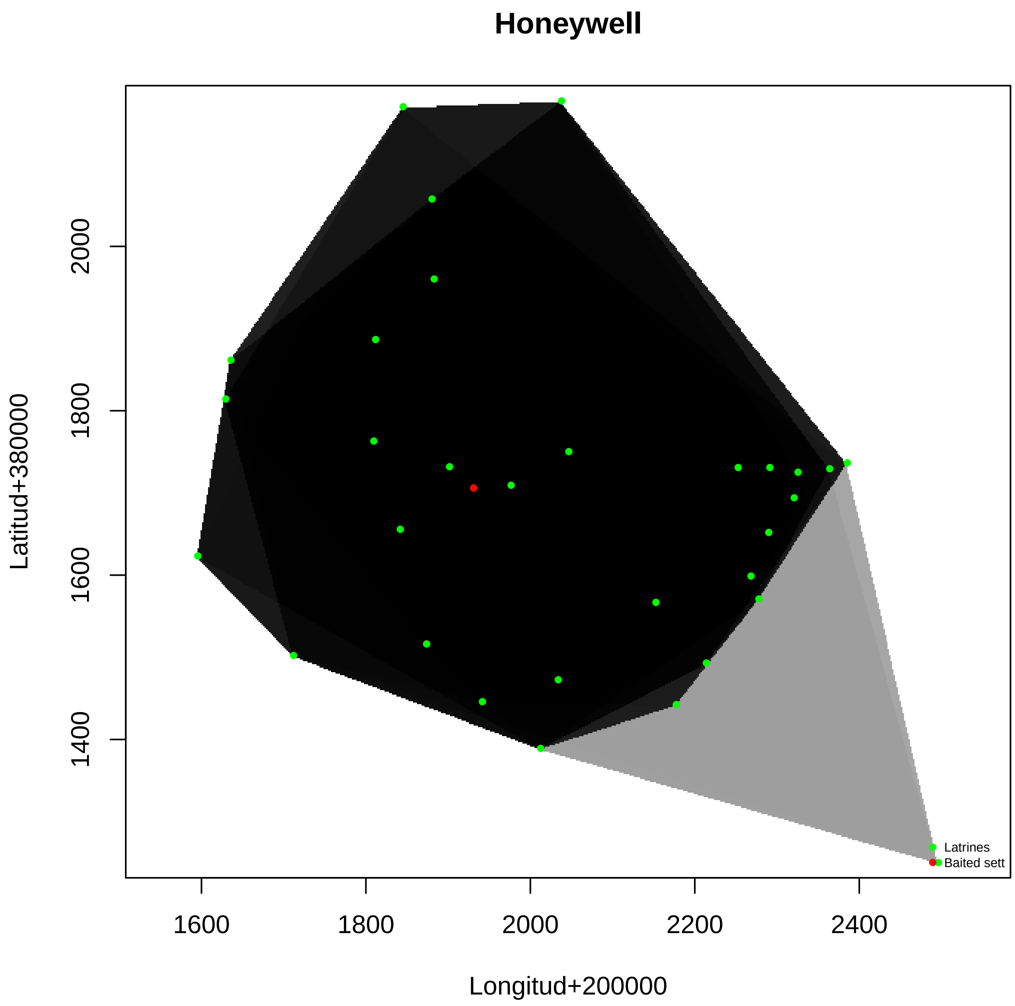


Figure B.9: Probability of inclusion of the Honeywell territory using the Conditional Outlier Prediction Model

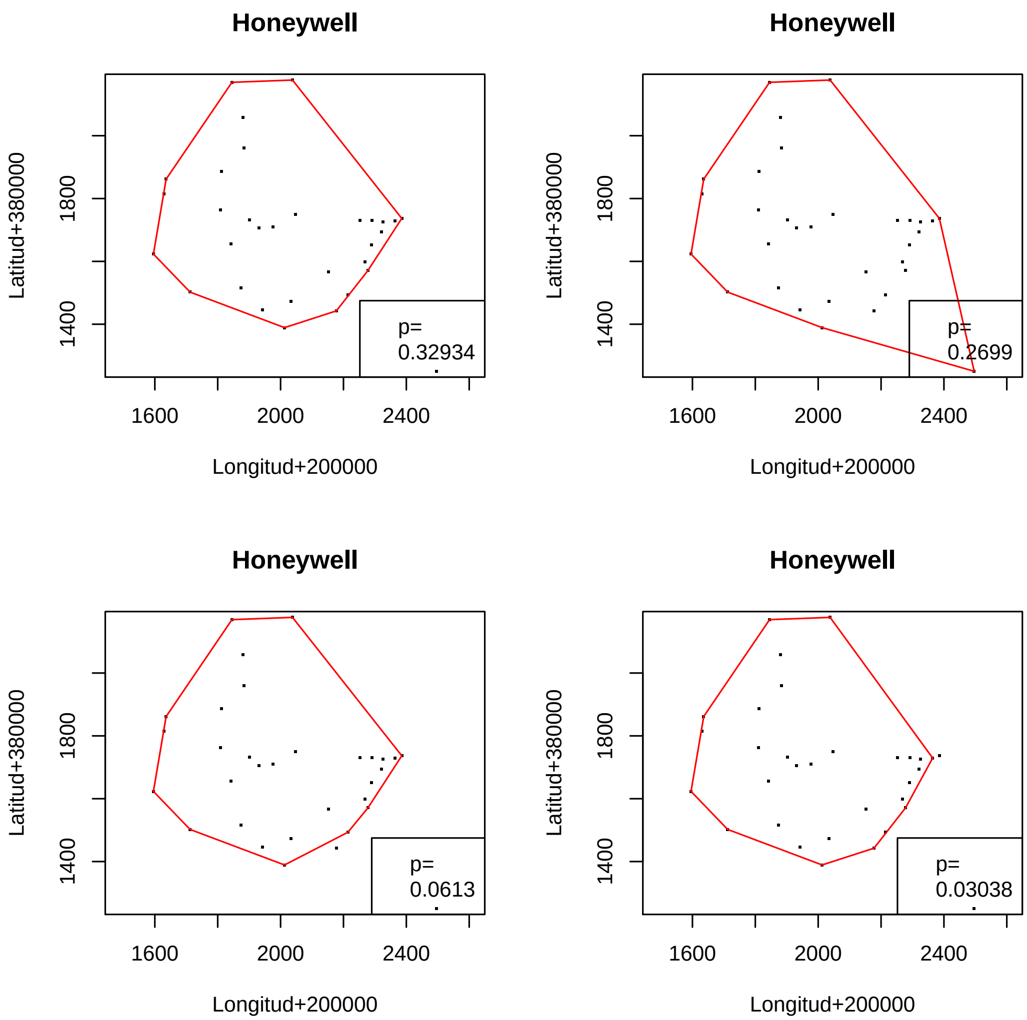


Figure B.10: Most frequent reconstructions of the Honeywell territory using the Conditional Outlier Prediction Model

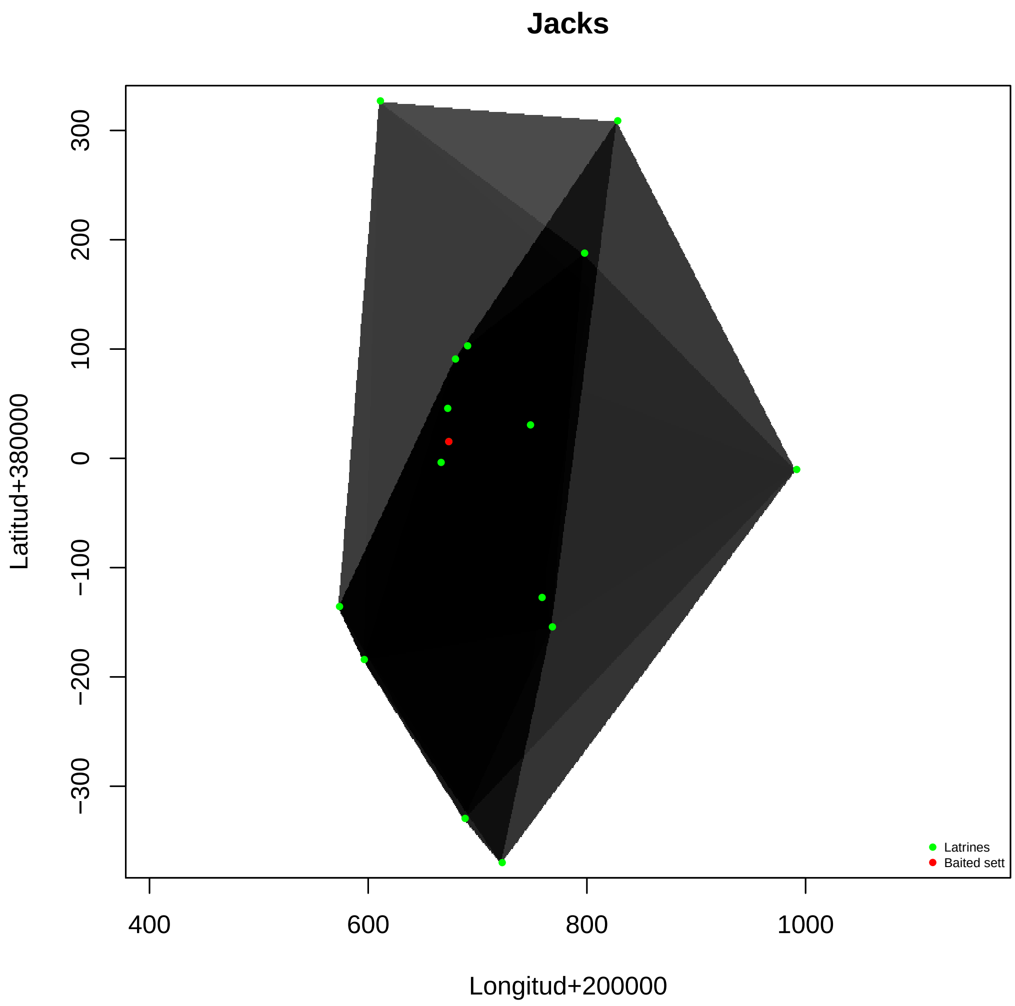


Figure B.11: Probability of inclusion of the Jacks territory using the Conditional Outlier Prediction Model

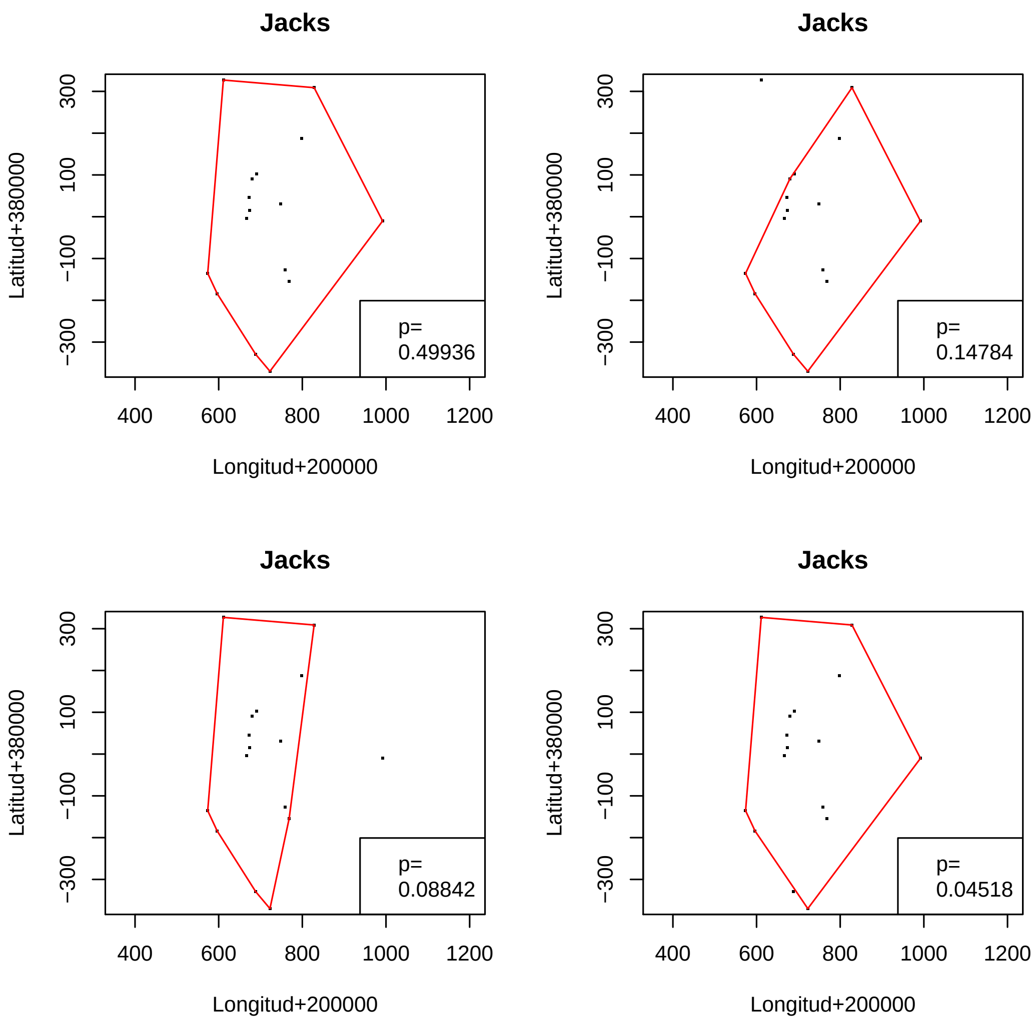


Figure B.12: Most frequent reconstructions of the Jacks territory Unconditional Outlier Prediction Model

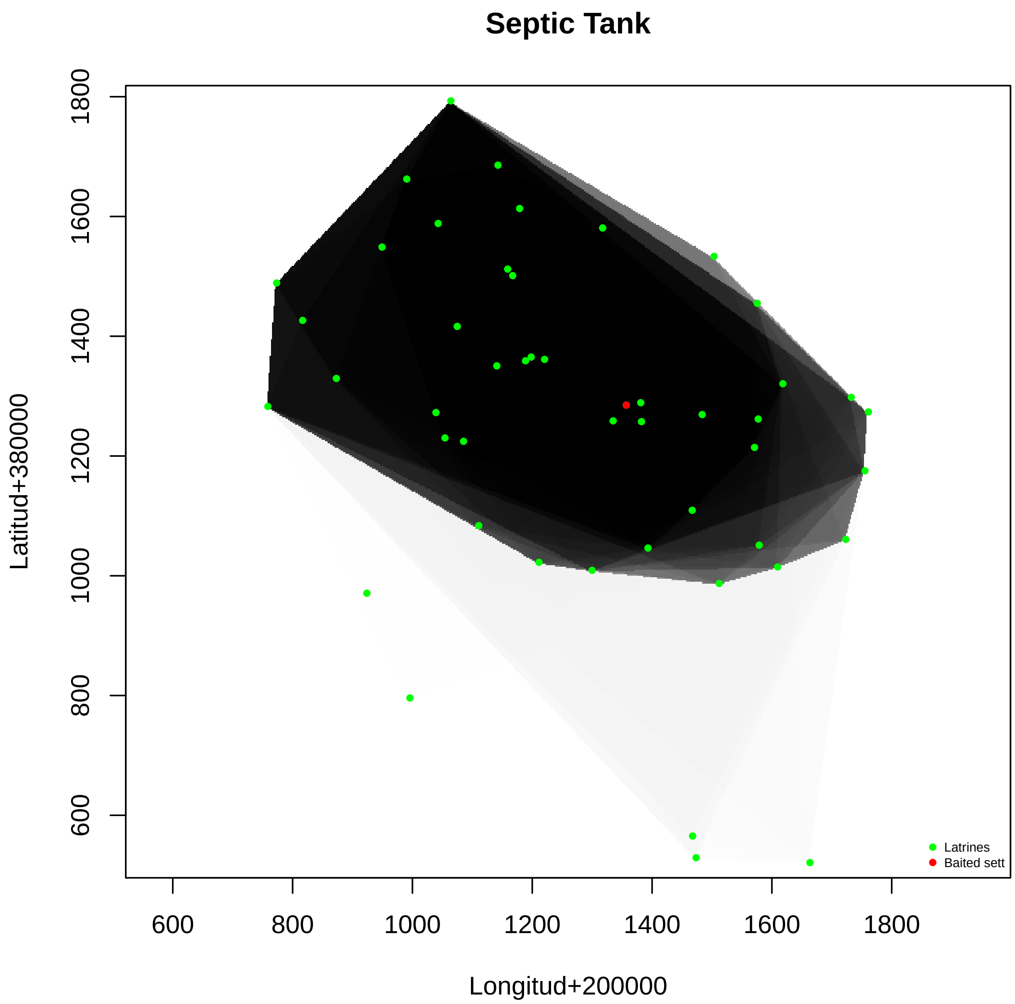


Figure B.13: Probability of inclusion of the Septic Tank territory using the Conditional Outlier Prediction Model

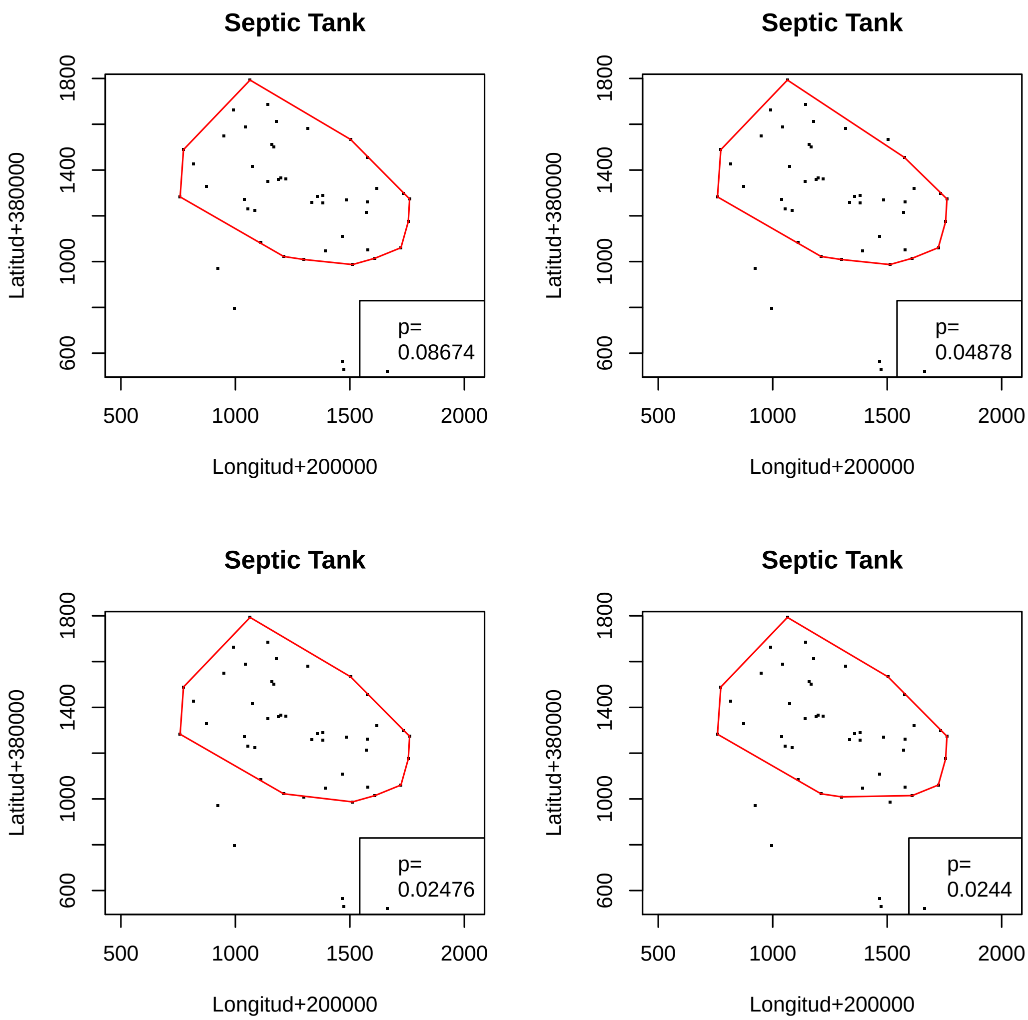


Figure B.14: Most frequent reconstructions of the Septic Tank territory using the Conditional Outlier Prediction Model

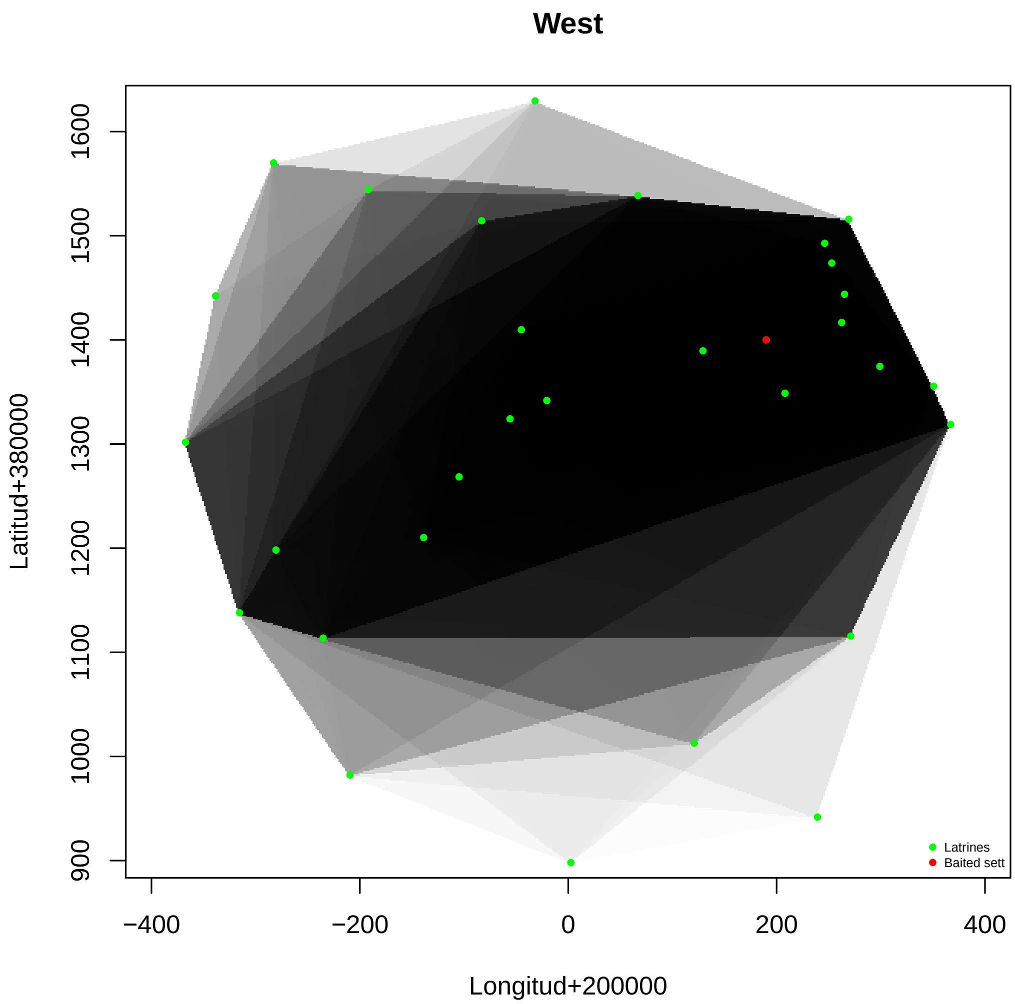


Figure B.15: Probability of inclusion of the West territory using the Conditional Outlier Prediction Model

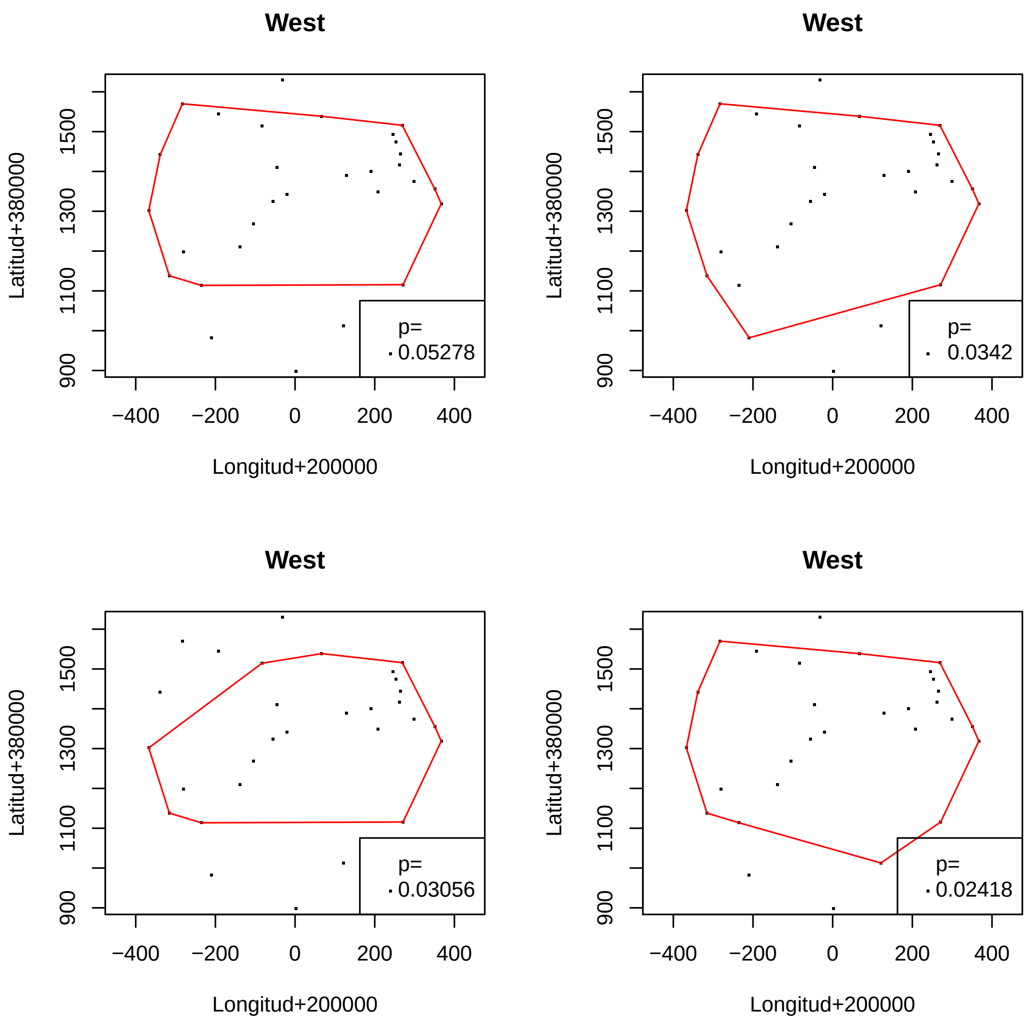


Figure B.16: Most frequent reconstructions of the West territory using the Conditional Outlier Prediction Model

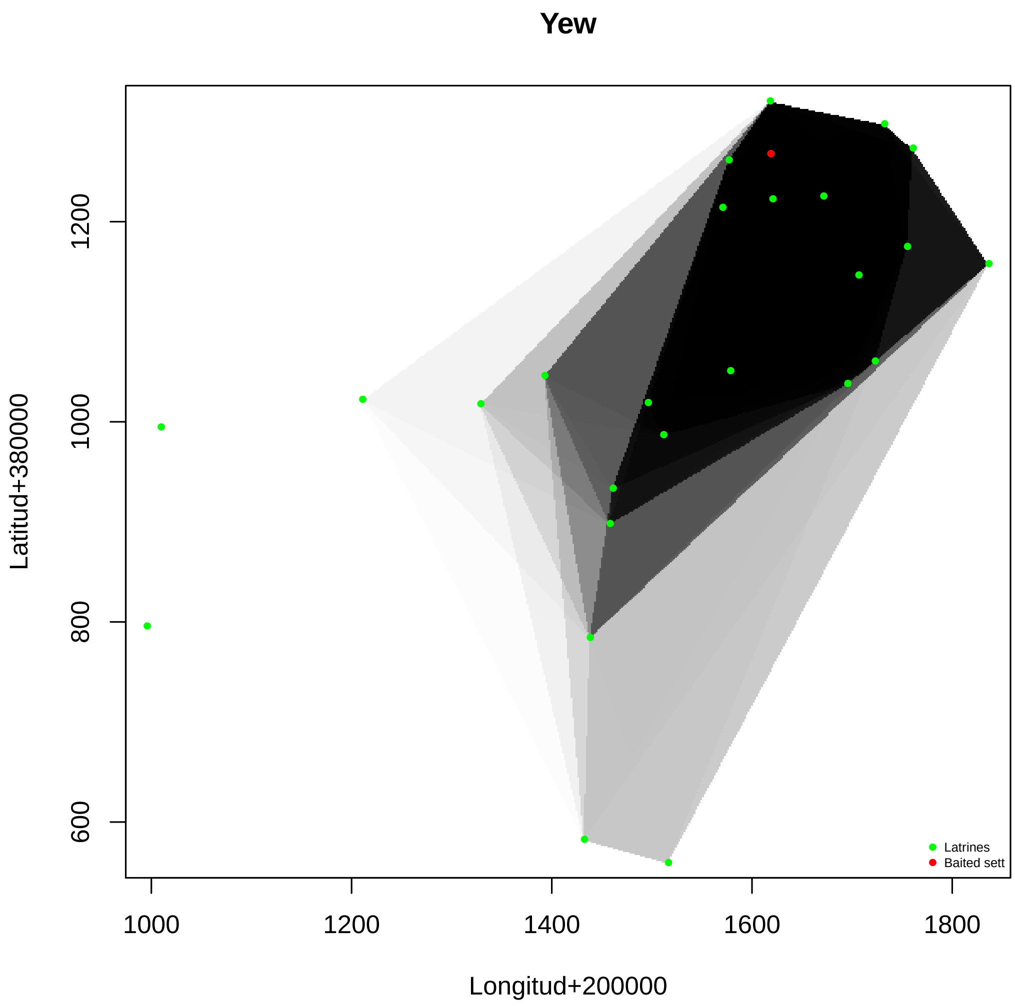


Figure B.17: Probability of inclusion of the Yew territory using the Conditional Outlier Prediction Model

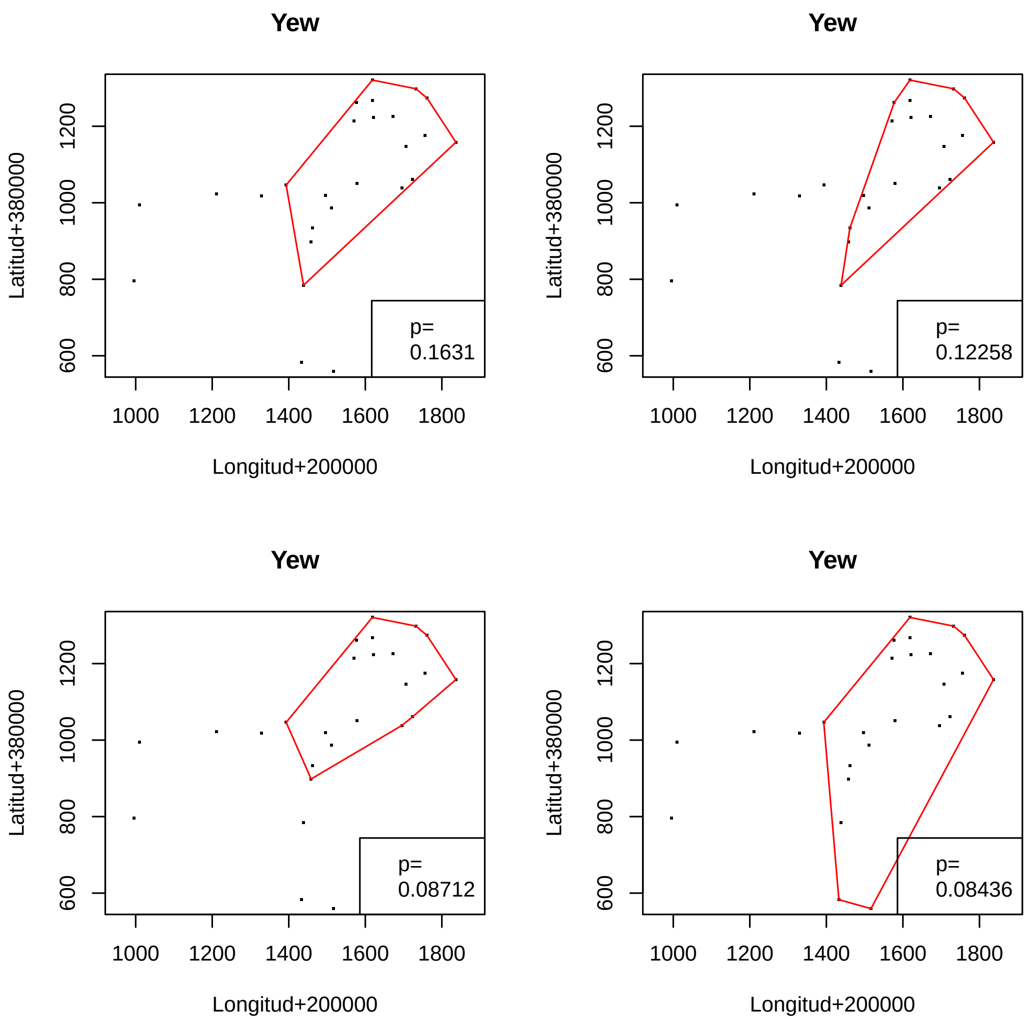
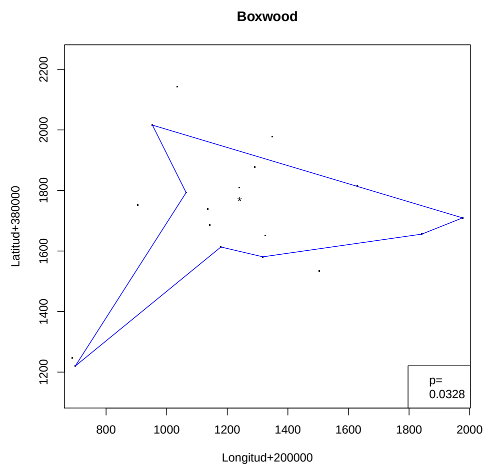


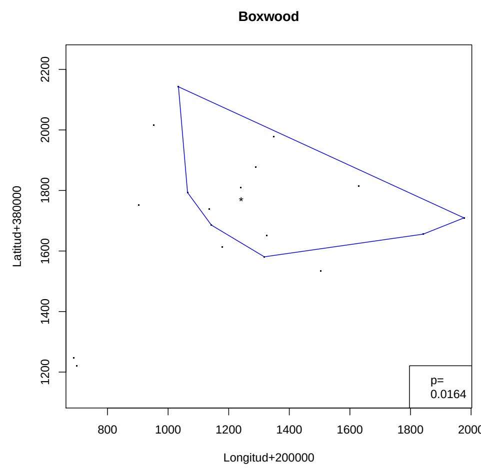
Figure B.18: Most frequent reconstructions of the Yew territory using the Conditional Outlier Prediction Model

Appendix C

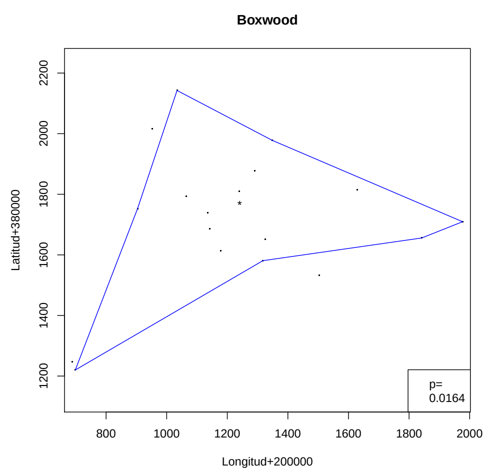
Appendix Unadjusted Ordinal Model



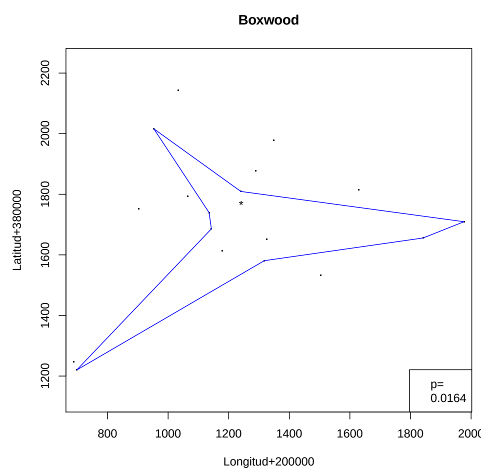
(a) Configuration 1



(b) Configuration 2

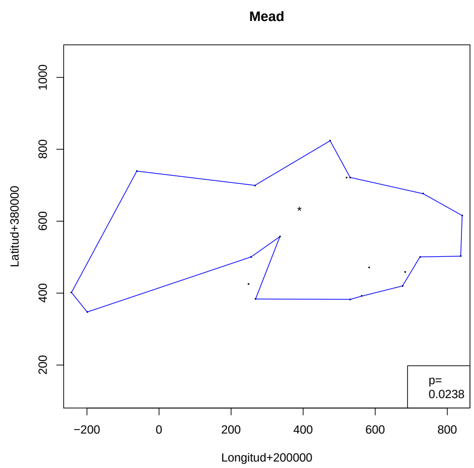


(c) Configuration 3

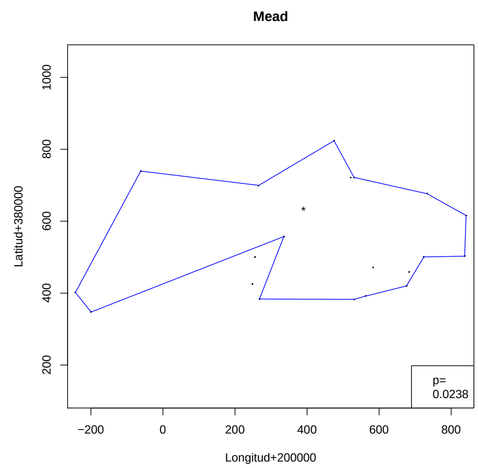


(d) Configuration 4

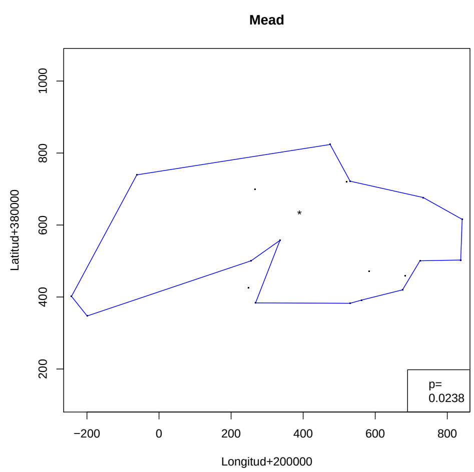
Figure C.1: Reconstructions of the territory Boxwood with higher frequency using the Unadjusted Ordinal Model.



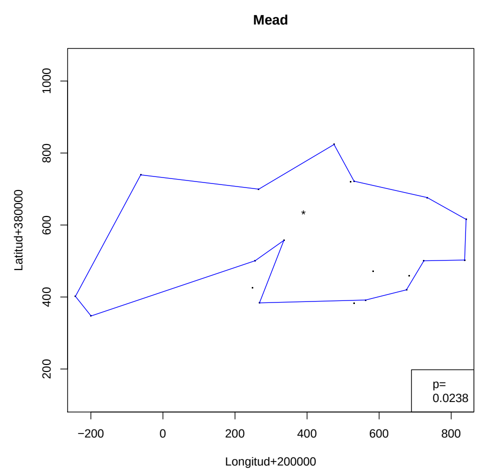
(a) Configuration 1



(b) Configuration 2

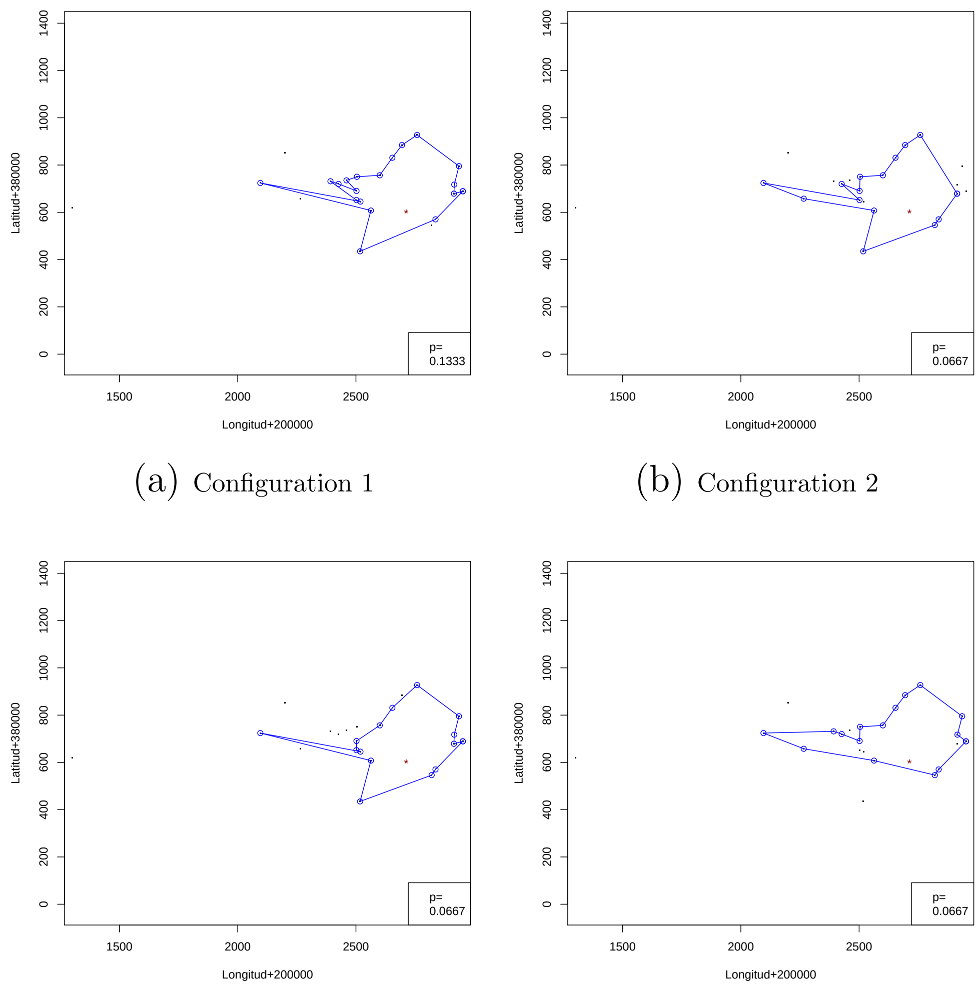


(c) Configuration 3



(d) Configuration 4

Figure C.2: Reconstructions of the territory Mead with higher frequency using the Unadjusted Ordinal Model.



(a) Configuration 1

(b) Configuration 2

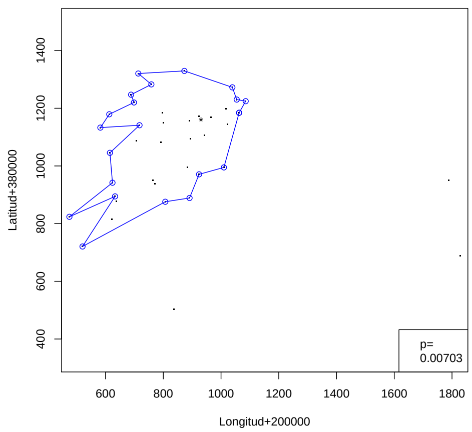
(c) Configuration 3

(d) Configuration 4

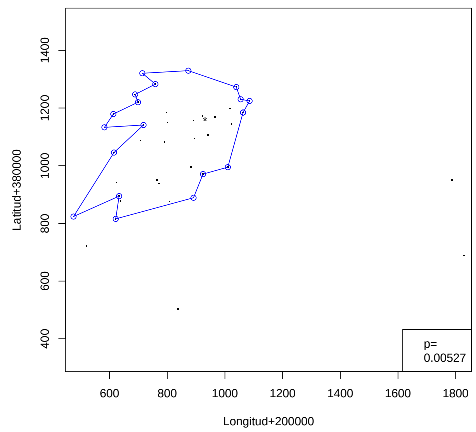
Figure C.3: Reconstructions of the territory Woodrush with higher frequency using the Unadjusted Ordinal Model.

Appendix D

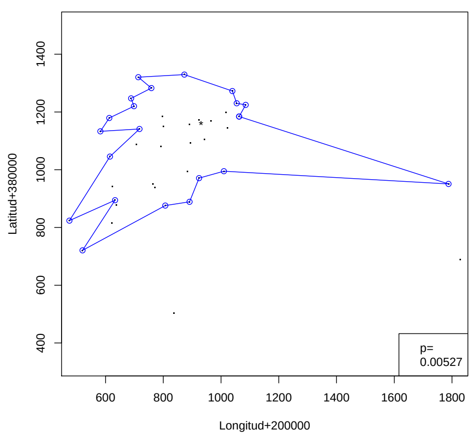
Appendix Adjusted Ordinal Model: Most Frequent Territories Figures



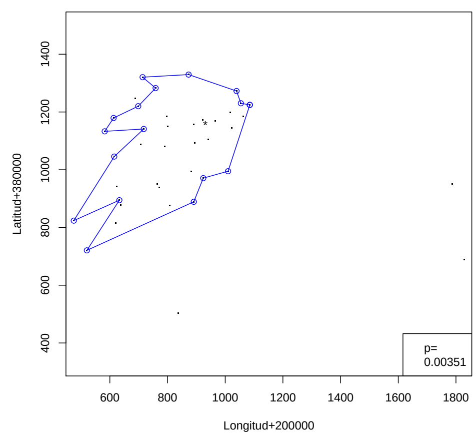
(a) Configuration 1



(b) Configuration 2

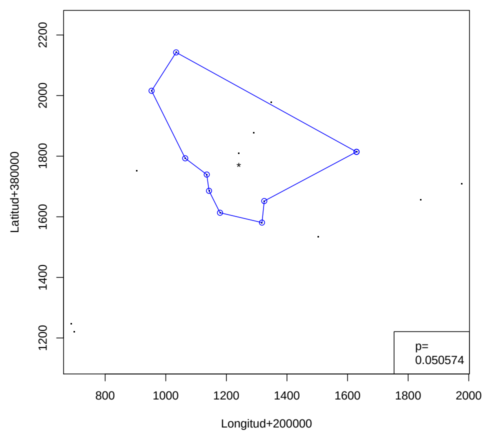


(c) Configuration 3

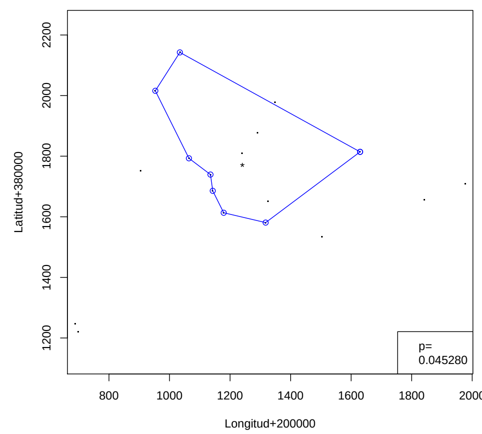


(d) Configuration 4

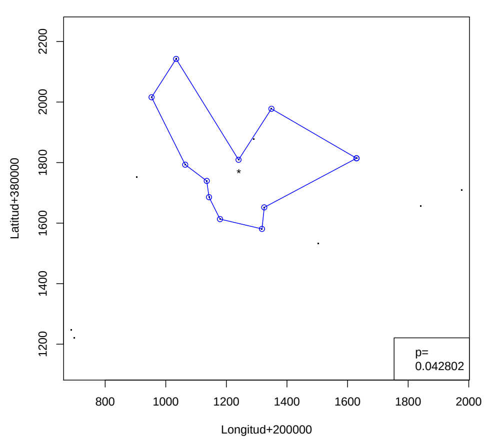
Figure D.1: Reconstructions of the territory Beech with higher frequency using the Adjusted Ordinal Model.



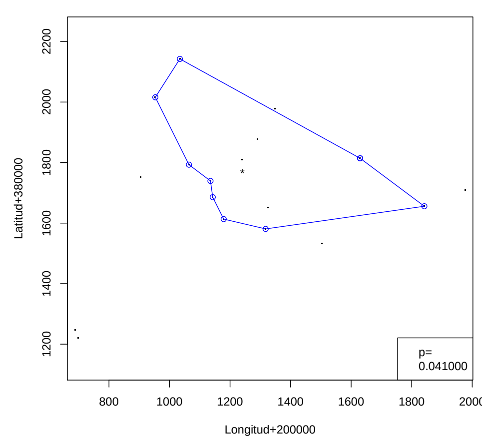
(a) Configuration 1



(b) Configuration 2

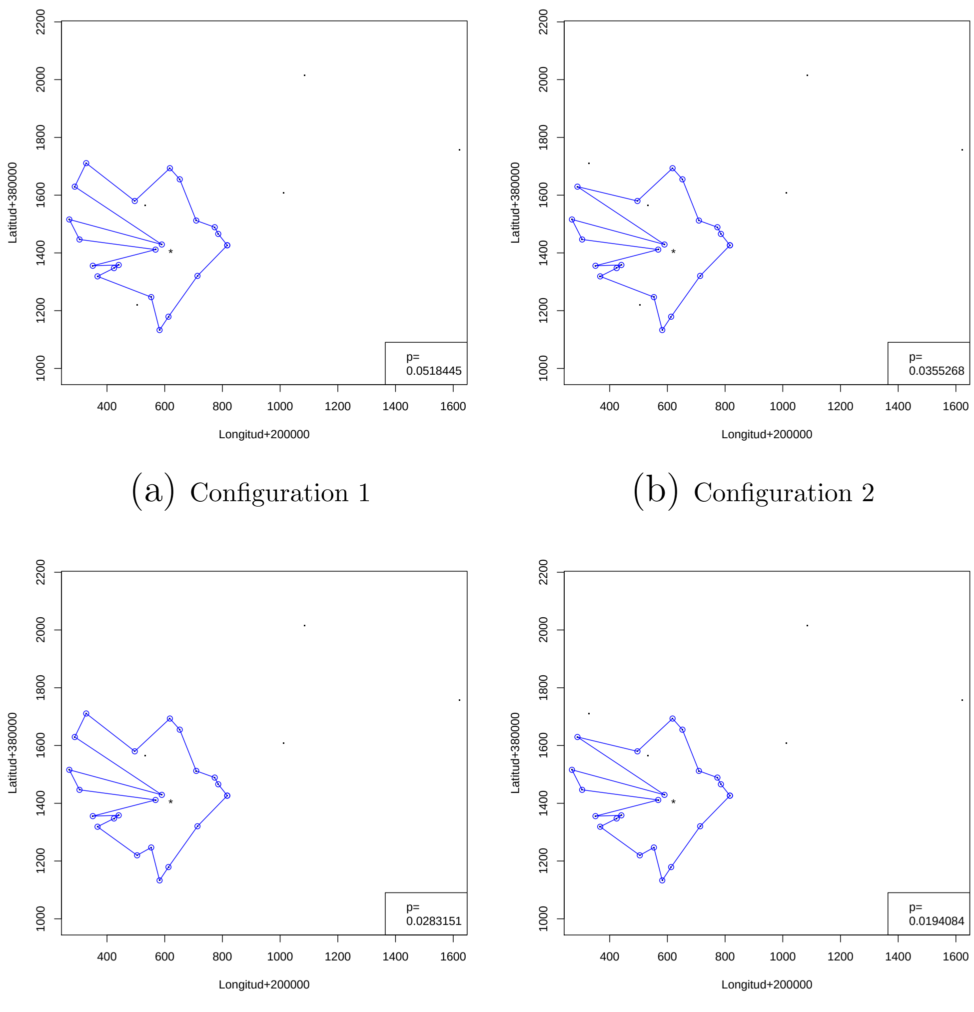


(c) Configuration 3



(d) Configuration 4

Figure D.2: Reconstructions of the territory Boxwood with higher frequency using the Adjusted Ordinal Model.



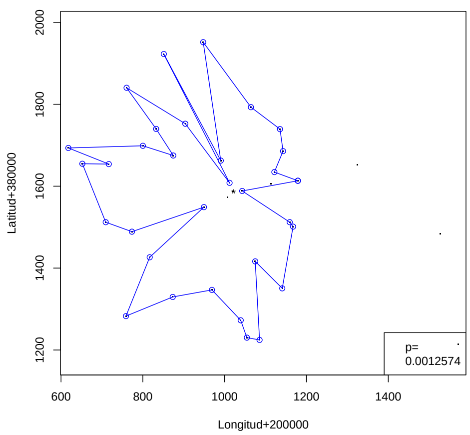
(a) Configuration 1

(b) Configuration 2

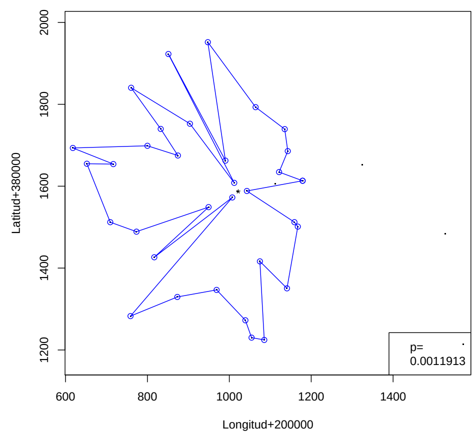
(c) Configuration 3

(d) Configuration 4

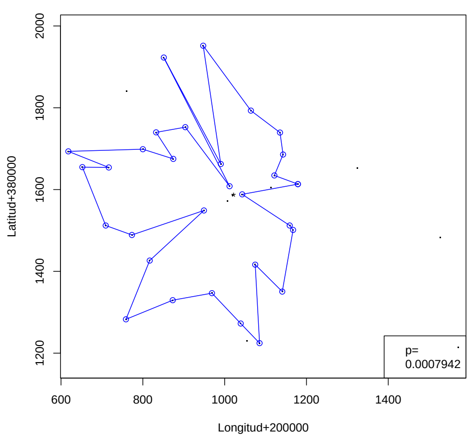
Figure D.3: Reconstructions of the territory Box with higher frequency using the Adjusted Ordinal Model.



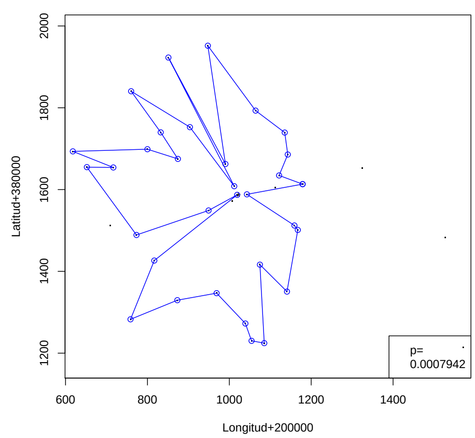
(a) Configuration 1



(b) Configuration 2

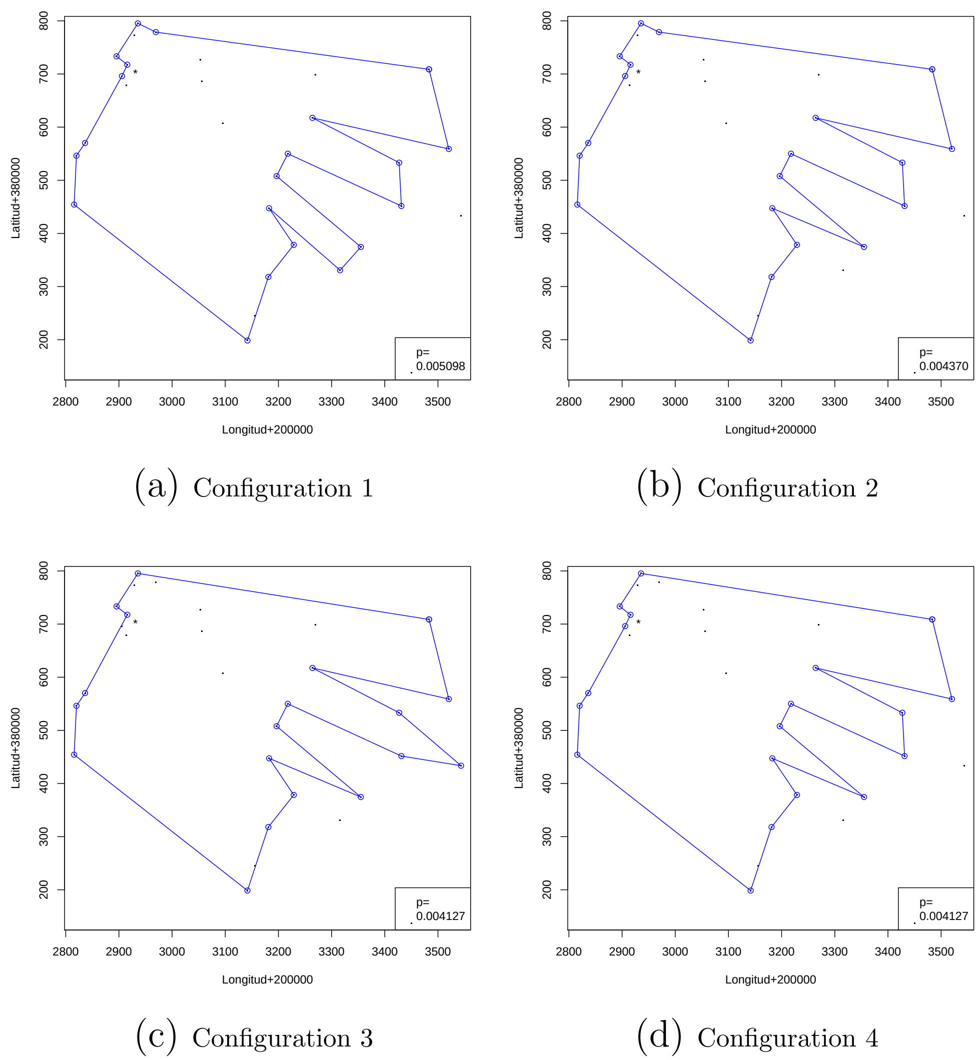


(c) Configuration 3



(d) Configuration 4

Figure D.4: Reconstructions of the territory Cedar with higher frequency using the Adjusted Ordinal Model.



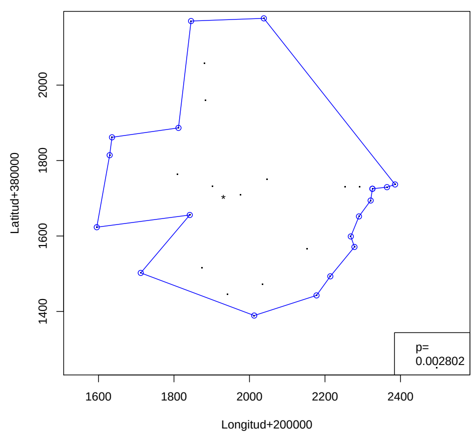
(a) Configuration 1

(b) Configuration 2

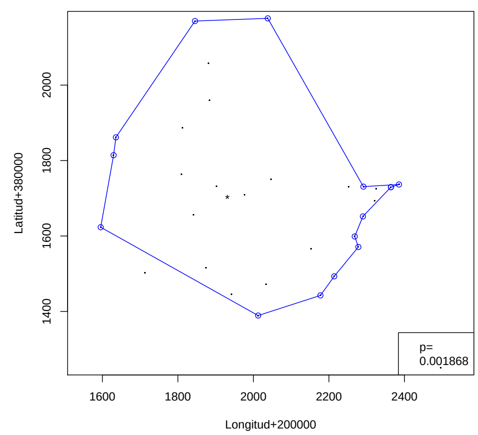
(c) Configuration 3

(d) Configuration 4

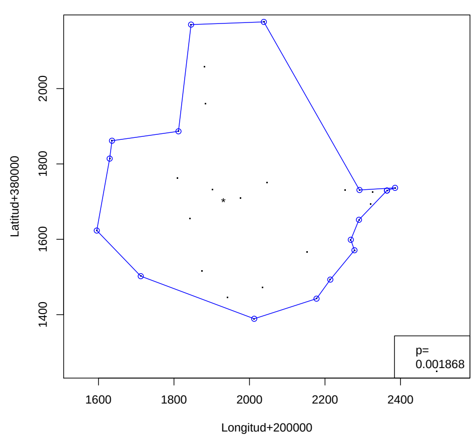
Figure D.5: Reconstructions of the territory Colliers with higher frequency using the Adjusted Ordinal Model.



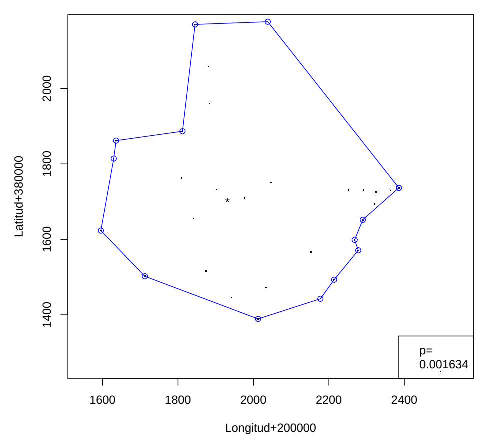
(a) Configuration 1



(b) Configuration 2

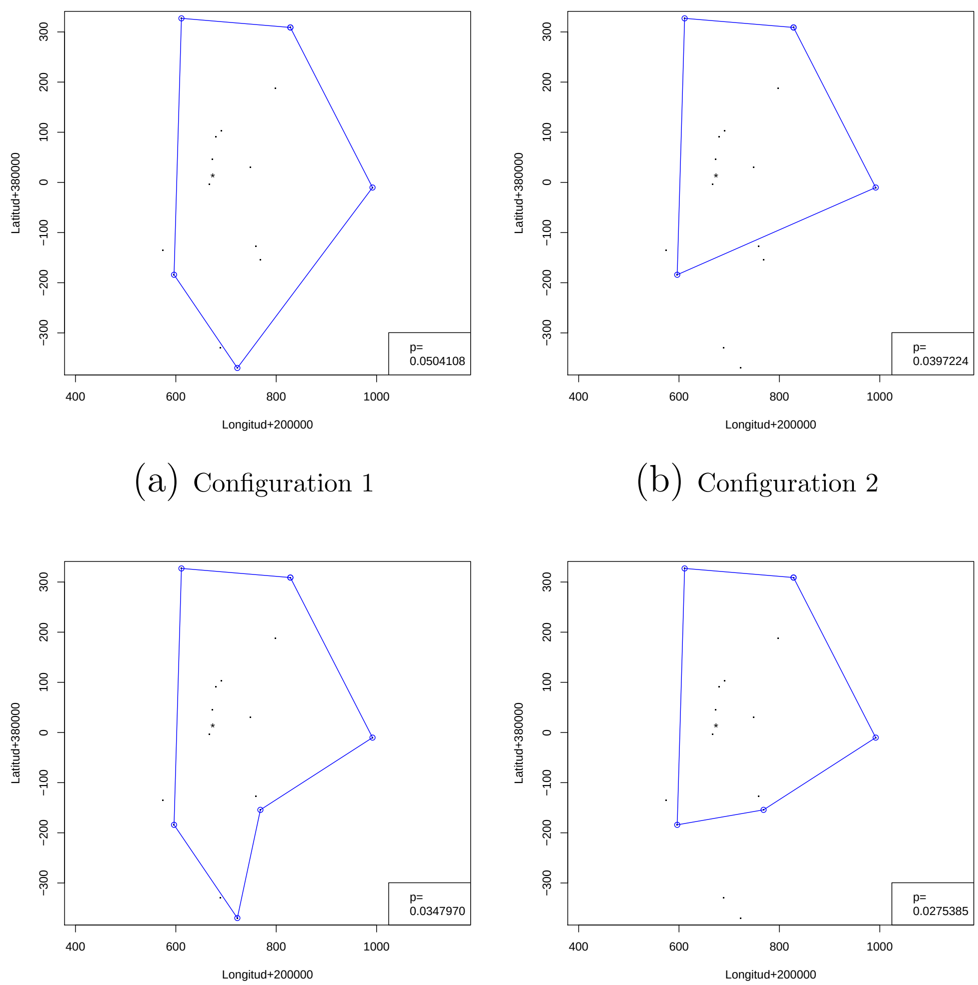


(c) Configuration 3



(d) Configuration 4

Figure D.6: Reconstructions of the territory Honeywell with higher frequency using the Adjusted Ordinal Model.



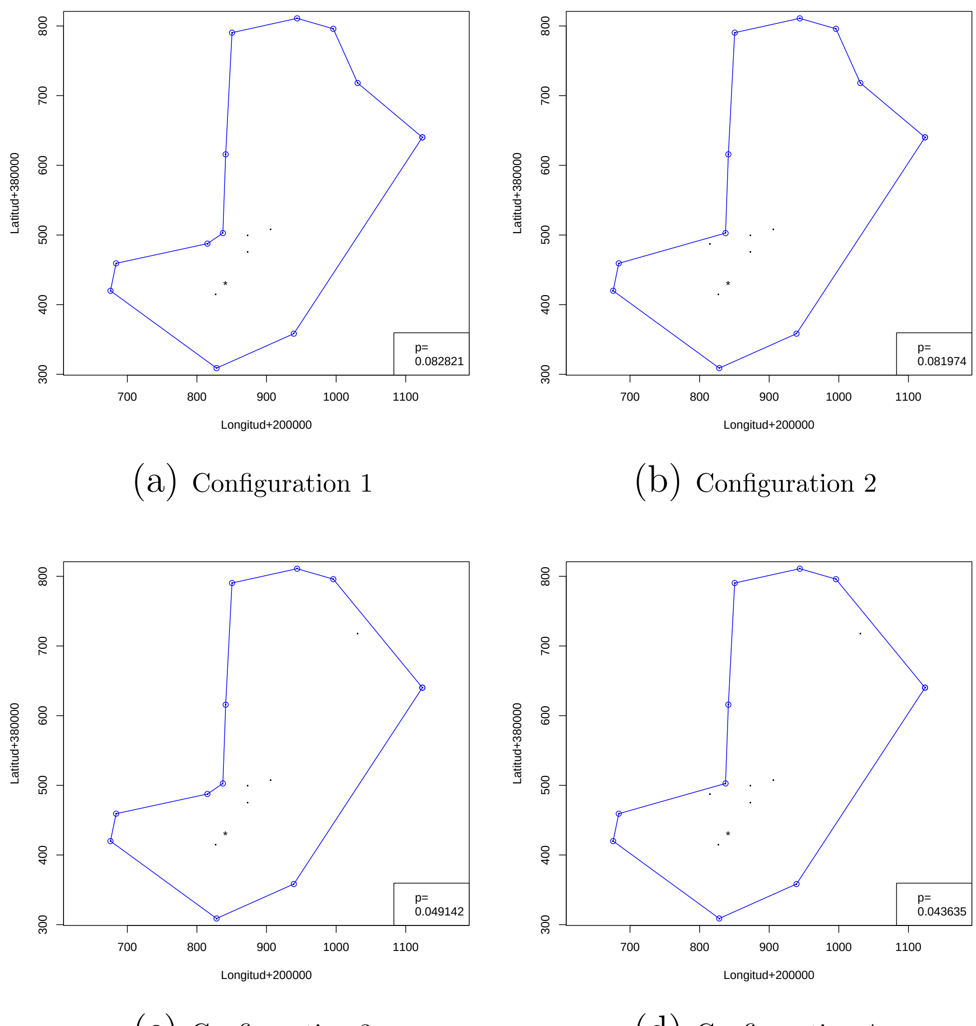
(a) Configuration 1

(b) Configuration 2

(c) Configuration 3

(d) Configuration 4

Figure D.7: Reconstructions of the territory Jacks with higher frequency using the Adjusted Ordinal Model.



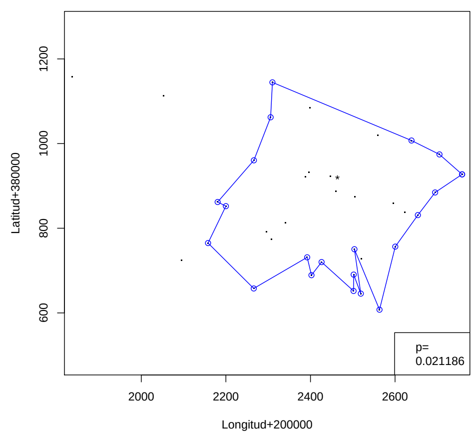
(a) Configuration 1

(b) Configuration 2

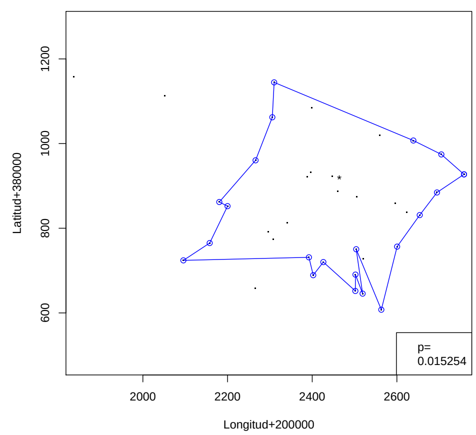
(c) Configuration 3

(d) Configuration 4

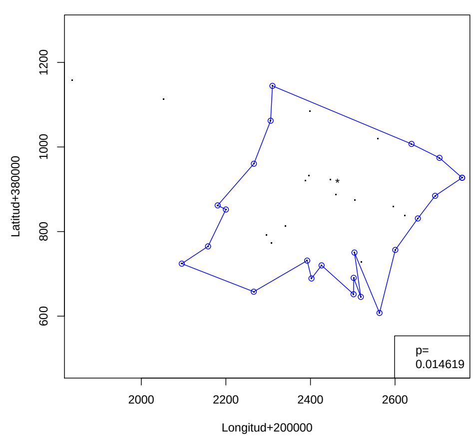
Figure D.8: Reconstructions of the territory Junction with higher frequency using the Adjusted Ordinal Model.



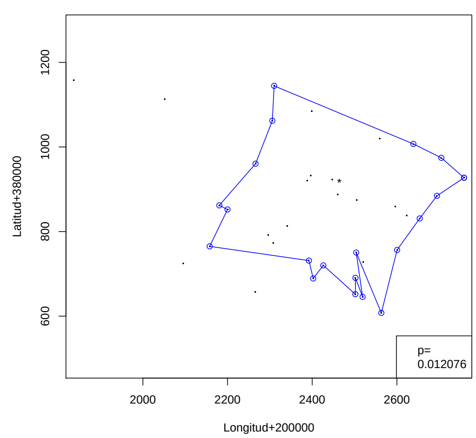
(a) Configuration 1



(b) Configuration 2

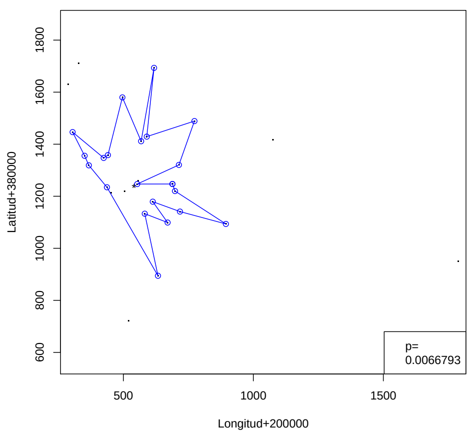


(c) Configuration 3

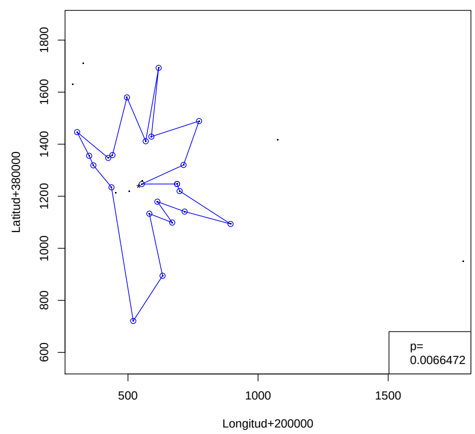


(d) Configuration 4

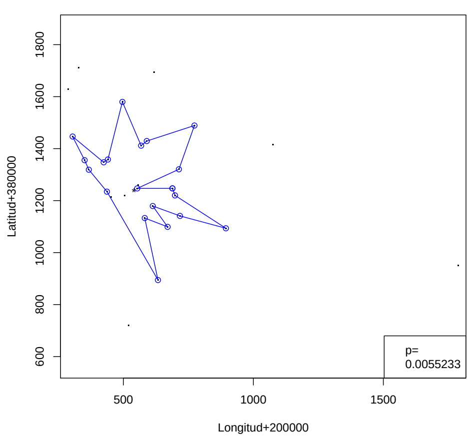
Figure D.9: Reconstructions of the territory Kennel with higher frequency using the Adjusted Ordinal Model.



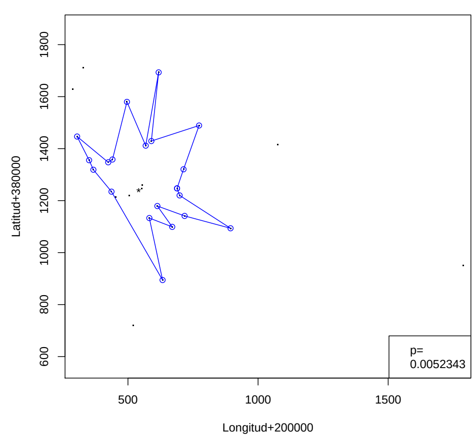
(a) Configuration 1



(b) Configuration 2

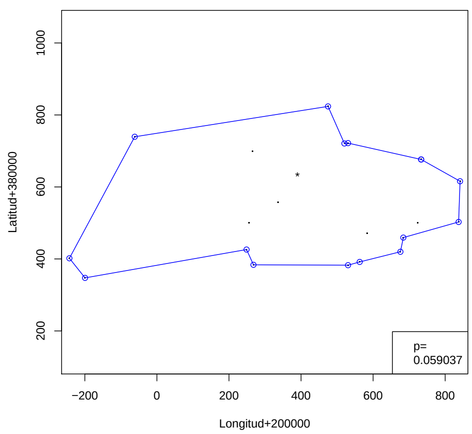


(c) Configuration 3

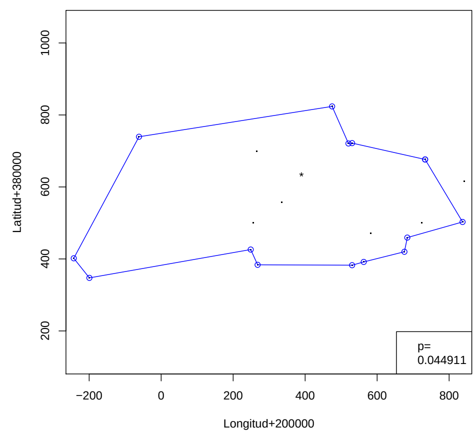


(d) Configuration 4

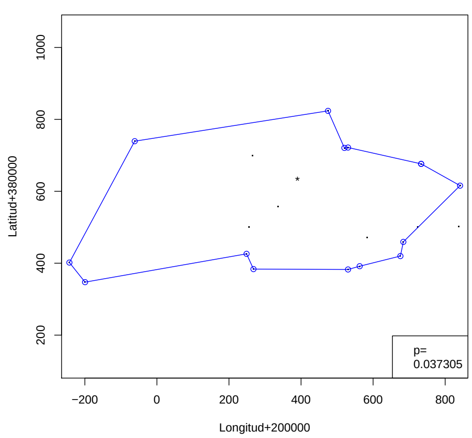
Figure D.10: Reconstructions of the territory Larch with higher frequency using the Adjusted Ordinal Model.



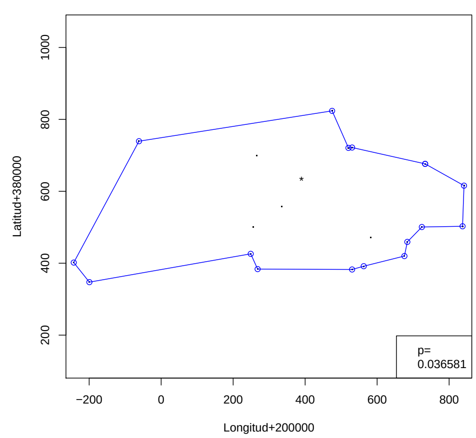
(a) Configuration 1



(b) Configuration 2

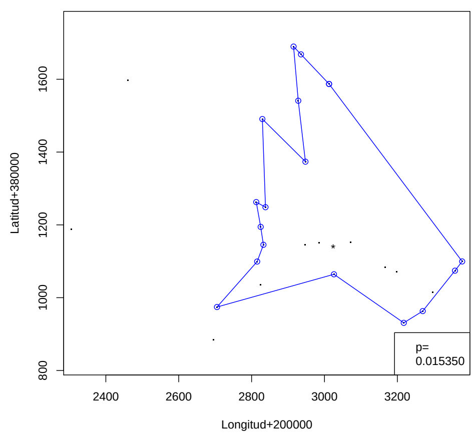


(c) Configuration 3

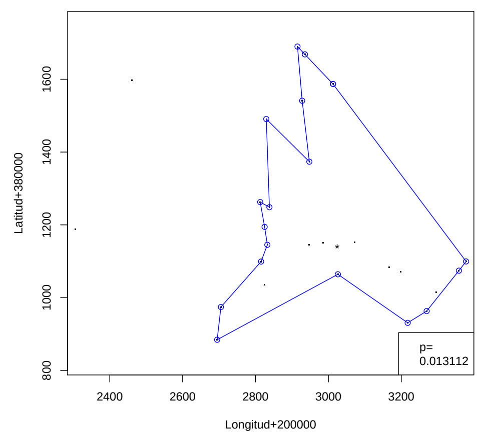


(d) Configuration 4

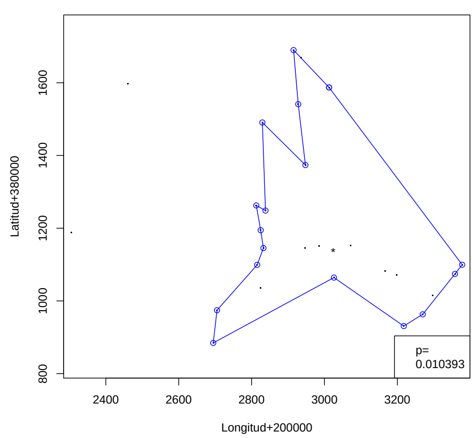
Figure D.11: Reconstructions of the territory Mead with higher frequency using the Adjusted Ordinal Model.



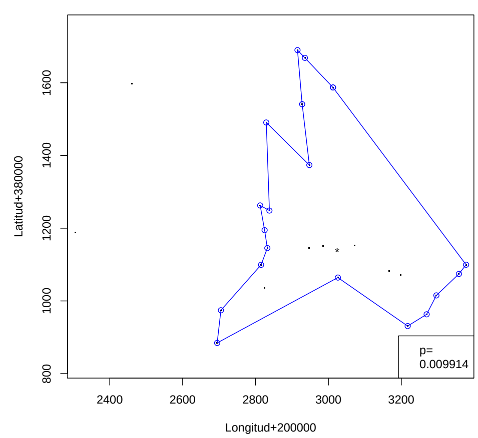
(a) Configuration 1



(b) Configuration 2

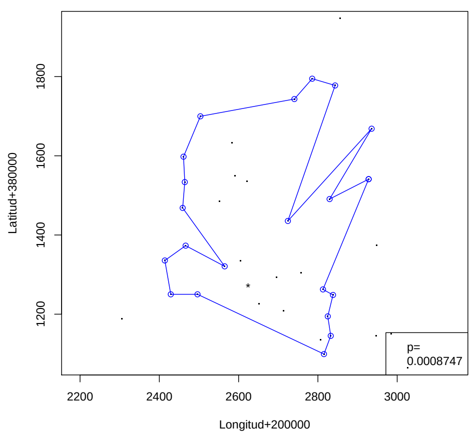


(c) Configuration 3

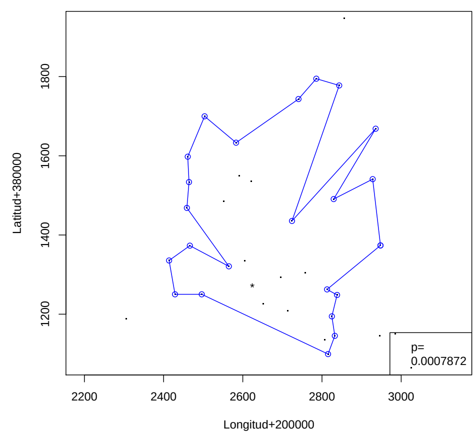


(d) Configuration 4

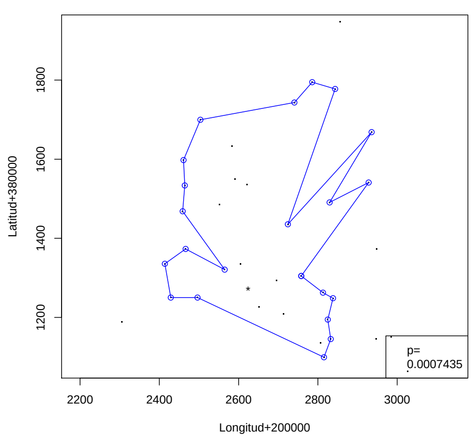
Figure D.12: Reconstructions of the territory Nettle with higher frequency using the Adjusted Ordinal Model.



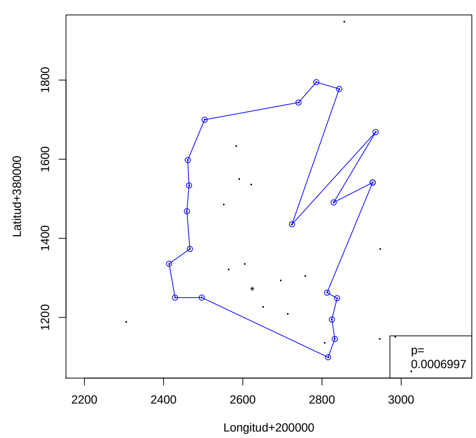
(a) Configuration 1



(b) Configuration 2

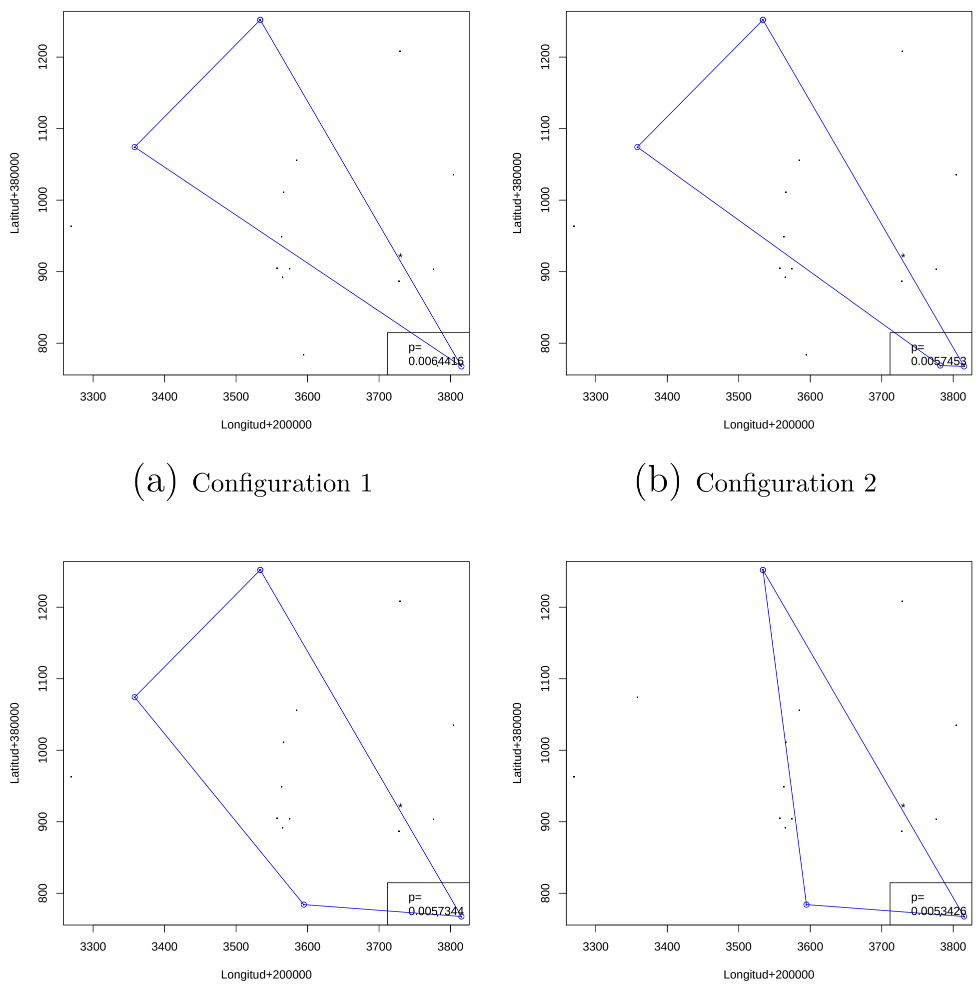


(c) Configuration 3



(d) Configuration 4

Figure D.13: Reconstructions of the territory Old Oak with higher frequency using the Adjusted Ordinal Model.



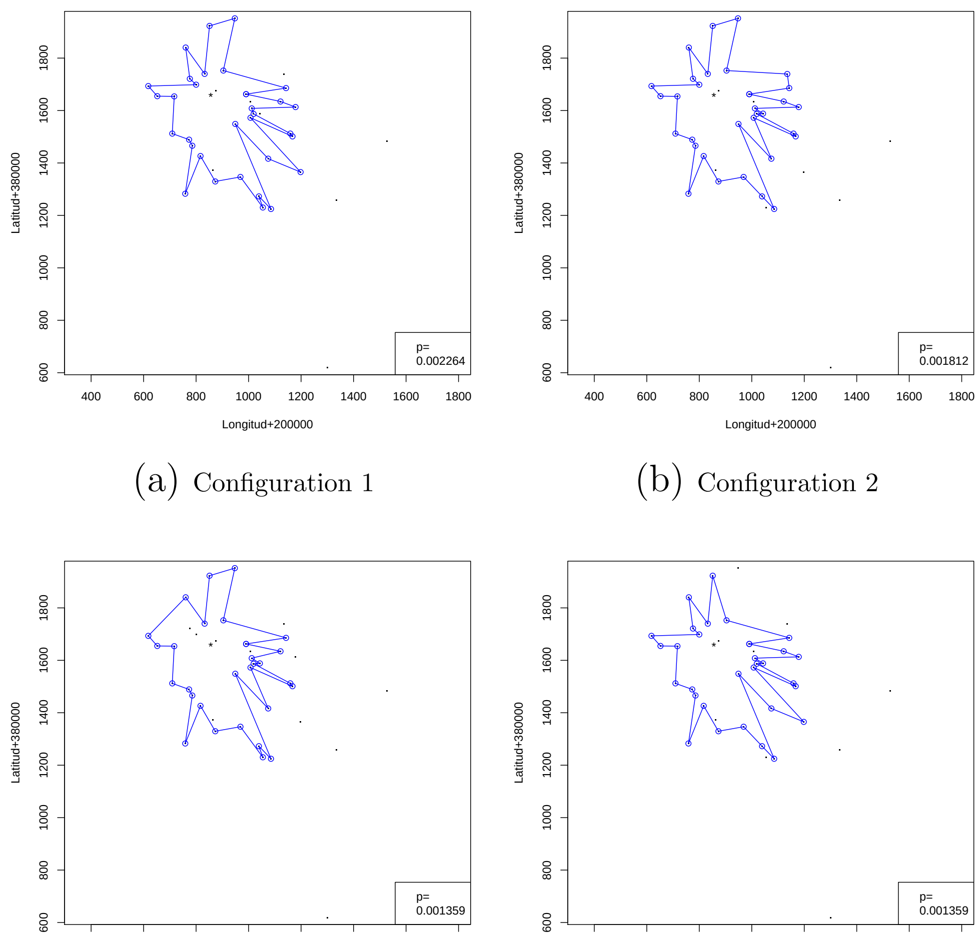
(a) Configuration 1

(b) Configuration 2

(c) Configuration 3

(d) Configuration 4

Figure D.14: Reconstructions of the territory Parkmill with higher frequency using the Adjusted Ordinal Model.



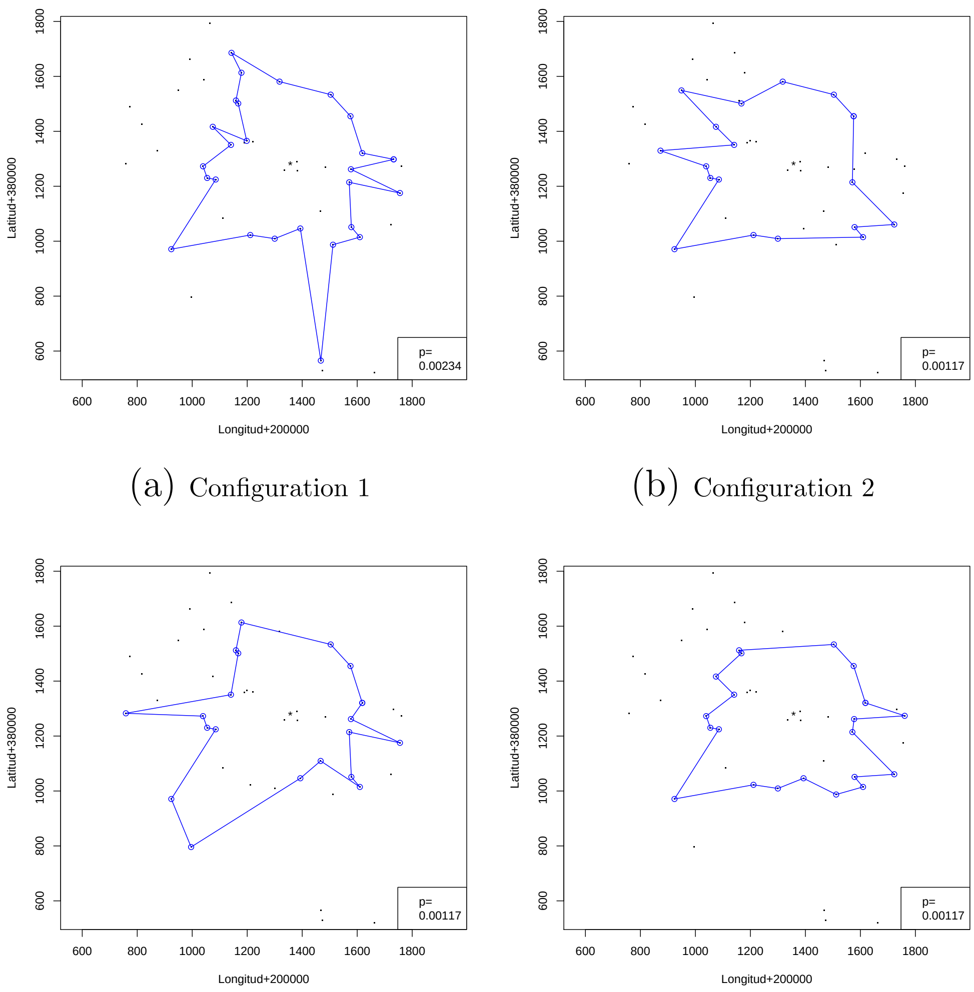
(a) Configuration 1

(b) Configuration 2

(c) Configuration 3

(d) Configuration 4

Figure D.15: Reconstructions of the territory Rabbit with higher frequency using the Adjusted Ordinal Model.



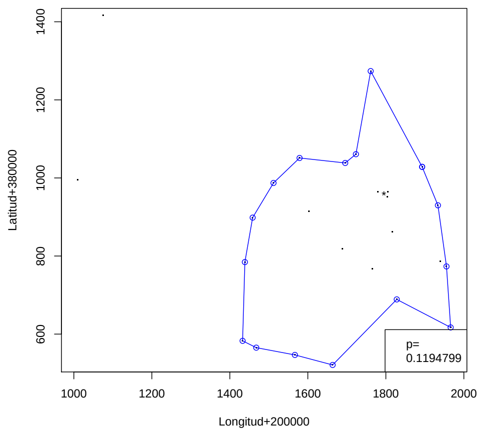
(a) Configuration 1

(b) Configuration 2

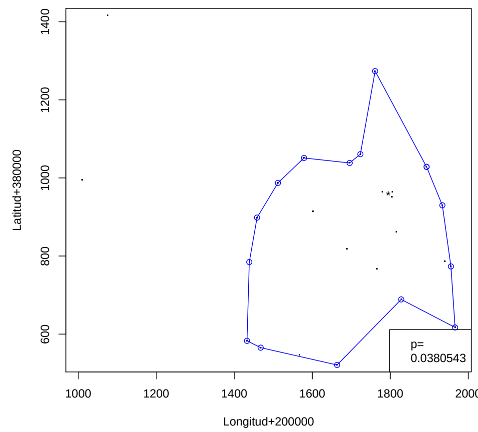
(c) Configuration 3

(d) Configuration 4

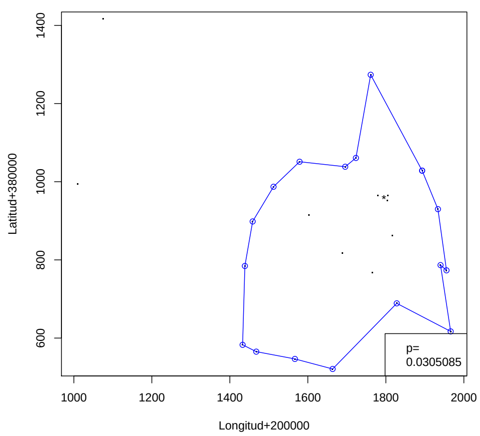
Figure D.16: Reconstructions of the territory Septic Tank with higher frequency using the Adjusted Ordinal Model.



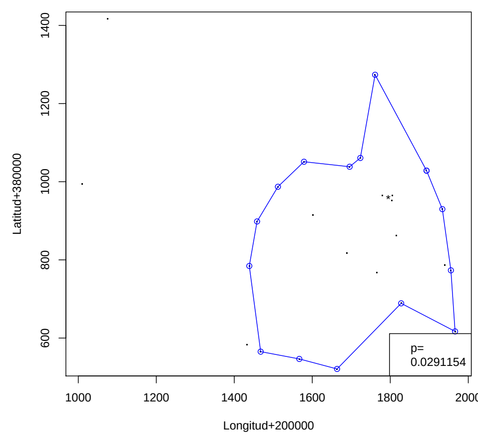
(a) Configuration 1



(b) Configuration 2

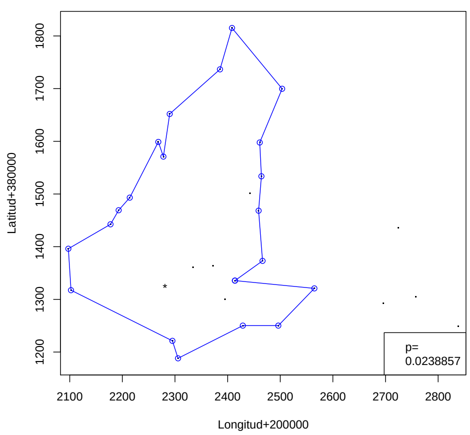


(c) Configuration 3

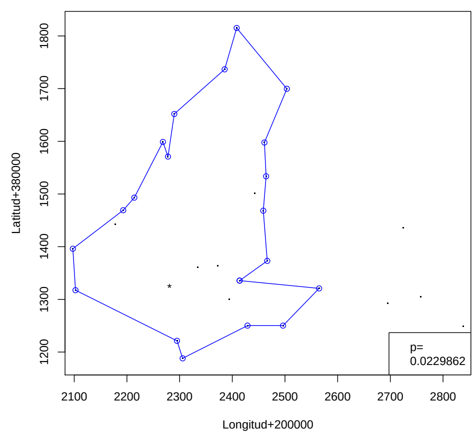


(d) Configuration 4

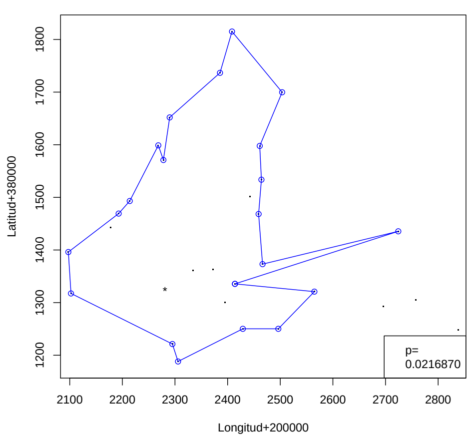
Figure D.17: Reconstructions of the territory Top with higher frequency using the Adjusted Ordinal Model.



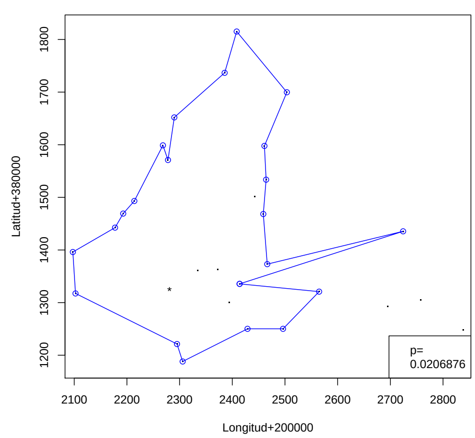
(a) Configuration 1



(b) Configuration 2

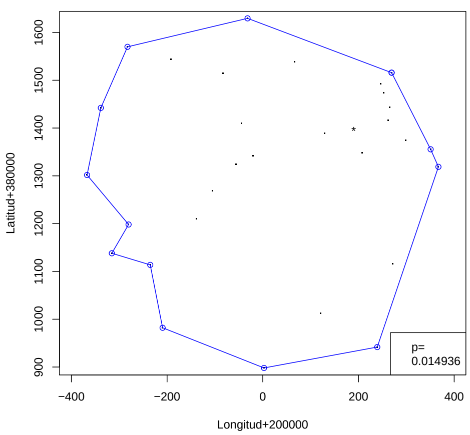


(c) Configuration 3

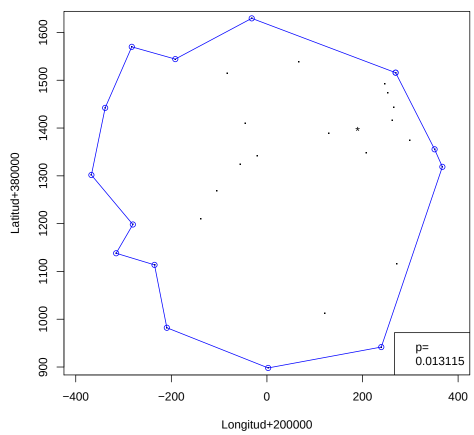


(d) Configuration 4

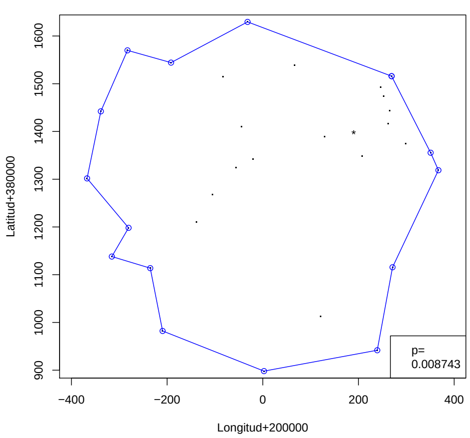
Figure D.18: Reconstructions of the territory Trackside with higher frequency using the Adjusted Ordinal Model.



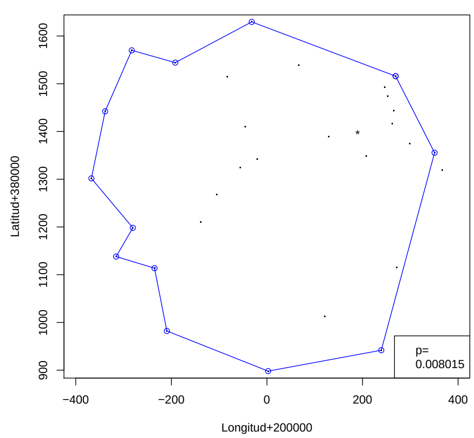
(a) Configuration 1



(b) Configuration 2

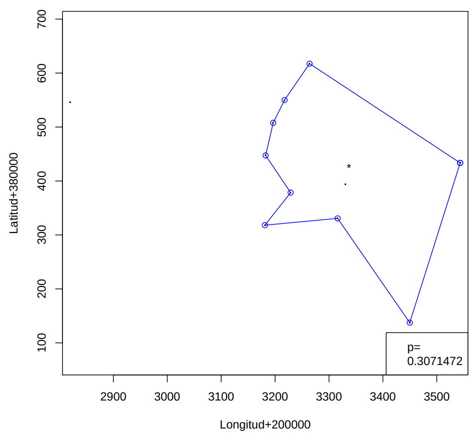


(c) Configuration 3

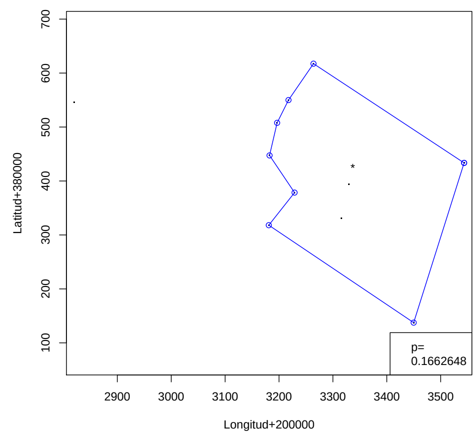


(d) Configuration 4

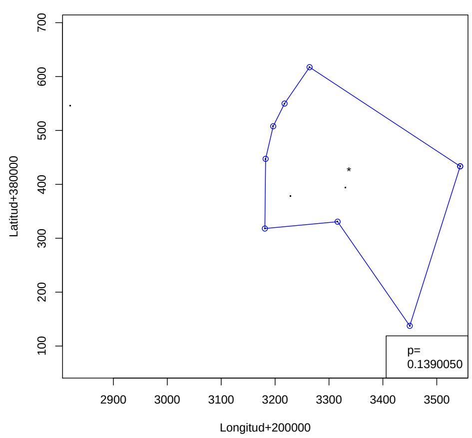
Figure D.19: Reconstructions of the territory West with higher frequency using the Adjusted Ordinal Model.



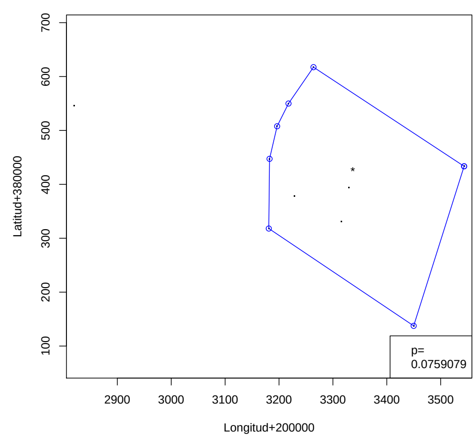
(a) Configuration 1



(b) Configuration 2

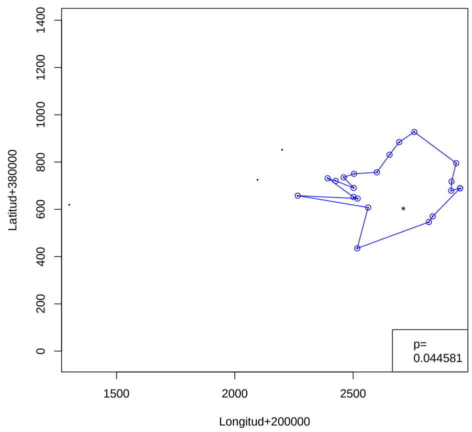


(c) Configuration 3

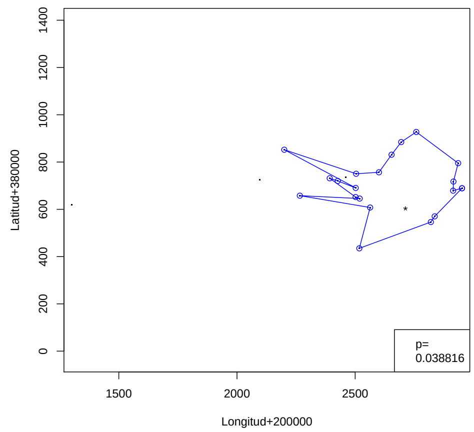


(d) Configuration 4

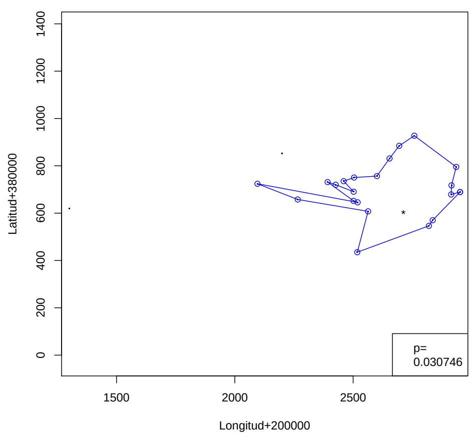
Figure D.20: Reconstructions of the territory Wood Farm with higher frequency using the Adjusted Ordinal Model.



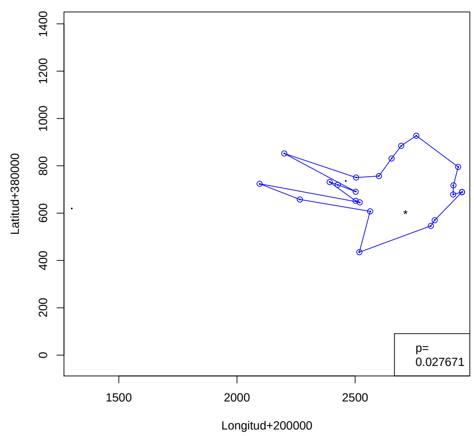
(a) Configuration 1



(b) Configuration 2

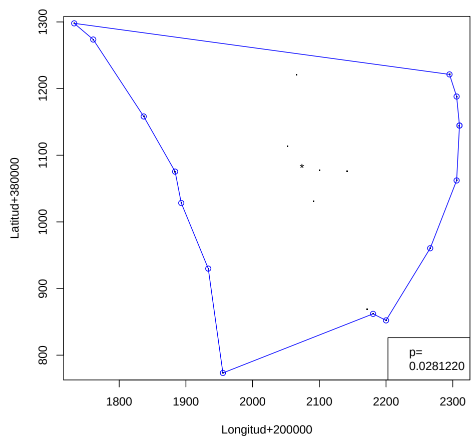


(c) Configuration 3

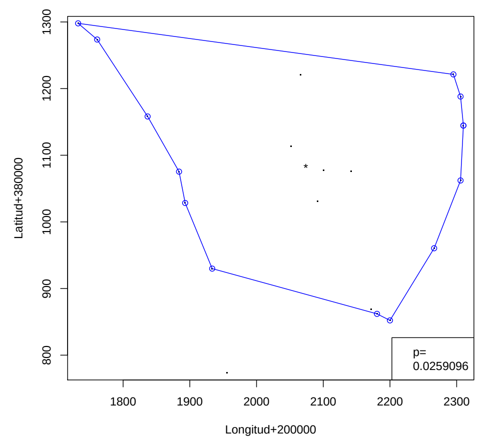


(d) Configuration 4

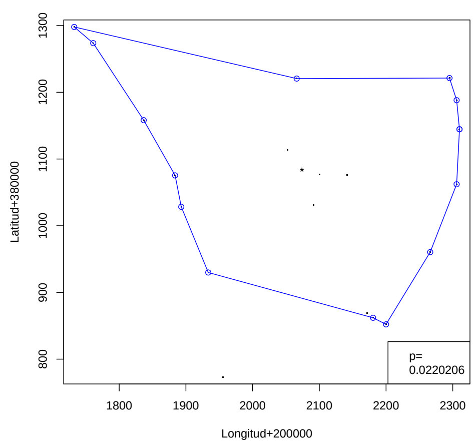
Figure D.21: Reconstructions of the territory Woodrush with higher frequency using the Adjusted Ordinal Model.



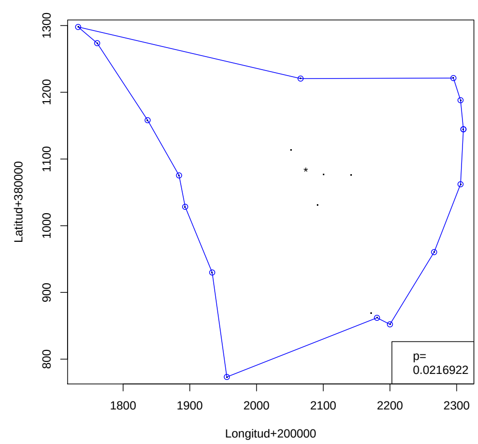
(a) Configuration 1



(b) Configuration 2

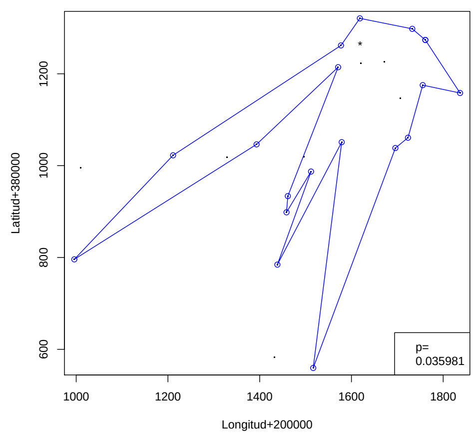


(c) Configuration 3

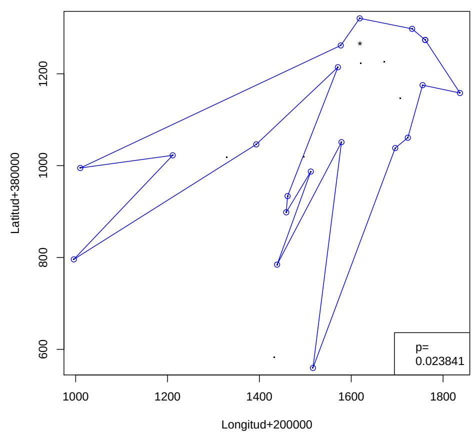


(d) Configuration 4

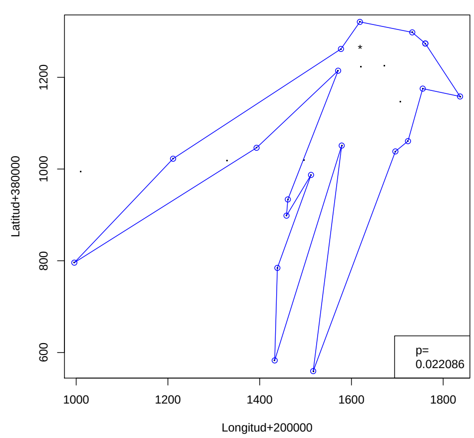
Figure D.22: Reconstructions of the territory Wych Elm with higher frequency using the Adjusted Ordinal Model.



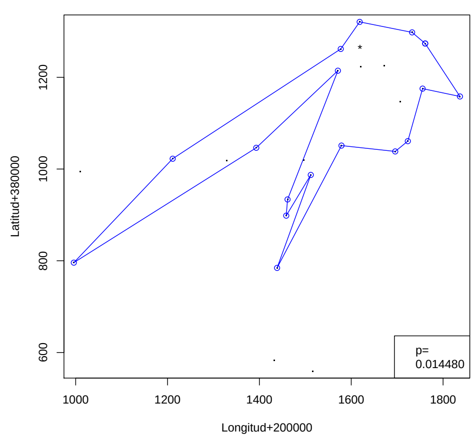
(a) Configuration 1



(b) Configuration 2



(c) Configuration 3



(d) Configuration 4

Figure D.23: Reconstructions of the territory Yew with higher frequency using the Adjusted Ordinal Model.

Appendix E

Appendix Adjusted Ordinal Model: Probability of Inclusion Figures

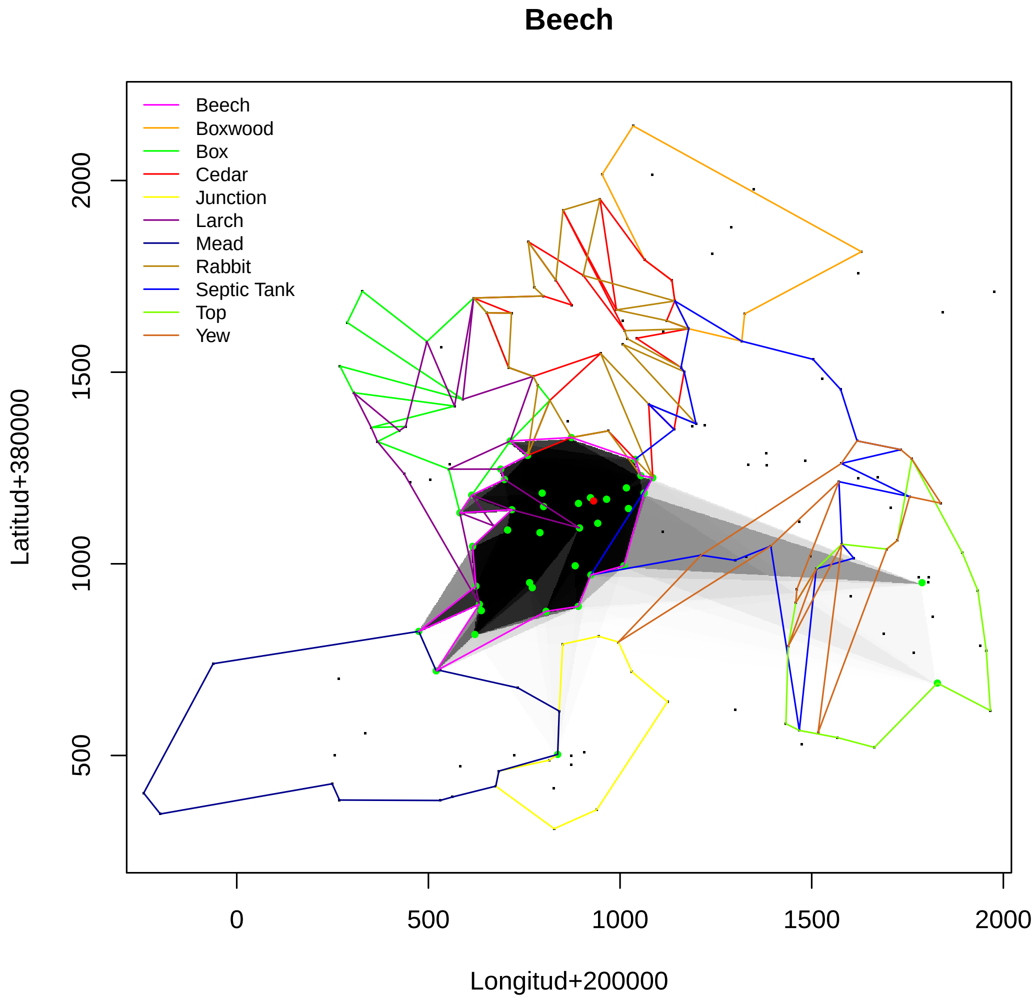


Figure E.1: Probability of inclusion of the Beech territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

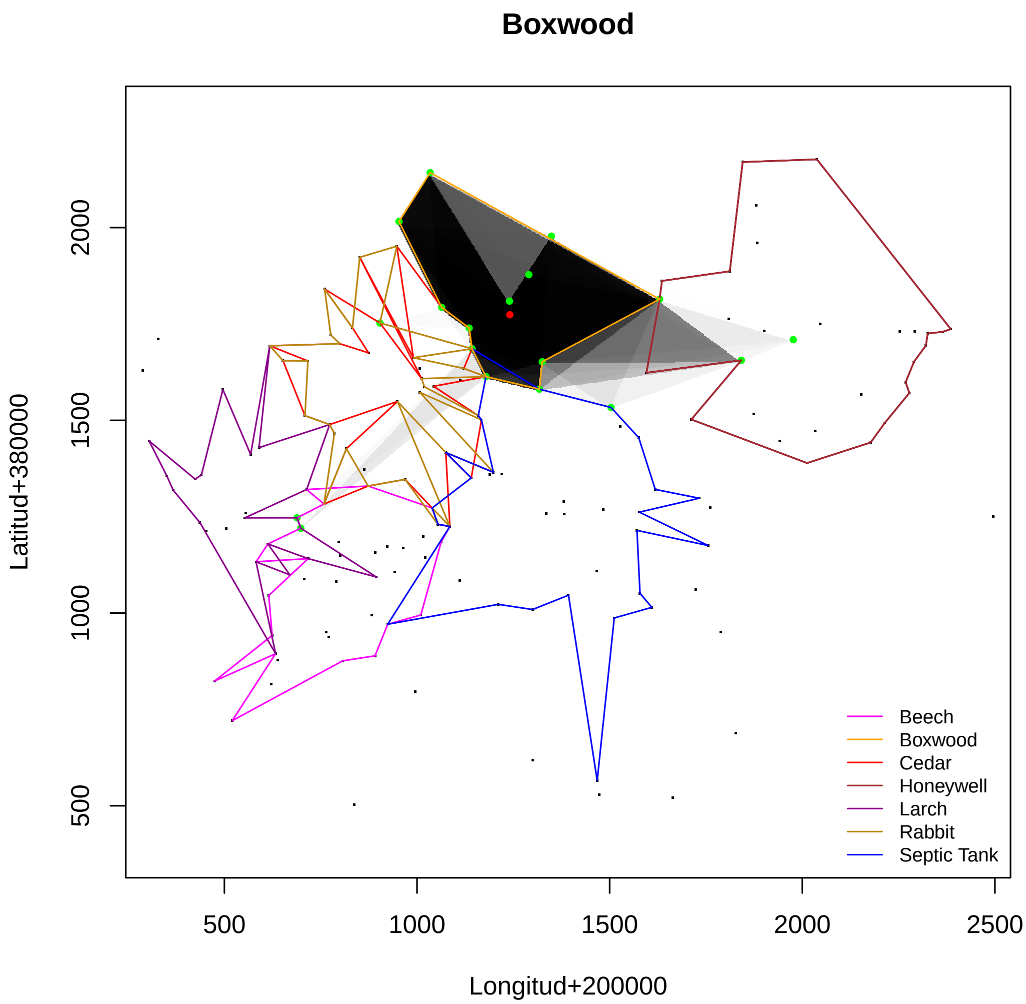


Figure E.2: Probability of inclusion of the Boxwood territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

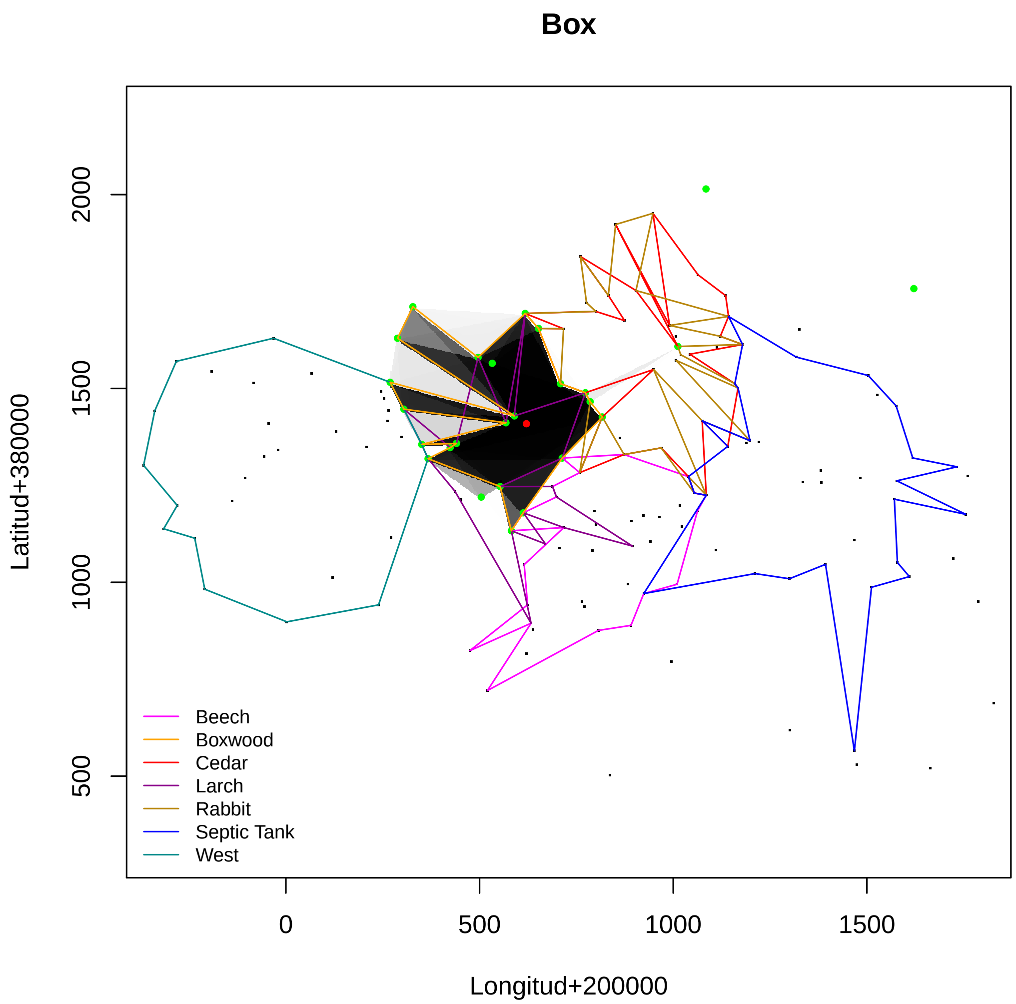


Figure E.3: Probability of inclusion of the Box territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

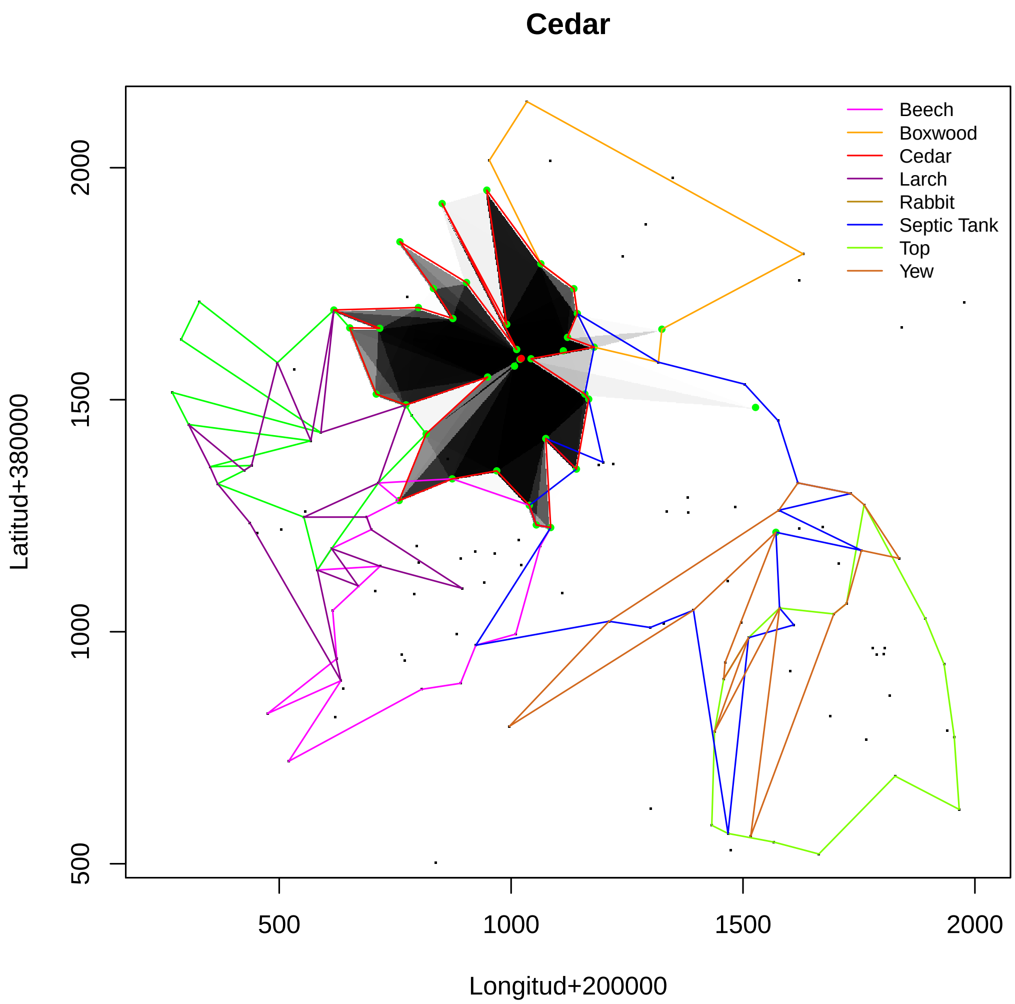


Figure E.4: Probability of inclusion of the Cedar territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

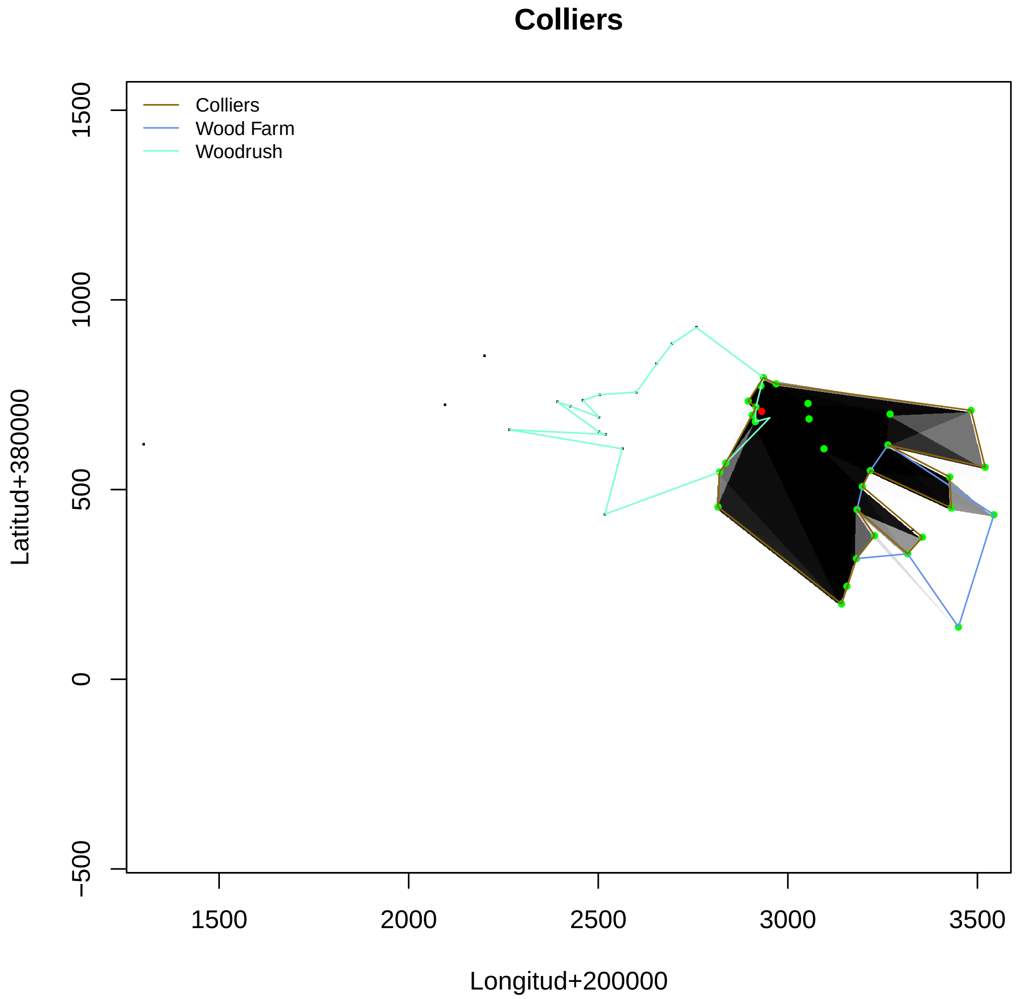


Figure E.5: Probability of inclusion of the Colliers territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

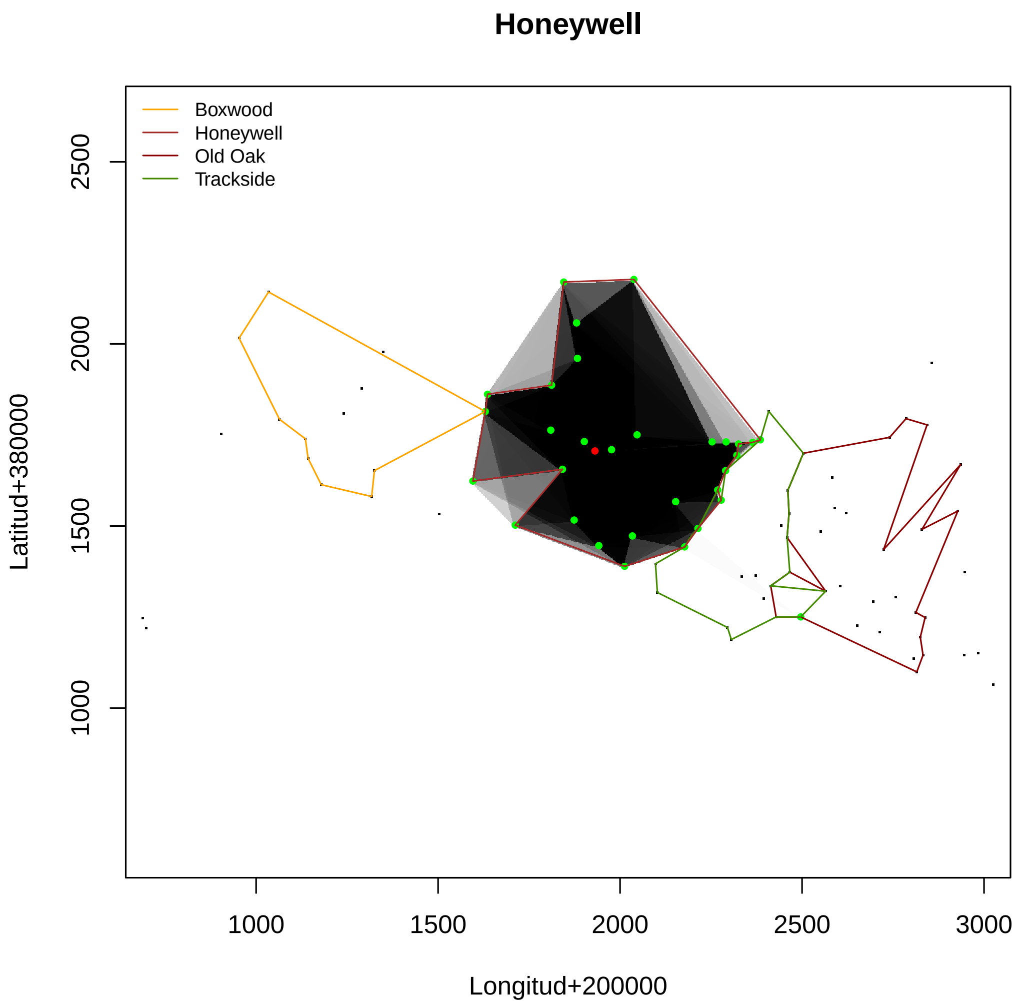


Figure E.6: Probability of inclusion of the Honeywell territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

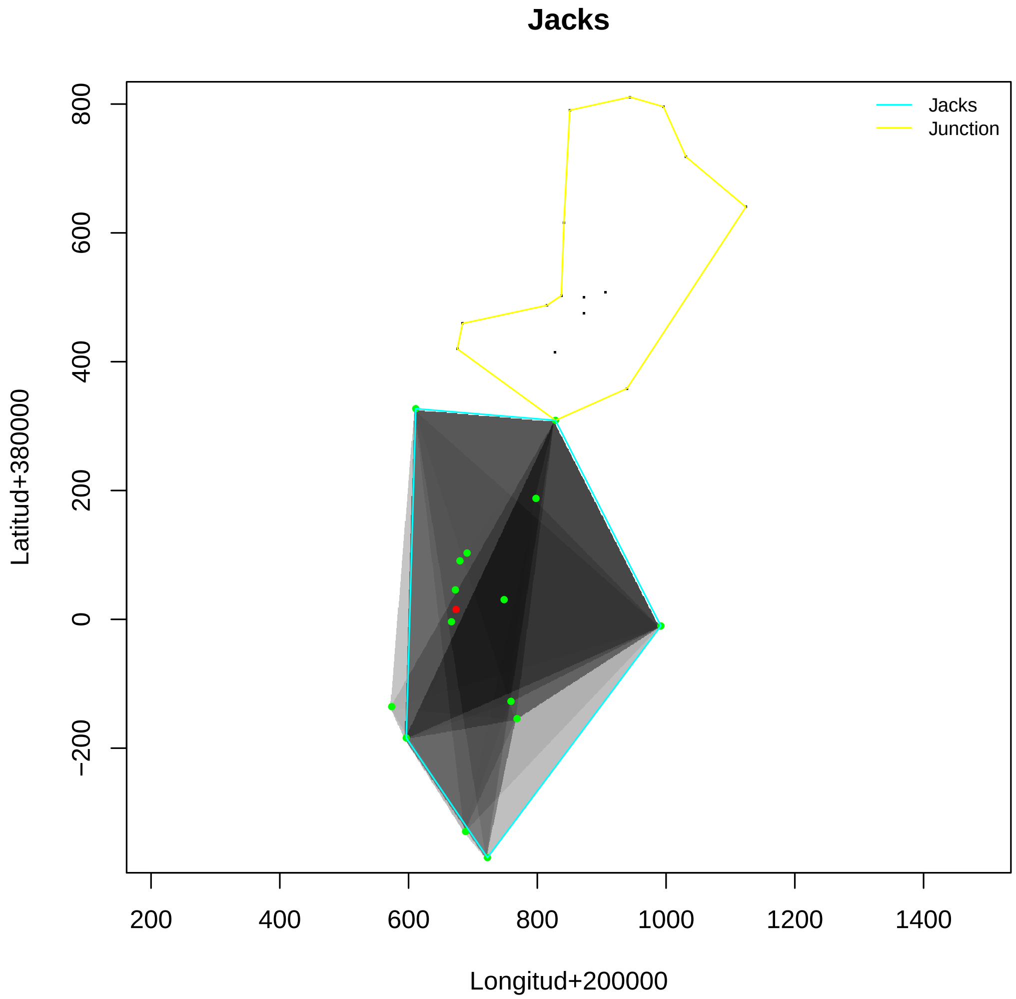


Figure E.7: Probability of inclusion of the Jacks territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

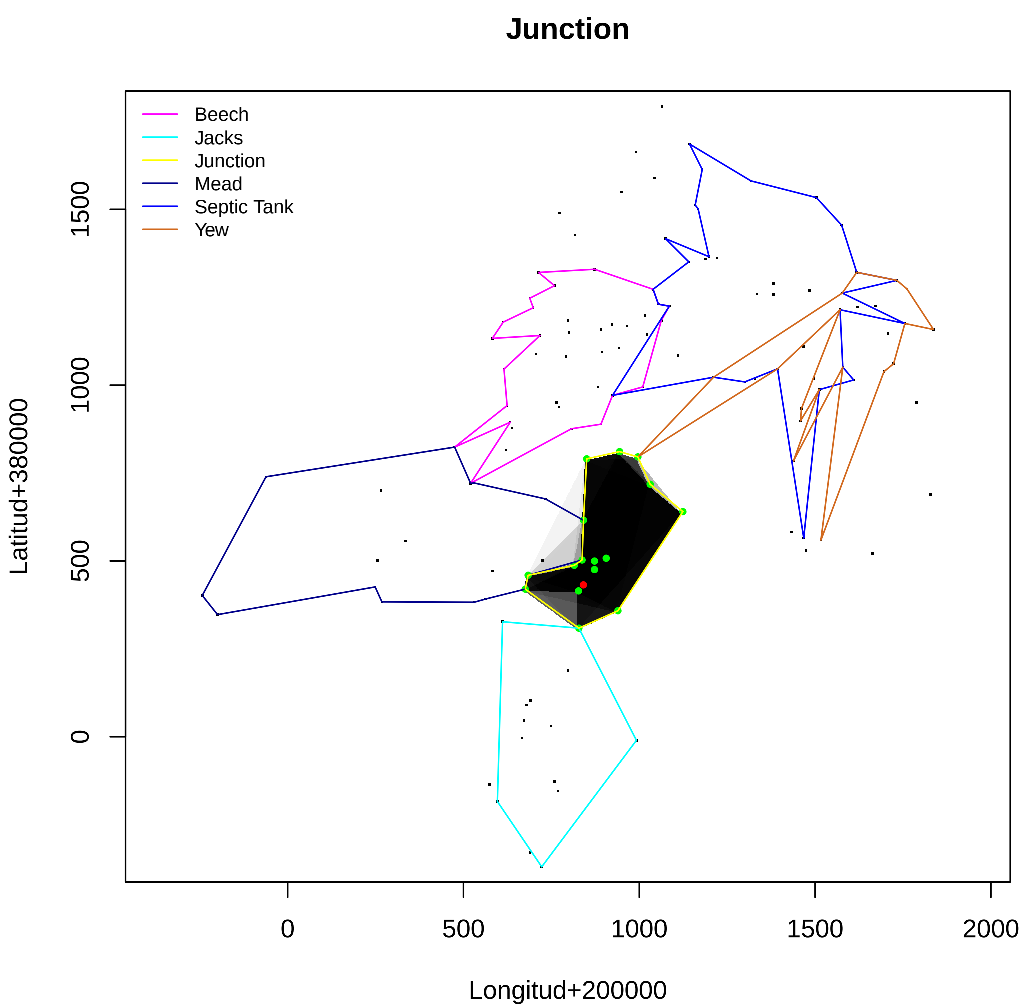


Figure E.8: Probability of inclusion of the Junction territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

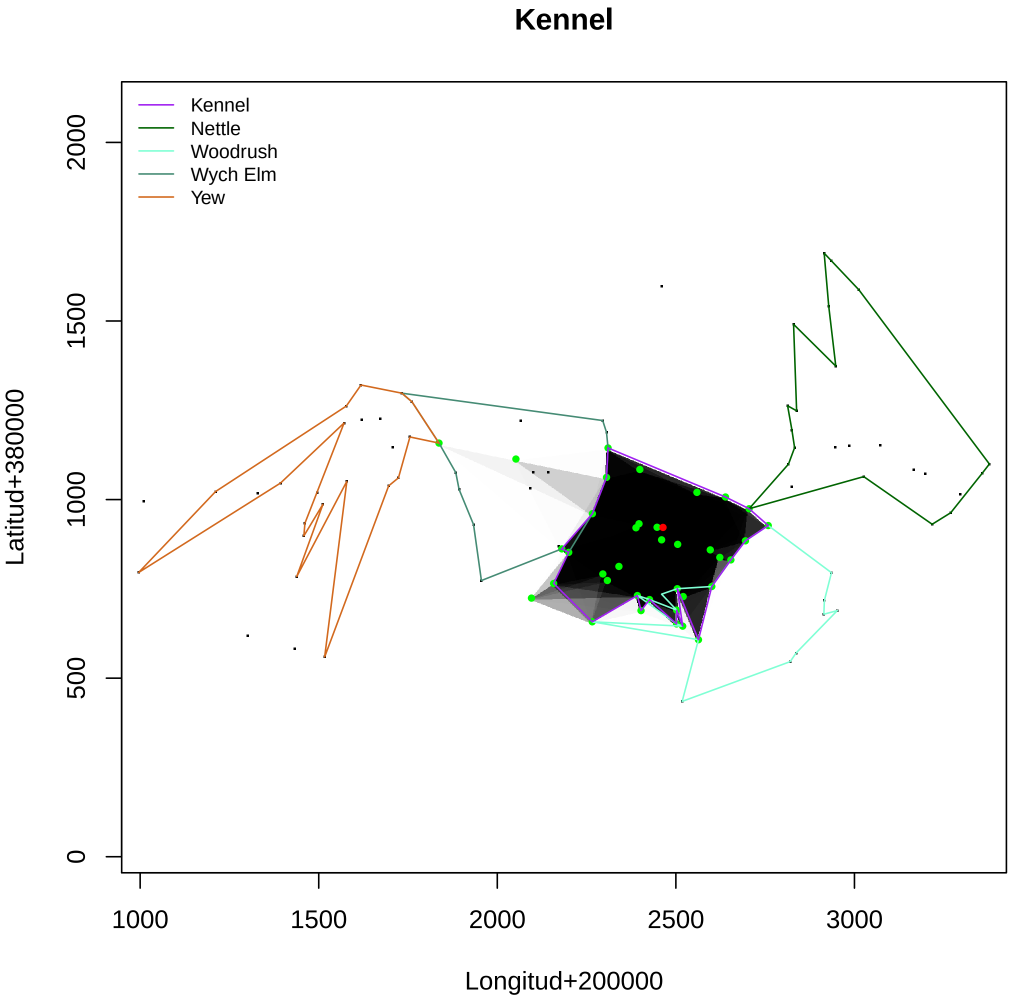


Figure E.9: Probability of inclusion of the Kennel territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

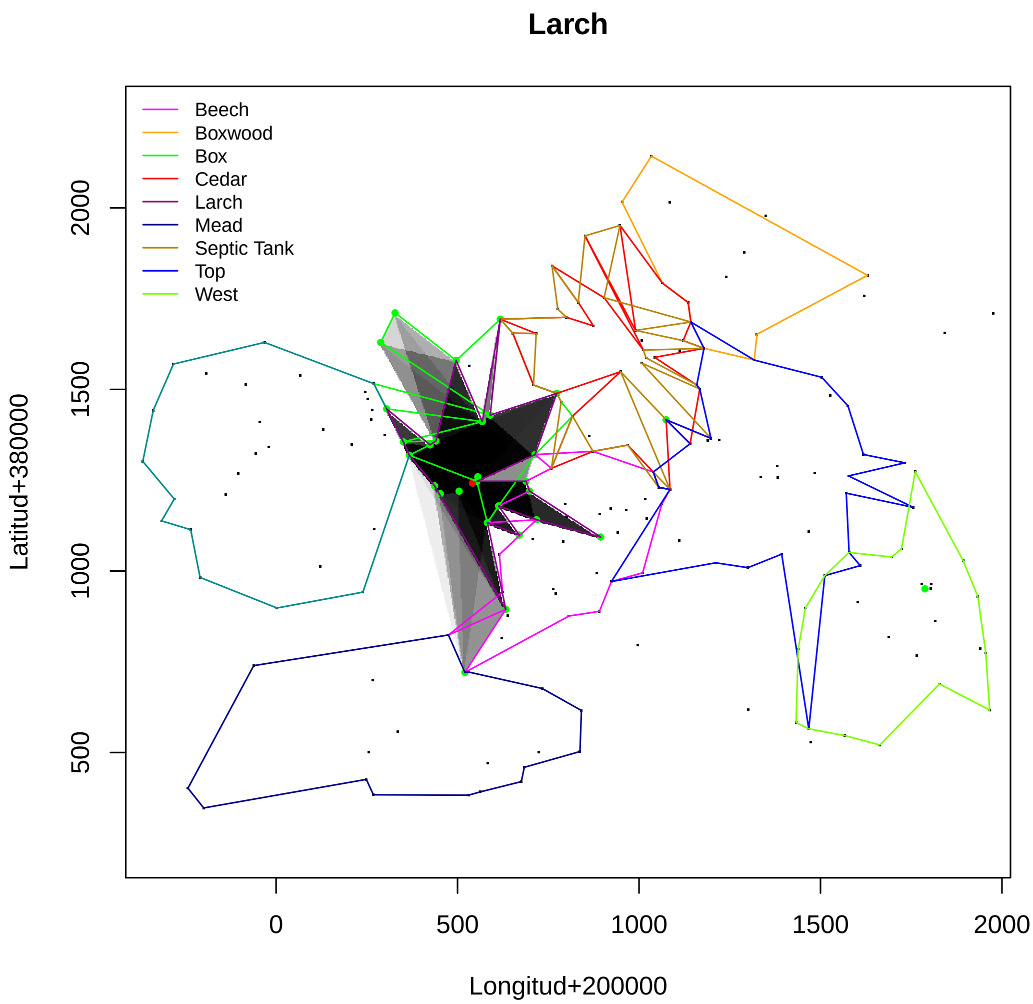


Figure E.10: Probability of inclusion of the Larch territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

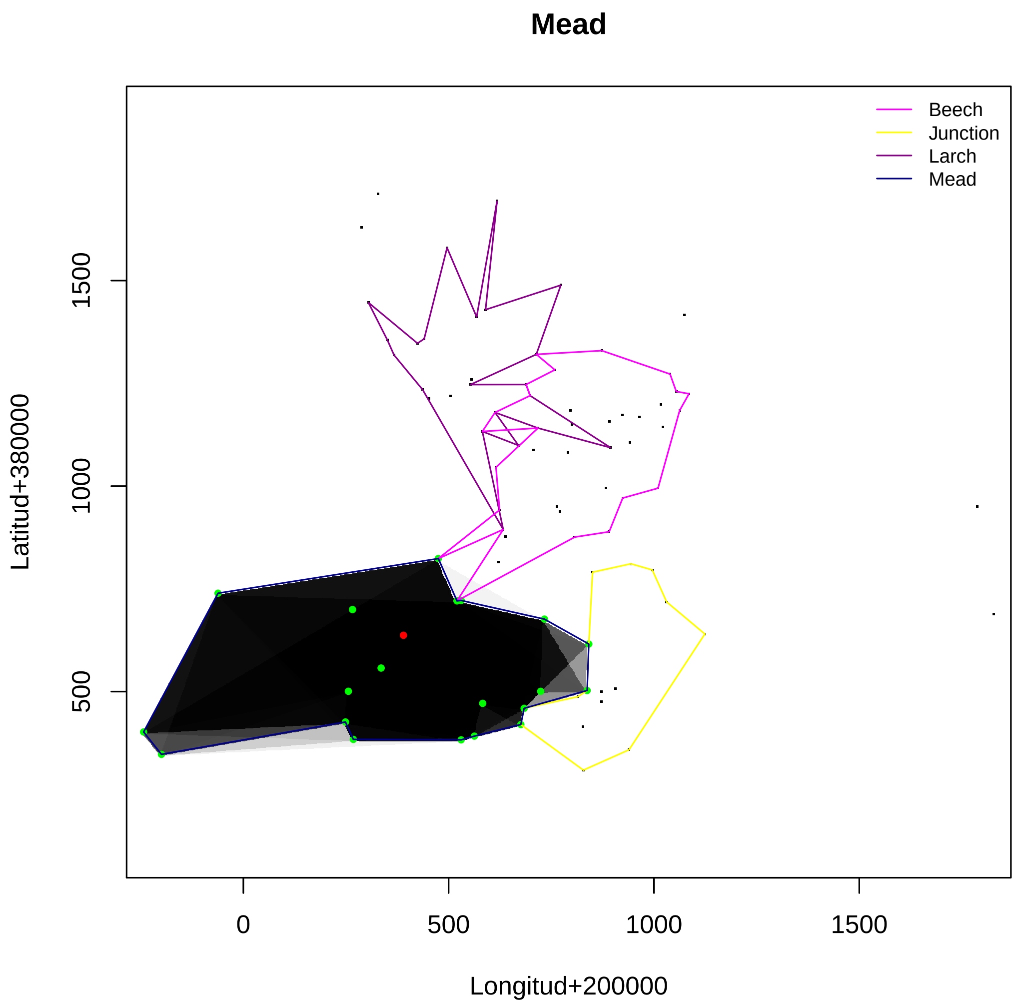


Figure E.11: Probability of inclusion of the Mead territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

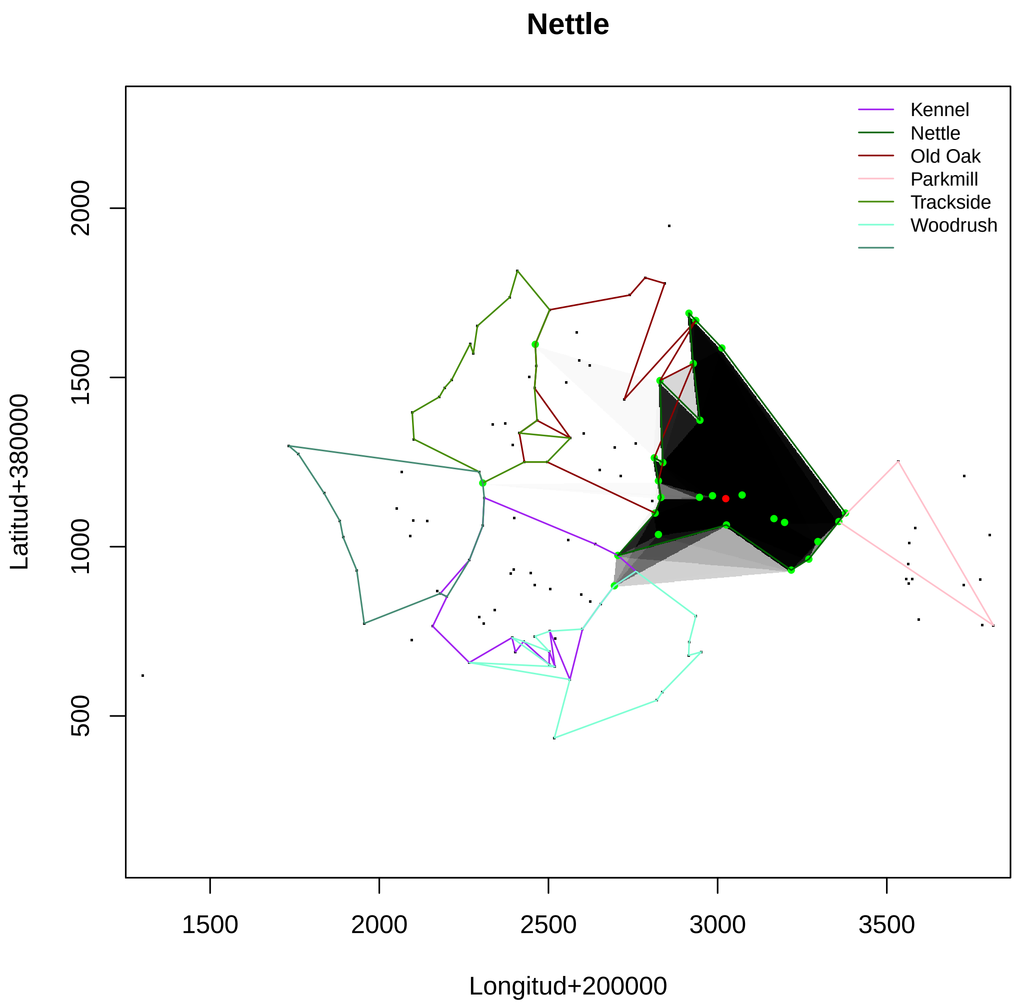


Figure E.12: Probability of inclusion of the Nettle territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

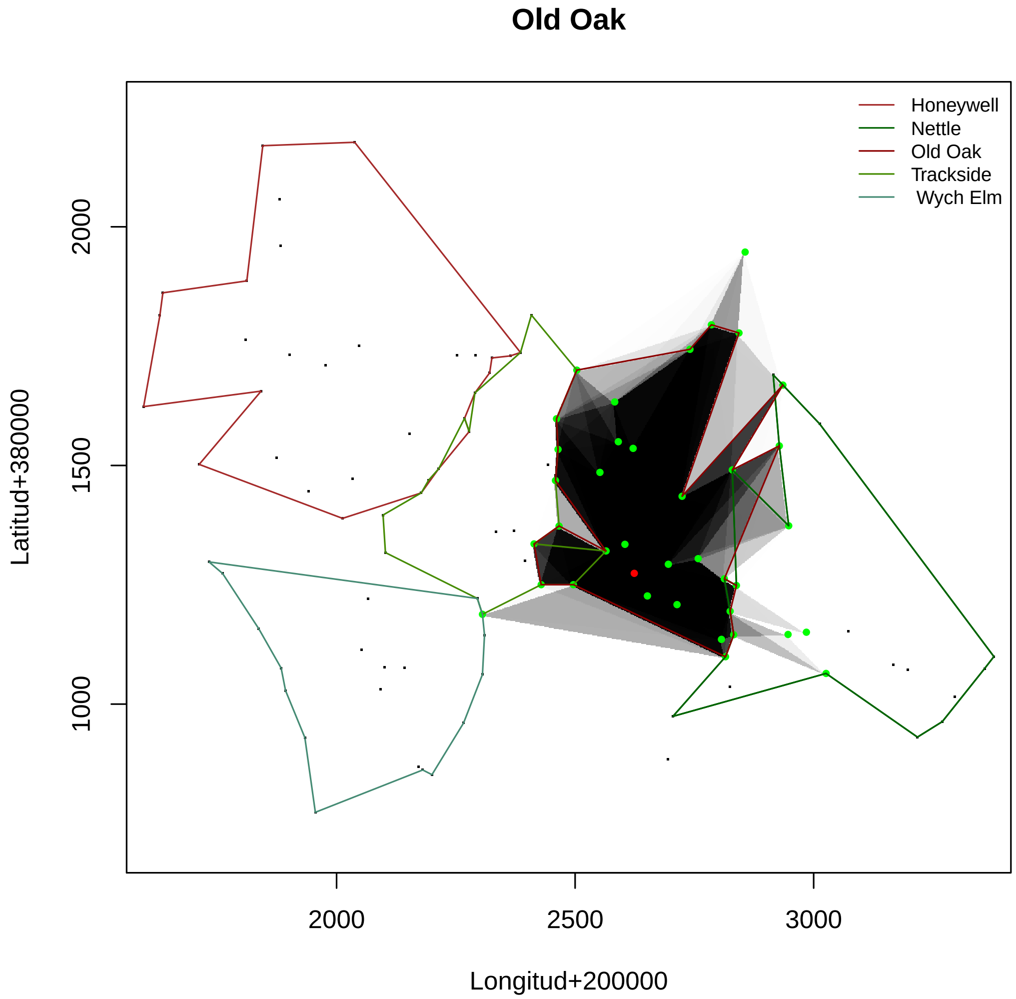


Figure E.13: Probability of inclusion of the Old Oak territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

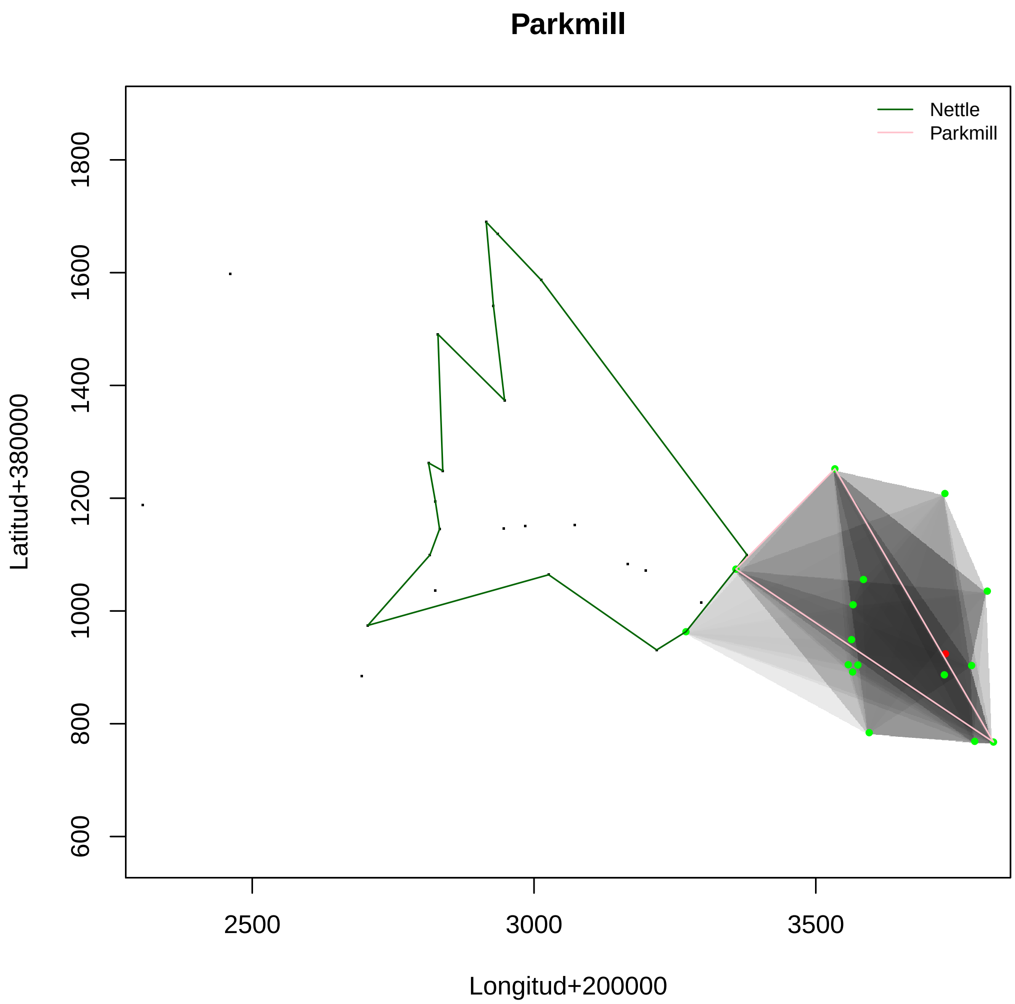


Figure E.14: Probability of inclusion of the Parkmill territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

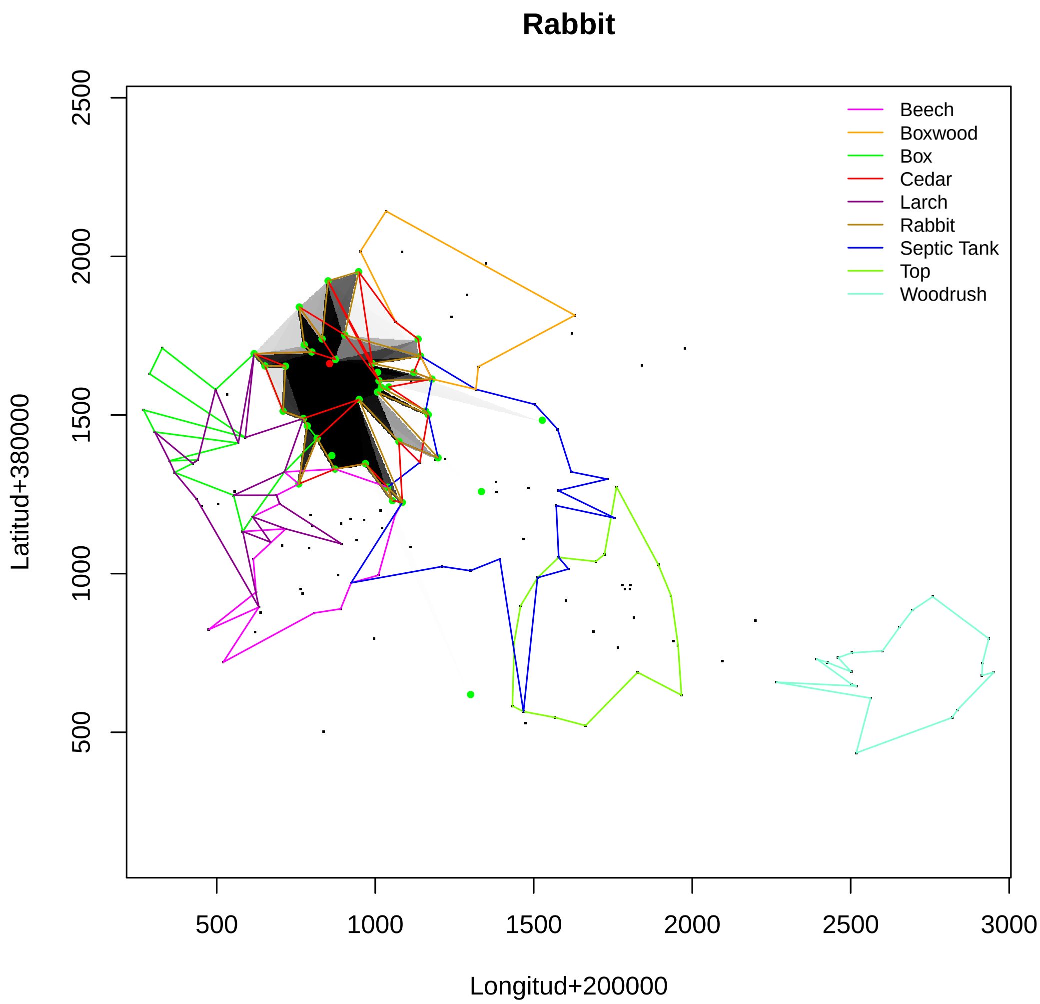


Figure E.15: Probability of inclusion of the Rabbit territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

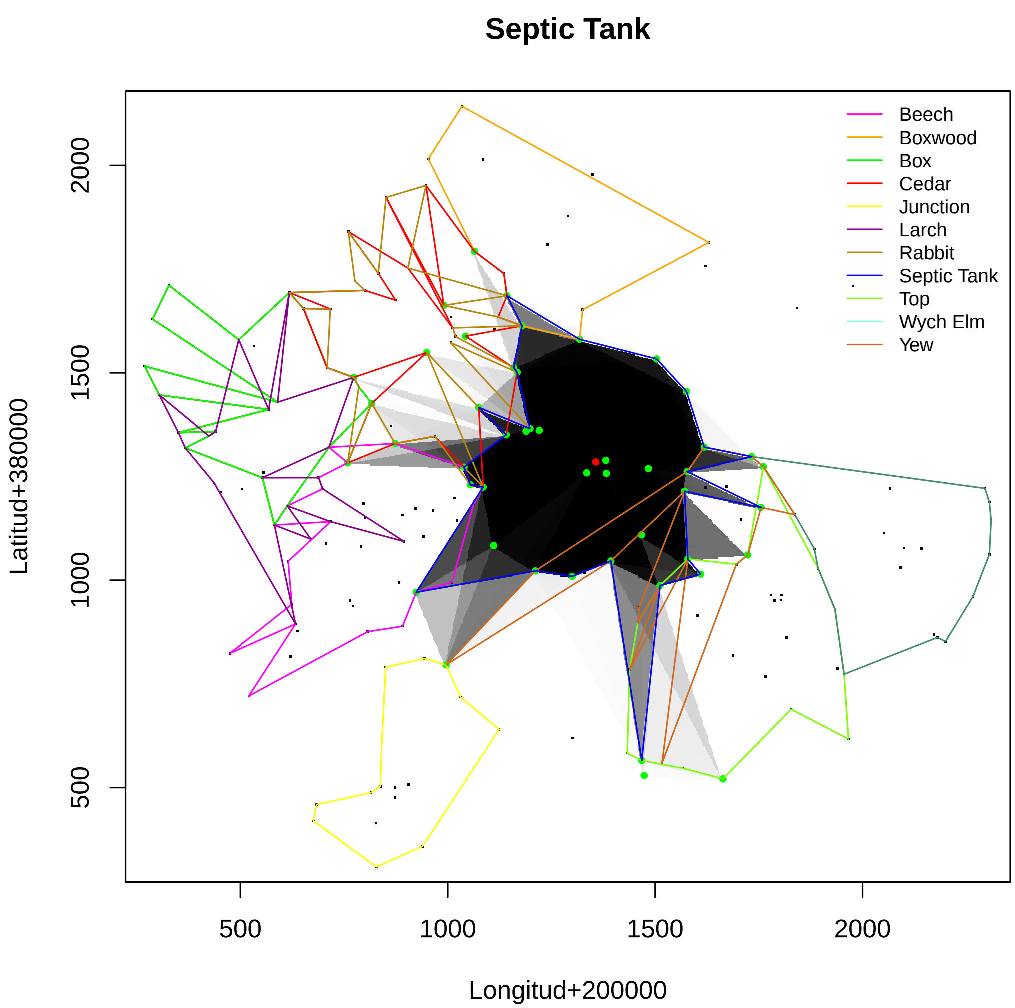


Figure E.16: Probability of inclusion of the Septic Tank territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

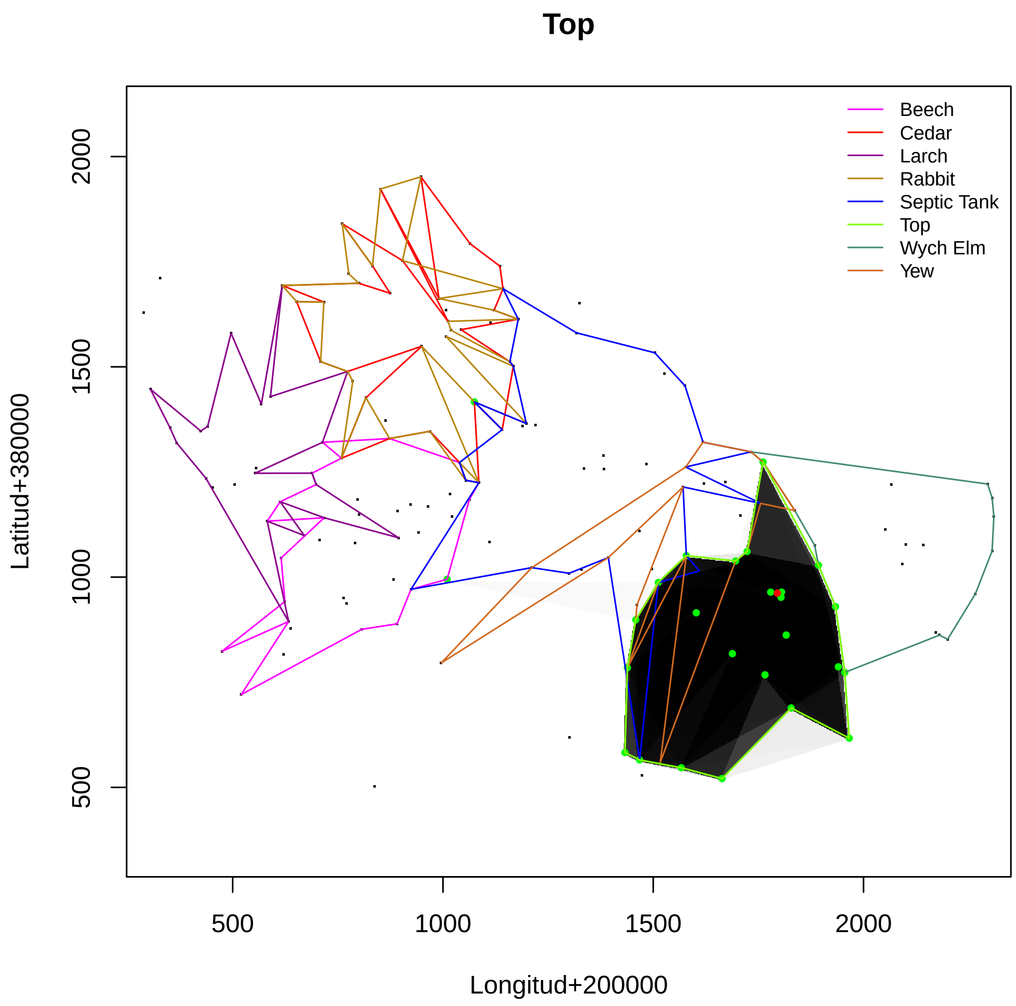


Figure E.17: Probability of inclusion of the Top territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

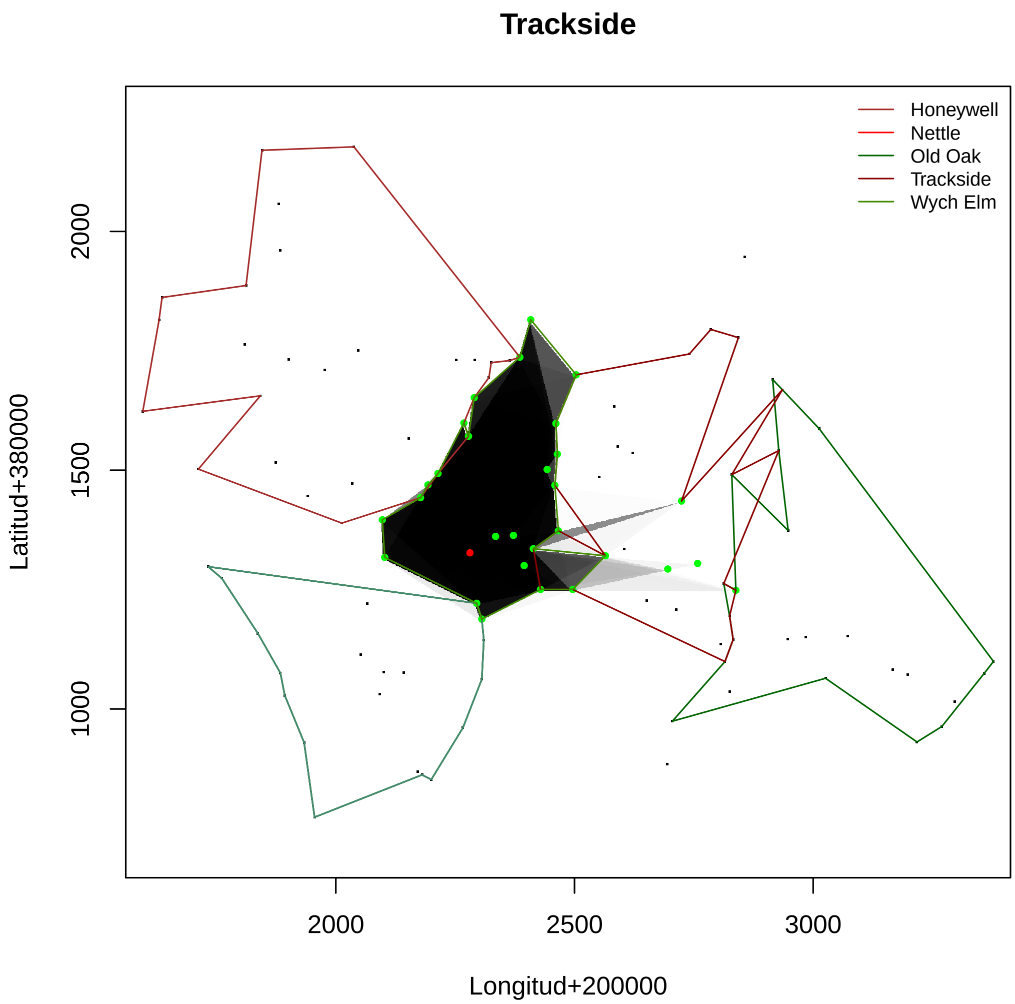


Figure E.18: Probability of inclusion of the Trackside territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

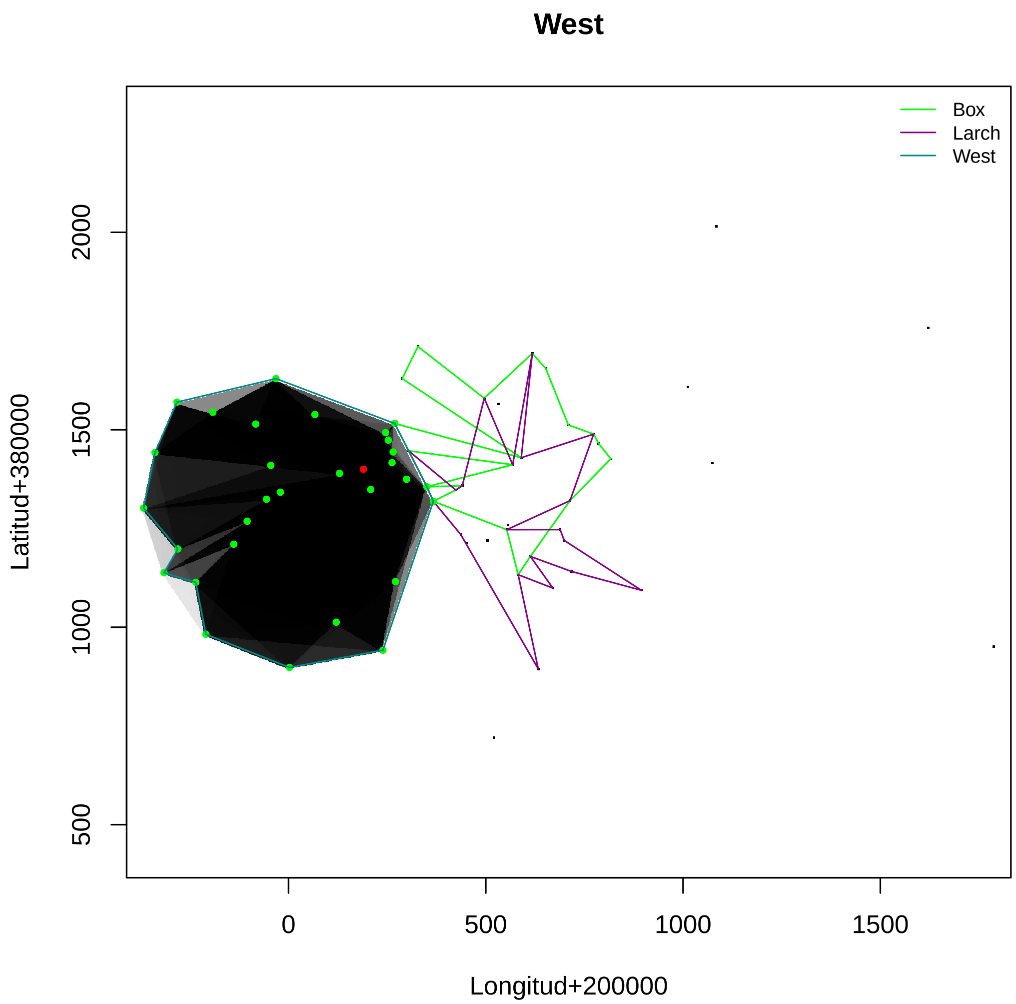


Figure E.19: Probability of inclusion of the West territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

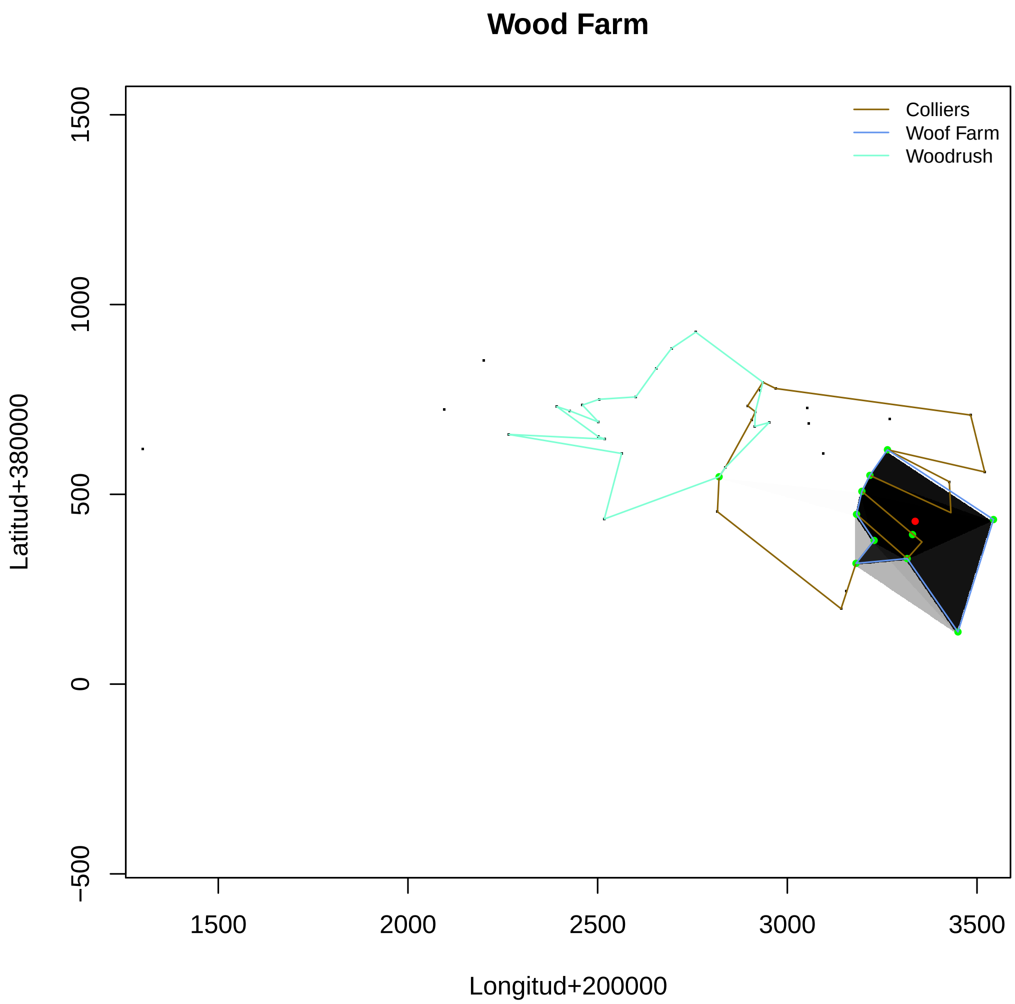


Figure E.20: Probability of inclusion of the Wood Farm territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

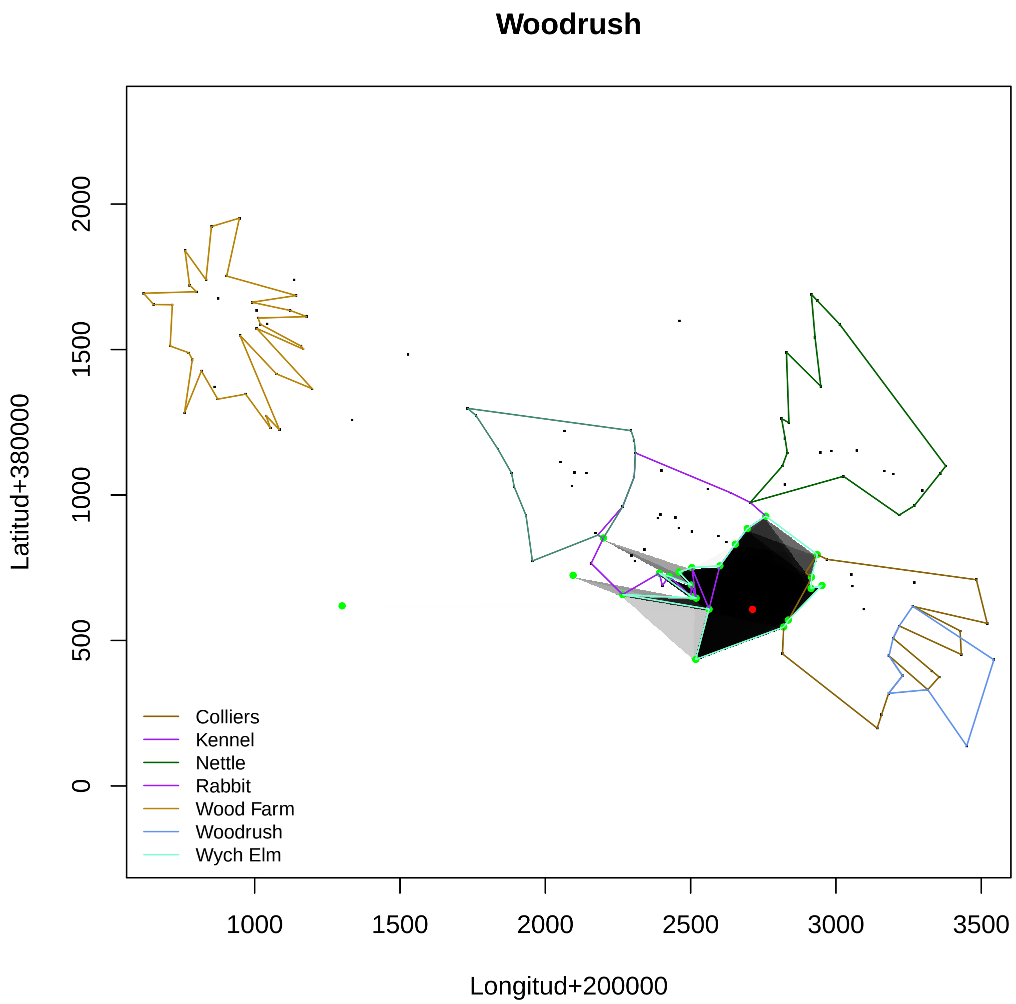


Figure E.21: Probability of inclusion of the Woodrush territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

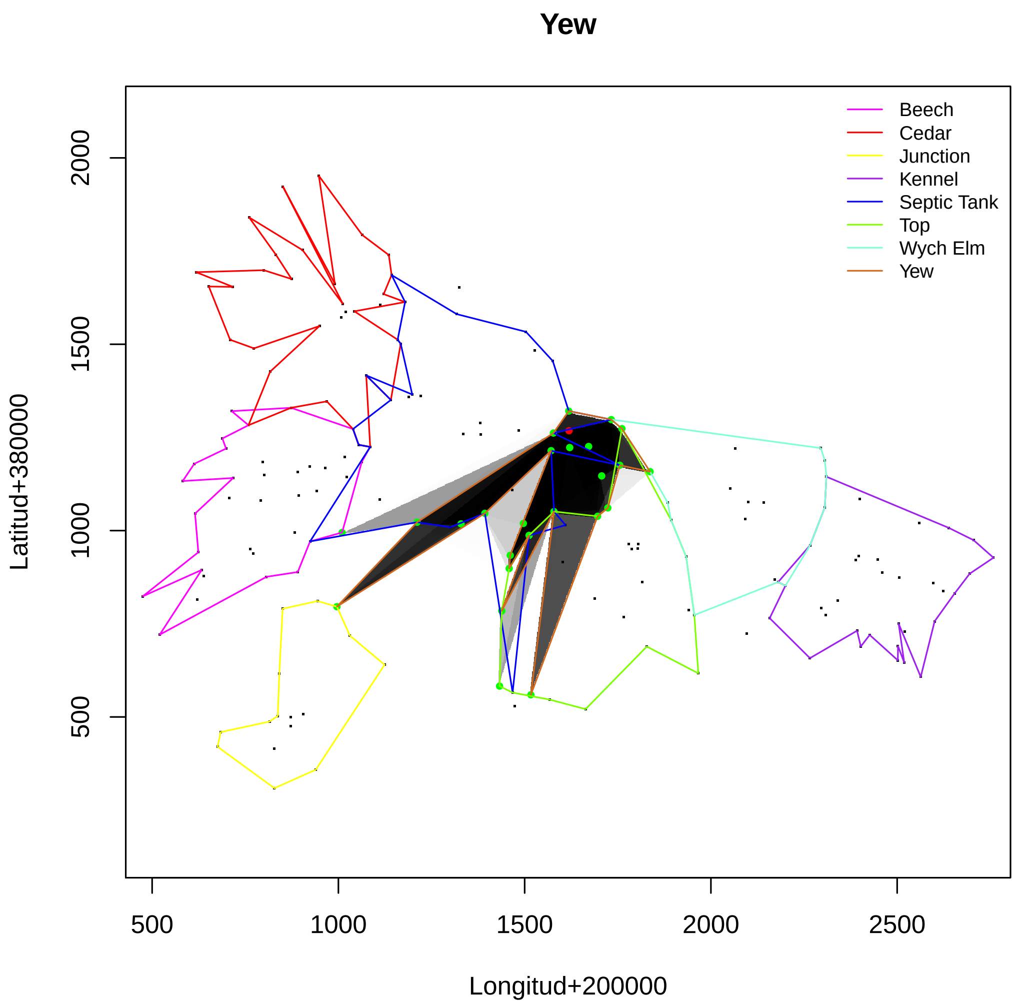


Figure E.22: Probability of inclusion of the Yew territory using the Adjusted Ordinal Model and the reconstruction with higher probability of the territories that share latrines with it.

Bibliography

Agresti, A. (2010). *Analysis of Ordinal Categorical Data.* John Wiley & Sons, Inc., Hoboken, New Jersey, 2nd edition.

Akaike, H. H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, 30(1):9–14.

Bearder, K. and Martin, R. D. (1980). The social organization of a nocturnal primate revealed by radio tracking. In Amlaner, C. J. J. and MacDonald, D. W., editors, *A Handbook on Biotelemetry and Radio Tracking*, pages 648–663. Pergamon Press, Oxford.

Blackwell, P. G. and Macdonald, D. W. (2000). Shapes and sizes of badger territories. *Oikos*, 89(2):392–398. Available from: http://www.jstor.org/stable/3547335.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):pp. 211–252. Available from: http://www.jstor.org/stable/2984418.

Broseth, H., Knutsen, B., and Bevanger, K. (1997). Spatial organization and habitat utilization of badgers *Meles meles*: Effects of food patch dispersion in the boreal forest of central Norway. *Zeitschrift Fur Saugetierkunde-International Journal of Mammalian Biology*, 62(1):12–22. Available from: <GotoISI>://WOS:A1997WP81700002.

Brown, J. A. (1993). *Transmission of bovine tuberculosis (*Mycobacterium bovis*) from badgers (*Meles meles*) to cattle.* Phd thesis, University of Bristol, Bristol, UK.

Burgman, M. A. and Fox, J. C. (2003). Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Animal Conservation*, 6(1):19–28. Available from: http://doi.wiley.com/10.1017/S1367943003003044.

Delahay, R. J., Brown, J. A., Mallinson, P. J., Spyvee, P. D., Handoll, D., Rogers, L. M., and Cheeseman, C. L. (2000). The use of marked bait in studies of the territorial organization of the European Badger (*Meles meles*). *Mammal Review*, 30(2):73–87. Available from: <GotoISI>://WOS:000087624700001.

Delahay, R. J., Carter, S. P., Forrester, G. J., Mitchell, A., and Cheeseman, C. L. (2006a). Habitat correlates of group size, bodyweight and reproductive performance in a high-density Eurasian badger (*Meles meles*) population. *Journal of Zoology*, 270(3):437–447. Available from: <GotoISI>://WOS:000241244000006.

Delahay, R. J., Walker, N. J., Forrester, G. J., Harmsen, B., Riordan, P., MacDonald, D. W., Newman, C., and Cheeseman, C. L. (2006b). Demographic correlates of bite wounding in Eurasian badgers, *Meles meles L.*, in stable and perturbed populations. *Animal Behaviour*, 71(5):1047–1055. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0003347206000479.

Delahay, R. J., Ward, a. I., Walker, N., Long, B., and Cheeseman, C. L. (2007). Distribution of badger latrines in a high-density population: habitat selection and implications for the transmission of bovine tuberculosis to cattle. *Journal of Zoology*, 272(3):311–320. Available from: http://doi.wiley.com/10.1111/j.1469-7998.2006.00271.x.

Elstub, S. (2006). Delineating Badger Territories from Bait-Marking Data. *Unpublished manuscript*.

Fischer, C., Ferrari, N., and Weber, J. M. (2005). Exploitation of food resources by badgers (*Meles meles*) in the Swiss Jura Mountains. *Journal of Zoology*, 266:121–131. Available from: <GotoISI>://WOS:000230023900002.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472. Available from: http://www.jstor.org/stable/2246093.

Green, P. J. (1981). Convex hulls in the analysis of bivariate data. In Barnett, V., editor, *Interpreting Bivariate Data*, pages 3–19. John Wiley & Sons Ltd, New York.

Harris, S. (1984). Ecology of Urban Badgers Meles meles: Distribution in Britain and Habitat Selection, Persecution, Food and Damage in the City of Bristol. *Biological Conservation*, 28(4):349–375. Available from: <GotoISI>://WOS:A1984SG54400003.

Hutchings, M. R. and Harris, S. (1999). Quantifying the risks of TB infection to cattle posed by badger excreta Quantifying the risks of TB infection to cattle posed by badger excreta. *Epidemiology and Infection*, 122(September 2000):167–174.

Jennrich, R. I. and Turner, F. B. (1969). MEASUREMENT OF NON-CIRCULAR HOME RANGE. *Journal of Theoretical Biology*, 22(2):227–237. Available from: <GotoISI>://WOS:A1969C802600002.

Johnson, D. D. P., Kays, R., Blackwell, P. G., and Macdonald, D. W. (2002). Does the resource dispersion hypothesis explain group living? *Trends in Ecology & Evolution*, 17(12):563–570. Available from: <GotoISI>://WOS:000179192000009.

Kass, R. and Raftery, A. (1995). Bayes Factors.

Kilshaw, K., Newman, C., Buesching, C., Bunyan, J., and Macdonald, D. (2009). Coordinated latrine use by European badgers, Meles meles: potential consequences for territory defense. *Journal of Mammalogy*, 90(5):1188–1198. Available from: http://www.jstor.org/stable/27755113.

Krebs, C. J., Gilbert, B. S., Boutin, S., and Boonstra, R. (1987). ESTIMATION OF SNOWSHOE HARE POPULATION-DENSITY FROM TURD TRANSECTS. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 65(3):565–567. Available from: <GotoISI>://WOS:A1987G708600017.

Kruuk, H. (1978a). Foraging and spatial organisation of the European badger, *Meles meles L. Behavioral Ecology and Sociobiology*, 4(1):75–89. Available from: http://dx.doi.org/10.1007/BF00302562.

Kruuk, H. (1978b). Foraging and spatial organisation of the European badger, Meles meles L. *Behavioral Ecology and Sociobiology*, 4(1):75–89. Available from: http://dx.doi.org/10.1007/BF00302562.

Kruuk, H. (1989). *The Social Badger: Ecology and Behaviour of a Group-living Carnivore*. Oxford University Press, Oxford.

Kruuk, H. and Dekock, L. (1981). Food and habitat of badgers (*Meles-meles L*) on Monte Baldo, Northern Italy. *Zeitschrift Fur Saugetierkunde-International Journal of Mammalian Biology*, 46(5):295–301. Available from: <GotoISI>://WOS:A1981MN85700005.

Kruuk, H., Gorman, M., and Leitch, A. (1984). Scent-marking with the subcaudal gland by the european badger, *Meles Meles L. Animal Behaviour*, 32(AUG):899–907. Available from: <GotoISI>://WOS:A1984TE92400030.

Kruuk, H. and Parish, T. (1981). Feeding Specialization of the European Badger *Meles meles* in Scotland. *Journal of Animal Ecology*, 50(3):773–788. Available from: http://www.jstor.org/stable/4136.

Lee, D. T. and Schachter, B. J. (1980). Two Algorithms for Constructing a Delaunay Triangulation. *International Journal of Computer & Information Sciences*, 9(3):219–242. Available from: <GotoISI>://WOS:A1980KK74200003.

Macdonald, D. W. (1983). The ecology of carnivore social behaviour. *Nature*, 301(5899):379–384. Available from: <GotoISI>://WOS:A1983QA49000031.

Macdonald, D. W., Mace, G., and Rushton, S. R. (1998). *Proposals for a National Mammal Monitoring Network: An Overview*. Department of Environment, Transport and the Regions, London, UK.

Macdonald, D. W., Mitchelmore, F., and Bacon, P. J. (1996). Predicting Badger Sett Numbers: Evaluating Methods in East Sussex. *Journal of Biogeography*, 23(5):649–655. Available from: http://www.jstor.org/stable/2846053.

Martín, R., Rodríguez, A., and Delibes, M. (1995). Local Feeding Specialization by Badgers (*Meles meles*) in a Mediterranean Environment. *Oecologia*, 101(1):45–50. Available from: http://www.jstor.org/stable/4220850.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):pp. 109–142. Available from: http://www.jstor.org/stable/2984952.

Mech, L. D. and Barber, S. M. (2002). A critique of wildlife radio-tracking and its use in national parks: a report to the U.S. National Park Service. Available from: http://www.npwrc.usgs.gov/resource/wildlife/radiotrk/index.htm.

Mohr, C. O. (1947). Table of Equivalent Populations of North American Small Mammals. *American Midland Naturalist*, 37(1):223–249. Available from: <GotoISI>://WOS:A1947YA68700006.

Moorcroft, P. R. and Lewis, M. A. (2006). *Mechanistic Home Range Analysis*. Princeton University Press, Princeton, New Jersey.

Neal, E. and Cheeseman, C. L. (1996). *Badgers*. Poyser Books. Available from: <GotoISI>://BIOSIS:PREV199800144901.

Neff, D. J. (1968). The Pellet-Group Count Technique for Big Game Trend, Census, and Distribution: A Review. *Journal of Wildlife Management*, 32(3):597–614. Available from: http://www.jstor.org/stable/3798941.

Palphramand, K. L., Newton-Cross, G., and White, P. C. L. (2007). Spatial organization and behaviour of badgers (*Meles meles*) in a moderate-density population. *Behavioral Ecology and Sociobiology*, 61(3):401–413. Available from: <GotoISI>://WOS:000242813000007.

Palphramand, K. L. and White, P. C. (2007). Badgers, *Meles meles*, discriminate between neighbour, alien and self scent. *Animal Behaviour*, 74(3):429–436. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0003347207002369.

Peng, R. D. (2013). General Polygon Clipping Library for R. Available from: http://www.cs.man.ac.uk/~toby/alan/software/;http://github.com/rdpeng/gpclib.

Pigozzi, G. (2010). Digging Behaviour while Foraging by the European Badger, Meles meles, in a Mediterranean Habitat. *Ethology*, 83(2):121–128. Available from: http://doi.wiley.com/10.1111/j.1439-0310.1989.tb00523.x.

Plumptre, A. J. (2000). Monitoring Mammal Populations with Line Transect Techniques in African Forests. *Journal of Applied Ecology*, 37(2):356–368. Available from: http://www.jstor.org/stable/2655916.

Putman, R. J. (1984). Facts from faeces. *Mammal Review*, 14(2):79–97. Available from: http://doi.wiley.com/10.1111/j.1365-2907.1984.tb00341.x.

R Core Team (2014). R: A Language and Environment for Statistical Computing. Available from: http://www.r-project.org/.

Rogers, L. M., Cheeseman, C. L., Mallinson, P. J., and Clifton-Hadley, R. S. (1997). The demography of a high-density badger (*Meles meles*) population in the west of England. *Journal of Zoology*, 242:705–728. Available from: <GotoISI>://WOS:A1997XR08300006.

Roper, T. J. (1994). The European badger *Meles meles*: food specialist or generalist? *Journal of Zoology*, 234(3):437–452. Available from: http://dx.doi.org/10.1111/j.1469-7998.1994.tb04858.x.

Roper, T. J. and Lups, P. (1995). Diet of Badgers (*Meles-meles*) in Central Switzerland - An Analysis of Stomach Contents. *Zeitschrift Fur Saugetierkunde-International Journal of Mammalian Biology*, 60(1):9–19. Available from: <GotoISI>://WOS:A1995QM17500002.

Rosalino, L. M., Loureiro, F., MacDonald, D. W., and Santos-Reis, M. (2005). Dietary shifts of the badger (*Meles meles*) in Mediterranean woodlands: an opportunistic forager with seasonal specialisms. *Mammalian Biology*, 70(1):12–23. Available from: <GotoISI>://WOS:000227236700002.

Schoener, T. W. (1981). An empirical based estimate of home range. *Theoretical Population Biology*, 20(3):281–325. Available from: http://dx.doi.org/10.1016/0040-5809(81)90049-6.

Schoener, T. W. and Schoener, A. (1982). Intraspecific Variation in Home-Range Size in Some Anolis Lizards. *Ecology*, 63(3):809. Available from: http://www.jstor.org/stable/1936801.

Scottish Natural Heritage (2003). Guidance on badger surveys. Available from: www.snh.gov.uk/docs/B957619.pdf.

Shepherdson, D. J., Roper, T. J., and Lups, P. (1990). Diet, food availability and foraging behavior of badgers (*Meles-meles L*) in Southern England. *Zeitschrift Fur Saugetierkunde-International Journal of Mammalian Biology*, 55(2):81–93. Available from: <GotoISI>://WOS:A1990DC26300002.

Smal, C. M. (1993). The National Badger Survey: preliminary results for the Irish Republic. In Hayden, T. J. and Murray, D., editors, *The badger*, pages 9–23. Royal Irish Academy, Dublin, Ireland.

Smith, B. J. (2007). boa : An R Package for MCMC Output Convergence. *Journal of Statistical Software*, 21(11):1–37. Available from: http://www.jstatsoft.org/v21/i11/paper?pagewanted=all.

Sutherland, W. J. (1996). *Ecological Census Techniques: A Handbook*. Cambridge University Press, 2 edition.

Thornton, P. S. (1988). Density and distribution of badgers in Southwest England - A predictive model. *Mammal Review*, 18(1):11–23. Available from: <GotoISI>://A1988M991000002.

Tinkle, D. W., Dana, S., and McGregor, D. (1962). Home range ecology of Uta Stansburiana-stejnegeri. *Ecology*, 43(2):223–229. Available from: <GotoISI>://WOS:A1962C145400031.

Tukey, J. W. . (1949). Comparing Individual Means in the Analysis of Variance. *International Biometric Society*, 5(2):99–114. Available from: http://www.jstor.org/stable/3001913.

Tuyttens, F. A. M., Delahay, R. J., MacDonald, D. W., Cheeseman, C. L., Long, B., and Donnelly, C. A. (2000). Spatial perturbation caused by a badger (*Meles meles*) culling operation: implications for the function of territoriality and the control of bovine tuberculosis (Mycobacterium bovis). *Journal of Animal Ecology*, 69(5):815–828. Available from: <GotoISI>://WOS:000089754900008.

Tuyttens, F. A. M., Long, B., Fawcett, T., Skinner, A., Brown, J. A., Cheeseman, C. L., Roddam, A. W., and MacDonald, D. W. (2001). Estimating group size and population density of Eurasian badgers *Meles meles* by quantifying latrine use. *Journal of Applied Ecology*, 38(5):1114–1121. Available from: http://www.jstor.org/stable/827248.

Tuyttens, F. A. M., Macdonald, D. W., Swait, E., and Cheeseman, C. L. (1999). Estimating Population Size of Eurasian Badgers (*Meles meles*) Using Mark-Recapture and Mark- Resight Data. *Journal of Mammalogy*, 80(3):950–960. Available from: http://www.jstor.org/stable/1383265.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S.* Springer, New York, fourth edition. ISBN 0-387-95457-0. Available from: http://www.stats.ox.ac.uk/pub/MASS4.

Virgós, E., Mangas, J. G., Blanco-aguiar, J. A., Garrote, G., Almagro, N., and Viso, R. P. (2004). Food habits of European badgers (*Meles meles*) along an altitudinal gradient of Mediterranean environments: a field test of the earthworm specialization hypothesis. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 82(1):41–51. Available from: <GotoISI>://WOS:000220803800005.

Wilson, G., McLaren, G., and Harris, S. (1997). *Changes in the British Badger Population, 1988-1997.* Kerenza J Ltd.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36.

Worton, B. J. (1987). A review of models of home range for animal movement. *Ecological Modelling*, 38(3-4):277–298. Available from: <GotoISI>://WOS:A1987K412600005.

Worton, B. J. (1995). Using Monte-Carlo simulation to evaluate kernel-based home-range estimators. *Journal of Wildlife Management*, 59(4):794–800. Available from: <GotoISI>://WOS:A1995RY58900022.

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34. Available from: http://www.jstatsoft.org/v32/i10/.