Spatial Analysis of Potential Habitat for the Dormouse in the Capital Region of Denmark

– A Filter Method for Habitat Detection –

Spatial Analysis of Potential Habitats for the Dormouse in the Capital Region of Denmark

- A Filter Method for Habitat Detection -

Major Thesis Report

Ruben López Vázquez

Aalborg University
Geoinformatics
Dep. of Surveying, Planning and Land Management

The front cover of this report was inspired by the distinctive layout of the publishing house O'Reilly. This cover uses a 400-year old depiction of a 'dormouse' by Edward Topsell, a cleric from Kent, author of "The History of Four-footed Beasts" published by William Jaggard in London, 1607. The image was extracted from a paper by Carpaneto and Cristaldi (1995) who caption it as "the first known drawing of a dormouse (Topsell, 1607): a Garden Dormouse with unusually prominent teeth." However, I have been advised by Lars Dalby (PostDoc researcher at Kalø) to mention the fact, that its facial mask identifies it as a different species than *Muscardinus avellanarius* (the subject of this report). The dormouse depicted is, as Carpaneto and Cristaldi (1995) quote, *Eliomys quercinus* (the garden dormouse). The cover layout was created in *Adobe Photoshop CS6*, *Adobe Illustrator CS6*, and *Adobe InDesign CS6* using Minion Pro font for the title and Palatino Linotype for the subtitle. The interior layout was designed in *LaTeX*, based on a template by Jesper Kjær Nielsen, professor at Aalborg University. The text font is URW Palladio, and the code font is Segoe UI default on the Github scheme on *PyCharm* interface, which used the Intelij user interface layout. The illustrations that appear in the report were produced as well by the author using *Adobe Photoshop CS6* and/or *Adobe Illustrator CS6*. The graphic figures have been produced using the statistical language R (v.3.2.3) in *R Studio* 1.0.136.

Aalborg University
Landinspektør (cand.geom.)
A. C. Meyers Vænge 15,
2450 København

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Spatial Analysis of Potential Habitats for the Dormouse in the Capital Region of Denmark.

Theme:
Major Thesis

Project Period:
Spring Semester 2017

Project Group:
Individual work

Participant(s):
Author: Ruben López Vázquez

Supervisor(s):
Supervisor 1: Jamal Jokar Arsanjani

Copies: 2

Page Numbers: 96

Date of Completion:
June 9, 2017

Abstract:

This thesis aims to expand the habitat of the hazel dormouse by detecting the signature of its local environment in prospective, far away locations by means of the automatised process of data.
The hazel dormouse is an endangered species (nowadays rare in Denmark), and still in apparent decline. In the face of extinction, this research seeks to expand the known suitable territory for the species. It understands the man-made degradation of the environment as unconscious and as negation but, chiefly, as a temporary condition.
If previous conservationist efforts have imagined a viable habitat as a physical structure that positively interconnects populations, this work attempts to conceive viable habitat as an enclave, as well as something other than purely spatial. Thus, habitat is defined here by a stable variation over time of land use, of climatic phenomena, of intensities of use . . . as a continued, favourable relationship.

AALBORG UNIVERSITET
STUDENTERRAPPORT

Aalborg University
Landinspektør (cand.geom.)
A. C. Meyers Vænge 15,
2450 København

Titel:
Rumlig Analyse af Potentielle Habitat for Haselmusen i Region Hovedstaden Danmark.

Tema:
Speciale

Projektperiode:
Forårssemestret 2017

Projektgruppe:
Individuelt arbejde

Deltager(e):
Forfatter: Ruben López Vázquez

Vejleder(e):
Vejleder: Jamal Jokar Arsanjani

Oplagstal: 2

Sidetal: 96

Afleveringsdato:
9. juni 2017

Abstract:

Dette speciale forsøger at udvide habitatet for hasselmusen ved at fastslå, i andre steder i Danmark, dets lokale miljøs kendetegn ved hjælp af automatiseret databehandling.
Hasselmusen er en truet art (i dag sjælden i Danmark), og stadig i tilbagegang. I lyset af udryddelse, søger denne forskning at udvide det kendte egnede område for arten. Den forstår den menneskeskabte nedbrydning af miljøet som ubevidst og som benægtende, men frem for alt, som en midlertidlig tilstand.
Hvor tidligere bevarende indsatser har forestillet sig et levedygtigt habitat som en fysisk struktur, der forbinder bestande, forsøger dette speciale at forstå et levedygtigt habitat, først som en enklave, og så som noget andet end rent rumligt. Således er habitat defineret her ved stabile forandringer over tid af arealanvendelse, af klimatiske fænomener, af brugsintensitet … som et positivt løbende forhold.

Contents

Preface

Geoinformatics arose in the pursuit of reliable, iterative and precise measurements for spatial analysis. Its competence for measuring spatial relations should allow this report to work with quantitative but also qualitative data. It could be used for analysing space, but also what actually resides in time. Such credentials were the reason for me to train in this field, but the actual reason for choosing this topic, more than to train on precision, it was to do with its suitability to assemble several methods and languages for automation (i.e. sql, python, R, opals, eCognition . . .). This report attempts to make use of such capabilities, and, if possible, shedding some light on the reported decline of a mammal species, which stands out as a sensitive indicator for land use changes. Here you should expect to find some measurements on the ethology of *Muscardinus avellanarius*, which should explain its relation to its surroundings to some extent. Ethology derives from 'ethos', the behaviour of the species, but of course this does not mean that I could tell any specifics. Despite the accuracy of some of the instruments used, and the power derived from the automation of the calculations, the best I can do is to inspect averages and measures of central tendency. A bit like the recount of a tracker, this text will work on the pattern of activity of this nocturne species. However, unlike the recount of a tracker, this will be very far from being detailed and incidental.

I became acquainted with the attentive research carried out by "Danmarks Miljøundersøgelser" (Aarhus Universitet) on *Muscardinus avellanarius* (i.e. the common hazel dormouse) during a seminar on *LiDAR* data. The effort had characterised space insofar as habitat, which meant recording significant traits on the nearest surrounding space to dormouse nests. However complete, their data could still gain from a broader assessment on space, this time as mere physical relation, an angle that later on will be developed in detail. Thus, the departure point for this research was a rich database, that had collected data on the types of trees, their variety and abundance, the density and the degree of cover provided by their canopy, their understory, the amount of natural light present and its soil moisture. It was a local approach that would help identify locations where dormice could be found. However, its practicality was still compromised, the size of the population was very small, and this is where geoinformatics had to help.

Common hazel dormice live in forests, where they find food and shelter. Nevertheless, this ’mouse’ will not reinforce a conception of nature as pure, distant and secluded other. Nests are found quite close to human activity, and there are accounts that dormice once set sail, and that for a while they were kept captive as pets. (Carpaneto and Cristaldi 1995) Of course, rather than the thrill underlying such accounts, this report is only interested on improving dormice detection. It will not try to prove historical facts, neither research the social ties of the species, nor it will be an ecological study. It will only use ethology in order to establish the suiting spatial relation that works as a signature, for filtering and/or evaluating the possible settlement of the species.

Such relationship should work in the same way as the evidence that allows the tracker to advance. Nevertheless, for a tracker, dormice would be the ultimate goal; and for this report, instead, the ultimate goal is their physical geography. By focusing on the physical geography of the species, the present research builds from a scale that should improve the search. Instead of directly probing a local environment, which implies the need to employ cumbersome ’computing force’ at a very small scale, here the strategy is layered. It starts at very large phenomena like soil and climate. I use this layered approach of selection and filtering, crucially because I assume that it is the most effective. It narrows down the search across scales, by only paying attention to the ethology and phenology of the species.

Early on it was brought forth the decline of the species. The dormouse is an endangered species, but, what is the source of that danger? It is currently not well known. One could think of pollution, the instability of key environmental variables, but, from the angle of former research, as well as projects like the BioGrenzKorr-Project (Schulz et al. 2013), and from various bibliographical sources (Fedorowick 1993; Bailey, Haines-Young, and Watkins 2002; Bennett, Radford, and Haslem 2006), it is often attributed to habitat loss and fragmentation. The approach of this research understands that the avoidance of habitat fragmentation is paramount, however, it also looks at Denmark as the fringe of the common dormouse’s geographical distribution. Thus it wants to underscore, that such condition of fringe species might be due to something other than the purely material. Camille Fløjgaard et al. report that the common dormouse is widely distributed in the warmer Central and Southern Europe, and citing the Office for Official Publications of the European Communities they also report that in these areas the common dormouse is considered a species of Least Concern on the IUCN Red List. (Fløjgaard et al. 2009) This is an important piece of information, because it tells that dormice are not equally endangered in its entire distribution. However, and to illustrate the point I try to make, this very same report adds right in the next sentence, that "[h]abitat destruction must therefore be considered the most immediate and main threat to the continued occurrence of common dormouse in Denmark". (Fløjgaard et al. 2009) Habitat loss is not the same as habitat fragmentation. Let us find habitat!

The reason why the decline of this species is not well known, is on its composite nature. It "appears to be linked to a complex set of factors including climatic changes, fragmentation, and the deterioration and loss of specialized habitat." (Root and Schneider 2002) Even finding 'habitat' is never so simple. If it is well established that dormice populations are declining across much of its northern range (Mills 2012), there are as well publications in areas like Lithuania, where, though the flora differs from what some consider the optimal habitat for this species, researchers report on very healthy and strong communities. (Juškaitis 2007)) In light of this, and in order to phrase in a precise manner the aim of this exercise, I should state that, because *LiDAR* and satellite imagery are inadequate to detect dormice, the aim is to improve their habitat detection, which is not the same. Dormice might not be found on a perfectly suitable piece of land. Climatic fluctuations have been found to correlate with the decline of the common dormouse. Some have placed the North Atlantic Oscillation phenomenon (NAO) as a player on making such communities only more precarious. (Sanderson 2004) A more precise and systematic habitat detection might help evaluate the extent to which climate participates. Thus, in spite of the role of continuity in enabling the migration of individuals, the constitution of Denmark as an archipelago inspired this exercise to disregard any spatial strategy for interconnecting habitat, and, instead, to begin to search for new potential enclaves. Then, encouraged by the fact that the species has long benefited from a degree of human proximity; holding the hypothesis that the current decline of the species under an Atlantic climate might be due to seasonal bottlenecks (driven by climatological fluctuations); this project set out to evaluate how inadequate is the most developed part of the territory, for the purpose of noticing (if ever so faintly) the climatological involvement in the absence of the species.

To summarise, this is a project that relies on a limited population of dormice, which tries to use their spatial relation as a filter, and which cannot establish by itself the causes for the reported decline. But the purpose of this filter is, on the one hand, the evaluation of the suburban environment as a viable space for placing new dormouse enclaves; while, on the other hand, to take a step toward establishing the relevance of climatological phenomena, in relation to the effect of physical loss of habitat, on the decline of this species, here, at its north-western fringe.

Aalborg University, June 9, 2017

Ruben López Vázquez
<rlopez15@student.aau.dk>

I would like to thank senior researcher at Aarhus University, Geoffrey Brian Groom, for his trust and his willingness to make not just the data available, but also the indispensable licenses to such useful software as *eCognition* and *OPALS*. Also, I would like to extend this gratitude to the companies, that handed me the opportunity to gain some further experience with their software, as a student. It is in the line of their work that I would like my career to continue.

Chapter 1

Introduction

Muscardinus avellanarius is a very small, nemoral and nocturnal rodent. This means that it lives in forests, in very specific types of forests more than in any forest, and this already shows something important about the character of the dormouse. It is not the usual 'mouse' one has in mind. It is not everywhere, it is in specific places, and this allows for an specific spatial signature. It does not reproduce to the size of plague proportions, but regulates its population. This means that it is not common, but also that *Muscardinus avellanarius* has, from the very point of view of biology, a very different strategy. It is known to be, to some extent, a K-strategist. (Skok and kryštufek 2012) A K-strategist is defined as a species that makes an strategic use of food as a way to survive. It relies also on efficiency, instead of just relying on fast population growth, as the r-strategists do. (Margalef 1993)
These are the first important facts about the dormouse. His/her behaviour is selective, thus follows a pattern and it is placeable; and his/her survival, instead of relying on the availability of energy in order to explode in numbers, relies on establishing an efficient relation to its environment, while adapting population numbers to the avail of territory (i.e. the avail of food, energy). Mice are somewhat r-strategists. Like plankton, wherever there is enough energy, they may appear in large numbers to turn the situation into fast growing population numbers, without any regard for entropy. They survive by ensuring their presence in numbers, which relies on that steady injection of energy, wherever it happens. Dormice are instead like trees, they are to a certain extent K-strategists. This means that they regulate their numbers and store energy to adapt to a very distinct energy cycle. They balance a budget by being efficient at using the available resources. Thus, although both (K and r) strategies can be successful, they crucially differ in their level of structure, in their use of information. This is very important, for it makes dormice a species with a distinct spatial signature. Just like trees, their spatial density matters. Distances matter, and it enables spatial analysis to have some, perhaps not much, but some predictive power.

1.0.1 Habitat loss

"The common dormouse is the only representative of the rodent family Gliridae, in Denmark. Its postglacial immigration probably occurred c. 10.000 B.C. (Aaris Sørensen 1998) with the extension of deciduous forest (including species such as oak, alder, lime and hazel), influenced by a warmer climate." (Wilhelmsen 2003) The numbers of dormice are reported to be declining, same as their distribution. Many have identified this decline with habitat loss, but that says as much as 'not much'. Often, the size of their territory is thought as a limiting factor in their reproduction, and this is because dormice "are vulnerable to habitat fragmentation as they have low population densities, limited dispersal and low reproductive potential" (Bright 1993, Mills 2012). In Germany the mean dormouse population density was found to be of 0,9 individuals/ha. (Ehlers 2012) That is quite spacious for such a tiny animal, and Rimvydas Juškaitis and Sven Büchner, confirm in their book "Die Haselmaus", that, for its size, dormice are settled rather sparsely. As quoted by Keckel, Büchner, and Ansorge (2012), in comparison to other small rodents the dormouse appears in low densities, even in good habitats. Still, other authors have found their home range to be of smaller size. Juškaitis (1997) reports in more detail, differences between the home range of males and females. He found the home range of males (throughout their active season) to be 1.0 ± 0.05ha, whereas for females it was smaller 0.8 ± 0.05ha. Also that their ranges overlap partially between males and females. Here it starts to be evident that, even if a species like *Muscardinus avellanarius* is more suited to spatial analysis, setting its spatial variables in stone is not possible. These fluctuate in relation to the area under scrutiny. Not just because they depend on gender, but also (and most importantly) because these are in function of their environment and even probably the season (i.e. the year). Fischer (2014) reports that Bright and Morris found their home range to be significantly smaller than the previous: in 1991 they found it to be 0.45 ha for males, and 0.19 ha for females, depending on their habitat composition (which, I assume, it responds to food availability within it). Furthermore, they noted changes in size between one year and another. In this way, in 1992 they found the nest range to be 0.68 ha for males and 0.22 ha for females.

Although my work is not a biological research, and though I am not after finding or validating these values, it is important to understand these fluctuations. I need to work with these values as a signature for filtering and classifying space as 'suitable for dormice'. With regard to this one must realise that, if such annual fluctuations exist, they establish variation as well as a certain vagueness to the behaviour of the species. It perhaps forces dormice to travel across longer distances. However, their very survival strategy demands them to put up weight before hibernation, up to a 50% (Lloyd et al. 2015), in order to damper the larger variation imposed by winter.

Moreover, the dormouse competes directly with larger animals like the grey squirrel for most of the diet. That, coupled to their very small range of action, forces this exercise to be pessimistic, and adopt the largest home ranges as an act of caution. In this way, I take as a home range a 1 ha circle, which has a radius of 54.62 m. Since Fischer's previous master thesis found by radio-tracking dormice in a relatively near community (in the Danish island of Fyn), that the average distance travelled by common dormouse per night was 205 m for males and 197 m for females, which gives an average distance of 201 m, (Fischer 2014) a distance that could comprise with ease a deviation of up to 55m from the nest. Even Bright and Morris in 1992 found, that the dormice kept under their surveillance, in the UK, had travelled a mean total distance of 143.38 m. (Fischer 2014) From this last datum, once again in compliment with the most restrictive figure for the sake of caution, the definition given in this exercise to a cluster of nests was established at a radius of 75 m (i.e. a round trip of 150 m).

By the name of cluster I try to define an habitat area, which groups together those inhabited nests that are within the average distance range (allowing mice to go and come back) in a single night. The concept of habitat is then not limited to a foraging area, it has to enable as well passage for genes to be transmitted, it has to offer shelter. But of course then there cannot be a clear cut, a migrating dormouse can travel further and even meet another travelling within its territory in the opposite direction. Even though a rare occasion it should be contemplated, since it still establishes a bond between nests. This bond was named community, and it took into account that "[A] previous survey of Büchner (2008) showed that individually marked hazel dormice crossed open ground with distances up to 500 m . . . ". (Keckel, Büchner, and Ansorge 2012) It was was calculated by doubling the largest distance plus twice the standard error, and it was generously rounded up at 600 m, for it will only be used for contrast on the nest surroundings.

In this way, after having defined habitat as an enclave, now I should move to the concept of habitat loss. It should be understood as the deterioration of areas, formerly capable of sustaining a population of dormice, to the point of these becoming inviable. But, what does it mean such inviability? Some studies have quantified "that dormice are more likely to be present in woods larger than 20 ha. This implicates a minimum habitat size of 20 ha for viable populations." (Keckel, Büchner, and Ansorge 2012) In other words, a much more reduced forest size begins a process that leads to inviability by isolation of a reduced group of mice. Yet again, things are not that straight forward. This was in an study in East-Saxony, and even there "[i]t seems further that up to 20 m no effect of habitat isolation exists because dormice were present in wood lots smaller than 20 ha when the distance to the next wood with dormice was lesser than 20 m." (Keckel, Büchner, and Ansorge 2012) From this perspective habitat lost and fragmentation seem to coincide, but so far this is a purely 'geometrical' account that ignores the productivity of the territory.

A richer environment could enable bigger densities of dormice, and thus the enclave to remain viable. In this line, the very same German article points that the most isolated case on record to a wood of "2 ha in exten[t] and in a distance of 669 m to the next wood." (Keckel, Büchner, and Ansorge 2012) Such productivity, or let us call it suitability of the environment, takes into account the composition of tree species and the structure of the understory. The common dormouse "prefers a forest environment of high plant diversity, glades with regeneration, deciduous trees of different age groups and a rich understory supplying berries and herbs." (Wilhelmsen 2003) In consequence, habitat loss should not only refer to the encroaching of a population into a an ever smaller area, but the failure of that area to support dormice. Even if its boundaries remain where they are, and even if fragmentation does not exist. Thus we see new variables come into play, which are not lesser than spatial continuity, and the extent of forests.

1.1 An Introduction to Variables for a Dormouse Filter

What indicators are relevant for modelling are also relevant for this process of filtering space. It is basically a very similar process. The main difference is, that in this process there is one filter per variable and, instead of using them to create a multivariate regression formula, each indicator (or parameter) is used independently and overlaid upon the others. Each layer (or parameter) will be then adding or subtracting some value to the suitability of space, which is a multilayered, composite map. This is nothing new. It is in a way what planners like Ian McHarg (chair of landscape architecture department at University of Pennsylvania, and known for developing the role of environmental planning not through visual aspects of form, but from the interwoven relationships of our actions in an environment), were already putting forward in the 60's. His method, explained in a book like "Design with Nature" (1969), is a proposal that, even if at that point unconsciously, was already re-evaluating the concept of 'the other'. His terms were 'man' and 'nature'. He overlaid geology with its aquifers, and nature with its forests and marshes, on top of man made cities with its infrastructure in order to establish a relation that, he thought, needed to be balanced.

1.1.1 Climate change

One of the first sets of variables that come to mind is climatological. Climate change "is likely to have implications for dormouse over-wintering survival. Interannual variation in abundance of wild populations of dormouse in the UK indicates that cold, dry winters are correlated with higher dormouse abundance the following summer, suggesting warmer winter temperatures decrease dormouse winter survival (Sanderson 2004). Our findings provide an explanation for this:

increasing ambient winter temperatures result in a higher frequency of periodic arousals that would in turn increase individual energy requirements (Thomas et al. 1990)." Mills (2012)) But it is not just the slow increase of planetary average temperature, that which is more forbidding for the dormouse in a climate change scenario. The fact that CO_2 alone is capable of storing more energy than the average chemical mixture of gasses known as air in our planet, is something well established since the XIXth century. Starting with the work by Joseph Fourier in the 1820s, which put forward that gases in the atmosphere could trap sun radiation, and later the work by John Tyndal and then by Svante Arrhenius, which led to understand that human emissions could result, eventually, in an increased absorption of energy by the atmosphere. [1] That increase of available energy, and its role on accentuating atmospheric events, is perhaps (in my view) the most destabilising phenomenon for a small mammal, whose strategy is based on efficiently following a distinct and familiar energy cycle.

When I followed the report on the dynamics and the peculiarities of *Muscardinus avellanarius* in Lithuania done by Juškaitis (1995), I found very interesting data on the mortality and behaviour of the common dormouse. Despite their steady decline, even in locations where dormice seem to be doing well, their mortality does not follow a steady pattern. "According Likhachev, mortality of adult animals varied from 58% to 80% (an average 71% over seven years). Mortality of young animals in their first year was only 60% (with variations from 48% to 81% in different years)." (Juškaitis 1995) Thus begins habitat loss to have perhaps multiple grounds, when we see that such irregular mortality begins to be associated to climatological parameters. Bright and Morris (1996) argue in their article "Why are Dormice rare? . . . ", that "[t]he Dormouse is also sensitive to climate, both directly and probably indirectly through the effects of weather on the timing and abundance of food (insects, flowers and fruits)." Henceforth should this exercise beware that habitat loss is also the result of the local weather disallowing perfectly fitting forests to be populated by dormice.

> *"Evidence is presented which indicates that declining range and numbers are due to a complex interplay of factors which include fragmentation, deterioration and loss of specialized habitat. The Dormouse is unusual in being a relatively k-selected small mammal, with exacting ecological requirements which render it very vulnerable, particularly to habitat fragmentation. The Dormouse is also sensitive to climate, . . . not just to absolute climatic measures (e.g. temperature, rainfall), but especially to climatic stochasticity, particularly at the edge of its range. There are strong associations between the distribution and changing status of the Dormouse and various climatic parameters, and clear parallels with other climate-sensitive taxa, notably bats and butterflies."* (Bright and Morris 1996)

---

[1] Consult the American Institute of Physics at http://history.aip.org/climate/co2.htm

In light of this I did right the same bibliographical research as I have done with spatial parameters: when I established the definition of home range according to Juškaitis (1997), and of cluster and community based on the ′mean total distance travelled′ measurements by Fischer (2014). In terms of climatological events, the two most important measurements that I could use for this filter, were the aforementioned temperature and rainfall. With regard to the stochasticity of the weather patterns, and their sharper increase in strength, not much more can be done but to uphold proof that *Muscardinus avellanarius* dwells in areas where these parameter seem to remain stable.

A good indicator should be the difference between diurnal and nocturnal temperatures when being normalised the yearly thermal amplitude of the locations where dormice nest. Such indicator is called Isothermality, and it is mapped in figure1.1. In such map one can see, that the distribution of the known nests of *Muscardinus avellanarius* to this exercise relates to areas, that enjoy very stable temperatures (i.e. a small thermal amplitude). However, with the current available data, I cannot reach a more general and final conclusion on this fact. This is because the nest data available to this exercise is scarce and poorly distributed. It only covers a very small fragment of the total extent of dormouse habitat. However, from nest data portrayed on other researches like the on the "Forvaltningsplan" Naturstyrelsen, Miljøministeriet 2011 (the management plan 2011 from the Danish Ministry of Environment); or the 2012-2014 NOVANA-project it is possible to ascertain a tendency on isothermality. This parameter is characterised by showing very broad (about 10 km wide) bands representing a change of 0.1°C, and, if the highest value in the mapped area reaches 5.1°C, the location of nests within Denmark reaches a maximum at 2.8°C in historical Jutlandic enclaves; 2.6°C in the currently strongest dormouse communities on the island of Funen; 2.3°C in the island of Langeland, and 2.4°C in Sjælland. These maxima for Denmark are all above the maximum within my nest data, which reaches only 2.2°C, although it is clearly lying within the lower half of the amplitude mapped in figure1.1. Thus it seems still very safe to establish, that dormice thrive in areas with a settled climatological cycle, as opposed to a very volatile one, which is important (as much as knowing its spatial relationship) for introducing it into the research area. But first I need a more detail description of the climatological habitat, based on other studies, on its northern fringe.

Description of the climatological habitat

Since dormice hibernate, it becomes apparent that there is a most relevant season: which is right before hibernation (i.e. the months of September and October). Temperature, and then precipitation, became the main references in the consulted bibliography for the rate of survival and environmental assessment. The areas for reference where Britain to the west, and Lithuania to the east of the northern fringe.

Figure 1.1: Map depicting Isothermality. Isothermality was calculated as the division of the Mean Diurnal Range (i.e. the mean of the monthly (max temp - min temp)) by the Annual Temperature Range (i.e. the maximum temperature of warmest month minus the minimum temperature of coldest month). The location of nests on south-west Sjælland was highlighted in red, as well as the location of the research area in the north-east.

The difference between both selected zones is just an exaggeration on the transition existing in Denmark between the west coast of Jutland, which has a West European maritime climate, and the east coast of Sjælland, which has a climate closer to continental. Thus, the difference is on the amount of rain, as well as on the annual range of temperature. A maritime climate is wetter (in detriment to dormice) and, because of the Gulf Stream, the seasonal temperatures tend to be buffered by sea temperature, leading to mild winters (in detriment to dormice), and fresh summers (yet again in detriment to dormice). It makes only sense that a hibernating rodent should prefer steadily cold winters. As Mills (2012) argues, mild winters result "in a higher frequency of periodic arousals" and thus in higher energy requirements. This is confirmed by empirical measurements in literature. Juškaitis (2008) reports that "ambient temperature . . . was one of the main factors determining prevalence of torpor among *M. avellanarius* in spring and early summer." Moreover in winter:

> *"Arousals occur mainly during the day (not at dusk as in the summer) suggesting that they are induced by daytime warmth. At a ground temperature of 9^{o}C, arousal is 2.5 times more likely than at 3^{o}C (P. W. Bright, unpublished data), and variable temperatures have an even greater effect on stimulating arousal ... "* (Bright and Morris 1996)

All this suggests that hibernation is just an strategy to adapt to long periods of low available energy, by lowering energy consumption. What does this means for spatial analysis? Well, basically the fact that, even if not aggressive, dormice survival depends on competing for finding and storing (i.e. fattening up) a certain amount of calories. That is equivalent to say, that each individual has a nest range, which only overlaps between members of the opposite sex (Juškaitis 1997). Furthermore, that energy must be readily available in their near surrounding space (they only travel up to 200 m in a night), or else they must reduce their numbers to match the available energy in their environment. This has also been proven:

> *"...M. avellanarius is distinguished by having relatively stable abundance without great inter-annual population fluctuations (see chapter 4.1.2). Stability in population numbers is interpreted as the result of density-dependent processes acting on the species (Flowerdew 1987). Self-regulatory internal mechanisms reduce population growth before resources become limiting.* (Juškaitis 2008)

In terms of spatial analysis, this apparently inane piece of information gave a very important tool to distinguish the quality of specific spatial arrangements. Since I have no data that displays the evolution in time of the known nests in south-west Sjælland, I could not tell if their location or any of their spatial metrics could indicate a healthy cluster, a growing community or the opposite. Now this 'frozen' image in time has some interpretability. Just like the distance between footprints and their depth tells a tracker the speed at which the target was moving, the density of nests combined with their numbers can be used to interpret the strength of the community (i.e. how supportive and suitable their surroundings are). Now it is possible to weight the information gathered on their surroundings and combine it into a signature that prefers those more advantageous conditions.
Figure1.2 represents an idealised spatial depiction of a dormouse community, using circles to delimit their foraging range. This figure is a starting point for analysing their spatial arrangement, and this one just spells trouble. The individuals hardly ever meet and, at the scale that I now know they conduct their lives, these three inhabited nests are not clustered. The three concentric circles give a sense of scale, though they would be elongated (foraging grounds for dormice are often located at margins, where vegetation grows thick: along the edge of high forests and forest paths, surrounding forest bogs, ditches or brooks, on hedges connecting woods ... (Bright and Morris 1996)). The three circles represent 25 m, 50 m and 75 m, which establishes the maximum distance that defines a cluster.

Straight dashed lines indicate the interface between overlapping territories.

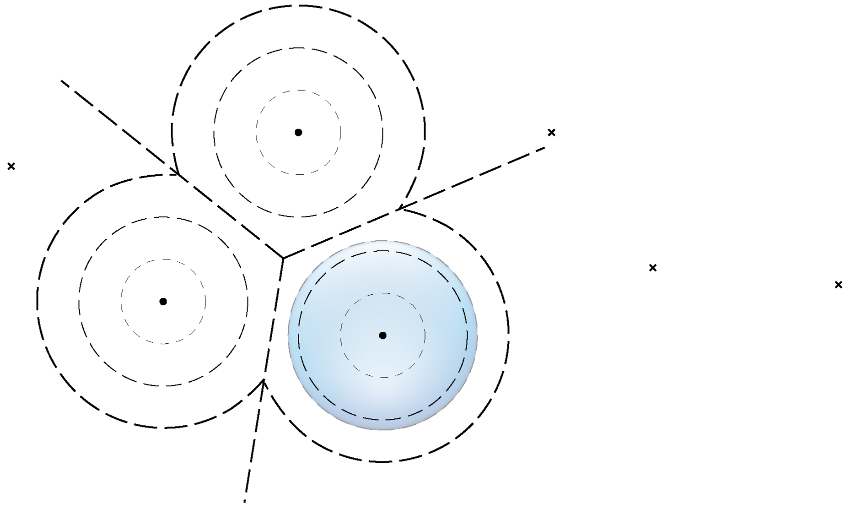


Figure 1.2: Idealised spatial structure, where dots represent inhabited nests. A 1 ha range is drawn in solid colour, and dashed circles give a sense of scale at 25, 50 and 75 m (which defines a cluster).

If we take a look at figure1.3, it shows a more common distribution of nests in range, yet this cluster is extremely dispersed as well as spelling trouble.

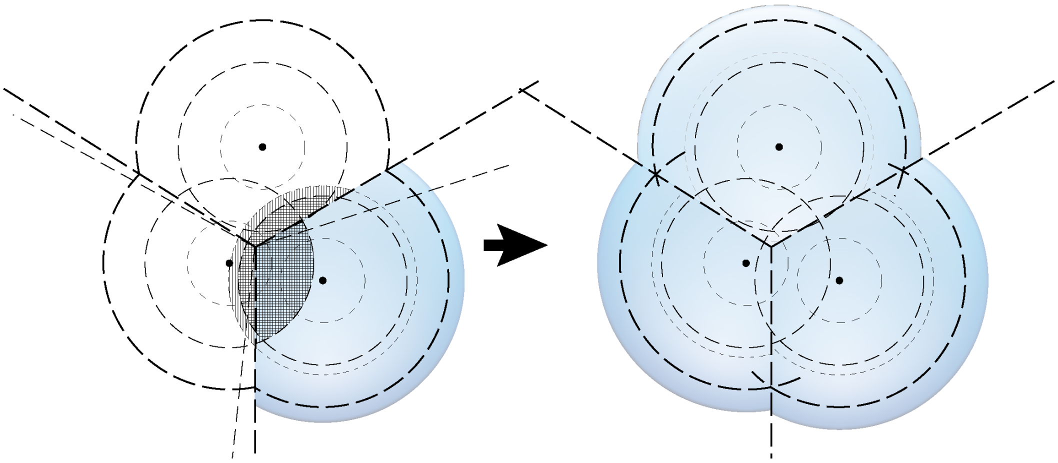


Figure 1.3: Idealised cluster on very unproductive grounds. The hatching would show overlapping ranges, if these where perfect circles, and dominance could be expressed by a shift in their interface.

If figure1.2 represents a community of three isolated (non-clustered) nests, then figure1.3 represents a troubled cluster. The sign for this would be, the expansion of their home range (in solid fill). Let us image that perhaps intense weather phenomena have reduced the productivity of their environment; perhaps human activity has removed far too much biomass (including food) from surrounding vegetation.

Either way this very sparse settlement in low numbers indicates an unfit location for dormice. Even if I did not have data substantiating their increased home range, finding a sparsely inhabited cluster, containing few inhabited nests should indicate a troubled community. Be it an incidental situation or part of an evolving trend, such arrangement, and specially the combination (low number and low density inhabited nests) marks an unlikely situation for a new forming and growing settlement. In Lithuania, a land that sustains healthy dormice communities, densities have been recorded as low as 1 individual/ha, which is why I drew a theoretically oversized home range (in the above figures) at 56.4 m. Such density is considered low, and it is found after winter mortality has thinned the spring population, which endorses the diagnostic. A typical population entering hibernation in autumn has a density of 3 individuals/ha in Lithuania. (Juškaitis 2008) Having nuts in the diet, dormice should not roam farther to find sustenance.

Therefore, in an opposite way, I should argue that wherever clusters are found containing a large number of nests arranged compactly, the area must be suitable for dormice, since its pressure upon natural resources is uncommon in declining environments. In those situations the population of mice would have self-regulated its numbers in terms of density. (Juškaitis 2008) Thus, in highly compact clusters either the land is very productive, or its suitability extends further than the minimum boundary of the cluster, which allows the home range of certain individuals to extend farther, but at an energy cost that is still balanced out by their caloric intake. After this digression into the self-regulatory mechanism that should allow spatial analysis to weight the data on the base of cluster spatial characteristics (see section2.2.3), there is still more useful climatological traits to describe.

We know that temperature is one of the main parameters determining the torpor state needed for hibernation. "Dormice were found torpid when ambient temperature was $\leq$ 14–15°C. This coincides closely with the results of Eisentraut (1956), who established that the critical temperature for *M. avellanarius* to fall into dormancy was about 15–16°C." (Juškaitis 2008) We also know, even if we cannot give it value to how significant this fact is, that they rather live in areas with small values of isothermality, which means that the daily thermal range is reduced. This piece of information (which I cannot establish as a fact for the lack of time data) does not conflict with the fact that dormice need rather cold winters. "According to Likhachev (1966), M. avellanarius abundance is determined by mortality during hibernation. This in turn is dependent on several factors: feeding conditions in autumn and especially winter climatic conditions." (Juškaitis 1995) Consequently steadily low temperatures in winter (i.e. with low variability during the day) are a key to search for. To give it a sense of scale where conditions a suitable to dormice, in Lithuania the average annual temperature is 6.2°C, and the lowest temperature in January averages –5.1°C. In other latitudes of *M. avellanarius* range, it is found at higher altitudes seeking winter rest.

> *"M. avellanarius can reach an altitude up to almost 2000 m above sea level. The highest record of M. avellanarius was in Macedonia at the altitude of 1980 m a.s.l. (Kryštufek & Petkowski 1990). In the Austrian Alps, three localities were at the height of 1801 m, 1850 m and 1920 m a.s.l. (Spitzenberger & Bauer 2001). M. avellanarius was found up to 1900 m a.s.l. in the Tatra Mountains (Anděra 1987), up to 1700 m a.s.l. in Croatia (Tvrtković et al. 1994), and up to 1600 m a.s.l. in Sicily (Sarà & Casamento 1994)."* (Juškaitis 2008)

Yet again an apparently disconnected piece of information from Denmark but, even though Denmark does not have mountains, this fact led me to investigate for the relevance of elevation in the Danish communities, and even became relevant to the selection of the research area.
However a successful survival of dormice does not only come down to favourable winter conditions. As aforementioned, the feeding conditions during the autumn are as well of relevance. This refers to the quality in calories of the diet, but also to its availability, and most importantly to the climatic conditions in autumn, and previously during the summer. Remember that a dormice population is in function of the climate directly and indirectly, through the effects of weather on the timing and abundance of insects, flowers and fruits. (Bright and Morris 1996)
Juškaitis (2008) found that favourable environmental conditions in autumn translate as experiencing a dry, warm and long autumn. "The average temperature was 10.4°C in September, 10.3°C in October, and 6.0°C in November; rainfall amount was 22.7 mm, 1.9 mm and 47.4 mm, respectively." Under such conditions not only mice are capable of finding food, food is available and they remain longer active, but also for a very tiny rodent of up to 20 g of weight, it is a very different task to climb through thin wet stalks rocked by the wind. Weather conditions must have a direct impact on mice. As Fischer (2014) argues in her master thesis, rain and wind have a direct effect by rapidly cooling the mammal, as well as by making the journey more costly. In Lithuania "the annual precipitation averages 675 mm with variations from 520 mm to 900 mm in different parts of the country." (Juškaitis 2008) This range is also true for their fellow dormice living in Britain. Bright and Morris (1996) describe their distribution in Britain to be consistent with location that have less than 200 rainy days per year, with the only exception of the Cumbrian population. Furthermore, they add, "[d]ormouse sites lie in areas which receive an average of less than 100 mm of rain in August (Woodley, 1980)." All this climatological data will inform the filter as a mask. Both precipitation and average temperatures will be used on the selection of suitable grounds for the dormouse. Even though Bright and Morris (1996) argues that "a measure of accumulated temperature might be more relevant to the Dormouse", a count of 'degree days' above a defined baseline instead of averaged data. This will be perhaps the task of yet another Major Thesis. For now, seeing how well the data fit, I will have to do with measure of central tendency. Dormice in Britain dwell within the 16.5°C isotherm

for July. There are few exceptions, in the localities of Northumberland and Cumbria, but this datum is also confirmed in Lithuania, where the average temperature in the month of July is 16.7°C. Furthermore, this is also the case of the data available in Denmark, as we can see from the map in the next page, figure1.4
In the actual selected area, there has been given some leeway to the important climatological variables rain and temperature. If in principle temperatures are more favourable in North-east Sjælland than in the South-west during the summer (there is an increase of Average Mean temperature for the month of July of 0.5°C, which is confirmed in average for the aroused period between April and August), still autumn temperatures are lower in north-west Sjælland. The averaged temperature for the vital months of September and October, when mice fatten up on most energetic nuts like hazel, is the time when temperatures are 0.5°C lower, by 11°C. This is an important difference, for the summer, even though it would see an increase in mice activity, is not the time of the year when mice put up most of their weight. The calorie-rich diet that they use to build winter reserves, would now be found in more disadvantageous months, on average, on temperatures lower than their torpor threshold. Furthermore, the amount of precipitation in summer will also be unfavourable, due to its significant increase (15mm during the summer moths with respect to the South-west). Under such circumstances dormice might remain torpid during colder nights in autumn, meaning fewer hours available to collect nuts. At the same time, if they were supposed to gain weight during the summer, that might not work. Not just because of being a less efficient way of gathering calories, due to the type of food, but because such food could also be thwarted. The summer presents increased amounts of rain, which could cause a lack of pollination and pollinators. North-west Sjælland might turn up to be a challenge.

1.1.2 Other valuable traits for selection and filtering

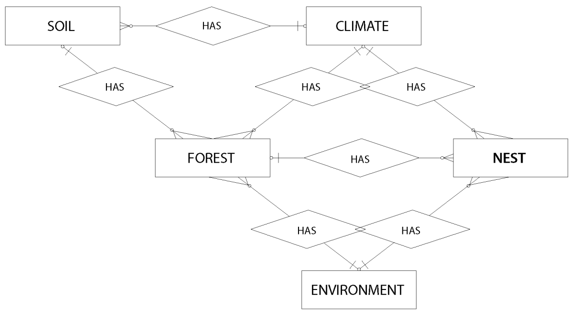
Survival is not that simple. It cannot be completely explained by spatial analysis. Therefore I will use a spatial filter, and I will query variables individually. In order to do a proper modelling job, one should know much more on the ecology, even the physiology of the species, besides having extensive data on *M. avellanarius*. For instance, after I had taken into account all possible spatial parameters for individual survival (which surely I may have not); after I had calculated fat reserves and their probable exposure to climate-driven fluctuations, which I have not, despite the fact that these compromise dormice survival by inducing a metabolic increase (Pretzlaff and Dausmann 2012), then I would still be left with the fact that "age structure of the autumn population is also very important for dormouse mortality during hibernation. Winter mortality was correlated to age structure in the autumn, i.e. percentage of young in the population ($r = 0.84$; $p < 0.001$)." (Juškaitis 1995) It is therefore, that this is not a model on survival, but a filter on the spatial signature of the species. It should only be asked to deliver truthful measurement.

Figure 1.4: Map depicting isotherms for the averaged temperature in the month of July. Given that "Dormice were found torpid when ambient temperature was $\leq$ 14–15°C." (Juškaitis 2008) and that this is a nocturnal species, this map in conjunction to fig1.1 explains the strong climatological reasons for being Denmark at the fringe of the dormouse distribution. Eisentraut (1956) established the critical temperature for *M. avellanarius* to fall into dormancy at 15–16°C. Thus areas with fresh climate and large thermal range, make for a very bad combination, where the amount of hours for feeding during summer and autumn months are very limited.

In the direction of spatial analysis, there are still some other important spatial traits, which depend on soil and the composition of the vegetation. For instance, "three plants identified by the Mammal Society survey (Hurrell and McIntosh, 1984) as being most strongly associated with the Dormouse are Hazel Corylus avellana, Honeysuckle *Lonicera periclymenum* and Bramble *Rubus fruticosus agg.*" (Bright and Morris 1996) Also the structure of the canopy and understory matters.

> *"They are nocturnal in habit and wholly arboreal, spending more than 90% of their time at 2 metres or more above the ground in spring and summer. In the autumn, they will come lower, particularly to exploit bramble (Rubus spp) fruits, but still avoid activity on the ground (Morris & Bright, 1993). They prefer to move among trees having plenty of near horizontal branches, not in thin tall trees growing at high densities, with a predominantly upright structure."* Bright and Morris 1995

If the ERD was to be drawn as close as possible to the natural environment, it would have the following relationships: Nests would have to be at its centre, and their success would be in function of chiefly two independent variables (i.e. climate and the environment). They would relate directly to the climate, but also indirectly, for the climate intervenes in the formation of soil, as well as it directly defines the species present in a forest. Nests would also relate to their environment, which basically here would be left to the already discussed parameters on the physical geography surrounding them. This can be seen down below in figure1.5, which represent a conceptualisation of the relationship established in the real world.



Actual ER-Diagram

Figure 1.5: Entity Relationships Diagram (ERD). It is a diagram that defines the structure of the database. This would be its closest layout to actual reality, if data was available and other questions, like the kind of questions to be made on the database did not affect it. As I will reveal in the following chapter2, the final adopted ERD looks substantially different.

1.2 Data Exploration

- Data, Coordinate Ref System, Authority ID, Raster Resolution:
  DTM, ETRS89 / UTM zone 32N, EPSG:25832, 0.4 m
  DSM, ETRS89 / UTM zone 32N, EPSG:25832, 0.4 m
  ALS, ETRS89 / UTM zone 32N, EPSG:25832
  Aereal Orthophoto, ETRS89 / UTM zone 32N, EPSG:25832, 0.125 m
  Land Parcels, ETRS89 / UTM zone 32N, EPSG:25832

- Data from Styrelsen for Dataforsyning og Effektiviserings. AWS Suite (Adresse Web Services), Coordinate Ref System, Authority ID:
  Address Data, ETRS89/ETRS-TM32, EPSG:3044
  HWSD Soil Data, WGS 84, EPSG:4326, 0.008333 deg
  Climate Data, WGS 84, EPSG:4326, 0.008333 deg

- FOT (Fælles Objekttype) data, Coordinate Ref System, Authority ID:
  np2b2013fredskov, ETRS89 / UTM zone 32N, EPSG:25832
  Kommune Grænse, ETRS89/ETRS-TM32, EPSG:3044
  Municipal Grid, ETRS89/ETRS-TM32, EPSG:3044
  Kommunekort, ETRS89/ETRS-TM32, EPSG:3044
  Bykerne, ETRS89/ETRS-TM32, EPSG:3044
  Erhverv, ETRS89/ETRS-TM32, EPSG:3044
  Lav Bebygelse, ETRS89/ETRS-TM32, EPSG:3044
  Høj Bebygelse, ETRS89/ETRS-TM32, EPSG:3044
  Gartneri, ETRS89/ETRS-TM32, EPSG:3044
  Overflade, ETRS89/ETRS-TM32, EPSG:3044
  Vejmidte, ETRS89/ETRS-TM32, EPSG:3044

- Satellite Imagery, band, Coordinate Ref System, Authority ID, Resolution:
  Sentinel-2, Band 2 (Blue), ETRS89 / UTM zone 32N, EPSG:25832, 10 m
  Sentinel-2, Band 3 (Red), ETRS89 / UTM zone 32N, EPSG:25832, 10 m
  Sentinel-2, Band 8 (NIR), ETRS89 / UTM zone 32N, EPSG:25832, 10 m
  Sentinel-2, Band 8 (SWIR), ETRS89 / UTM zone 32N, EPSG:25832, 20 m

- Inherited dormouse data:
  dMouseData_subsets, ETRS89 / UTM zone 32N, EPSG:25832

Thus there were initially three reference systems, although two could be considered compatible. These must be unified to (EPSG:25832) a Universal Transverse Mercator projection for zone 32, expressed in metres:
EPSG:25832 (+proj=utm +zone=32 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m)
EPSG:3044 (+proj=utm +zone=32 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m)
EPSG:4326 (+proj=longlat +datum=WGS84 +no_defs)

1.3 Research Questions

This exercise was thought to have two goals: The first one is to establish a spatial suitability signature for the Dormouse, driven by geometrical and climatological data, which should inform a suitability filter; the second is to identify potential areas for dormouse habitat in Hovedstaden, the capital region of Denmark, by using that filter. In order to achieve these goals, I need to keep receptive to the natural limits the species has, specially because these shift, are difficult to measure, and could be described as vague. The species has been reported to have benefited from being near humans. There is well documented evidence (Bright and Morris 1996, Juškaitis 2007), that the dormouse is associated with forest coppicing. This bond is explained by the fact that coppicing rejuvenates the forest, which provides a denser understory, i.e. it provides shelter. However, the amount of disturbance created by humans has a limit, and the current decline of the species might attest to that. Not only its habitat has shrunk since the time of the industrial revolution, but that and its following technological revolutions have changed forest management practices, our socio-economic culture, the degree of technology and energy mobilised in our endeavours, as well as the very atmospheric chemistry. Therefore, the causes for its decline might be varied, but it is currently clear, that its relationship to the environment is nuanced and out of balance. Furthermore, we should also mind, that its current distribution makes of the dormouse a fringe species in Denmark, which also speaks to its vulnerability. Not only habitat loss impacts dormice at the margins of its distribution, as it would any where else, but environmental fluctuations perhaps even more so. Thus, what I ask in this study is:

- Under the current degree of development, and their inherent intensities, would there be space for a dormouse enclave within the metropolitan area of Copenhagen?
- Given that this is a bioGeographically marginal area for the distribution of the species, a cause for the decline of the species could be seasonal bottlenecks associated to, for instance, the North Atlantic Oscillation phenomenon (NAO), or, in the near future, the disturbance of climate change. Is there the slightest hint in my data, that dormice could have been impacted by global climatological phenomena, or that they could be impacted by global climate changes in Sjælland?
- And finally, which traits of their environment are key, or part of the signature that best characterises a habitat filter for *M. avellanarius*?

Chapter 2

Material and Methods

2.1 Data

As aforementioned, this exercise follows a previous research that had collected data on the types of trees surrounding nests, their variety and abundance, the density and the degree of cover provided by their canopy, their understory, the amount of natural light present and its soil moisture. It had created a rich database from the point of view of a biologist, on existing conditions at a local level. However this is not very efficient from a geographical perspective. So, this report is a very different endeavour. It does not pivot on the dormouse, but on the relationships that characterise its habitat. So, before I enter into details on the data used, I will dedicate some time to explain the actual Entity Relationship Diagram used.

The method has sorted the data into three stages, which are carried out at three different scales, due to the spatial resolution of the available data. These three scales were chosen as multiples of each other, in order to allow the data sets to fit in an overlay. Figure2.1 depicts the network of relationships that were used to inform the actual database in this exercise, which notably differs from figure1.5. Among the most significant changes is the insertion of the climatological parameters as attributes of the entity 'environment'. Their insertion into the environment table has to do with the lack of an specific primary key (i.e. a single identifying attribute), which prevented them from meeting the normalisation requirements, and from standing in their own table. Moreover, this meant that, even though its spatial resolution was around a 700 m pixel, it was introduced at this stage at 300 m, for the purpose of assembling them into the environmental and soil data.

So the first stage is driven by selective metrics, named because of their use in an early selection of potential new areas. A second stage uses filtering metrics, which are a high resolution (pixel size of 0.4 m) set of measures on spatial relationships. Finally there is a set of prescriptive metrics (at 20 m pixel), since these are non-antagonistic, adaptable parameters, toward the habitat of *M. avellanarius*.

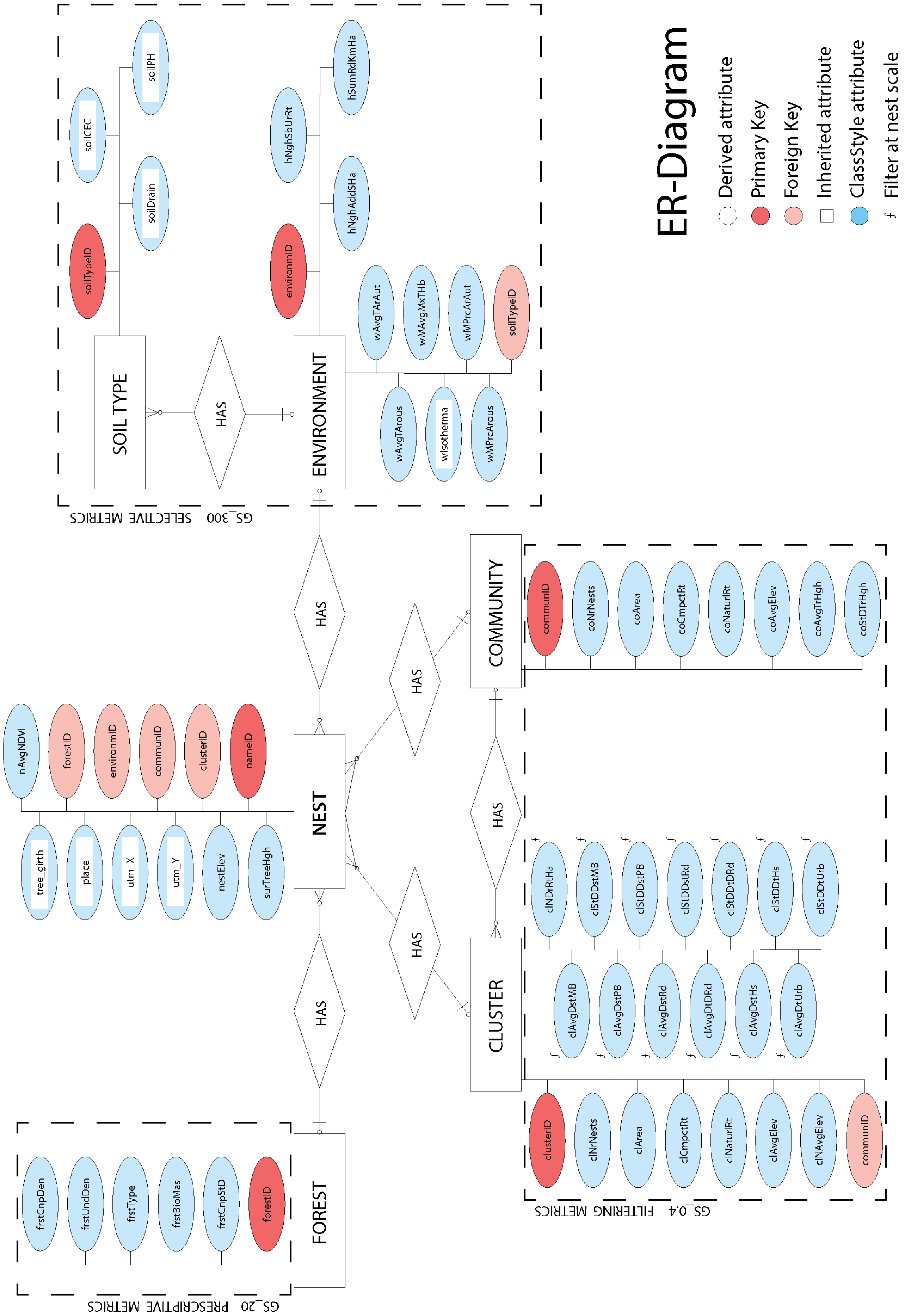


Figure 2.1: ER Diagram used to model the actual network of relationships found between the most basic entities at play, which include the dormouse nests and the variables that explain their location.

2.1.1 Climate

The climatological data was downloaded from a website that specialises on ecological modelling[1]. WorldClim provides free Global Climate Data on the Current conditions (down to 1960-1990). It is offered as Creative Commons License and the highest resolution global data available is at 30 arc-seconds (below 1 km) resolution, which can be downloaded by tile. Other available resolutions are 10, 5 and 2.5 minutes.

For this exercise, or any other 'local' study, the best suited resolution was the 30 arc-seconds, which downloads a .zip file containing a tile of WorldClim at 30 x 30 arc-seconds. One must first select the appropriate tile (which in this case here it was the number 16), and then select a variable (or parameter) to download and a file format. The available variables are the monthly minimum temperature (ºC * 10), the monthly maximum temperature (ºC * 10), the monthly average temperature (ºC * 10), the monthly precipitation (mm) and other various bioclimatic variables. The available file formats are two: one is a generic geoTif raster file, and the other an ESRI grid.

The Climatic Variables

The raster layers that can be downloaded include variables like monthly total precipitation, and monthly mean, minimum and maximum temperature, as well as 19 derived bioclimatic variables. These are bioclimatic ready-made raster variables, which are 'derived' in the sense that these result from the monthly temperature and rainfall values. They are meaningful variables, for they are often used in species distribution modelling, but the time frame in which these compound the climatological values does not satisfy this exercise. The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) aspects of seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). (Hijmans et al. 2005) These variables are then packed in 'quarters' or periods of three months, instead of according to the year cycle of the dormouse. Therefore, from all nineteen coded variables I only used one, namely isothermality, which was of interest only in the very early selection process of suitable research areas, but was not introduced in the filtering process. It would have been meaningless, because of its very strong correlation. As commented earlier.

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) (* 100)

BIO7 = Temperature Annual Range (year absolute max - year absolute min)

---

[1] Check http://www.worldclim.org/

2.1.2 Climatological Preprocessing

The way I used this data to build the climatological attributes in the database was, by taking the monthly precipitation and average temperature and compounding them into meaningful periods of time for *M. avellanarius*. The cycle of this species is marked by hibernation, as well as by the phenology of the plants that sustain it. For example '*Arous' stands in the database for 'arousal period', which comprises the vegetative season from April to the beginning of September. Then the database has a different attribute '*ArAut', which stands for 'arousal period in autumn', and comprises data from both months of September and October, for, as I have argued before, these months are crucial in preparing the dormouse for hibernation. It is at this point in time that hazel and other nuts, very rich in calories, are available.

With respect to the downloaded data layers, one must be aware of its quality with respect to the scale at which it will be used. The data were generated through interpolation of average monthly climate data from weather stations on a 30 arc-second resolution grid (which almost corresponds to 1 km^2 resolution). (Hijmans et al. 2005) This is almost obvious from the display of variables like Isothermality in figure1.1, specially in Sweden more than in Denmark. The jaggedness of the edges of each isothermal band is a product of that interpolation.

According to Hijmans et al. (2005) the WorldClim interpolated climate layers were made using: Major climate databases compiled by the Global Historical Climatology Network (GHCN), the FAO, the WMO, the International Center for Tropical Agriculture (CIAT), R-Hydronet, and a number of additional minor databases . . .

The very same website states, that the averages calculated for the 1960-90 period were only carried out for those stations which had records for multiple years. Plus, from that 30 year period, the average was only carried out where there were at least 10 years of data. The fact that reaching a spatial resolution of 700 m required interpolation did already create some artefacts, and the lack of data might have reduced the quality of the measurements, since these are measurements of central tendency. However, having climatological data at a 700 m resolution and from a period extended in some cases to the 1950-2000, where data was available is as good as one can wish. On the other hand there was the possibility of centring the data locally, on the extent of the research areas, but this would not have improved the resulting accuracy of attributes. First of all, DMI data is not freely available, which meant I had to copy the data myself, from also discontinuous measurements, from three stations per area, which shifted some times, and with very little support for precipitation data. Not only I had a long task in front of me, where I would as well need to interpolate between stations; I did not have the time or patience to remove stations with errors, as they did. Though it covers the entire world, this database was far better than what I could produce, consisting of precipitation records from 47,554 locations, mean temperature from 24,542 locations, and minimum and maximum temperature for 14,835 locations. (Hijmans et al. 2005)

2.1.3 Soils

Soil data was downloaded from the International Institute for Applied Systems Analysis (IIASA), an institution that conducts scientific research into environmental, economic, technological, and social issues. The data downloadable from their page[2] is supposed to be accessed from their own software, the HWSD Viewer (which stands for the Harmonized World Soil Database Viewer). It is freely available for use and dissemination of its information only for educational or other non- commercial purposes. The actual property of the data is held by several other institutions: the Food and Agriculture Organization of the United Nations (FAO), the very same host (the International Institute for Applied Systems Analysis), the ISRIC-World Soil Information, the Institute of Soil Science – Chinese Academy of Sciences (or ISSCAS), and the Joint Research Centre of the European Commission (JRC).

Again this is a global dataset, in which information from different databases have been pieced together, and can be downloaded as a .zip file. It contains a .bil file (i.e. Band interleaved by line) that I opened in Quantum GIS (v. 2.18.4) to then be exported as a raster in a geotif file. Such file format is nothing other than a binary file that holds the image data, in this case, a single band. The spatial resolution of the data is similar to the climatological data. It corresponds to a 1 km^2 resolution in some parts of the globe. Furthermore, similarly to the climatological data set, this one has also been harmonised: All fields have been checked for outliers or data errors. Both errors and missing data were corrected by using neighbouring units consisting of the same soil type. FAO et al. 2012

The data has a primary key, which allows it to be inserted into the database of this exercise as an independent table. Also, because soil data tends to have a very high correlation, same as climatological data, this table does not need for too many rows. Soils are a continuous phenomenon that remains very stable for very long distances. This is, once it has been generalised to classes. Of course, soil presents small variations and is not perfectly uniform in thousands of hectares, yet, for the purpose it serves, texture or pH can be considered to remain stable and thus generalisable for very large swaths of the territory. Soils do present significant changes, or rather abrupt boundaries, when geomorphological processes create distinctive features, like the succession of terraces by a river. The main explanation to why soil characteristics are so constant is, that soils are the combination of parental material (i.e. local rock), plus the action of weather and living organisms on that parental material.

Because soils impact *M. avellanarius* indirectly, through the kind of vegetation they support, and its productivity, despite the wealth of attributes on these data, I will

---

[2] The Harmonized World Soil Database is available at http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/HWSD_Data.html

only use three hand picked attributes, which relate to the availability of nutrients and water on the ground: pH, drainage, and CEC or Cation Exchange Complex (which relates to the ability to retain nutrients on the soil). Furthermore, the HWSD attribute database establishes two subdivisions on the soil profile. The database provides standardised soil parameters for top- and subsoil. (FAO et al. 2012) I will only use top soil parameters, for only one reason: the time available for this research. Although the fertility of a soil cannot be only established by its top soil, the real difference comes only down to the root system of the established vegetation. Being central the figure of the dormouse in this exercise, being hazel nuts and beech seeds its main high-calorie staple, and having both trees a very superficial root system, I took the decision to only contemplate the top soil.

Moreover, in this exercise soil will not be at the core of the filtering, for it only has a much smaller impact than weather and the environment. It will only be use in the preliminary selection of research areas, which simply need to be similar enough. That will not be very difficult. The HWSD attribute database provides information on each soil unit (there are 15773 soil mapping units), but, because of being such a stable phenomenon, between both research areas appear only four different classes of soil. Three of them are very similar: two cambisols and a luvisol, which have similar characteristics. Cambisols are still weakly developed soils, which means that their horizon differentiation just begins to be evident[3]. Their main difference from Luvisols is, that these have an accumulation of high activity clays and high base saturation deep below the surface, which have been washed down the profile of the soil[4]. Both have a medium to fine texture and a neutral pH. The fourth soil, appears only marginally in the north-east research area. It is a podzol, which has a very different chemistry, for it has a much lower pH due to the presence of organic acids. Because Podzols tend to be capable of retaining less nutrients and water[5], this will be the only rejected soil class.

2.1.4 Parcels

The data set consists of parcel data extracted from the Geodatastyrelsen Matrikel map. The data has been corrected to show coincident mutual boundaries. However, their geometric accuracy may vary. The parcel map is updated on an ongoing basis, as part of the registration of parcel changes, but the data containing such corrections is only made available for download every 3 months, which is more than sufficient for the intent of this exercise. In overall, although the accuracy of boundaries is of relevance, it will not present a big problem the use of this dataset. It is structured according to the INSPIRE object types "CadastralParcel", "CadastralBoundary" and "CadastralZoning".

---

[3] See http://www.fao.org/docrep/003/Y1899E/y1899e08.htm

[4] See http://www.fao.org/docrep/003/Y1899E/y1899e12.htm

[5] See http://www.fao.org/docrep/003/Y1899E/y1899e12.htm

2.1.5 Municipal Boundaries

This is yet another set that does not have a big impact on the project. The data was collected from the FOT data, and the main reason to gather this data is to test the significance of humans in the location of dormice clusters. The idea to use this data comes from a conversation I had with a biologist working at Kampinos National Park. Kampinos is a national park on a flood plain of the river Vistula, right after passing by Warsaw. It is a few kilometres close to the capital and, despite the idea we might hold about the purpose of a National Park and the way fauna use it, there was a goshawk population (*Accipiter gentilis*) that lives in the woods of the park, but tends to have their nest very close to the border neighbouring Warsaw. This was not a coincidence. I was told that the number of eggs laid did correlate with the numbers of city pigeons in the surrounding suburban environment. This is a bird, which we have for a wild species, and by wild we also imply its condition of being the other, yet showing a certain level of synanthropy, though being wild. For such species intangible borders like those of the National Park do not play a role, and the actual continuum from forest into urban is just that, continuous. Other species do tend to notice the protection that borders have in the human psyche.

2.1.6 Forests

Although one would think otherwise, this data set had very little impact on the project. The data was collected from the FOT data, and it referred mainly to what is called 'fredskov', or protected forest. In the main research area, where dormice live, close to 100% of the forested area has been protected. So there was not much one could do with this piece of information. The main reason to gather this data was to test the accuracy of the forest classification done in *eCognition*, see 2.1.9.

2.1.7 Environment

The label environment stands for very disparate data that describe a specific part of the dormouse's milieu. That is, alien elements like infrastructure, urbanity and houses. These element represent an influence of the dormouse surroundings that could be beneficial, in the case of rural labouring like the coppicing of hazel trees; or mostly and impact like the intensity of use of surrounding surfaces.

Some papers have reported on the ability for dormice to cross barriers like roads up to 15m wide; (Chanin and Gubert 2012) others even report on their suitability as nesting areas (Schulz et al. 2012). In our very research area two clusters of mice appear at the embankments of the railroad. However, urban density proves to be problematic for their settlement, which is the main research topic in this thesis. Thus the value of these datasets. The roads belong to, yet again, the FOT data; and the municipal addresses dataset, which is a point dataset, belongs to the AWS Suiten, downloadable at Kortforsyningen.

For this study I used the line features that depict the mid-axis of roads. The relationship between roads and the nesting areas for dormice is established by the fact that 10m wide road can already be considered an edge of the forest. A paper by Williams et al. (2013) reveals by conducting a regression, that the distance from edge of the wood is a very significant measurement ($p < 0.001$) for finding nests. So roads fragment territories, though not for mice when these are smaller, countryside roads. Conversely these can be an opportunity to create an edge. Thus this data had to be supplemented by the attributes that establish what kind of road we are handling. Better than road width, or much better than road class, the best attribute to establish road impact was their surface finishing: dirt roads/ paved roads.
With regard to the probable intensity of use of the territory by humans, houses was not the best indicator. Polygon vectors on urban centres and suburban areas were downloaded, plus data on addresses for all the involved municipalities (including those that had their boundaries overlapped shortly).

2.1.8 Environmental Preprocessing

Previous to using the data, it is important to perform a table join, in order to obtain the attribute for road surfaces into the downloaded middle of road line data. Then, with regard to the use of urbanity or sub-urbanity, one must define them. There is no such data available. Thus I used FOT data on land uses, where urban centres (i.e. bykerne) and high-rises (i.e. høj_bebygelse) are available. The merged polygons, corresponding to these two classes, were then used to configure areas of high human activity as well as low presence of undisturbed vegetation.
Sub-urbanity was then defined similarly, by merging other land uses. In this case were low-rise buildings (i.e. lav_bebygelse), which includes houses, plus garden centres (i.e. gartneri) and businesses (i.e. erhverv) , which include logistics, quarrying and other activity.

2.1.9 Imagery

For this exercise I downloaded two sets of Sentinel imagery. Both correspond to the satellite number A (i.e. Satellite Sentinel-2A), which is the one equipped with a multispectral camera (i.e. dedicated to 13 different bands of the electromagnetic spectrum) that records from visible light up to Short Wavelength Infrared (SWIR). ESA 2000 There is also available data from an older satellite of the sentinel mission. Both have been launched and are operative in orbit, however Sentinel-1 retrieves data by a Synthetic Aperture Radar (SAR) instrument. ESA 2000 Such data serves no purpose to this exercise, since the very part of the electromagnetic spectrum it uses, radio waves, have a far longer amplitude, which establishes that the spatial resolution cannot be of any use.

Sentinel 2A

The data was downloaded from the Copernicus Open Access Hub, hosted by Scihub[6], after selecting Sentinel-2A data only, and after searching for a maximum ratio of cloud cover of 12% through the months of April to August. The reason to select those months is, that they cover the months of vegetative growth that correspond to the same period of arousal in dormice, meaningfully separated from the autumn period of arousal.

One important piece of information to keep in mind is, that the sentinel data is classified into levels. The raw data is a level 0, which contains all information required to develop it from a compressed status (in which it is transmitted to the ground), into all subsequent levels. ESA 2000 The Payload Data Ground Segment (PDGS) is responsible for the systematic processing and archiving of Sentinel-2 up to Level-1C, which is the lowest distributed level of data. Until this point the data has been uncompressed and corrected radiometrically and geometrically. So Level-1C product provides orthorectified Top-Of-Atmosphere (TOA) reflectance, with sub-pixel multispectral registration. ESA 2000 At this point it is required to process the data to a level 2A, which introduces atmospheric correction, to finally obtain the Bottom-Of-Atmosphere reflectance (BOA) in each of its 13 recorded bands. This step can be carried out in the Sentinel-2 Toolbox can be downloaded from the Science Toolbox Exploitation Platform (STEP)[7]; or even in QGIS, as I did, after installing the Semi-Automatic Classification Plugin. This plugin allows for the classification of remote sensing images by means of using the metadata attached to the downloaded imagery.

In all I downloaded two images. The first one was selected from the early spring. The reason is that an early spring image would help to discern between broadleaved, coniferous and mixed forests. Nevertheless, Sentinel-2 is a rather new satellite in service, so there was few years of data available. The best chance there was for these areas to be covered was April 2016 (i.e. Date: 2016-04-22T10:20:25.000Z), which had a cloud cover percentage of 3.75168%. However small, this percent was not enough to prevent an entire cluster to be covered by a very small, local cloud. So, the second image was taken from the month of July (i.e. 2016-07-07T04:07:44.000282Z), which is good for two reasons, it is a month of high vegetative activity, but also being from 2015, set the time very close to the generation of the LiDAR data used in this exercise, as ell as the time of the survey on south-west Sjælland for dormice, which informs my database. Even though the cloud cover percentage was higher than the previous (i.e. 10.614828%), this time the entirety of the area was free of clouds.

---

[6] Visit https://scihub.copernicus.eu/dhus/#/home

[7] Visit http://step.esa.int/main/download/

2.1.10 Imagery Preprocessing

In order to be able to use the data, it has to be preprocessed to BOA reflectance values, which means that, at least using a transformation that takes into account the effects of the atmosphere on the actual reflectance at the ground. A second correction, known as DOS1 (i.e. the Dark Object Subtraction) can as well be performed on demand. It is an image-based atmospheric correction, which means that has a lower accuracy than the physically-based correction. So the correction is not based on data like atmospheric transmittance, or other known atmospheric phenomena like diffusion. Instead:

> "the basic assumption is that within the image some pixels are in complete shadow and their radiances received at the satellite are due to atmospheric scattering (path radiance). This assumption is combined with the fact that very few targets on the Earth's surface are absolute black, so an assumed one-percent minimum reflectance is better than zero percent". (Chavez 1996, Congedo 2016)

After correction the images are ready to be used. In this case the main reason to download these images was to calculate factors like NDVI (i.e. the Normalised Difference Vegetation Index), which would help describe and quantify habitat from the point of view of its vegetation. The use of images with low cloud cover is because of the fact that these do not only have low NDVI values, which block the actual ones on the ground, but also that their shadows impact the NDVI values of exposed ground, since they block light (better said, they absorb most infrared radiation) differently for the Near Infrared and the red spectrum.

Further uses of these images could be the assessment of the vegetation's health, and the classification of vegetation. However, the fact that the platform is in orbit conditions very much its spatial resolution. Sentinel-2 instruments have three different spatial resolutions, depending on the length of the wave used to measure. It goes from 10 m to 60m, which is very coarse for the type of classification I intend. It is therefore, that for this purpose I downloaded some orthophotographs containing a NIR channel besides the usual RGB visible spectrum.

2.1.11 Orthorectified Aereal Imagery

From Korforsyningen I downloaded two sets of orthorectified images, one for each research area. The main reason for this is that the platform is now an aircraft, flying very low in comparison to a satellite, which enables for a spatial resolution of 0.125 m. With that kind of resolution I can use OBIA classification software (i.e. Object Based Image Analysis, which does not use just pixel calculations, that is only for an initial step of segmenting the image into objects, or groups of pixels on which to use textural and spectrometric analysis) like *eCognition*.

The data is now packaged into 10km squares, which means that one must make sure that the research areas are completely covered. In case these are split between two, four, or even six tiles, then it is necessary to create a mosaic in a geodatabase (when using *ArcMap*), as a preliminary step to the clipping of the research areas. Next, and before carrying out any classification, it is useful to evaluate the spectral distance, i.e. the radiometric separation between training signatures. What kind of materials are we trying to categorize? and what pixel values these present? If different classes of land use/materials were too similar, this could cause classification errors. In order to avoid them and sometimes enhance or increment that spectral distance, it is recommendable to weight differently some bands (e.g. the Near Infrared band has often a bigger weight when classifying vegetation against hard pavements). In this exercise this is very present, since I try to classify vegetation between coniferous-deciduous-mixed. For this, given that there is always some correlation between visible bands, the NIR band will be weighted double.

2.2 Methods

From all the data in the new database only about one in seven attributes has been inherited (8/55). All other attributes have been calculated for this specific database, having in mind the specific requirements a dormouse has on its environment. This means that most of the previous work on the dormouse has not been taken into this exercise, only the nest locations, for this exercise has a very different character. I almost exclusively use spatial methods like spatial correlation, overlay, distance, size, slope, compactness . . . Yet this does not mean, that the way in which the former data was collected, would not affect the current methodology. It did.

2.2.1 The Work Area

The first step on was to establish a research area encompassing the available nest data, plus another area within the metropolitan belt of Copenhagen. Both areas were supposed to be similar in size, similar values in selective metrics, but they were supposed to have a sizeable difference in urban metrics. The metropolitan area contains 898 grid cells containing at least some suburban land use, while the one on the country side only 632 cells. That is a difference in suburban extent of a 30%. Thus there is some difference, although such difference should also be of urban character. It should incorporate the presence of varying intensities (typical of urban environments) to that of extent (typical of suburban areas). The high-rise tenements, the old urban cores, the suburbs . . . The goal is to find out, if such heterogeneous growth still allows the presence of *M. avellanarius* in its interstices. Thus the selection of a work area must encompass a minimum degree of difference, while its filtering must accept not just central tendencies, but also extremes.

Both areas have in common a certain degree of suburbanity, though the amount of urbanity is very limited on the country side. From the preliminary results on suburbanity (*"hNghSbUrRt"*), the metropolitan area contains up to 30 cells fully packed with suburban development, while the country side counterpart contains only nearly three. That shows a ratio that would suggest the metropolitan area is 10 times more dominantly suburban. Furthermore, the 632 cells on the country side have a Mean value of sub-urbanity equal to 0.0808%, with a standard deviation of 0.1593. In this area even sub-urbanity is very light footed. Nevertheless the distance to urbanity in the country side (*clAvgDtUrb*) would almost clog the size of metropolitan interstitial spaces. Such differences will do for the required contrast.
To this methodology it is important to select a contrasting area, but one that still establishes a contact (by means & StD). One that can be filtered by the impact of surrounding areas on them, and still show interstitial spaces to be analysed by their size. For instance, within the metropolitan area more than a half of the cells containing some suburban development are below the mean value (483 of 898 cells have a value below 0.0808%). This indicates, that even metropolitan areas contain interstitial spaces in contact with lightly developed cells. The missing piece of information is whether the size of such spaces would still do.

2.2.2 The Dormouse Nest Data

The way in which the data was collected, and the purpose of the previous research, heavily influences this one. I do not know of processing errors, or any reason to reject extreme values. To me there is as much information in any of its values, and therefore I had to work with indicators of central tendency, in order to allow extreme values to partake in the filter: Cells with a value on the positive side of the estimator will be awarded a value of 1 and, as their value deviates away from the median, their value will decrease up to a distance of 1.5 the value of the interquartile range (i.e. IQR), where the awarded value will become zero. Median and IQR are robust descriptive indicators (i.e. less sensitive to outliers).
This is a geographical filter, in which several layers will be fuzzified and overlaid, with the purpose of finding new habitat. The attitude is to give more space to this species, even if we are talking of enclaves, instead of acting on their actual habitat fragmentation (as projects like BioGrenzKorr (Schulz et al. 2013) have done in the past). This exercise demands to study the significance of certain spatial characteristics, but the best way to do that is by contrasting areas inhabited by dormice against uninhabited areas . . . yet how do we know that empty areas actually contain no nests? The data was collected from transects in areas where dormice had been found. There was no comprehensive combing of the entire research area. Thus I cannot be sure that uninhabited areas actually are a negative; nor whether they have been uninhabited in time. Thus, how can we be sure that a particular space cannot sustain dormice? and, what is the meaning of finding an empty nest?

I do not have any time sequence. I do not know if a single-nest-area is an area into which the species is moving, or one that no longer supports dormice. Furthermore, some of the nests have been placed by researchers on areas with pre-dating occupancy, and some of these are still vacant. Could such nests introduce any deviation from where mice would have chosen naturally to nest? Could their vacancy be due to its recency? From the methodology point of view, this means that I cannot use tools like logistic regression, in order to explore these questions as a binary suitable / unsuitable, because of errors of type two (i.e. concluding that a suitable relationship does not exist, when in fact it does, and mice live there). Moreover, those empty nests are often at a very close range from occupied nests, in a cluster, and the correlation of spatial values would make contingent such differences in occupancy. So I decided to use only occupied nests, and their spatial clustering means that their values do only reiterate their membership to a suitable relationship. This introduces the idea of weight of the gathered data. The more a relationship is reiterated, the bigger value it should have.

2.2.3 Weights

The entire filtering method relies on the assumption that the spatial relationships established by dormice in south-west Sjælland are good, in the sense that these locations allow them to thrive. Thus follows that these very same relationships, if replicated on another location, should still be valid, as long as some precautions are taken. These are accounted by the selective phase, previous to filtering.

Nevertheless, how good are these relationships actually? That we do not know a priori. One way of of grasping their qualities is by devising a measure of success. A very simple measure is their population. A large population should indicate, that its spatial relationships must have some virtue. Another measure would be the density of the population, which is specially pertinent, because it is based on a spatial dimension. Density tells us about the availability of resources, as well as the absence of disturbance. Not only it seems interesting to use this value because it relates to space, but also because a ratio of cluster nest occupancy has had to be dismissed.

Thus I had two different metrics to weight the value of the spatial information extracted from each particular cluster: a simple count of nests (i.e. population size), and a density, which, put in spatial terms, was understood as the compactness of the spatial distribution of nests. In order to capture this as a scale-independent ratio I decided to use the isoperimetric quotient, which evaluates compactness by comparison to the most compact shape possible: a circle. But not just any circle, it does so in relation to a circle that comprises the same area as the shape under scrutiny. The formula (Bogaert et al. 2000) is written right here in equation(2.1):

$$f_{circ} = \frac{4\pi A}{P^2} \quad \text{where} \quad A = Area, \text{ and } P = Perimeter \tag{2.1}$$

Afterthoughts on Weights

However, something has been said that calls for checking the validity of these weights. It was not considered (at the time I adopted density as an appropriate measure of success) that dormice might dwell on lineal features in the landscape. Furthermore, the method establishes compactness in relation to a circle that encompasses the same area as the cluster, and the cluster had been defined as a buffer of 75 m in radius. Because of this, a one-nest cluster would be considered compact, which counters the idea that the larger the population, the bigger the success of a cluster. A single-nest cluster would be supposed to be successful by this metric, and therefore, both measures could not be used together.

Compactness is a measure that incorporates a spatial dimension, and that appealed to me. So I tried again without buffers. Excluding the single-nest clusters by the fact that these are not even a cluster, could I use compactness to rate clusters according to the theory explained in section1.1? Theoretically, a two-point cluster would always have a very bad fit, for two points can only form a line. Thus, on the one hand, it made only sense from a geometrical point of view. The larger the amount of points, the bigger the potential for a cluster to resemble a circle. However, on the other hand, the problem was that, under the conditions in place, a larger amount of points would only make it more difficult for a dormouse cluster to resemble a circle. First, because what seems to guide dormice settlements are physical, linear elements, or boundaries in the landscape, which offer the light and moisture conditions for the understory to develop. Dormice would just follow the availability of shelter, and thus would they rarely find themselves forming a circle, which is the reason why the coefficient of compactness is far from being a circle (See below. Table2.1).

Table 2.1: Descriptive statistics of the cluster compactness ratio data (0 = a line, 1 = a perfect circle).

> summary(ClCmpctRt)

Min.	*1st Qu.*	*Median*	*Mean*	*3rd Qu.*	*Max.*
0.3201	*0.4276*	*0.5184*	*0.4962*	*0.5672*	*0.6251*

In this way, a larger amount of nests in a cluster might make it more likely to deviate from a perfect circle. Aware of a potential problem, I had to explore the existence of a correlation between response variable (i.e. compactness ratio) and its new explanatory variable (i.e. the number of nests in the cluster), see Table2.2. There is strong correlation between both variables, and it is an inverse correlation: A bigger number of nests does produce less compactness. Thus I had to drop the compactness metric, for being counter-intuitively unfit for its purpose. Further test results are given in sectionB.0.1.

Table 2.2: Correlation between Cluster Compactness Ratio and the number of nests in the cluster.

> cor(dat)		
	ClCmpctRt	*ClNrNests*
ClCmpctRt	*1.0000000*	*-0.8369477*
ClNrNests	*-0.8369477*	*1.0000000*

2.2.4 Outliers

In my data there are some values that could be considered outliers, just because of their extreme value in comparison to the mean and the standard deviation. Another way to describe them is, that an outlier is a value with a large residual in linear regression. It is an unusual observation, like nests closer than 100 m to a house. One might argue about their fitness as a predictor, yet prediction of the location of a nest might have several variables, a degree of fuzziness on each of them, and a sense of probability. The 'peculiarity' of such observations might be comprised by the total prediction model, including the weights or fuzzy membership functions applied to the supposedly independent variables.

On the one hand, I have no idea on the accuracy of the measurements, I cannot simply remove them from the database based on the assumption that they are errors. Neither do I have enough observations to be in the position to drop a cluster of nests, due to its peculiarity. Since this exercise attempts to use simple but robust calculations, it has placed more emphasis on averaged measurements than on individuals. Measures of central tendency should be reinforced by weights. Least squares regressions should be aided by robust regression techniques, in case they be affected by influential observations. As aforementioned, this filtering method must also encompass extremes.

Robust regression is an alternative to least squares regression whenever these have to deal with outliers. It is a method available in *R*, and depend on existing statistical packages. One must install the *'MASS'* package. The purpose of this method is to weight down values that are considered outliers, or better said, those values that have leverage and influence on the estimation of the regression coefficients. The *'rlm'* (robust lineal model) command in the MASS package, which implements two strategies: bisquare and Huber weighting. On the one hand, bisquare uses the residuals to weight the regression. The UCLA Institute for Digital Research and Education divulges in its website the equation that estimates weights, applied by its own software *'Stata'*. Bisquare uses $\sum_{i=1}^{n} w_i(y_i - x'b)x'_i = 0$. (IDRE 2014) The weights on those values depend on their residuals, and the residuals depend ultimately on the weights that transform the regression model. Therefore, the only way to solve these dependencies is by iteration, which must go on until its values converge. On the other hand, Huber weighting takes a different approach and just classifies the type of value by how big it is in relation to the smallest. (IDRE 2014)

Thus the weight function is defined as:

$$w(e) = \begin{cases} 1 & \text{for} \quad |e| <= k \\ \frac{k}{|e|} & \text{for} \quad |e| > k \end{cases} \tag{2.2}$$

2.2.5 Filter Variables

Now I will start to explain the core of this exercise. I call it 'exercise' instead of Major Thesis, because there is no major work here. As aforementioned, there is only a concatenation of simple tests, which provide context, and then a series of layer of information to be overlaid, very much like Ian McHarg used to work through the 60's[8]. Also it is for me a first attempt at using fuzzy logic and fuzzy sets, envisioned by Zadeh in 1965, which are an old contribution to today's high performance computers and large capacity storage devices. What this exercise is about, is a Multi-criteria Analysis, which, due to the vague complexity of trade off decisions, between aroused and hibernation variables, between shelter and food provision (young-thick and mature-productive forest) . . . requires the help of fuzzy membership functions, in order to classify regions of space in relation to what suits *M. avellanarius*.

Cognitive processes are characterized by being vague or fuzzy. Are they? or is it reality itself already fuzzy? Anyone trained on soils has had that experience. A sharp line on a map does not quite respond to the reality in place. Fuzzy sets can handle qualitative data, or deal with statements that give way to interpretation, anything that has a comparative background, anything that is in relation to . . . It does not simplify an already nuanced world. In my case, one could say that the data on which this exercise relies is under sampled, biased (by being geographically limited), and vague. Perhaps its greatest challenge is its spatial vagueness. For instance, dormouse density has been found to be in average 1.1 ha per individual in Germany by Ehlers (2012), while in Lithuania Juškaitis (1997) reports it to be different, and even dependent on the sex of the individual: having males about 1.0 ± 0.05ha, while females would be 0.8 ± 0.05ha. This vagueness is basically what brought a "fuzzy" approach to this exercise.

Besides the fuzzy quality of its filter variables, this layered approach is characterised by a initial phase of selection. The purpose of this selective step is to begin narrowing down the area to cover with the filter. Therefore, it is a first step across scales, taken at 300m pixel spatial resolution. At this stage, the main source of data covers soil types, relevant climatological variables for the phenology of the species, and other environmental variables that relate to the use of space by humans.

---

[8] Read 'Design with Nature', a book published by Ian McHarg in 1969.

Soil Selective Variables

There are three main characteristics for the selection of soils within a certain range of values. All three plus elevation will mask the cartography with noData:

1. s_1: *soilCEC* responds to the cation exchange capacity of the topsoil. This metric gives an idea of the fertility of the soil, for it reports on its ability to hold nutrients (i.e. cations), and exchange them with surrounding water. Thus it depends on the presence of water in the porous fraction of the soil, but also on the soils pH.

2. s_2: *soilDrain* is a class attribute, which measures the ability of the soil to evacuate excess water. It also expresses the soil texture.

3. s_3: *soilPH* is a measure for the acidity and alkalinity of the topsoil. It is actually measured in the soil-water solution. It conditions fertility and even the apparition of certain toxic salts dissolved by low values.

These three selective metrics are supposed to be of significance for defining the type of plant community and its fertility, which impact indirectly *M. avellanarius*.

Environmental Selective Variables

There are four main metric variables that have been used for selecting a suitable environment, i.e. held within a range of values. All are based on the discussion above.

1. w_1: *wAvgTArAut* is the averaged temperature of the months of September and October during the last 50 years, approximately. Autumn is a period of high activity, as reported by Fischer (2014) (fig.4 in her report), coinciding with the availability of the most calorie-rich food before hibernation.

2. h_1: *hNghAddSHa* is the sum of neighbouring addresses per hectare. Its values have been derived from laying a sampling grid, which cells are 1Ha in size, and give an account of the density of residences.

3. h_2: *hSumRdKmHa* is the sum of road kilometres per hectare, derived from a sampling grid. It measures the potential human access to each of the cells.

4. h_3: *hNghSbUrRt* is a ratio of area occupied by suburban land use (i.e. low-rise residences combined with businesses and gardening areas) per hectare. It gives an account of the amount of space left for other land use types, to be used when density values are low.

These four selective metrics are supposed to have significance directly on the dormouse population, as well as indirectly through expressing habitat loss.

Geometrical Filtering Variables

There are six main geometrical variables that have been used for the selection of a suitable environment in relation to features in the landscape at 0.4 m spatial resolution. These relate to the cluster formations of nests. Thus these are averaged weighted distances, in accordance to the population size of their cluster:

1. d_1: *clAvgDstPB* is based on the minimum distance measurement made from each nest location to its parcel boundary. Parcel boundaries often imply a change of land use, thus an opening in the canopy (i.e. road, path or open field). Some collinearity might be expected and checked.
2. d_2: *clAvgDtDRd* is based on the minimum distance measurement made from each nest location to a dirt road (i.e. unpaved path or track).
3. d_3: *clAvgDstHs* is based on the averaged minimum distance measured from each nest location to its five closest buildings.
4. d_4: *clAvgDtUrb* is based on the minimum distance measurement made from each nest location to land use classified as urban (i.e.bykerne + hoj_bebyggelse).
5. d_5: *clAvgDstRd* is based on the minimum distance measurement made from each nest location to roads (i.e. paved roads of any class).
6. d_6: *clAvgDstMB* is based on the minimum distance measurement made from each nest location to the municipal boundaries.

These six filtering metrics are supposed to have significance directly to dormice. In fact, there is literature that confirm the importance of distances to the edge of the wood. (Williams et al. 2013) All of them will be coded as 1 or 0, on the base of their membership to a fuzzy class, and constitute the actual filter.

Finally there are some other variables that should (a priori) take part on the filter, and these are not geometrical but qualitative. These variables describe the surrounding space to the nest, and they were prompted by the fact that dormice live in singular locations, within old established forests.

1. q_1: *nDrVarRt* is a ratio on the variation of slopes surrounding the nest. It is a classification on a focal calculation (a 3x3 pixel neighbourhood, with 0.4 m spatial resolution), that is later averaged to a 20 m pixel. It has to do with the fact, that nests for hibernation are most often built on the ground. So I developed it, in order to quantify the degree of drainage pattern present.
2. q_2: *nAvgNDVI* is a measurement of the average vegetation index (an indicator of green vegetation) surrounding the nest in a 20m pixel.
3. q_3: *nTrHghVrRt* is a ratio on the variation of tree heights surrounding the nest, which indicates variations in canopy cover and tree age.

2.2.6 Variables on the Selection Mask

The selection mask has 'rigid' values, as in opposed to fuzzy boundaries. It serves the purpose of making a preliminary filtering, which in this case allowed to find a testing area within the metropolitan region of Copenhagen, but, of course, it could have been applied to the entirety of Sjælland, whenever the purpose had been to establish a comprehensive map of suitable areas for *M. avellanarius*. At this stage, we are often working at 300 m spatial resolution, which allows for that.
The masking is carried out by giving 'noData' values to those pixels that fail to meet the values given by the variables. There are three types of variable: the ones relating to soil characteristics, the ones relating to land use (as explained on urbanity), and elevation. As an example, the cut value for terrain elevation was selected at 40 m above sea level. The values for the soil are detailed in table2.3.

Table 2.3: Range of soil values used for masking the working areas with a suitable (within range) and unsuitable/noData (out of range) values.

Location:	*Variables:*		*Value range:*
Slagelse (W)	*HWSD_SoilData.T_PH_H_2O*	*6.4*	$6.0 < pH < 7.5$
Soroe (E)	*HWSD_SoilData.T_PH_H_2O*	*7.1*	
Slagelse (W)	*HWSD_SoilData.T_CEC_SOIL*	*6*	$5 < CEC < 8$
Soroe (E)	*HWSD_SoilData.T_CEC_SOIL*	*7*	
Slagelse (W)	*HWSD_SoilData.DRAINAGE*	*4*	$3 < d < 6$
Soroe (E)	*HWSD_SoilData.DRAINAGE*	*5*	

Obviously slightly larger or smaller values would still not be critical to the existence of *Corylus avellana* (i.e. hazel tree) or *Fagus sylvatica* (i.e European beech), which actually tends to appear on basic soils (with pH values higher than 7).

2.2.7 Variables on the Fuzzy Multilayered Filter

The filtering process will involve the assignment of continuous values, ranging from 0 to 1, as a form of normalisation of each variable. These new continuous values will be assigned at the pixel level, and will indicate the strength of membership to a set of pixels, which exhibits an specific event or attribute. Pixels assigned a zero value were not members of the set, while those with value one had full membership, which is characterised by a central value, a range or spread of values, and a rate of decline (a membership function), only linear was used. In this case, the event that defined the set, was the presence of an occupied dormouse nest. The fuzzification of the data was carried out by the tool 'Fuzzy Membership' in ArcMap. Finally, all layers (i.e. variables) were combined according to an overlay type by the tool 'Fuzzy Overlay'. A value threshold results in the filtering process.

The filtering steps goes as follows:

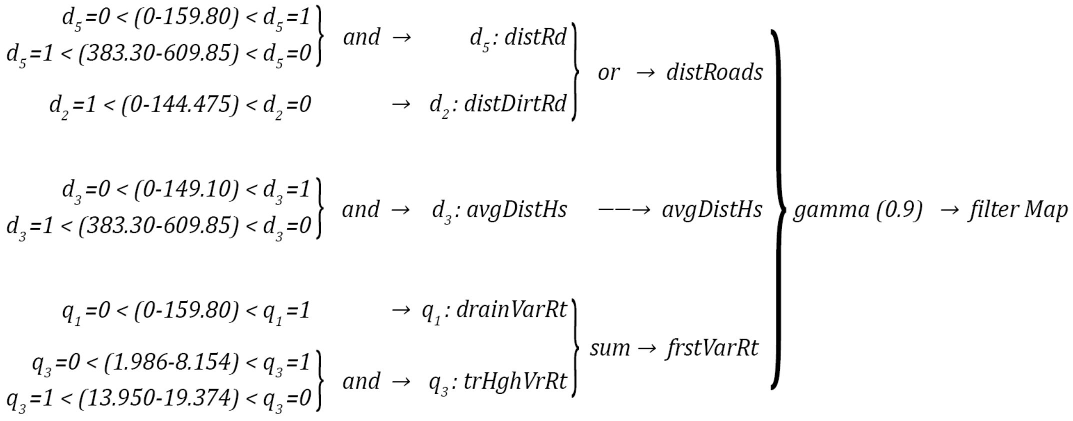


Figure 2.2: The steps on the filter calculation comprise different raster files (variables on distance to paths, roads and houses, plus forest characteristics), which are overlaid following a certain method.

The creation of a cartographic filter is a different process than the creation of a statistical model. For instance, in cartography there is no clear selection of the variables involved in the filter, neither there is a guiding procedure, nor the way to combine them is given. The earliest selection of variables was obtained by an statistical multivariate regression model. Yet it picked 10 out of the 13 variables enunciated above. Later they came down to be only five. For more on this, read on the results (chapter3.1.1). While some variables can still reinforce a distribution of observations in an statistical model, those same variables only add visual noise to a cartographic filter. Moreover, in the preliminary study it became clear that, in order to revealing, any study on dormice should preferably use very small scales. Nevertheless, for the cartographic filter there must also be a congruence on spatial resolution. Every spatial resolution must be a multiple of the lowest spatial resolution, in this case 0.4m. In consequence, given that two variables describe forest characteristics, and that, in order to expose these, I needed to work at a larger scale (tree crowns are easily 5m in diameter), I had to increase the pixel size by a 50 fold. So the fuzzification of all data was done on a 0.4 m pixel by *'Fuzzy Membership'*; yet forest variation was generalised to a 20 m pixel by focal statistics.
The actual fuzzy membership functions were drawn out from the statistical description of each variable by *Median* and *Interquartilic Range* (i.e. IQR). The actual values are on tableB.11. Furthermore, the slope sign of the estimators (β) on each variable was used in order to assign to the function a positive and a negative side. All functions were decided to be linear, but some of the data had more than just one slope. For instance, to walk away from roads increases the chances of finding dormice, yet walk farther than a distance, and those chances will start to decrease.

That might be partly because a road establishes an edge to the forest, and dormice nest in lighter areas closer to edges. So, unless there is a glade, dormice are never found deep in the forest, farthest from roads. In that way, there is an initially ascending function that becomes, later, a descending slope. All functions were corroborated visually by plotting observations on each variable against distance. The density of observations can be confirmed to cluster toward an intermediate distance. Thus, in those cases with triangular or trapezoidal functions, the membership used overlay (always selecting type *and*) on two linear functions of opposed sign.

After talking on the choice of linear functions, the sign of their slope in accordance to the slope of the estimator, and on the extraction of intermediate values for delimiting the slopes, there is only one main ingredient missing. Now I am able to use the '*fuzzy membership*' tool, yet I still need to ultimately combine these variables. That will depend on the nature of the variables to be combined.

Significant Variables

From the list of variables in tableB.11, two of them stand out for having a very different scale than the one at which dormice operate. Distances to municipal boundaries is a clear one. The climatological data has a spatial resolution of about 700 m pixel, which disqualifies its use for a graphical output, in spite of its significance, and even though the slope of its estimate is the steepest, thus the most sensitive to the event under scrutiny.

Let us remember that here the event is the presence of an occupied nest, and that the quality of variables as predictors of such event is, in relation to scientific research, that these repeatable and transparent, so that it can monitor the event over time and for different locations; that the variable be quantifiable and simplifying, so that it can reduce uncertainty and complexity; and last but not least, that it be sensitive to change, for that is the point of a filter, a model, any act of surveillance. (Fjellstad and Frederiksen 2004)

Furthermore, there are some variables that seem to be redundant. The ratio of kilometres of roads per hectare is already treated at a finer grain by establishing the distance to roads and dirt roads. What is more, it seems it has a too big spatial resolution, for the pixel is at 100 m, which is larger than an already oversized nest range (buffer of 56.4 m). Also the distance to parcel boundaries seems redundant, when most often those boundaries coincide with roads or paths, are often set back by the presence of buildings, or just fail to coincide with the edge of forests, which is what it stands for, as a proxy. Selecting variables in accordance to their scale, significance, sensitivity (β estimate), and simplicity took the filter to finally comprise five, not just extensive (geometrical) but also qualitative (intensive), spatial indicators. Still there is a last aspect to be taken into account.

Non-disqualifying Variables

There is a very important distinction to be made on how variables constrain the viability of the population, and are helpful to identify habitat for *M. avellanarius*. There are those costly to remove or attenuate, and those that can be modified or altered without that supposing a burden to urban development, or a large expense of resources. These are non-disqualifying variables, from the point of view of habitat identification. Sometimes we even do not know the mechanism in detail, by which they affect the population, but distance to municipal boundaries might be non-disqualifying, whilst averaged temperature in autumn is.
Another good example is the density of the understory. It can be inexpensively increased by practising light maintenance tasks like thinning. What about the richness? (or diversity of species present in a forest) It can be increased by planting anew and by extracting wood selectively, etc…

$$NDVI = \frac{NIR - R}{NIR + R} \quad \text{where} \quad R\text{: red band and } NIR\text{: Near Infrared Band} \tag{2.3}$$

Among such variables there is one that is paradigmatic. It is a very useful calculation, yet also, for the purpose at hand, unnecessary for the filter: the averaged NDVI values (Normalized Difference Vegetation Index) of a forest per area unit. It is a simple calculation, see equation (2.3), which uses two bands of the electromagnetic spectrum (i.e. the red and NIR bands). Its utility to this exercise is double. On the one hand it can be used for the automatic segmentation of land uses (which will be carried out on specialised software like *eCognition Developer*, by Trimble). It will be used to differentiate between broadleaved and coniferous forest. On the other hand, NIR values can be used to assess the health of a stand of vegetation, which could impact on the diet of dormice. The first one will be of interest to the 'forest' entity in the ERD. The second it is of interest in order to describe the habitat for the filter. This means that the first purpose will require of quality spatial resolution, in order to achieve an accurate segmentation. For this method orthophoto imagery resolution will be required. For the second purpose, in which obtaining an accurate measurement gains relevance, free of atmospheric distortion, satellite data should be used. Satellite data cannot have as much spatial resolution as aerial photography. Sentinel NIR bands (i.e. bands 8/8a) achieve up to 10 m resolution, which is still useful enough for the interpretation of healthy vegetation. However, for both purposes, and in spite of its great contribution, this is here a non-disqualifying variable, i.e. one that allows the detection of habitat, but that is not of much significance when assessing if an area is suitable for dormice.

2.3 Computation of the Fuzzy Filter

We have five initial variables for the filter. Three are distances to man-made structures, and two are generic characteristics (i.e. measurable and descriptive anywhere) on forests. Both forest qualities are defined by variety, because distinct richness is what makes dormice locations to stand out: forest biodiversity, miscellaneous tree ages, a manifold of soil features ...
Three of these variables have a trapezoidal membership function. This means that their membership function has at least three sections (the first section on *nTrHghVrRt*'s function has constant zero value up to x= 1.986 m of height, then it ascends, remains constant at value 1, and finally descends back to zero value). This trapezoidal shape, plus the fact that I have chosen to stick to linear functions, means that these variables need to be graphically created in two independent instances, where the *fuzzy membership* tool is applied, and then be put together by *fuzzy overlay* type *'and'*. This type is equivalent to an overlay with gamma-value=0.5. It selects locations that have at least a 0.5 membership strength or greater, for it will return the minimum value of the sets the cell location belongs to. A kind of least common denominator for the membership of all the input variables in the overlay. (Esri 2017) Thus one other thing becomes clear. It is going to be crucial to know how many steps or overlay instances we need, for $f_1 + f_2 + f_3 \neq (f_1 + f_2) + f_3$.
Following figure2.2 I identify three main indicators, that derive from the five selected variables:

- Roads are man-made structures that fragment the landscape and introduce a degree of disturbance but also access to resources, which in the case of dirt roads and paths are clearly in relation to forest management, as in contrast to local or regional traffic. It is due to the clear predominance of one use above the other, according to one or the other variable, that both cannot merge by losing individual strength. Because I want to identify the highest membership values on both variables after the procedure, I used type *'or'*.

- Houses are also man-made structures that indicate an active management of resources, specially when found isolated. Nests tend to cluster close to some of them, thus they also have a decreasing membership function. However, their ascending membership function makes clusters of houses into places of small values. In short range occupied nests are seldom found.

- Forest qualities are represented by two very different measurements, although both are based on variety. Also, both need to generalise their spatial resolution to 20 m, in order to capture such variation. That means for the height of trees a very lengthy focal calculation (a neighbourhood of 50 by 50 cells), before which data was multiplied by 100, in order not to lose information.

Remember that *LiDAR* data has a vertical accuracy up to 5cm.
The type of operator used in this case needed some thoughtful consideration. After examination of the results from the variation of surface slopes, one can identify large depressions on the ground (see image2.3), which, with a high degree of certainty, correspond to the imprint of fallen-uprooted trees. The facts that these depressions contribute to infiltration of rain water, that the drainage patter creates a diversity of moisture conditions, only encourages a bigger diversity of species, thus a greater variation in tree height, which also connects both variables back, by having different root systems and proclivity to be toppled. In a way, both variables seem to, even though weakly, have an increasing effect on each other, and such effect does not call for an *or* overlay type operator, but for a *sum* type, which is not an algebraic sum. (Ebadi et al. 2004) To overlay variable by *sum* does not mean that the location will become more suitable. Overlay cares for the relationships and the interaction established between variables, and will only introduce an amplifying effect between their values.

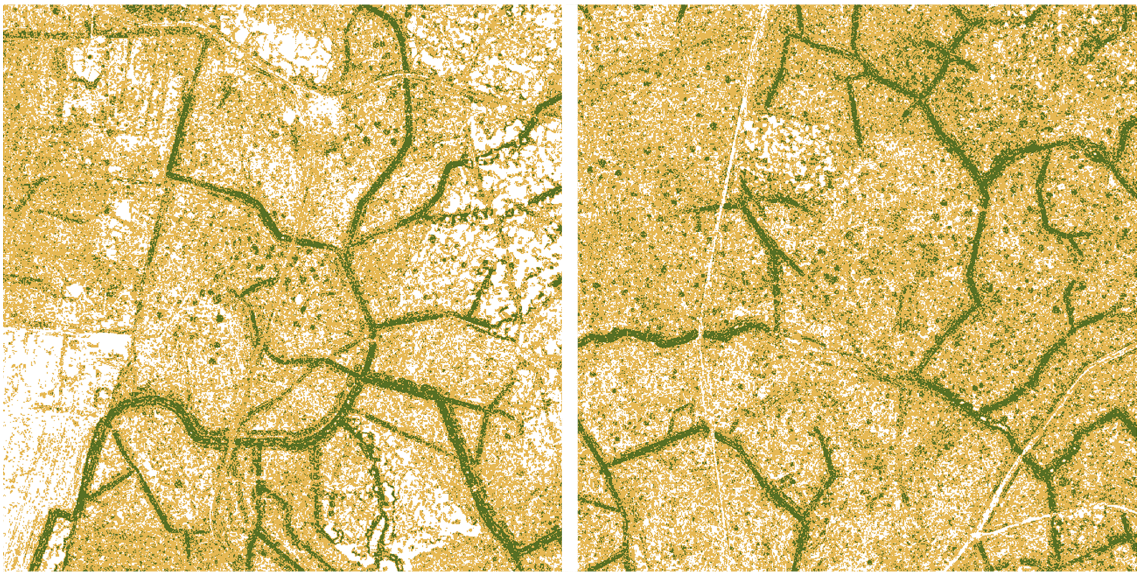


Figure 2.3: Drainage pattern on forest soils, both in north-east (left) and in south-west Sjælland (right). In both cases circular marks appear to be the root zone of already fallen trees. Also in both cases, the drainage pattern not only provides the adequate conditions for the tree species present and their variability, but such variability creates different densities and sizes of circular marks. Both are images at scale 1:2500.

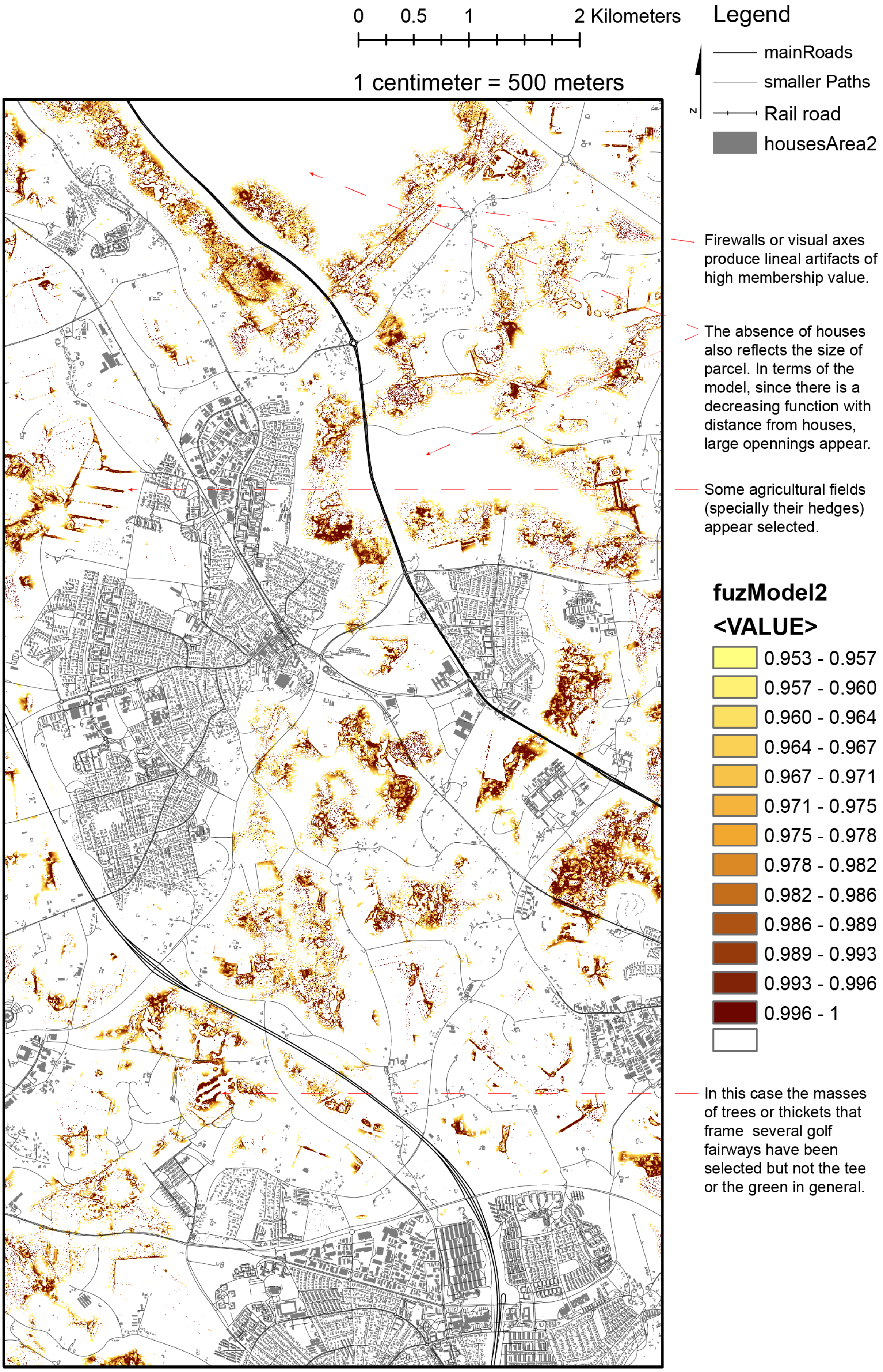
The final result is overlaid using the *gamma* type, which establishes a relationship between the multiple variables, instead of returning the value of a single one. (Esri 2017) In this case, 0.9 is closer to *sum* (1), than to *product* (0).

2.3.1 Statistical Methods

Let us start this section where it is most interesting. Once I have chosen and used the different statistical tools available, how do I interpret the results? What I have achieved is to put things into relation. I will have found the pertinent variables, or indicators, and I will have quantified those relationships. Normally when someone filters something, that person's interest is on the ability to retain or separate a part of that something. Some would think of analysing the residuals. The smaller the standard error of the residuals, the lower discrepancy between measurement and prediction, the better the accuracy of the screening. They are right. Such residuals, because of the way the model has been built, are supposed to be optimal. An OLS (Ordinary Least Squares) regression attempts to minimise these residuals, and it should produce residuals that pivot around a mean error equal to zero.

However, in this case we are one step removed from the outcome. We are not filtering gravel, being gravel already defined. We have to define what fraction is useful first. If you follow the analogy, you will understand that, here, it is identifying which variables are useful, that comes first. It is their estimators and their conjunct accordance with spatial variation, what must be identified as a preliminary step. Thus some will use the variation between observations, in relation to the variability in the dependent variable, in order to establish a dynamic relation of fit between model and phenomenon. It is not just the error reported, it is variability as a way to assess sensitivity. Such measurement would be an attempt to reveal the proportion of the total variability unexplained by the model. The inverse of which is what the R^2 statistic is supposed to disclose. Yet statisticians often object.

Willett and Singer (1988) gave caution about R^2 as a measure of fit for a model. They write that "unthinking reliance on an R^2 statistic . . . can lead the unwary data analyst to an overly optimistic interpretation of the proportion of variance accounted for in the regression." Again, we are looking at a relationship. It is the association between the dependent variable and its explanatory variable. In order for that correlation to exist, there must be first and foremost variation, and in that variation (which is difference), there must be repetition. It is the consistent similarity in such a dynamic character between both variables, it is that repeated accordance in variation that we are trying to seize. R^2, which can be calculated by several definitions, depends on the actual distribution of the event. Is it a lineal relationship or not? What is the context of that relationship? It is because of the variety of contexts in which the model must predict, it is when different variables have different weights according to such context, that the coherent and impartial mathematics fail. The model will then be reported from both sides: from its errors (using the RMSE) and from its fit to variation (using R^2).



Firewalls or visual axes produce lineal artifacts of high membership value.

The absence of houses also reflects the size of parcel. In terms of the model, since there is a decreasing function with distance from houses, large opennings appear.

Some agricultural fields (specially their hedges) appear selected.

In this case the masses of trees or thickets that frame several golf fairways have been selected but not the tee or the green in general.

Figure 3: Unmasked result of the fuzzy filter applied on north-eastern Sjælland.

Chapter 3

Results

This chapter will describe statistical results, queried on the database built for this thesis, and present the graphic results of the filter. Thus, it is structured into two main blocks: Results on variable significance (selection phase), and on the filter.
Within the selection phase three variables were to be homogeneous: elevation, climate, and only chosen intensities from the human environment. So I will start by detailing the findings on them. You will find tables and graphs joining the text. Some think they should be on the appendixes, in order for the narration to flow. However, I prefer to present them, because of their deeper explanatory power on the account. Later on, I will move to describe the geometrical variables used for filtering, and any other further variables related to dormouse habitat.

3.0.1 Provision on Elevation

The first significant results to be reported are on elevation. There is a positive correlation between elevation and the number of nests that compose a cluster in south-east Sjælland, see table3.1. Although such correlation is only capable of explaining a 20% of the variance (R^2: 0.21), and the standard error indicates that data fits loose, it is significant with high probability (p-value: 3.6e-07).
It is not perfect, but it is enough to raise a flag in my mind. The scattered cloud of residuals denote that it might not be just elevation on its own, but perhaps a causal vector composed of several dimensions. If we step out of the averaged elevation of the cluster nests, and consider the surface: All nests are found well above 40 m of elevation above sea level, at an average altitude of 60.07m (StD 9.96m), while Sjælland has a low relief, with heights that very seldom surpass such elevation.
However, the causal effect of elevation is still unclear. In order to find more on the possible causes behind this significance, further entities were set in relation to elevation. Thus, I tested for a lesser human disturbance in altitude, and I did it by using data on municipal addresses, which can be easily measured. These should be a proxy for inhabited houses, and even the presence of agriculture.

Table 3.1: Correlation between cluster population and elevation. A linear OLS regression has been fitted through data explaining the number of nests within a cluster by each cluster average elevation (i.e. nestElev)

> summary(lm(clNrNests~nestElev))					
Residuals:					
	Min	1Q	Median	3Q	Max
	-9.911	-3.815	1.551	3.217	4.749
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-0.57880	2.33805	-0.248	0.805	
NestElev	0.21023	0.03879	5.420	3.6e-07	***
Significance codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '
Residual standard error: 3.864 on 109 degrees of freedom					
Multiple R-squared: 0.2123,		Adjusted R-squared: 0.2051			
F-statistic: 29.38 on 1 and 109 DF,		p-value: 3.604e-07			

On table3.2 are the results of trying to explain the amount of addresses by their elevation. It is in fact, to be precise, the sum of addresses per hectare in relation to the averaged elevation of that very same hectare of terrain.

Table 3.2: Correlation between number of addresses and elevation. A linear OLS regression has been fitted through data explaining the number of addresses in a hectare by the averaged elevation of their terrain (i.e. DTM100mA1).

> summary(lm(st$addr100mA1~st$DTM100mA1))					
Residuals:					
	Min	1Q	Median	3Q	Max
	-1.644	-0.824	-0.639	-0.393	104.105
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.79307	0.21519	8.332	< 2e-16	***
DTM100mA1	-0.02223	0.00420	-5.293	1.24e-07	***
Significance codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '
Residual standard error: 4.788 on 7018 degrees of freedom					
Multiple R-squared: 0.003977,		Adjusted R-squared: 0.003835			
F-statistic: 28.02 on 1 and 7018 DF,		p-value: 1.237e-07			

The result says, that there is significant but a weak, negative correlation. The sign means that the higher we go, the farther apart will addresses be found, although there will not be a noticeable difference in dispersal, for the range of existing elevation is very reduced, and the slope that links both variables is close to level. So higher grounds contain a consistently lower concentration of addresses, yet this does not say anything about the actual amount of disturbance.

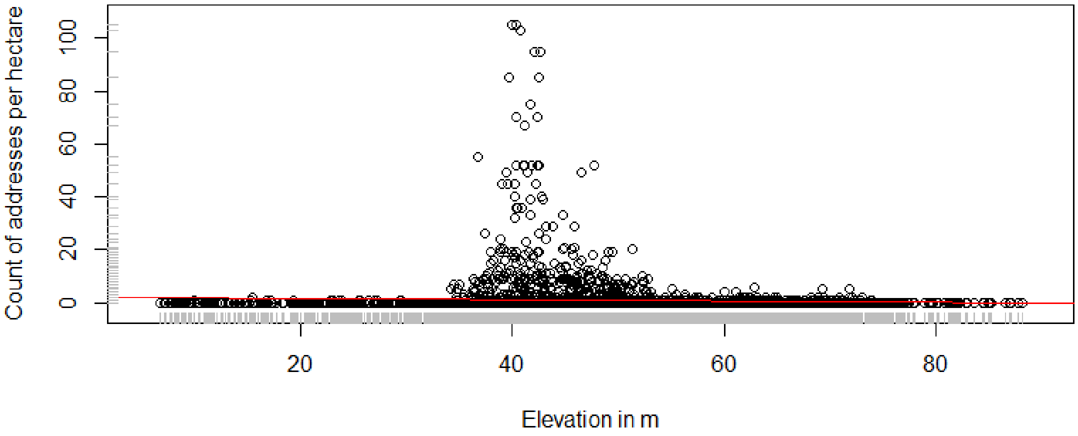


Figure 3.1: In the above graph can be seen the count of addresses per hectare against elevation. The distribution seems to accumulate around the value of 40 m, which is where the town of Sorø is located.

In order to clarify things, plotting one variable against the other in figure3.1) allows to see the only urban area in our data, located at about 40 m above sea level. All nests are placed above that elevation, but about half of them are located below 60 m (average nest altitude: 60.07 m). This shows that half the population of dormice are much more likely to be closer to human dwellings and their disturbance, and one would be tempted to consider the other half farther from disturbance. However, reality does not follow along, because people travel in order to recreate. Forests are one of those types of land use often chosen for recreation, and distances are, perhaps, much more important than such a small difference in elevation. Thus, although the sum of addresses could have something to say about the degree of disturbance experienced in an area, it is also common sense to think that it will not be enough to properly account for it, which is corroborated by the amount of variation explained by this correlation (R^2: 0.004).
Moreover, reality does not follow along, because, seeing the disproportion of the maximum value of the residuals, when I tried to verify all the relevant model assumptions, which allow me to use a simple linear regression model, I got into trouble. The plot of the explanatory variable against the residuals does not show a random scattered, as it should; as they deviate farther toward the positives, the width of the plot is absolutely not constant. Thus the constant variance assumption does not hold. If I check the normality assumption, the QQ-plot does not align either, and the histogram of residuals has an eminent tail and asymmetry (see figure3.2). Thus, the distribution of addresses cannot be explained by elevation.

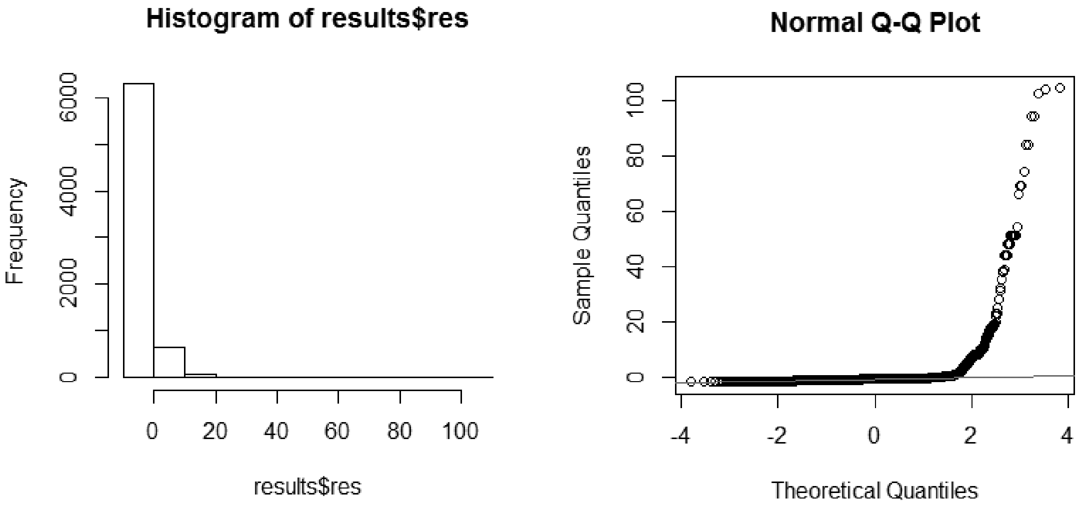


Figure 3.2: The here above plots check on the statistic assumptions of a lineal regression model. The residuals of such a model must tend to average zero (i.e. balance out), and their quantiles should be equivalent to those of a normal distribution (i.e. they should not show a skewed distribution). These residuals do not uphold such assumptions.

That conclusion might become obvious to anyone that points out that a sum of addresses per hectare can be better explained by urbanity, or nuclei of human wealth, as geographers like Ash Amin and Nigel Thrift put it in 2002. That is probably the cause of trouble to the regression residuals. Figure3.1 illustrates that very clear. However, from these thoughts still one question rose in my mind, in relation to the correlation with elevation: Do forests dominate higher grounds?

For now, I will keep on trying to grasp the significance of elevation on its own. I looked back at the descriptive statistics on elevation in south-west Sjælland, yet this time at different scales. I mentioned that nests are found at an average elevation of 60.07m (StD 9.96m). Nevertheless, this averaged elevation tends to decline as the area considered into calculation around them is enlarged. Only two of the seventeen clusters inverse that trend. Clusters have an average nest altitude of 58.78m. Communities have an even lower average elevation at 56.94m. So, where are these nests located on the topography? and, does the correlation between the number of nests and elevation decline by enlarging the area in consideration?

The answer to the last question lies in appendixB, in the tables B.1 and B.2. These convey the piece of information, that the correlation does decrease. It correlates worse, although it is still significant. This can be seen on the steady increase of the residual standard error (from RSE: 3.854 at the actual nest elevation, to RSE: 3.875 and p-value: 4.914e-07, with elevation averaged to cluster extension; to RSE: 4.032 and p-value: 4.456e-05, when averaged to community extension).

So elevation values surrounding the nest seem to sink, as we move to larger scales, and their power to explain the number of occupied nests within a cluster also seems to weaken, as confirmed by the ever increasing residual error. Then I moved to consider dormice on a slope, and attempted to analyse their location on the surrounding topography. Is there significant evidence that nesting on higher ground is for the dormouse a pattern of some relevance?

I used R to carry a couple of one sample t-tests, and the answer was no. There is no significant evidence, that nesting on higher ground is, for the dormouse, a pattern of some relevance. The t-values obtained in the t-tests with 6 and 16 degrees of freedom, correspond to p-values of 0.07411 and 0.083 respectively (see in ChapterB, tableB.3). This results are not statistically significant for any of the usual significance levels < 0.05. Thus values do not deviate significantly from their mean elevation. Therefore, I must uphold the null hypothesis that the mean elevation of clusters, and to a lesser extent the mean elevation of communities, is not significantly different than that of dormice nests (60 m).

Therefore, although dormice tend to be more commonly found at higher elevations, and although it tends to locate its nests on above average elevations, the results of these tests do not support the idea that dormice locate their nests consistently above the mean, or consistently on the upper half of the slope.

At this point, just in order to establish the causal relation that makes dormice live at higher elevations, once that the hypothesis of a lesser disturbance was inconclusive, I went back to the question on forest. Could there be any noticeable change in forest productivity where dormice live, that is, with elevation. I needed first to know the average NDVI value for forests, and, for those forests where dormice are found in south-west Sjælland, NDVI is 0.74088 (with a StD of 0.09494). First of all, that is a normal value. Second, something that deviates by a 10%, after making 428619 measurements, results to be quite stable. In terms of range (NDVI values: -1 to 1) that tells us, that the NDVI values of vegetation deviate very little, in comparison to the variation found between different materials.

However, I performed the OLS regression, in order to correlate elevation directly to NDVI values. In order to do that, one must take care of the fact that both elevation and NDVI rasters have different spatial resolution. Thus the elevation and NDVI values used in this regression have been averaged to a hectare (i.e. a 100m pixel, which is a multiple of both a 0.4 m pixel and a 10 m pixel). The results can be seen in table3.3. They turned to report a positive and very significant correlation between the vegetation index and elevation. Still, this is not the same as to say, that productivity of forest increase up slope, or with the slope. In fact this began to answer the question I previously put. There might not be a difference in forest productivity, but a probable difference in land use with elevation. In a plot of NDVI values against elevation, figure3.3, it possible to see how land use with high NDVI values extends further into higher ground.

Table 3.3: Correlation between Normalised Difference Vegetation Index values and elevation within a hectare. There seems to be a significant correlation, although of very little sensitiveness: the slope estimator is very flat, meaning that changes in elevation result in very small changes in productivity.

> summary(lm(st$NDVI100mA1~st$DTM100mA1))					
Residuals:					
	Min	*1Q*	*Median*	*3Q*	*Max*
	-0.6722	*-0.2317*	*0.0801*	*0.1913*	*0.5947*
Coefficients:					
	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*	
(Intercept)	*0.1502738*	*0.0113039*	*13.29*	*<2e-16*	***
DTM100mA1	*0.0078791*	*0.0002206*	*35.72*	*<2e-16*	***
Significance codes:	*0 '***'*	*0.001 '**'*	*0.01 '*'*	*0.05 '.'*	*0.1 ' '*
Residual standard error: 0.2515 on 7018 degrees of freedom					
Multiple R-squared: 0.1538,		*Adjusted R-squared: 0.1537*			
F-statistic: 1276 on 1 and 7018 DF,		*p-value: < 2.2e-16*			

The point cloud clusters around two different values, one that might average ≈ 0.7, and the other ≈ 0.2. That points to their membership to two different land uses (forest-prairie/crops), standing apart due to their materiality. Also, notice that the top NDVI values do not increase with elevation. They remain about flat.

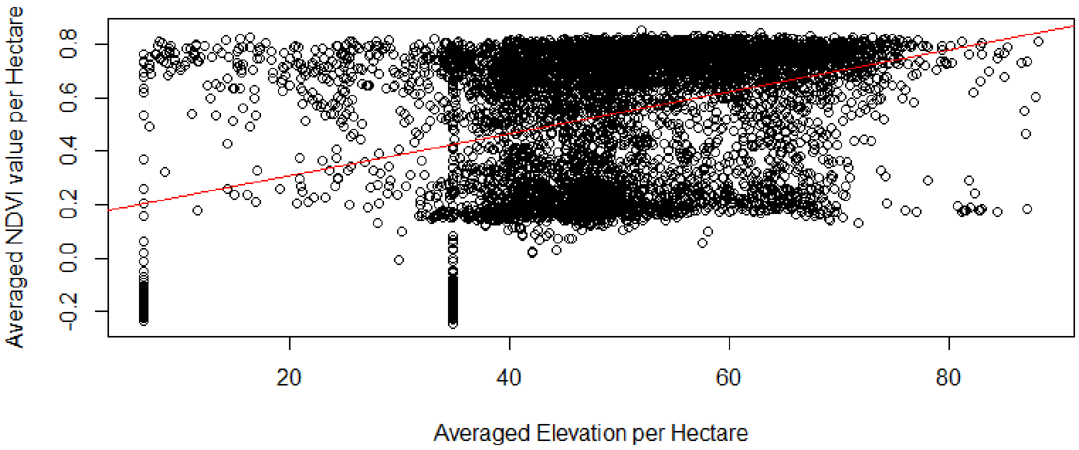


Figure 3.3: This graph shows NDVI values in relation to elevation. It makes easy to spot two large flat features (same elevation) due to their range of values (from the negatives up). Those are the two main lakes. Then it is also evident the fragmentation of the cloud of values around an average value of 0.7 (forests), and another around 0.2, which corresponds to the values of open grass/bushes vegetative growth. Lower values (about 45 m a.s.l.) reveal large open soil/hard surfaces (urban).

3.0.2 Provision on Climate

Climate is a variable that affects dormice in several ways. The most direct is described by their phenology. Unlike common mice, *M.avellanarius* does hibernate. One of the possible hypothesis for them to be found at elevation, that now seems weakened, relies on the idea that harsh winters are good for a process of hibernation. Elevation has turned to be subtle, and the climate ever more so (mind that the isothermal ratio on this part of Sjælland is at 2.2°C). Most climatological variables have failed to correlate with the number of occupied nests. However, there is one specific variable 'wAvgTArAut' (i.e. the average temperature while the dormouse is still aroused in autumn, i.e. the months of September and October) that has managed to give surprising and significant results. If we correlate the size of cluster population directly to the average temperature in the months when dormice fatten up, previous to going into hibernation, we get table3.4:

Table 3.4: Correlation between the number of occupied nests, and the averaged temperature of the months of September and October from the past 50 years. What is surprising is that it indicates that milder autumn months do not encourage larger populations. Keep in mind that the data varies from 11.6°C to 12.10°C, measured in 0.05°C intervals.

> summary(lm(ClNrNests~wAvgTArAut))					
Residuals:					
	Min	*1Q*	*Median*	*3Q*	*Max*
	-9.5777	*-3.2860*	*0.0056*	*4.2973*	*4.4223*
Coefficients:					
	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*	
(Intercept)	*315.823*	*82.240*	*3.840*	*0.000207*	*****
wAvgTArAut	*-25.833*	*6.991*	*-3.695*	*0.000346*	*****
Significance codes:	*0 '***'*	*0.001 '**'*	*0.01 '*'*	*0.05 '.'*	*0.1 ' '*
Residual standard error: 4.105 on 109 degrees of freedom					
Multiple R-squared: 0.1113,		*Adjusted R-squared: 0.1032*			
F-statistic: 13.65 on 1 and 109 DF,		*p-value: 0.0003455*			

The relation is significant, and the residuals seem to be doing well, with their mean close to zero. R^2 is not very high at all. Nevertheless, the most remarkable feature to report is, that it displays a negative correlation. It tells the tale backward. It goes against expectations, and the estimate of the lineal regression model shows a good degree of slope (β = -25.833). This is without doubt a remarkable result. Still, could such correlation and estimate be circumstantial? i.e. could it be caused by other variables that determine the nest population? Is it unlikely?

The results of other tested climatological variables without significant results were: "wMPrcArous" (mean precipitation while aroused from April through August) has a p-value = 0.5909; "wMAvgMxTHb" (mean of the average maximum temperatures while in hibernation) has a p-value: 0.4617.

3.0.3 Provision on the Environment

So far, the most salient observations on their habitat are, that elevation plays a role on the distribution of dormice in south-west Sjælland; and that dormice choose distinctive areas for nesting: Usually the area covered by their nest range is on average above their surroundings, but we know little about how those areas look. Going back to the Normalised Difference Vegetation Index data, I decided to try to characterise the area by the presence of vegetation. In this case I used an averaged NDVI value to a 25m radius from the nest. This distance was chosen in relation to the results of the previous research using radio-tracking devices, (Fischer 2014) which established the nest range of individuals by minimum bounding geometry, and by their total travelled distance. 25 m is also the average distance between nests in my dataset. As a result this averaged NDVI value is 0.7557 (StD = 0.07578), which is higher than the average value for the entire forest extension. Furthermore, in order to provide a sense of scale, I have to report that the average NDVI value of July in Sjælland is around 0.10, when I include all land uses (mind you, values for NDVI range from -1 to 1). Thus the results report that forests belong to a land use with high NDVI values, but most importantly at this point, that *M. avellanarius* turns up above those values. Could this mean that they willingly place themselves in areas of thick vegetation and foliage? Do these NDVI values result from a significant deviation from average NDVI values in the forest?
I carried out yet another t-test (table3.5)in the surroundings of dormouse nests, in relation to the average NDVI value of south-western forests in Sjælland:

Table 3.5: One sample t-test. NDVI values surrounding nests are significantly greater than the mean.

```
> t.test(NDVI, alternative="greater", mu=0.7408763795)
        t = 41.904,               df = 3313,            p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0.7408764
95 percent confidence interval: 26.87151    Inf
sample estimates: mean of x = 27.93942
```

The one-sample t-test indicates that there is strong evidence, that values in the locations where dormice nest deviate above the average NDVI values. The null hypothesis has been rejected with great significance (p-value < 2.2e-16). It seems to transpire that they willingly place themselves in areas of thick foliage. So what else can we say to describe the forest that surrounds them. Well, the heights of trees surrounding a nest are not representative of the trees within the community. R-squared already elucidates how little does one know about whether dormice live in a broad area. One must only consider our surrounding tree height (R^2: 0.00979). The regression variable cannot explain more than a 1% of the variation. Moreover, the regression model seems to have real difficulties with fitting the point cloud: residuals are off balance, their standard error is not reassuring . . . see table3.6.

Table 3.6: Correlation between the surrounding tree height on occupied nests and the averaged tree height within the community grounds.

```
> summary(lm(SurTreeHgh~CoAvgTrHgh))
Residuals:
                        Min        1Q       Median      3Q       Max
                      -10.040    -6.498     -4.298     7.003    12.996
Coefficients:
                      Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)           -0.7442    14.0265     -0.053    0.958
CoAvgTrHgh             0.8580     0.8265      1.038    0.302
Significance codes:      0 ‘***’   0.001 ‘**’   0.01 ‘*’   0.05 ‘ . ’   0.1 ‘ ’
Residual standard error: 8.4 on 109 degrees of freedom
Multiple R-squared:  0.00979,              Adjusted R-squared:  0.0007053
F-statistic: 1.078 on 1 and 109 DF,        p-value: 0.3015
```

So, there is no correlation between the height of the trees surrounding nests and the average tree in an area. This actually very informative, for it supports the idea, that nests are placed most often on singular locations. Trees that host nests are not representative of their environment. They might deviate far enough from the averaged (found to be 16.97 m on the most exhaustive calculation on the normalised canopy model). The results of this t-test are reported in table3.7. Both average tree heights are significantly dissimilar.

Table 3.7: One sample t-test. Tree height on trees surrounding occupies dormouse nests deviate significantly and are smaller than the mean tree height within their community grounds.

```
> t.test(n$SurTreeHgh, alternative="less", mu=16.97)
      t = -4.2863,            df = 127,             p-value = 1.78e-05
alternative hypothesis: true mean is less than 16.97
95 percent confidence interval: -Inf 15.12225
sample estimates: mean of x = 13.95787
```

This t-test establishes strong evidence, that the average tree height surrounding occupied nests (11.9881 m, with StD 5.653) is smaller than the average tree height found in the community (i.e. 16.970 m, with StD 8.650). Dormice live in areas where smaller trees grow, but they do not necessarily live on small trees. In fact, these results neither prevent them from living in an isolated old tree (which is more likely to have cavities apt for nesting), surrounded by a much thicker an younger forest. I must remind the reader, that I was never capable of finding the hosting tree, due to discrepancy with the recorded coordinates. This is the reason why I attempted to get, instead, an averaged height of trees surrounding the nest location (i.e. about the maximum expected margin of error on the GPS: 5 meters).

When I tried to correlate the results of this search with the only field measurement that I should be sure about its precision, the girth of the host tree could be easily measured with a tape, the results were completely negative.

Table 3.8: Correlation between measured tree girth and its averaged surrounding tree height in a five meter buffer.

```
> summary(lm(Tree_Girth~SurTreeHgh))
Residuals:
                      Min         1Q      Median        3Q        Max
                 -110.904    -34.984      -9.904    38.186    139.070
Coefficients:
                 Estimate  Std. Error    t value   Pr(>|t|)
(Intercept)      116.7827     10.6642     10.951     <2e-16    ***
SurTreeHgh        -0.2218      0.6645     -0.334      0.739
Significance codes:     0 '***'   0.001 '**'   0.01 '*'   0.05 '.'   0.1 ' '
Residual standard error: 59.54 on 126 degrees of freedom
Multiple R-squared:  0.000883,          Adjusted R-squared:  -0.007046
F-statistic: 0.1114 on 1 and 126 DF,    p-value: 0.7392
```

It is not only insignificant, the residuals are far from showing an aligned point cloud, and also the slope of the model is inverse, which is a catastrophe. Furthermore the R-squared value is ludicrously low (R^2: 0.000883). As good as null variation in the independent variable has found explanation on the dependent variable. Basically, the results show that both sets of data behave independent from each other. Their relationship is very close to random.

Other relevant characteristics of the immediate physical surroundings of the nest can be analysed by studying the slopes. In this case the variation of slopes, after being classified into rough-medium-flat, and normalised to 0-100%.

Table 3.9: This one sample t-test establishes that soil disturbance within dormouse clusters does not deviate significantly from the average disturbance found within their communities.

```
> t.test(ClNaturlRt, alternative="greater", mu=61.5630918334)
      t = 0.34482,            df = 16,            p-value = 0.3674
alternative hypothesis: true mean is greater than 61.56309
95 percent confidence interval: 58.82432     Inf
sample estimates: mean of x = 62.23714
```

This t-test tries to discern whether the degree of disturbance of the ground within the nest range of dormice differs significantly from the average. The test does not prove to be significant enough to reject the null hypothesis (p = 0.3674), in spite of the fact, that the actual average value within clusters (where dormice live) shows a greater slope variation than in the rest of the territory.

Then I try a very similar variable, yet in this case, I am using nests as the base for observations. It is then important to keep in mind, that the pixel size I work with is now 20m, which is good enough to encompass its nest range. The extraction of data is done from those pixels onto the nest point, which allows for the following statistical summary on soil disturbance much closer to the nest's location:

Table 3.10: Statistical summary defining data on soil disturbance by a 20 m spatial unit or clNDrRtHa. This data is continuous, as it results from the average of the slope analysis carried at a spatial resolution of 0.4 m. It ranges from 1 to 3, and it shows, immediately, not just a bias from 1.5, but also a small Standard deviation.

```
> summary(ClNDrRtHa)
        Min.            Mean        Standard Deviation        Max.
       1.3252         1.997503           0.215297            2.3932
```

This metric ranges now, theoretically, from values 1 to 3 and it is continuous data, for it has averaged 5000 initial 0.4 m pixels. In reality it adopts a minimum value 1 on perfectly levelled surfaces like lakes, and a maximum recorded value of 2.7032 when I trained on data in Sjælland, where terrain was very uneven. After a simple eye inspection of such terrain, the reason for unevenness was a well developed degree of drain features on the ground. Forests seem to have and even denser network than agricultural fields.

Table 3.11: This one sample t-test establishes that soil disturbance on a 20 m spatial resolution deviates significantly from the average disturbance found within the research area.

```
> t.test(nDrVarRt, alternative = “greater”, mu=1.642798069515683)
      t = 17.262,            df = 110,            p-value < 2.2e-16
alternative hypothesis: true mean is greater than 1.642798
95 percent confidence interval: 1.961195     Inf
sample estimates: mean of x = 1.9975
```

From these results it is easy to infer that nests (with mean=1.9975 and StD=0.2153) are significantly located on grounds presenting a well developed network of drainage features. Not only dormice look for young, thick forests, and grounds above the average elevation to its surroundings, they also choose forest floors with a well developed drainage. There is very strong evidence that values deviate from the mean in the area (p-value < 2.2e-16). Moreover, all local variables (i.e. ground features, elevation, tree height, HDMI values ... at a spatial resolution of 5-20m) result in excellent degrees of significance.

Finally, in order to introduce directly the human factor into the filter on their environment, two simple correlations on the population size of dormouse clusters, in relation to their minimum distance to houses (taking a sample of the five closest houses), and in relation to the extension of suburban clusters of houses.

Table 3.12: Correlation between the number of nests in a cluster and the ratio of suburbanity in its grid cell. The grid cell used, same as with the address count, is 1Ha in size.

```
> summary(lm(ClNrNests~hNghSbUrRt))
Residuals:
                        Min         1Q          Median      3Q          Max
                        -8.917      -2.608      -1.288      3.707       5.823
Coefficients:
                        Estimate    Std. Error  t value     Pr(>|t|)
(Intercept)             12.2930     0.4743      25.916      <2e-16      ***
hNghSbUrRt              -38.9252    26.2299     -1.484      0.141
Significance codes:     0 '***'     0.001 '**'  0.01 '*'    0.05 '.'    0.1 ' '
Residual standard error: 4.311 on 109 degrees of freedom
Multiple R-squared:  0.0198,            Adjusted R-squared:  0.01081
F-statistic: 2.202 on 1 and 109 DF,     p-value: 0.1407
```

Although the resulting slope of the fitted lineal model is negative, as one would expect, this correlation does not have significance enough (p-value: 0.1407).

Table 3.13: Correlation between the number of nests in a cluster and the distance to their five closest buildings. The reason to choose for five building was to use measures of central tendency instead of just a single measurement, which often could be considered an outlier.

```
> summary(lm(ClNrNests~DistHs))
Residuals:
                        Min         1Q          Median      3Q          Max
                        -10.7764    -2.8282     -0.7548     4.1066      6.1389
Coefficients:
                        Estimate    Std. Error  t value     Pr(>|t|)
(Intercept)             11.656425   0.810356    14.384      <2e-16      ***
DistHs                  0.001362    0.003384    0.402       0.688
Significance codes:     0 '***'     0.001 '**'  0.01 '*'    0.05 '.'    0.1 ' '
Residual standard error: 4.351 on 109 degrees of freedom
Multiple R-squared:  0.001483,          Adjusted R-squared:  -0.007678
F-statistic: 0.1619 on 1 and 109 DF,    p-value: 0.6882
```

Neither seems to be of relevance the fact of being surrounded by farms and single family residences (p-value: 0.2825). Still these results should not disqualify the data, since their values could address problems on other variables and increase predictability of the whole model. Furthermore, if anything has been proven by this analysis, that is the role of scale at which the analysis is conducted, and the importance of measures of central tendency. It is thus, that I then queried occupied nests in relation to an average distance to the five closest buildings, as well as other geometrical distances to different human boundaries. You will find the table of results in appendixB, in tablesB.5 to B.9.

3.1 Variables

The values extracted from the training clusters define already a crude way of filtering. These are values on geometrical properties of the dormouse environment, and are summarised on table3.14. They could be part of the filter by informing fuzzy membership thresholds, and could later be overlaid to produce a cartographic result on physical requirements. These variables are meant to complement the disclosed provisions above, on climate, elevation and environmental traits.

Table 3.14: Weighted average distances and standard deviations on the location of dormouse nests. These are cluster averages, extracted from the training data before realising that a smaller spatial resolution would help locate nests better. They were weighted by cluster population count.

clAvgDstUr	*clStDDstUr*	*clAvgDstRd*	*clStDDstRd*	*clAvgDtDRd*	*clStDDtDRd*	*clAvgDstPB*	*clStDDstPB*	*clAvgDstHs*	*clStDDstHs*
4881.064 m	862.430 m	285.877 m	64.571 m	65.477 m	29.274 m	79.618 m	24.979 m	283.084 m	45.719 m
	1724.860 m		129.143 m		58.549 m		49.959 m		91.438 m
(3156.204 / 6155.324) m		(156.734 / 415.02) m		(6.982 / 124.026) m		(29.659 / 129.577) m		(191.646 / 374.522) m	

Therefore, besides these extensive variables, which respond to the presence of human features (in the final model represented by *DistRd*, *DistDRd* and *DistHs*), it has been found that there is a need for intensive variables: one responds to the presence of a drainage pattern on the forest ground (*nDrVarRt*), and a final one that responds to the variation of tree heights (*trHghVrRt*) at a pixel size of 20 m (i.e. 400 m^2). None of the values of these variables can be easily modified, and thus these determine the identification of new habitat. Nevertheless, whenever the use of the filter is meant to just identify existing habitat, the model could be enriched by two extra variables on vegetation characteristics like *nAvgNDVI* or *frstBiomass*. For more on key variable refer to chapterB tableB.10.

3.1.1 The Filter Model

The filter model was built by iterating several combinations of variables, starting with a blend of geometrical variables (i.e. *DistPB* + *DistDRd* + *avgDistHs* + *DistUr*). Only these four extensive variables seem to explain somewhere close to a 40% of the variation in occupied nests (see table3.15). After several iterations, where I would sometimes remove the variable that fitted worse, those that showed strong collinearity, and after changing spatial resolution to distances closer to their nest range, I arrived to the best combination, which can explain more than twice that initial value. The final model (see table3.16) reports an R^2: 0.9244, and the standard error of the residuals have improved to the point of being less than a third.
All variables are significant. Including variables like *ClAvgDstMB* (i.e. the distance to municipal boundaries). The incorporation of this variable might have not improved substantially the R_2 at that point in time, but it did reinforce its sensitivity.

Table 3.15: Steps in the process of building the filter model. Observe the drop of the residual error, and the rise of R^2. The model benefited from mingling extensive and intensive variables.

Regression Model:	RSE:	R^2
> lm(ClNrNests~ClAvgDstPB + ClAvgDtDRd + ClAvgDstHs + ClAvgDstUr)	3.835	0.3956
Extended Model:		
> lm(ClNrNests~ClAvgDstPB + ClAvgDtDRr + ClAvgDstHs + ClAvgDstUr + clNDrRtHa + hSumRdKmHa + hNghSbUrRt)	2.353	0.686
> lm(ClNrNests~ClAvgDstPB + ClAvgDtDRr + clAvgDstRd + ClAvgDstHs + ClAvgDstUr + ClAvgDstMB + clNDrRtHa + hSumRdKmHa + hNghSbUrRt + wAvgTArAut + hNghAddSHa)	1.576	0.8595

That step allowed to reduce the standard error to 1.576 (<1.668), and to increase R^2: 0.8595 a 2% (>0.8425). However, by introducing that new variable the model did not only not loose explanatory power, but it became better at varying together with the independent variable (appearance of occupied nests).

Table 3.16: Summary of the final regression model.

> summary(lm(ClNrNests~DistPB + DistDRd + avgDistHs + DistMB + DistRd + wAvgTArAut+ hSumRdKmHa + nDrVarRt + nTrHghVrRt + Richness))					
Residuals:					
	Min	1Q	Median	3Q	Max
	-3.2359	-0.6478	-0.0582	0.6543	3.1890
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-5.184e+02	4.654e+01	-11.138	< 2e-16	***
DistPB	-1.608e-02	2.122e-03	-7.580	1.80e-11	***
DistDRd	-1.426e-02	2.765e-03	-5.158	1.27e-06	***
avgDistHs	2.280e-02	2.099e-03	10.862	< 2e-16	***
DistMB	-2.373e-03	3.454e-04	-6.872	5.50e-10	***
DistRd	-1.284e-02	1.760e-03	-7.293	7.26e-11	***
wAvgTArAut	4.576e+01	4.011e+00	11.410	< 2e-16	***
hSumRdKmHa	-1.146e+01	1.092e+00	-10.496	< 2e-16	***
nDrVarRt	-1.532e+00	7.604e-01	-2.015	0.0466	*
nTrHghVrRt	-5.447e-02	2.363e-02	-2.305	0.0232	*
Richness	1.361e-01	6.388e-02	2.131	0.0355	*
Significance codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’	0.1 ‘ ’
Residual standard error: 1.25 on 100 degrees of freedom					
Multiple R-squared: 0.9244,		Adjusted R-squared: 0.9168			
F-statistic: 122.3 on 10 and 100 DF,		p-value: < 2.2e-16			

Furthermore, the final model has a very small residual median (-0.0582), their values are not skewed and, most importantly, the RMSE is 1.2068. I only need to check for multicollinearity and, most importantly, its real impact on its estimators. In figure3.4 can be seen, that there is no impact in the variance of the estimators.

Because I only care about prediction (i.e. the cartography to be produced), check figures3.5 to 3.8, I can simply live with the detected levels of collinearity, which are frankly moderate. I just need to make sure that I restrict the scope of the model to coincide with the range of predictor variables. After calculating the VIF, only three variables exhibit probable collinearity (*avgDstHs*, *DistRd*, *DistMB*). See table3.17.

Table 3.17: Table representing the Variance Inflation Factors (VIF) calculated by the package *rms* on Rstudio. The calculation is carried out on the covariance matrix of parameter estimates, using the method of Davis et al. 1986. (Harrell 2017)

> vif(lm(clNrNests~DistPB + DistDRd + avgDistHs + DistMB + DistRd + wAvgTArAut+ hSumRdKmHa + nDrVarRt + nTrHghVrRt + Richness))									
DistPB	*DistDRd*	*avgDistHs*	*DistMB*	*DistRd*	*wAvgTArAut*	*hSumRdKmHa*	*nDrVarRt*	*nTrHghVrRt*	*Richness*
3.540196	*1.377129*	*5.902518*	*6.115112*	*5.753409*	*3.548926*	*2.336245*	*1.881533*	*1.358666*	*2.661525*
(3156.204 / 6155.324) m		(156.734 / 415.02) m		(6.982 / 124.026) m		(29.659 / 129.577) m		(191.646 / 374.522) m	

In fact, that level of multicollinearity should not be important, because after calculating a ridge regression, none of the variables presented any effect on its estimators due to correlation to other variables. I multiplied the covariance matrix by several lambda values, and the results were simply a flat (unmovable) value of the estimators.

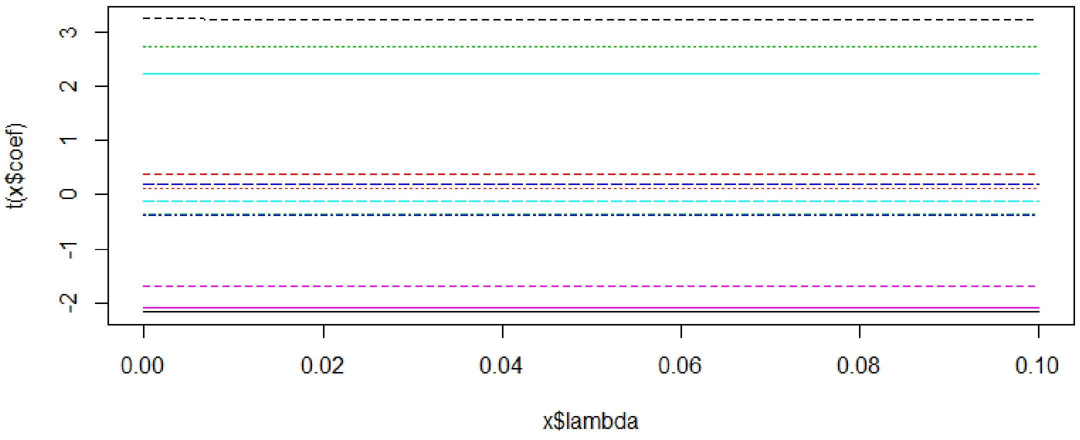


Figure 3.4: Ridge regression of the variables final statistical model. The estimators remain unaffected by collinearity.

In spite of the goodness of the model, it can still be calculated by a robust method, as explained in chapter2.2.4, in order to calculate another R^2, following the advice given in by Willett and Singer 1988. The RMSE of this robust regression was 1.24.

Table 3.18: Robust regression model results, as calculated by the package *rms*.

> summary(rlm(clNrNests~DistPB + DistDRd + avgDistHs + DistMB + DistRd + wAvgTArAut+ hSumRdKmHa + nDrRtHa + nTrHghVar + Richness, method = c("M", "MM", "model.frame"), wt.method = c("inv.var", "case"), model = TRUE, x.ret = TRUE, y.ret = FALSE, contrasts = NULL))

Residuals:

Min	1Q	Median	3Q	Max
-4.09176	-0.59179	-0.05792	0.66911	3.79142

	Value	Std. Error	t value
(Intercept)	-594.0241	39.9481	-14.8699
DistPB	-0.0158	0.0018	-8.6867
DistDRd	-0.0151	0.0024	-6.3445
avgDistHs	0.0206	0.0018	11.4205
DistMB	-0.0029	0.0003	-9.7399
DistRd	-0.0103	0.0015	-6.8196
wAvgTArAut	52.2566	3.4428	15.1786
hSumRdKmHa	-12.7914	0.9376	-13.6433
nDrRtHa	-1.4711	0.6527	-2.2538
nTrHghVar	-0.0352	0.0203	-1.7373
Richness	0.1080	0.0548	1.9690

Residual standard error: 0.9898 on 100 degrees of freedom

The calculation of the Root Mean Squared Error for the robust regression model went as follows in *Rstudio*:

```
> rmse <- function(error){sqrt(mean(error^2))}
> rmse(model$residuals)
rmse = 1.237546
```

The calculation of R^2 for the robust regression model, following Willett and Singer 1988, went as follows in *Rstudio*:

```
> SSe <- sum((model$w*model$resid)^2)
> observed <- model$resid+model$fitted
> SSt <- sum((model$w*observed-mean(model$w*observed))^2)
> r2 <- 1-SSe/SSt
r2 = 0.9726302
```

Thus finally, not only the statistical model has a good level of sensitivity to the phenomenon, but the sign of its estimators makes sense. On validation of the filter, the membership strength of the validation set (median: 0.9507, with IQR: 0.1146) fits closely that of the training set (median: 0.9464, with IQR: 0.0983), and an ANOVA test has shown no difference in variance between both sets (see TableB.4).

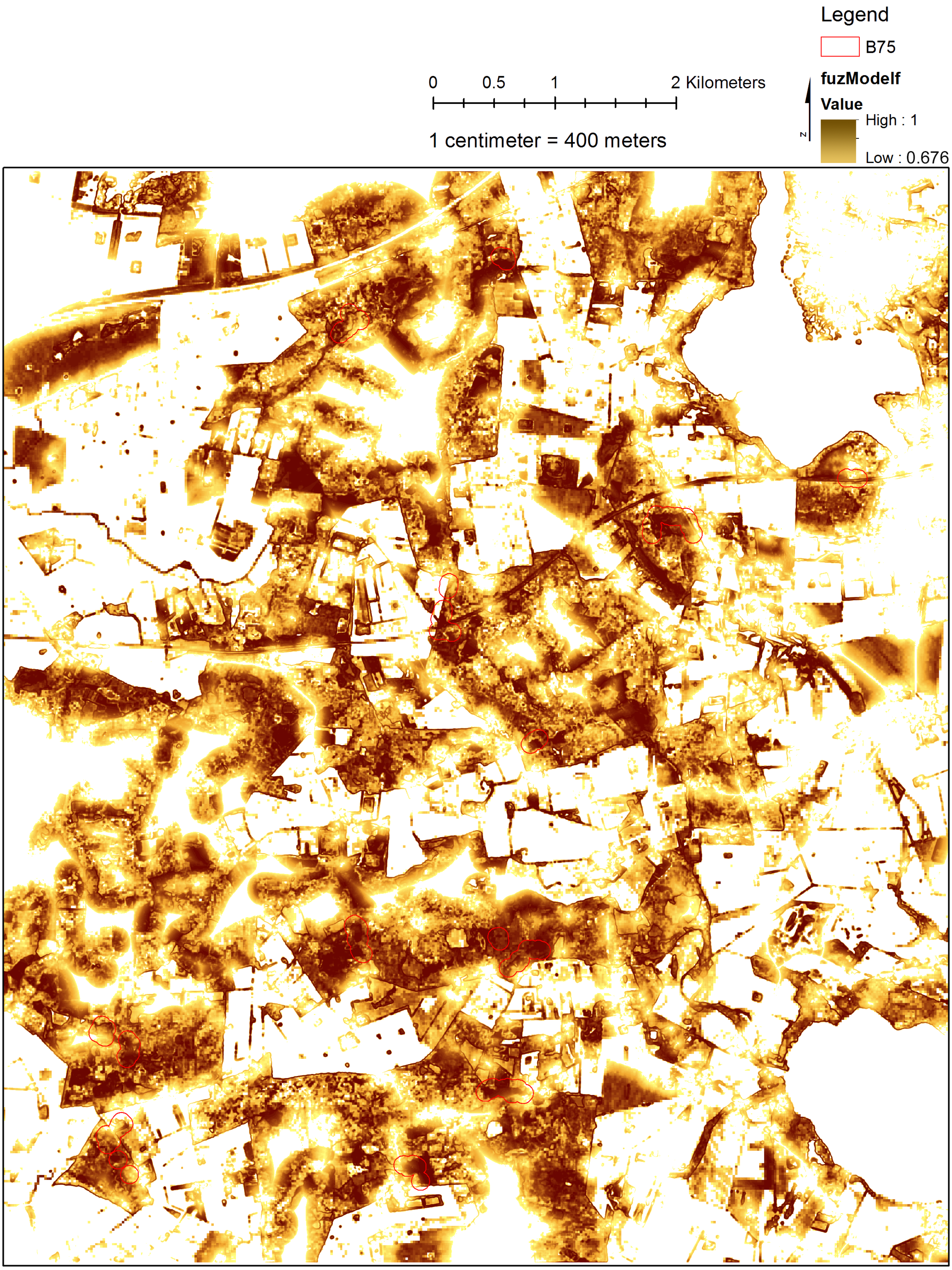


Figure 3.5: Cartographic result of the fuzzy filter applied on south-western Sjælland. Values have been stretched between those above the minimum detected on an occupied nest (0.676). Mind that the median membership value is 0.9464 (IQR:0.1). Nest cluster areas (buffers) lay on top.

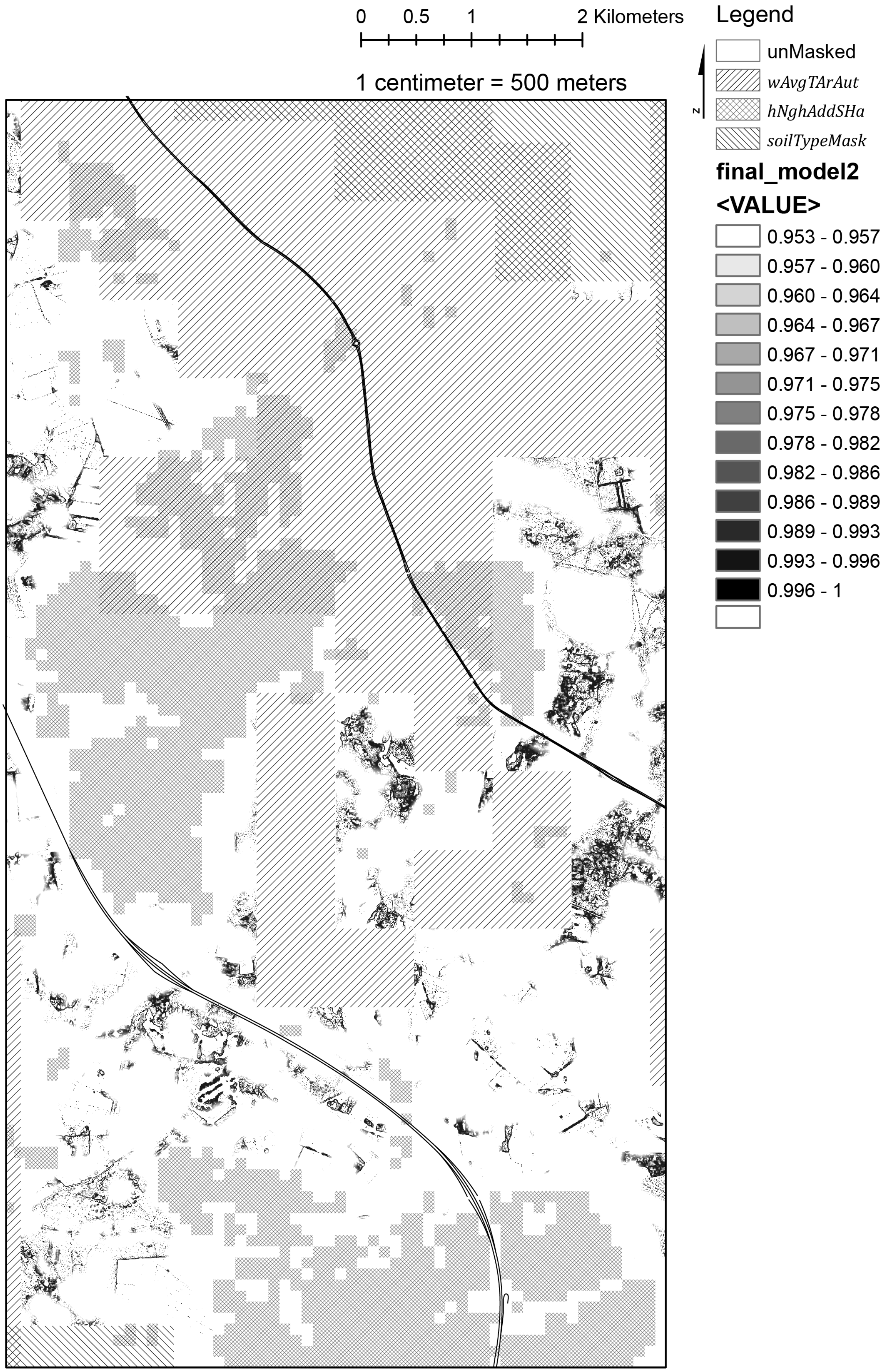


Figure 3.6: Hatched overlay of masks applied to the fuzzy filter on north-eastern Sjælland.

Figure 3.7: Final result of the fuzzy filter, masked by selective metrics, on top of an orthophoto.

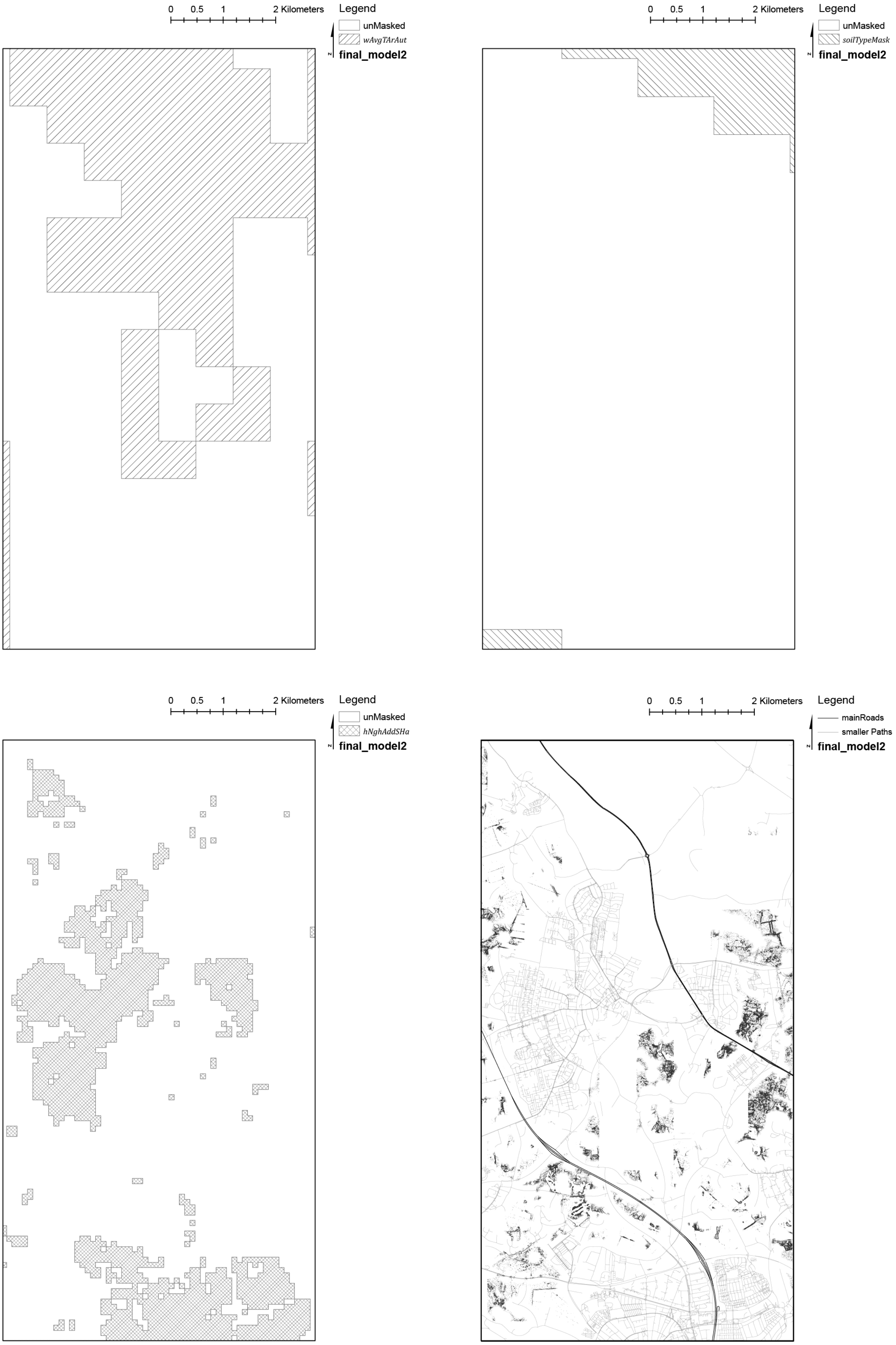


Figure 3.8: Three separate masks applied to the filter (weather, soils, urban) & a reference to roads.

Chapter 4

Discussion

In this chapter you will find an analysis and evaluation of the results, in accordance to the repeatability, simplicity, and sensitivity of both, methods and results, toward the phenomenon. You will also find a rationalisation of the results, a juxtaposition of possibilities, as well as a critical detachment from the theoretical train of thought, needed for a proper counterargument.

4.1 On Methods

The acquisition and pre-processing of the data was carried on *ArcPy*, and it was fully automatised by python. This method proved to be effectual when downloading tiled raster datasets, as well as when the data has been stored in the FTP server under a meaningful naming convention. The creation of the database was scripted in SQL, which is far more efficient and reliable than actually editing fields in a spreadsheet. Both languages (python & SQL) have as their strong suit an improvement in consistency and reliability, although, on the other hand, sometimes the implementation of certain tasks is not straight forward.

For instance, *PostgreSQL* is a spatial relational database very suited for spatial measurement. However, it is not advisable to do any sort of statistical calculation on it, like producing standard deviations. These procedures had to be performed in R, since, for some reason, an error would accrue in the calculation. Standard deviation involves raising data to the power of two, which increases the error, right before performing the addition of several measurements. The larger the amount of points and their measurement value (in some instances up to several kilometres), the larger the distortion introduced by the coded calculation.

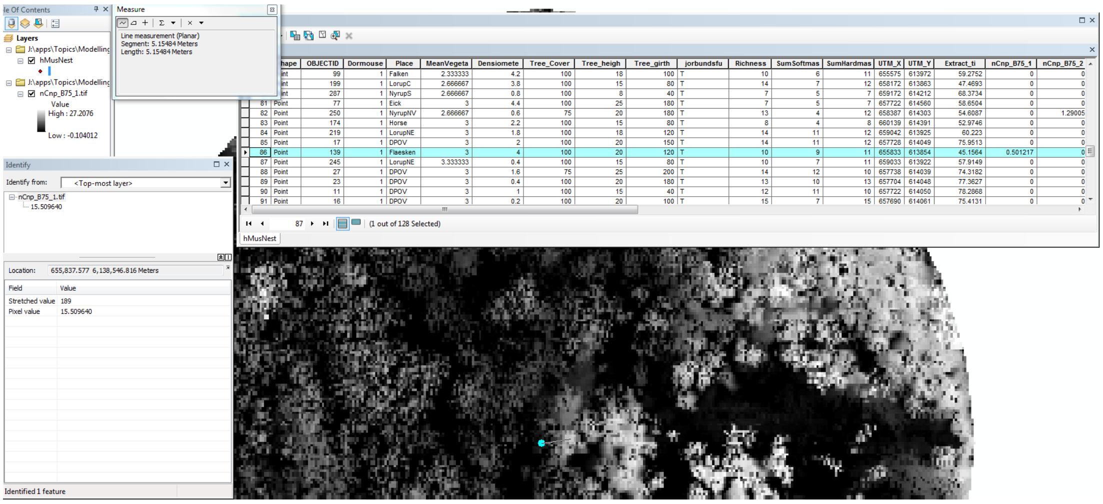
Furthermore, single calculations like obtaining the distance to the nearest house could be implemented without problem in *PostgreSQL*, however, due to the certainty of the presence of outliers in those measurements, the method to try to keep as many measurements as possible demanded to use measures of central tendency.

This increment of measurements, ordered by their neighbourhood to a nest in order to produce an average distance, could not be directly implemented. I decided to average the five closest building to a nest. This meant that I had to code in plpgSQL (a procedural language for *PostgreSQL*) a loop that ordered buildings by nearness, and executed the measurement up to five times. Unless one has a good command of SQL, getting this procedure done can cost some time delay.
With regard to python, which used *ArcPy* as interpreter, one must be aware that certain libraries or tools available in ArcMap, have not been implemented in *ArcPy*. In such cases run time errors will pop up. One specific task that was difficult to implement on python, was the translation of non-ASCII characters from table name, attribute heading, and data rows. This implied, that I had to code a subroutine, a 'def' to take care of Unicode characters. It looked as follows:

```
def translate(infield):
data = ''
for i in infield:
print i
if u'\xc6' in i:
data = i.replace(u'\xc6', "Ae")
elif u'\xe6' in i:
data = i.replace(u'\xe6', 'ae')
elif u'\xc5' in i:
data = i.replace(u'\xc5', 'Aa')
elif u'\xe5' in i:
data = i.replace(u'\xe5', 'aa')
elif u'\xd8' in i:
data = i.replace(u'\xd8', "Oe")
elif u'\xf8' in i:
data = i.replace(u'\xf8', 'oe')
else:
data = i
return data
```

Yet again, simple things like the use of regional characters can make it difficult and impractical to use a database, and cause to spend some time and effort, since these are not directly supported on python. It is much simpler to use a standard nomenclature, that also facilitates the management of the files.
In relation to the spatial analysis, perhaps the most difficult task was abridging the big mismatch between recorded tree height on the field, and the measurement derived from *LiDAR* data by using the UTM coordinates on the database. The probable origin of such mismatch could be on the actual inaccuracy of commercial GPS, which goes up to meters, and makes difficult to pinpoint the hosting trees.

According to the normalised canopy data (i.e. the subtraction of the DTM from the DSM dataset), hosting trees showed a much lower height than the declared by field data, often below a meter, which meant that they laid on shrubs. It is difficult to say how much, but even though the previous example talks of an inaccurate location, the source of the mismatch could also stem from the method used to measure the hosting trees. It is not easy to measure an entity 20 metres tall or more. It calls for either high-end equipment, i.e. a theodolite and enough space to make a comfortable measurement, or the kind of equipment and complex procedure to climb the tree. From reading the report by Fischer (2014), she mentions a measure tape among their tools, but not much more required to make this measurement.



Line measurement (Planar)
Segment: 5.15484 Meters
Length: 5.15484 Meters

	hape	OBJECTID	Dormouse	Place	MeanVegeta	Densiomete	Tree_Cover	Tree_heigh	Tree_girth	jorbundsfu	Richness	SumSoftmas	SumHardmas	UTM_X	UTM_Y	Extract_ti	nCnp_B75_1	nCnp_B75_2
	int	99	1	Falken	2.333333	4.2	100	18	100	T	10	6	11	655575	613972	59.2752	0	0
	int	199	1	LorupC	2.666667	3.8	100	15	80	T	14	7	12	658172	613863	47.4693	0	0
	int	287	1	NyrupS	2.666667	0.8	100	8	40	T	7	5	7	659172	614212	68.3734	0	0
81	Point	77	1	Eick	3	4.4	100	25	180	T	7	5	7	657722	614560	58.6504	0	0
82	Point	250	1	NyrupNV	2.666667	0.6	75	20	180	T	13	4	12	658387	614303	54.6087	0	1.29005
83	Point	174	1	Horse	3	2.2	100	15	80	T	8	4	8	660139	614391	52.9746	0	0
84	Point	219	1	LorupNE	3	1.8	100	18	120	T	14	11	12	659042	613925	60.223	0	0
85	Point	17	1	DPOV	3	2	100	20	150	T	14	11	12	657728	614049	75.9513	0	0
86	Point	139	1	Flaesken	3	4	100	20	120	T	10	9	11	655833	613854	45.1564	0.501217	0
87	Point	245	1	LorupNE	3.333333	0.4	100	15	80	T	10	7	11	659033	613922	57.9149	0	0
88	Point	27	1	DPOV	3	1.6	75	25	200	T	14	12	10	657738	614039	74.3182	0	0
89	Point	23	1	DPOV	3	0.4	100	20	180	T	13	10	13	657704	614048	77.3627	0	0
90	Point	11	1	DPOV	3	1	100	15	40	T	12	11	10	657722	614050	78.2868	0	0
91	Point	16	1	DPOV	3	0.2	100	20	100	T	15	7	15	657690	614061	75.4131	0	0

(1 out of 128 Selected)

Location: 655,837.577 6,138,546.816 Meters

Field	Value
Stretched value	189
Pixel value	15.509640

Identified 1 feature

Figure 4.1: This screen print shows in bright cyan a nest location. From the table can be read that its height is 0.5 m, far lower than what it was meant to be. About 5 m east, stands a tree 15.5 m tall.

Both could be valid explanations. There could be some inaccuracy on measuring the height of tree or on recording their location. Specially since I do not know the model of GPS used. If I had access, I could try to select unequivocally identifiable targets on location, and then carry calibration measurements against the high-precision of *LiDAR* data, in order to check the device's accuracy when surrounded by trunks. Nevertheless, the solution to this problem did not come from assessing accuracy, but from understanding what matters to the dormouse from its environment. If we take some steps back on what led us to this predicament, I could hypothesise, that dormice habitat is not so much characterised by the height of the hosting tree, but by the diverse heights within its nest range. This is a simple solution: instead of focusing on a datum, I searched for canopy elevation in a 5m radius to the given location. I did that and looked for the tallest tree, after having dismissed shrubs and small trees up to 2.5m, by applying a command like 'fill' on sinks (i.e. minor peaks), which did also lighten the data processing.

In figure4.1 is depicted the normalised canopy data of a cluster (i.e. 75 m in diameter). It is plain to see that the point location of the nest falls into dark pixel value, indicating low elevation, in contrast to its surrounding grey and up to white pixels, indicating higher elevation. About 5m east stands a tree measured as 15.5 m of height. Five metres is too far to be the actual location of a nest, when the tree girth is established at 1.20 m. However, the actual measurement of that very same host tree is declared at 20 m, and that does not fit any of the elevations that surround this location. Thus, in terms of simplicity, this solution played well. It was probably not just a matter of shifting the datum of the projection by so much, nor about finding the actual tree itself. This approach delivered an average height of 12m, while field data from previous research had it in average at 18.42m. Both data were lower than the cluster average height, as well as the average height of trees comprised by the community. However, the new measurement did not only sharpen the result, but brought a problem that seemed of accuracy, into a problem of optima for dormice. This exercise is about finding the optimal relations that characterise their habitat.

4.1.1 Accuracy

The way I conducted the measurements did not raise a problem of accuracy, which is present even more when it is focused on a single entity, but, as it introduced the neighbourhood and measures of central tendency, it focused on the whole. Accuracy is also an issue when the gear and methodology used for obtaining the data is not very demanding. The NOVANA guidelines (Søgaard, Elmeros, and Vilhelmsen 2013) only mention to bring measuring tape to the location, probably accurate enough for smaller measurements like girth.
There is always the possibility that something went wrong on the geoprocessing of the data, however, unless the geotagging of the nests was done with expensive equipment, as in backpack dual-band GPS with signal enhancement, the error is most likely to have occurred when the signal bounced from a nearby trunk. According to governmental data[1], GPS-enabled smartphones are accurate to within 4.9 m, and are at their worst when buildings or dense trees are present in the area, because user range error depends on things like the satellite geometrical location, signal blockage, receiver design . . . Enhancing the receiving signal requires battery power, as well as using a couple of antennas in charge of picking different frequencies. It is not only a much more expensive piece of equipment, but also much more cumbersome. In contrast, I decided to rely on the tendency expressed by *LiDAR* data, which has today an unsurpassed accuracy. It produces four to five points per square metre, and their accuracy is established to be of 15 cm in the horizontal, while only of 5 cm in the vertical dimension. (SDFE 2016)

---

[1] visit http://www.gps.gov/systems/gps/performance/accuracy/

4.2 On Results

Dormouse Nest Distribution

Dormice cluster into specific locations in the landscape. Being in some degree a K-strategist, makes him/her rely on efficiency, be selective, and a good candidate to be studied by spatial analysis. It is a species that makes an strategic use of food, follows patterns, adapts population numbers to the avail of territory, and it is placeable, for it relies on establishing an efficient relation to its environment.

To be a species fit for spatial analysis does not come exempt of difficulties and folds. The choice for a weighting methods is prove of that. In figureB.1 I presented that dormice nests are not arranged following a priority for compactness, but by following lines and edges in the landscape. There is an inverse correlation between the number of nests and their compactness ratio. An efficient adaptation to the territory seems to outweigh a sense of social membership that, we, primates, assume as the most efficient strategy to survival. So it seems, that the process of finding reliable methods and proxy variables has not been exempt of doubtfully K-strategist postulates. The spatial distribution of the tiny, light-weight dormouse seems to be defined mostly by a question of shelter. That is, it is defined by specific light conditions and varied plant communities, that result in dense foliage (i.e. shelter) and higher productivity of a varied understory (i.e. food).

There is a significant difference between their mean cluster compactness (mean = 0.5184, IQR = 0.1396), very influenced by such landscape lineal features, and their somewhat rounder community compactness (mean = 0.6097, IQR = 0.1634), thus closer to 1. Such difference comes to uphold the idea, that dormice are influenced by the landscape differently at different scales. The value of scale is of great significance not just for ecology, but for any qualitative measurement. Most geometrical variables of the dormouse environment tend to be of significance, when observed at the closer scale of its nest range, lesser in that of the cluster, and least at that of the community. Thus, what matters for the dormouse, is what happens at close range, which also upholds the idea that *M. avellanarius* can be very adaptable to the geometry of a confining space, at a scale close to its neighbourhood.

4.2.1 Elevation

The salient elevation values of most nests in a rather flat island like Sjælland, prompted to investigate this variable right away. The first and perhaps the most clear piece of information reported, was a significant correlation between the amount of nests found and their elevation (p-value = 2.653e-07). This could be explained by several hypothesis.

One could be, that with elevation there is a reduction in average temperature, which roughly corresponds to one average degree every 100 meters of ascension.

This could have an impact on how long dormice can hibernate without experiencing an arousal, which is not a little. Hibernation is not devoid of risks, and dormice survival is not guaranteed. Another hypothesis could involve some other lurking variable, which might correlate with elevation. For example, I tried to research if a reduction of households at altitude could be of significance, and it had a positive answer. There was a significant correlation between the number of addresses and elevation (p-value = 1.24e-07, see table3.2). However, the estimate of the slope was flat (β = -0.02223), indicating very little variation with respect to the chosen explanatory variable, and the percent of explanatory power of such variable was near to none (R^2 = 0.004). The difference in altitude is not more than 90m in the area, which is very little to really have an impact, let alone be explanatory. It becomes apparent that such difference in height could not translate per se into a lower intensity in landscape use, neither in economic activity, nor in plant productivity. However, the choice of land use has been more explanatory, and 'choice' should be better rephrased as human, cultural, tradition-bound practice. There are some pieces of evidence to think that way, see figureB.2. If elevation is not salient enough to deter human activity, it is however human activity itself, that seems to have deterred change, as it is rooted in time. The concentric lay out of parcels around inhabited nuclei in the agrarian neighbouring area (to the south-west of te forest), is a very old pattern. It suggests that a stable cultural, socio-economic phenomenon has been allowing and shaping the existence of dormice. That is when habitat, or the presence of forest at 'altitude' became of relevance. More than a shift in intensities (i.e. temperature, cultivation, moisture ...), it is probably the cultural allocation of forests that makes dormice nest at such 'altitudes'.

It became clear that dormice do not attempt to attain the highest elevation, because they are not even found where the hill summits. Neither there is significant evidence, that nesting on the highest grounds is a pattern of behaviour, the t-test does not show a significant deviation of nest locations from the averaged height that surrounds them (see tableB.3). From these results one could take, that the role of elevation in the ability to sustain the population is not significant. The difference in elevation does not comprise a range of variation big enough to produce changes in vegetation. Case in point is the homogeneity of forest production recorded in terms of NDVI values (see table3.3). The presence of forest seems the only explanation. Forests where dormice are found not only have a high NDVI value (0.74088), as it would be expected, but also an unremarkable standard deviation (0.09494). Their NDVI values might oscillate randomly within their extension, on the slopes . . . However, any affect of elevation on productivity was conclusively dismissed on figure3.3, for the top NDVI values remain flat, and do not decline along the x axis (i.e. along with elevation). If anything, the variation (i.e. standard deviation) of values only becomes smaller with altitude. A continuous presence of forests at higher altitudes was then demonstrated by the continuity of NDVI >0.7.

Nevertheless, it is important to notice that the locations where dormice are found are significantly above average NDVI values (see table3.5), with a level of significance of p-value < 2.2e-16. It is also important the fact that these locations have an even lesser standard deviation than the naturally found on their environment (StD = 0.07578 < 0.09494). That could indicate that dormice choose distinctive areas for nesting: Dormice willingly place themselves in areas with a great deal of foliage density, and thus also vegetative production. The results on average tree elevation surrounding nests are also coincident with the idea that nests are placed on singular locations. The heights of trees surrounding a nest are not representative of the trees within the community. They deviate consistently enough from the average tree height to test positive (see table3.7). Tree elevation surrounding nests is in average 11.988 m (with a StD: 5.653), which is smaller than the average tree height found in the community (i.e. 16.970 m, with StD: 8.650). The choice of smaller trees might have a probable connection to the development of low laying biomass, and thus with shelter from predators. A dale or an edge on the forest might provide a thicker branches and foliage.

So, when I researched into the importance of differences in elevation at different scales, not only average elevation declines as the area around the nests is enlarged (60.07 m for nest locations, 58.78 m for clusters, and 56.94 m for communities), but the increasing difficulty on fitting a lineal model through the data is an indication of the relevance of local scales, and the scattering of values. Any larger scale only weakens the correlation, which makes sense from already the very simple fact, that a larger area might introduce a larger standard deviation (whenever there is actually some relief). Nevertheless, nests still have an average elevation (about a meter in average) above their surroundings. If their environment had to be analysed on a smaller scale, such different on elevation could only refer to surface features.

Dormice live where habitat (i.e. forest) is available, in selective locations. That indicates that the environmental variable (depicted in the ERD diagram, figure1.5) is very strong. However, this does not mean that climate does not play any role.

4.2.2 Climate

The climatological variables have been surprising. Previous research have noted the possible relevance of phenomena like NOA (i.e. the North Atlantic Oscillation), and most reports on the habitat of *M. avellanarius* make reference to variables like temperature and rain. Yet, given the spatial resolution of the meteorological data, plus, most importantly, given the confluence of its always present and stark auto-correlation, and given that the current data on *M. avellanarius* is strictly local (i.e. few kilometres across), my results might be far from what could be expected. Any correlation to rain or average temperature might be difficult to show, when there is nearly no variation in the data. Figure1.1 shows an isothermal value for the research area close to 2°C.

However, there is one significant relationship to the climate. It takes place when dormice are about to switch their metabolic rate, in order to begin hibernation. The average temperature of both September and October has, as reported, a negative correlation (β = -25.833) to the number of occupied nests. This is a surprising result, because these are months when dormice fatten up, and when quality, high-calorie nuts are available. Mostly hazel, but also chestnut or beech seeds become gradually available during September and October. A milder autumn should give dormice more chances to fatten up but, yet again, I must advice caution. Even though the level of significance is indisputable (p-value: 0.000346), it is not unlikely that there could be other, lurking variables.

There are probably a couple of ways to make sense of this situation. One goes down the path already put forward. There is a climatological tendency marked by prevailing winds, a weather inertia resulting from the influence of larger processes, like the North Atlantic Oscillation (NAO). This could mean that warmer autumns would correlate with milder winters. It is in the character of NAO to send prevailing winds, from either the west or the east, over long periods of time. Westerlies would bring the buffering influence of the gulf stream, thus higher average temperatures, and a rainier weather. Both are potentially disruptive.

However, in order to follow this argument, I must precise that the climatological data available averages the temperatures from the last 50 years. The data does not capture the actual oscillation of atmospheric pressure, yet it captures its general trend. If we regard such phenomenon as a rigidly uniform whole, the negative sign of the estimator would mean that NAO has had, in all, a negative influence on the dormouse population for the past 50 years. That would agree with the species decline, although caution must be advised again, this is just mere speculation without having the actual data on the phenomenon.

The other way to make sense of this stems from relying on a combination of lurking variables. The variable used in the correlation does not have to explain everything. Just like the correlation between elevation and nests is not fully explained by elevation itself. The reason to believe on a set of lurking variables, alien to the purely climatological, is that there are actually other variables in literature that could explain this. Starting first with what was mentioned in section 1.1.2: Juškaitis reports that winter mortality could relate to age structure at the time of hibernation (i.e. the percentage of young in the population, which has a $r = 0.84$; $p < 0.001$). From that text transpires, that the bigger the proportion of juveniles, the higher the mortality over winter. In relation to the variable at stake a longer 'summer' could influence having a larger proportion of juveniles at the beginning of hibernation. Furthermore, there is forests and elevation, and this turns the argument around full circle. Forest (i.e. habitat) extends higher on elevation, and with that comes a decrease in average temperatures. This explains the sign of the estimator. It is negative for habitat increases with altitude, same as averaged temperatures descend.

4.2.3 Environment

So far, the most salient observations concerning the environment have been, that elevation plays a significant role on the distribution of dormice in south-west Sjælland, and that dormice tends to select singular locations, both in terms of their vegetation index and in the ages of the trees present. In addition, we have learnt that the climate is also shaping the area of distribution of *M. avellanarius*.
NDVI has proven to result in a very simple but effective way to analyse dormouse habitat. On the other hand, its values can be made to vary, for they depend on the very land use we often assign. By this I mean, that one must be careful about the contingency of environmental variables. Some have shown significance, but that does not take into account the degree of independence from human agency. NDVI values depend on land use, and land use is something that can change at our will, in contraposition to climatological variables, which do not. Thus NDVI values are significant but not independent from their context. Elevation seems to be significant, but, in this case, its significance still depends from the presence of a certain land use. All the same, its effect on average temperature and its degree of exposure do not. Variables can be that complicated in themselves.
So, while keeping all this in mind, I still have to honour the simplicity and sensitivity of some variables by still using them. For instance, the analysis on the forest ground, shows a consistently uneven surface where they nest. Such uneven character is to some extent contingent, manageable, yet important. Dormice tend to set their winter nest on the forest floor, on terrain above the elevation of its surroundings (about a meter in average). The t-test could not discern whether the degree of disturbance of the ground within the nest range of dormice differs significantly from the average (p = 0.3674), yet the mean of disturbance found was already higher than 50% of the values (62.34%). Yet again central tendency and smaller scales(an average above the actual value of clusters) becomes relevant. It indicates that dormice might prefer areas with a good system of drainage, i.e. areas that are not prone to flooding. One hypothesis could be, that winter nests benefit from the more stable microclimate offered by remaining close to the ground. Another would be, that such disturbance pattern is common to its natural habitat (old forests). The lack of machinery at work flattening the profile, allows for the surface to expose a natural drainage, created over centuries. So, perhaps what we can take from this, is that dormice live in the present and, contrary to what might be common among humans, they do not choose their residence because of its interesting historical background, nor are they aware of the ratio of unevenness. The variation of slope is indicative of an undisturbed natural terrain, and their habitat just coincides with that of old forests. Both hypothesis establish this variable as a good proxy, and it is still to be seen if land previously levelled by agriculture, featuring an efficient drainage, could still be repopulated by dormice, as long as a dense understory of *Lonicera*, *Rubus* . . . and a young, mixed forest including hazel were provided.

So this variable will be good at selecting virgin parcels, and deselect most agricultural parcels (depending on their drainage and if the cultivation technique uses ridges). Certainly it was helpful at dismissing any paved surface. Furthermore, even if the ground has never been altered by a plough, a soil developed by a sedimentary process would still have, naturally, large portions of flat surfaces. This is specially true when the average slope and the elevation of such area are low. Not many variables can ever be watertight.

Their notion of nesting place must then be expressed by a manifold or at list a combination of two variables. The other chosen variable for this combination was tree height variety. The main reason was that such variety indicates the presence of a diversity of species, or/and diversity of standing tree ages, or/and the presence of an edge on the continuum. Therefore, it takes care of three interesting traits at once. Also, just like before, any local variable (i.e. ground features, elevation, tree height, HDMI values ... at a spatial resolution of 5-20m) have been proven to have an excellent degree of significance. However, any smaller features will be gobbled up by a 20 m pixel. For example, there are small water surfaces, many small ponds and mini lakes, whenever one of its radii is less than 50 m could be reduced to a minimal expression of its initial surface. It would therefore be advisable to use a layer containing all water features in the masking selective process. Otherwise, big open waters are easily dismissed with great precision by the filter.

Finally, in order to introduce the human factor directly into their environment, one cannot simply use what correlates best. Already variables like drainage ratio and elevation have shown their parochial nature. Even though elevation tends to correlate with forests in many occasions, in many different countries, it is never a universal. It is more of a tailor-made proxy variable. In the same way, the filter (do not forget that uses membership values extracted from an agricultural area) will be tailor-made to agrarian environments. The fact that it is meant to be applied to a metropolitan area, does not allow the assumption that geometrical distances are the only reason for it not to be found there. That would defeat the purpose of seeing how the rest fare. Such distances must have a common ground on their measurement, between agrarian and metropolitan environments. Their units should have a small spatial resolution too. Even if distance to urban centres has a fantastic sensitivity and significance in south-west Sjælland, it cannot be allowed in a metropolitan area. Instead, distance to paths, houses, and roads brings that common ground, and already a significant difference in density.

In relation to the questions put, we need to be aware, that not only the successive socio-economic revolutions have reduced dormice habitat, but they have also changed forest management practices, our culture, the intensities we put our world through by means of technology and energy. What is important to understand, in order to get this study right, is that it does not depend only on the geometry of the spatial measurement, but on sensitiveness to certain key relationships.

Chapter 5

Conclusion

Some first conclusions to extract from this exercise are in relation to the species ecology. As one could expect, the impact and/or significance of most variables act on the dormouse at close range. In consequence one could entertain the idea that *M. avellanarius* might adapt to the geometry of a confining space. Not once but three times, have OLS regressions shown, that the significance of many variables turns on using a smaller scale. For a creature like *M. avellanarius* such result invites to think of metropolitan spaces at a certain scale. One that makes a metropolitan enclave much more likely to succeed. After all, dormice are found in literature nesting by motorways, or here, in Sorø, nearby railway tracks.

It is also important to notice, that the locations where dormice are found are significantly singular. Within the training area, these are above average NDVI values, above the average elevation of its surroundings, and their tree heights deviate toward that of younger and more diverse communities. Whilst these are probably reliable variables to look for in order to filter for habitat, it is useful to remember that these are, as well, labile and manageable. After applying the filter metrics obtained from the training data set, the rate of success at identifying forests in the metropolitan area dropped. However, most restrictive filtering follows hedges and embankments, the chosen locations are only at times agricultural parcels, and in one case a golf course. Thus it still manageable into potential habitat.

Having said that, the first question on this report has three ways to be answered. One way is to conclude that there is space for dormice. Pixels with a membership strength above 0.85 form 39 areas, all bigger than 1.7 Ha (i.e. capable of hosting a tiny cluster). The 15 largest average 55 Ha each (>20 Ha). They feature lineal spaces (20-50 m in width), but their scale should account for their fitness. Another way is to conclude yes, a 15% of existing protected forest overlaps with areas above 0.85 in membership strength (273 Ha). However, the third would conclude that there is no space left for an enclave, because of *clAvgDstUr* (i.e. distance to cities, see fig.B.3). An assessment on the role of an euclidean distance as proxy is needed.

In relation to the second question, the spatial resolution of the gathered data and the local extent of the data on *M. avellanarius* disallow any conclusive statement. However, the significant relationship to the average temperature of both the month of September and October (i.e. *wAvgTArAut*), and most importantly its negative correlation (β = -25.833), still upholds the conclusion, that dormice are impacted by climatological phenomena, and influenced by elevation.

In relation to the third and last research question, the key traits that signal the dormouse's habitat have been found to consist of a blend of extensive (geometrical) and intensive measures. These became anthropocentric only for the purpose of the subject of study. This report has chosen to probe a metropolitan area. Geometrical variables have been proven capable of explaining a good deal of the variation on the number of occupied nests. However, they need the right scale and common ground between the training area and the research area. For instance, both areas contain roads and houses, and it seems that the size of such entities in relation to forests is what can better analyse metropolitan interstitial spaces. Thus *avgDistHs*, *DistRd*, and *DistDRd* were selected as the extensive traits that best characterise dormice nests. On the other hand, besides *wAvgTArAut* (i.e. average temperature in the months of Sept.-Oct.) and *hSumRdKmHa* (i.e. accessibility), key intensive variables were found to be: the variation of ground slopes (i.e. *nDrVarRt*), and the variation of tree heights (i.e. *nTrHghVrRt*).

In general, the acquisition and pre-processing of the data by means of an automatised operation (i.e. *ArcPy*, *PostgreSQL*) have proved to be more efficient, consistent and reliable, when data is stored under a meaningful naming convention. Further work is very much recommended, specially in relation to adding to the filter the variables that this four-month project was not allowed to consider. That includes working on the vertical structure of the forest, by using standard distribution of the *LiDAR* point cloud, as well as the calculation of important proxy factors (see appendixA) like biomass.

Further lines of research could focus on the role of scale, and the actual value of euclidean distances as a proxy measurement for disturbance. Following these, one could research the connectivity of the newly found habitat in the metropolitan area. Other questions to be considered: Could land previously levelled by agriculture, and displaying an artificial but efficient drainage, still be repopulated by dormice? i.e. Is biomass and biodiversity alone capable of creating habitat for the dormouse? i.e. Can urban design be adapted for dormice or other endangered species?

Ruben López Vázquez
rlopez15@student.aau.dk
A. C. Meyers Vænge 15,
2450 København

Bibliography

Axelsson, Peter (1999). “Processing of laser scanner data—algorithms and applications”. In: *ISPRS Journal of Photogrammetry and Remote Sensing*. DOI: `10.1016/s0924-2716(99)00008-8`.

Bailey, S.A, R.H Haines-Young, and C Watkins (2002). “Species presence in fragmented landscapes: modelling of species requirements at the national level”. In: *Biological Conservation* 108.3, pp. 307–316. DOI: `10.1016/S0006-3207(02)00119-2`. URL: `http://dx.doi.org/10.1016/S0006-3207(02)00119-2`.

Bennett, Andrew F., James Q. Radford, and Angie Haslem (2006). “Properties of land mosaics: Implications for nature conservation in agricultural environments”. In: *Biological Conservation* 133.2, pp. 250–264. DOI: `10.1016/j.biocon.2006.06.008`. URL: `http://www.northernplainscmn.com.au/documents/20-Bennett_et_al_2006.pdf`.

Bogaert, J. et al. (2000). “Alternative area-perimeter ratios for measurement of 2D shape compactness of habitats”. In: *Applied Mathematics and Computation* 111.1, pp. 71–85. DOI: `https://doi.org/10.1016/S0096-3003(99)00075-2`.

Bright, Paul and Pat Morris (1995). “A review of the Dormouse (Muscardinus avellanarius) in England and a Conservation Programme to safeguard its future”. In: *Hystrix, the Italian Journal of Mammalogy* 6.1-2. ISSN: 1825-5272. DOI: `10.4404/hystrix-6.1-2-4043`. URL: `http://www.italian-journal-of-mammalogy.it/article/view/4043`.

Bright, Paul W. and Pat Morris (1996). “Why are Dormice rare? A case study in conservation biology”. In: *Mammal Review* 26.4, 157–187. DOI: `10.1111/j.1365-2907.1996.tb00151.x`. URL: `http://ptes.org/wp-content/uploads/2014/06/Bright-Morris-1996.-Why-are-dormice-rare-A-case-study-in-conservation-biology.pdf`.

Carpaneto, Giuseppe and Mauro Cristaldi (1995). “Dormice and man: a review of past and present relations”. In: *Hystrix, the Italian Journal of Mammalogy* 6.1-2. ISSN: 1825-5272. DOI: `10.4404/hystrix-6.1-2-4044`. URL: `http://www.italian-journal-of-mammalogy.it/article/download/4044/3980`.

Chanin, Paul and Leonardo Gubert (2012). “Common dormouse (Muscardinus avellanarius) movements in a landscape fragmented by roads”. In: *Lutra* 55.1, pp. 3–15. URL: http://www.chaninweb.co.uk/Chanin%20&%20Gubert2012_Lutra_55_1_LOWRES.pdf.

Congedo, Luca (2016). "Semi-Automatic Classication Plugin Documentation. Release 5.3.6.1". In: DOI: 10.13140/RG.2.2.29474.02242/1. URL: https://media.readthedocs.org/pdf/semiautomaticclassificationmanual-v5/latest/semiautomaticclassificationmanual-v5.pdf.

Davis, C. E. et al. (1986). "Modern Statistical Methods in Chronic Disease Epidemiology". In: ed. by S. H. Moolgavkar and R. L. Prentice. John Wiley and Sons. Chap. An example of dependencies among variables in a conditional logistic regression. 140–147. DOI: 10.1016/0025-5564(87)90067-8.

Ebadi, Hamid et al. (2004). "Evaluation of Indexing Overlay, Fuzzy Logic and Genetic Algorithm Methods for Industrial Estates Site Selection in GIS Environment". In: *XX th ISPRS Congress 2004. Commission V, WG V/2*. URL: http://www.isprs.org/proceedings/XXXV/congress/comm5/papers/564.pdf.

Ehlers, Sina (2012). "The importance of hedgerows for hazel dormice (Muscardinus avellanarius) in Northern Germany". In: *Peckiana* 8, pp. 41–47. ISSN: 1618-1735. URL: http://www.senckenberg.de/files/content/forschung/publikationen/peckania/volume_2012/issue_8/04_artikel_ehlers_s.pdf.

ESA (2000). *Sentinel-2 MSI*. URL: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi.

Esri (2017). *How Fuzzy Overlay works*. URL: https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-fuzzy-overlay-works.htm.

Fallah, Yashar and Mahmut Onur (2011). "Lidar for Biomass Estimation". In: *Biomass - Detection, Production and Usage*. InTech. Chap. 1. DOI: 10.5772/16919. URL: http://www.intechopen.com/books/biomass-detection-production-and-usage/lidar-for-biomass-estimation.

FAO et al. (2012). *Harmonized World Soil Database (version 1.2)*. webpage. URL: http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/HWSD_Data.html.

Fedorowick, Janis M. (1993). "A landscape restoration framework for wildlife and agriculture in the rural landscape". In: *Landscape and Urban Planning* 27.1, pp. 7–17. DOI: 10.1016/0169-2046(93)90024-8. URL: http://ac.els-cdn.com/0169204693900248/1-s2.0-0169204693900248-main.pdf?_tid=1f2898ba-d4d0-11e6-9e9f-00000aab0f6c&acdnat=1483790282_1e3043ace016143075e3fbd104ccfa0a.

Fischer, Michelle Fuller (2014). "Spatial behaviour of the Common Dormouse Muscardinus avellanarius in Denmark." MA thesis. Department of Bioscience, Aarhus University.

Fjellstad, Wendy and Pia Frederiksen (2004). "Deriving Indicators from Earth Observation Data - Limitations and Potential for Landscape Monitoring Workshop report". In: *Developments in strategic landscape monitoring for the Nordic countries*. Ed. by Geoff Groom. Nordic Council of Ministers, Copenhagen.

Fløjgaard, Camilla et al. (2009). "Potential 21st century changes to the mammal fauna of Denmark – implications of climate change, land-use, and invasive species". In: *IOP Conference Series: Earth and Environmental Science*. Vol. 8. 1. IOP Publishing. IOP Publishing, p. 012016. DOI: `10.1088/1755-1315/8/1/012016`. URL: `http://iopscience.iop.org/article/10.1088/1755-1315/8/1/012016/pdf`.

Harrell, Frank E. (2017). *VIF, Variance Inflation Factors*. URL: `https://www.rdocumentation.org/packages/rms/versions/5.1-1/topics/vif`.

Hijmans, Robert J. et al. (2005). "Very high resolution interpolated climate surfaces for global land areas." In: *International Journal of Climatology* 25.15, pp. 1965–1978. DOI: `10.1002/joc.1276`.

IDRE (2014). *Robust Regression | R Data Analysis Examples*. URL: `http://stats.idre.ucla.edu/r/dae/robust-regression/`.

Juškaitis, Rimvydas (1995). "The structure and dynamics of Common dormouse (Muscardinus avellanarius) populations in Lithuania". In: *Hystrix, the Italian Journal of Mammalogy* 6.1-2. ISSN: 1825-5272. DOI: `10.4404/hystrix-6.1-2-4039`. URL: `http://www.italian-journal-of-mammalogy.it/article/view/4039/0`.

— (1997). "Ranging and movement of the Common dormouse Muscardinus avellanarius in Lithuania". In: *Acta Theriologica* 42, pp. 113–122. DOI: `10.4098/at.arch.97-15`. URL: `http://rcin.org.pl/Content/12678/BI002_2613_Cz-40-2_Acta-T42-nr12-113-122_o.pdf`.

— (2007). "Peculiarities of habitats of the common dormouse, Muscardinus avellanarius, within its distributional range and in Lithuania: a review". In: *Folia Zoologica* 56.4, p. 337. URL: `http://www.ivb.cz/folia_zoologica/archive/56_337-348.pdf`.

— (2008). *The Common Dormouse Muscardinus avellanarius: Ecology, Population Structure and Dynamics*. Institute of Ecology of Vilnius University Publishers, Vilnius, 163 pp. DOI: `10.13140/RG.2.1.4566.7685`. URL: `https://www.researchgate.net/publication/256091906_The_Common_Dormouse_Muscardinus_avellanarius_Ecology_Population_Structure_and_Dynamics`.

Keckel, Marcus, Sven Büchner, and Hermann Ansorge (2012). "Does the occurrence of the Common Dormouse Muscardinus avellanarius in East-Saxony (Germany) dependent on habitat isolation and size?" In: *PECKIANA* 8, 57–60. ISSN: 1618-1735. URL: `https://www.researchgate.net/profile/M_Keckel/publication/258352050_Does_the_occurrence_of_the_hazel_dormouse_Muscardinus_avellanarius_in_East-Saxony_Germany_dependent_on_habitat_isolation_and_size/links/5570069908aeccd77741716b.pdf`.

Lloyd, Amanda et al. (2015). *Hazel Dormouse Hibernation Process.* web pdf.

Mandlburger, Gottfried et al. (2009). "Orientation and processing of airborne laser scanning data (OPALS)-Concept and first results of a comprehensive ALS software". In: *Laser scanning*. Ed. by F. Bretar, M. Pierrot-Deseilligny, and G. Vosselman. Vol. XXXVIII, Part 3/W8. ISPRS Journal of Photogrammetry and Remote Sensing. Paris, France: IAPRS. DOI: 10.1.1.559.400. URL: https://www.researchgate.net/publication/237437116_Orientation_and_processing_of_airborne_laser_scanning_data_OPALS-Concept_and_first_results_of_a_comprehensive_ALS_software.

Margalef, Ramon (1993). *Teoría de los sistemas ecológicos.* Barcelona: Publicacions de la Universitat de Barcelona.

Mills, Cheryl Anne (2012). "Conservation and ecology of the hazel dormouse, Muscardinus avellanarius". PhD thesis. University of Exeter. URL: https://ore.exeter.ac.uk/repository/bitstream/handle/10036/3907/MillsC.pdf?sequence=2.

Otepka, J., G. Mandlburger, and W. Karel (2012). "The OPALS Data Manager - Efficient Data Management for Processing Large Airborne Laser Scanning Projects." In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Vol. I-3. ISPRS. XXII ISPRS Congress, 25 August - 01 September 2012. URL: http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/I-3/153/2012/isprsannals-I-3-153-2012.pdf.

Pfeifer, N. et al. (2014). "OPALS - A framework for Airborne Laser Scanning data analysis." In: *Computers, Environment and Urban Systems* 45, pp. 125 –136. DOI: 10.1016/j.compenvurbsys.2013.11.002.

Pretzlaff, Iris and Kathrin H. Dausmann (2012). "Impact of Climatic Variation on the Hibernation Physiology of Muscardinus avellanarius". In: *Living in a Seasonal World*. Springer Nature, pp. 85–97. DOI: 10.1007/978-3-642-28678-0_8. URL: https://www.researchgate.net/publication/257697113_Impact_of_Climatic_Variation_on_the_Hibernation_Physiology_of_Muscardinus_avellanarius.

Rahman, M.Z.A. and B. G. H. Gorte (2009). "Tree Crown Delineation from High Resolution Airborne Lidar Based on Densities of High Points". In: *Laser scanning*. Ed. by F. Bretar, M. Pierrot-Deseilligny, and G. Vosselman. Vol. XXXVIII, Part 3/W8. ISPRS Journal of Photogrammetry and Remote Sensing. Paris, France: IAPRS. URL: http://www.isprs.org/proceedings/XXXVIII/3-W8/papers/123_laserscanning09.pdf.

Root, Terry L. and Stephen H. Schneider (2002). "Wildlife Responses to Climate Change: North American Case Studies." In: ed. by Stephen H. Schneider and Terry L. Root. Washington D.C.: Island Press. Chap. Climate Change: Overview and Implications for Wildlife.

Sanderson, FJ (2004). "The Population Ecology and Monitoring of the Dormouse Muscardinus avellanarius." PhD thesis. Royal Holloway, University of London.

Schulz, Björn et al. (2012). "Hazel dormice in roadside habitats". In: *Peckiana* 8, pp. 49–55. URL: http://www.senckenberg.de/files/content/forschung/publikationen/peckania/volume_2012/issue_8/05_artikel_schulz_b.pdf.

Schulz, Björn et al. (2013). *Cross border conservation of the hazel dormouse: Presence, genetics, management and perspectives*. URL: `http://naturstyrelsen.dk/media/nst/11826459/biogrenzkorr_dormouse_web.pdf`.

SDFE (2016). *Danmarks Højdemodel, DHM/Punktsky*. Tech. rep. Agency for Data Supply and Efficiency.

Skok, Janko and BoriS kryštufek (2012). "Dormice in small mammal assemblages in a mixed southern European forest". In: *PECKIANA* 8, 69–75. ISSN: 1618-1735. URL: `http://www.senckenberg.de/files/content/forschung/publikationen/peckania/volume_2012/issue_8/08_artikel_skok_j.pdf`.

Søgaard, Bjarne, Morten Elmeros, and Helle Vilhelmsen (2013). *Overvågning af hasselmus Muscardinus avellanarius*. Aarhus University and Dormiceconsult.dk. URL: `http://bios.au.dk/fileadmin/bioscience/Fagdatacentre/Biodiversitet/TAA02_Hasselmus_v2_godkendt-FKG.pdf`.

Wilhelmsen, Helle (2003). "Status of dormice (Muscardinus avellanarius) in Denmark". In: *Acta Zoologica Academiae Scientiarum Hungaricae* 49.Suppl 1, pp. 139–146. URL: `http://actazool.nhmus.hu/49/S1/vilhelmsen.pdf`.

Willett, John B. and Judith D. Singer (1988). "Another Cautionary Note aboutR2: Its Use in Weighted Least-Squares Regression Analysis". In: *The American Statistician* 42.3, pp. 236–238. DOI: `10.1080/00031305.1988.10475573`.

Williams, Rachel L et al. (2013). "Using Long-Term Volunteer Records to Examine Dormouse (Muscardinusavellanarius) Nestbox Selection." In: *PLoS ONE* 8.6, e67986. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0067986`. URL: `http://europepmc.org/articles/PMC3694881?pdf=render`.

Zhao, Kaiguang, Sorin Popescu, and Ross Nelson (2009). "Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers". In: *Remote Sensing of Environment* 113.1, pp. 182–196.

Appendix A

Project Workflow

Here is a summary of the steps taken, which will also make clear the processing tool used, the phases that built up the database, and the statistic results.

A.0.1 on Python

- Rename the files, attributes & values with ASCII characters
- Get FTP data:

  1. Terrain (DSM,DTM)
  2. LiDAR (ALS)
  3. Matrikel (parcels)
  4. FOT (Nature, Buildings, Infrastructure, Borders)
  5. Climate Data
  6. Soil Data

- Reproject all datasets
- Define "work Areas"
- Define "clusters" & "communities"

A.0.2 on SQL

- Create dataBase "hMus" (it will be based on three hierarchic schemas).

  1. public (will receive all incoming data into tables)
  2. hmus (will structure the data into types & into clusters and communities)
  3. DBhMus (will contain the tables of the finished database)

- Create table "hMusNests" in schema hmus.
- Merge all nests & create "hMusNests".geom
- Select occupied nests only & remove nests tagged in "Snarup"
- Create buffers & intersect points to create multipoint sets
- Create "clusters" & "communities" in schema hmus.

A.0.3 on Python

- (Mosaic &) clip the data to the "work Areas" and/or "Work Grid"
- Calculate the slope analysis (slopes, focal)
- Extract elevation on "nests", "clusters" & "communities"
- Calculate normalised Canopy (nCnp) & Sinks (>2.5m)
- Extract the slope analysis on "clusters" & "communities"
- Classify slope analysis & calculate "clNaturlRt" & "coNaturlRt"
- Extract the tree height on nests (within a 5m buffer)
- Extract the tree height on "communities" (buffer: 600 m)

A.0.4 on SQL

- Merge municipal addresses into "dbAdress"
- Count addresses on "addrgrid"
- Update "hMusNests" with new columns "nestElev" & "nTreeHgh"

A.0.5 Manual (Notepad++)

- Create txt file with key names to "clusters" & "communities".

A.0.6 on SQL

- Update "hMusNests", "clusters" & "communities" with new primary keys.
    1. intro foreign keys clusterid & communid in table "hMusNests"
    2. intro foreign key communid in table "clusters"
- Update "clusters" with "nAvgElev" from "hmusnest"
- Update "clusters" with "clArea" & "clCmpctRt" from "zonalGeom"
- Update "communities" with "coArea" & "coCmpctRt" from "zonalGeom"
- Update "clusters" with "clNaturlRt" from "clSlpStudy"
- Update "communities" with "coNaturlRt" from "coSlpStudy"
- Update "communities" with "coAvgTrHgh" & "coStDTrHgh" from "CoTreeHgh"
- Update "communities" with "coAvgElev" from "bAltStudy"
- Update "clusters" with "clAvgElev" from "bAltStudy"
- Count "clNrNests" & "coNrNests" on "clusters" & "communities"

A.0.7 on R

• Conduct OLS regression analyses on:

Elevation

• Correlate "clCmpctRt" by "clNAvgElev" (result ′none′)
• Correlate "clNrNests" by "clNAvgElev" (result +++)
⇝ If positive then correlate "clNrNests" by "clAvgElev" (result ++)
⇝ If positive then correlate "coNrNests" by "coAvgElev" +
• T-test significance of "clNAvgElev" deviates from "clAvgElev" (result ′none′)
• T-test significance of "clNAvgElev" deviates from "coAvgElev" (result ′none′)

Climate

• Correlate "clNrNests" by "wAvgTArAut" (result -)

Forest Maturity

• Correlate & plot "treeGirth" by "nTreeHgh" (result ′none′)
• Correlate & plot "surTreeHgh" by "coAvgTrHgh" (result ′none′)
• T-test significance of "clNaturlRt" deviates from "coNaturlRt" (result ′none′)
• T-test significance of "surTreeHgh" deviates from "coAvgTrHgh" (result +++)
• T-test significance of "nAvgNDVI" deviates from average NDVI (result +++)

Intangible Factors

• Correlate & plot "addr100mA1" by "DTM100mA1" (result +++)
• Correlate & plot "NDVI100mA1" by "DTM100mA1" (result +++)
• Correlate & plot "clNrNests" by "clAvgDstMB" (result +++)
• Correlate & plot "clNrNests" by "clAvgDstPB" (result +++)

A.0.8 Manual (QGIS/ArcMap)

• Extract to centroids kn_300 "soilPH" & "soilDrain" & "soilCEC"
• Extract to centroids kn_300 "wAvgTArous"
• Extract to centroids kn_300 "wIsotherma"
• Extract to centroids kn_300 "wMPrcArous"
• Extract to centroids kn_300 "wAvgTArAut"
• Extract to centroids kn_300 "wMAvgMxTHb"
• Extract to centroids kn_300 "wMPrcArAut"

A.0.9 on SQL

- Update "clusters" & "communities" with "clNrNests" & "coNrNests"
create column & select "soilPH"
create column & select "soilDrain"
create column & select "soilCEC"

A.0.10 Manual (ArcMap/QGIS)

- Create kn_300 and Union Grid to "hNghSbUrRt"
- Create subUrban areas by using "erhverv", "lav_bebygelse" & "gartneri"
- Extract by kn_300 on "hNghSbUrRt"
- Create kn_300 and Union Grid to "hSumRdKmHa"
- Clip and project "vejmidte"
- Extract by kn_300 on "hSumRdKmHa"
- Create kn_300 and dissolve to aggregate 3x3 "hNghAddSHa"
- Extract by kn_300 on "hNghAddSHa"
- Create kn_12 and Union Grid to "forestID"

A.0.11 on SQL

- Create column & select "hNghSbUrRt"
- Create column & select "hSumRdKmHa"
- Create column & select "hNghAddSHa"
- Create column & select "wAvgTArous"
- Create column & select "wIsotherma"
- Create column & select "wMPrcArous"
- Create column & select "wAvgTArAut"
- Create column & select "wMAvgMxTHb"
- Create column & select "wMPrcArAut"
- Create column & select "clAvgDstPB"
- Create column & select "clAvgDstMB" & "clStDDstMB"
- Create column & select "clAvgDstRd" & "clStDDstRd"
- Create column & select "clAvgDtDRd" & "clStDDtDRd"
- Create column & select "clAvgDtRRd" & "clStDDtRRd"
- Create column & select "clAvgDstHs" & "clStDDstHs"
- Create column & select "clAvgDtUrS" & "coStDDtUrS"

A.0.12 on eCognition

- Extract "clNDrRtHa" in a 20m grid
- Segment & classify orthorectified aereal photography of forests

A.0.13 Manual (ArcMap)

- Calculate NDVI values for the forest
- Extract NDVI values for nests (in a 25 m buffer)

A.0.14 on SQL

- Create column & select "clNDrRtHa"
- Create column & select "clAvgNDVI"

A.0.15 on R

- Conduct OLS regression analyses on:

Human Environment

- Correlate "clNrNests" by "clAvgDstHs" (result ++)
- Correlate "clNrNests" by "clAvgDtUrS" (result +++)
- Correlate "clNrNests" by "clAvgDstRd" (result ′none′)
- Correlate "clNrNests" by "clAvgDtDRd" (result +)
- Correlate "clCmpctRt" by "hNghAddSHa" (result ′none′)
- T-test significance of "clNDrRtHa" deviates from Total Average Mean

Climate

- Correlate "clNrNests" by "wAvgMxTHib" (result ′none′)
- Correlate "clNrNests" by "wMPrcArous" (result ′none′)

Diet

- Correlate "clNrNests" by "clAvgNDVI" (result +++)
- T-test significance of "nstAvgNDVI" deviates from Avg forest NDVI (result +++)

A.1 Further work

In this section I would like to delineate further work on this topic, and, in general, on the study of forests. The applications can be varied: from identifying habitat to calculating forest fuel availability, with regard to wildfires, or carbon sinks, with regard to climate change.

The main objectives for this exercise have been met. The model has achieved an acceptable level of accuracy (with RMSE: 1.2), and its results on the validation dataset are almost as good as on the training dataset. However, because scale matters, a good way to work with dormice involves using the precision and permeability of *LiDAR* point cloud data.

Further work could be and should be carried on this very topic, even though some of the data (used in this very report) derived from *LiDAR* data. The surface elevation models derived from *LiDAR* data, and that allowed them to have a 5cm vertical precision. (SDFE 2016) Yet *LiDAR* point clouds can deliver much more. The original data available had recorded the location of up to five returns per pulse (multi-return information), also the reflectance of each individual pulse, and other attributes like the echo number (order) and its recording time, plus, from the analysis and radiometric calibration of the echo waveform, the echo width, amplitude, and backscatter cross-section. (Mandlburger et al. 2009)
"[A]mplitude or reflectance registration gives radiometric information about the surveyed area . . . in a very narrow wave length band. It can be used in classification algorithms, e.g., separating paved areas from grass land." (Axelsson 1999) In addition, from the multi-return data it is possible to calculate the spread of returns (i.e. leaves and branches), or their characteristic distribution (clustered or dispersed at different heights). For such purpose, the point density required is already covered by the data available, which has a point density of 4-5 points/m^2. (SDFE 2016) Still, that is way below what state-of-the-art ALS can deliver right now >10 points/m^2. (Pfeifer et al. 2014) Still, such spatial density should be in relation to the actual purpose, and it can be increased on collection. "The point density is dependent on flying height and also on system dependent factors, such as platform velocity, field of view, and sampling frequency." (Axelsson 1999) Simple "algorithms such as the Normalized Difference Vegetation Index (NDVI) have been empirically correlated to structural parameters (Jonckheere et al., 2006; Solberg et al., 2009; Morsdorf et al., 2004, 2006) such as Leaf Area Index (LAI) of canopy-level." (Fallah and Onur 2011) Thus, even though this exercise has used derived data from *LiDAR* (specially on the surface and terrain models), there is much more to be explored in terms of forest structure characterisation.
For such work to reach completion, a much longer period of research and training will be required. Not everything is solved by acquiring higher precision. In fact, the obvious problem with rich data is its burden on computer resources, and its typically long run time periods. For that reason, software must be flexible and the algorithms they run must be built toward reducing complexity as well as the need for reducing user supervision. It is very inconvenient to have long computations on hold because of demanding user input. Thus, the chosen tool to carry these calculations within this exercise was *OPALS*, which has a data manager especially designed to meet such demands. It "provides both, fast spatial access to huge point clouds, as well as a flexible attribute schema to store additional point related quantities." (Otepka, Mandlburger, and Karel 2012) It is controlled by command line, as opposed to an always heavier GUI. It is also built by modules or tools, which enables the automatic building of further command line executables by coding on python or C++. (Otepka, Mandlburger, and Karel 2012)

In this way *OPALS* has been designed to accomplish all what is needed in relation to *LiDAR* point clouds. In fact, *OPALS* is flexible enough to provide "tools for all steps starting from full waveform decomposition, sensor calibration, quality control, and terrain model derivation, to vegetation and building modeling." (Pfeifer et al. 2014) That means that *OPALS* can be used on raw data to convert it into point clouds; as well as it can be used to manipulate such point clouds by several applications or tools in a stored in its repository. Further study into the dormice (or any other species) contemplated in a metropolitan environment might benefit from such competences.

At the same time, in order to deal with the reality of the data, *OPALS* "reduces the complexity of the entire point cloud by extracting meaningful subsets or a set of feature points." (Pfeifer et al. 2014) These subsets require classification, and in the case of spatial topology, an 'appropriate neighbourhood definition'. This means that topological neighbours are only added together when a certain definition or set of criteria is met. They must fulfil an "homogeneity requirement with respect to the current point. The homogeneity criterion is specified by the application and considers geometry (x,y,z) and possibly also the values of additional attributes." (Pfeifer et al. 2014) It is thus that some researchers consider that "original data should be used in the filtering and modelling process until an object dependent representation and generalisation can be made." (Axelsson 1999) Transforming the data to raster should only be done later at a different stage of the modelling.

However, that does not mean that after converting the data into raster no further information can be obtained. Single tree delineation and detection can be carried out by using the Canopy Height Model (CHM), the subtraction of the DTM from the DSM. There are studies like Rahman and Gorte (2009), which report that the CHM based approach performs best on coniferous trees. However, agreeing with Axelsson (1999) and Pfeifer et al. (2014), Rahman and Gorte (2009) also report as a shortcoming that post-processing on tree crown segments is necessary especially for CHM, and that tree delineation can be improved on broad-leaved trees (specially those with a rather flat crown), by using directly the point cloud. Their research is based on densities of high points (DHP). "The DHP method relies on the fact that the density of received laser pulses above a certain height is high at the centre of a tree crown and decreases towards the edge of the crown." (Rahman and Gorte 2009)

Beyond that there are 'composite metrics', which are those predictors that "tend to retain as much information as possible when reducing raw lidar point clouds into a format suitable as inputs to predictive models of canopy structural variables." (Zhao, Popescu, and Nelson 2009) Among these structural variables can be found canopy height, LAI (Leaf Area Index), canopy cover, biomass density, stem density, and basal area. All of them are usually collected by field measurements, in order to quantify the average canopy conditions at whatever scale needed. Instead Zhao, Popescu, and Nelson 2009 used some descriptive statistics of laser ranging data to do the job. The "typical lidar metrics that have been previously used include canopy densities, mean and percentile heights, second-order height statistics (Hudak et al., 2009; Lim & Treitz, 2004) and even some composite metrics such as the canopy height distribution and the canopy quantile function of Zhao et al. (2009)." It is certainly an enough stimulating time of fast development for this form of remote sensing. Simultaneously, R has become a very fruitful tool for modelling, as well as the automatising of processes and new techniques have added exponentially to these developments by combining, as Zhao, Popescu, and Nelson 2009 have tried, composite metrics and machine learning models.

Appendix B

Statistical Tables & Figures

Here is a collection of statistical tables and figures, that were omitted from the results, chapter3. They lack the relevance of the data selected to be displayed on that chapter. This means, that I considered that they would only contribute to clutter the account of results, yet they could still be of relevance to any other reader. Here can you find them, duly cross-referenced to their insertion in the text.

Table B.1: Correlation between number of nests in a cluster and average cluster elevation. A linear OLS regression has been fitted through data, in order to attempt to explain the number of nests that form a cluster by the averaged elevation of their surrounding terrain (i.e. ClAvgElev), which is extracted from the DTM by a 75 m merged buffer of all points.

> summary(lm(ClNrNrsts~ClAvgElev))					
Residuals:					
	Min	*1Q*	*Median*	*3Q*	*Max*
	-9.696	*-3.764*	*1.625*	*3.854*	*4.083*
Coefficients:					
	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*	
(Intercept)	*-0.50965*	*2.35556*	*-0.216*	*0.829*	
ClAvgElev	*0.21043*	*0.03934*	*5.350*	*4.91e-07*	*****
Significance codes:	*0 '***'*	*0.001 '**'*	*0.01 '*'*	*0.05 '. '*	*0.1 ' '*
Residual standard error: 3.875 on 109 degrees of freedom					
Multiple R-squared: 0.2079,		*Adjusted R-squared: 0.2007*			
F-statistic: 28.62 on 1 and 109 DF,		*p-value: 4 .914e-07*			

TableB.1 and the following tableB.2 belong to the account on results (chapter3, page46). They report on an increase residual standard error and declining explanatory power (R^2: 0.2 to R^2: 0.13) of the number of occupied nests in a cluster, by their surrounding elevation at increasing scales.

Table B.2: Correlation between number of nests in a cluster and average community elevation. A linear OLS regression has been fitted through data, in order to attempt to explain the number of nests that form a cluster by the averaged elevation of their surrounding terrain (i.e. CoAvgElev), which is extracted from the DTM by a 600 m merged buffer of all points.

> summary(lm(ClNrNrsts~CoAvgElev))					
Residuals:					
	Min	*1Q*	*Median*	*3Q*	*Max*
	-9.7543	*-2.9769*	*0.2457*	*4.6028*	*4.7556*
Coefficients:					
	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*	
(Intercept)	*-2.02182*	*3.30359*	*-0.612*	*0.542*	
CoAvgElev	*0.24824*	*0.05836*	*4.254*	*4.46e-05*	*****
Significance codes:	*0 '***'*	*0.001 '**'*	*0.01 '*'*	*0.05 '.'*	*0.1 ' '*
Residual standard error: 4.032 on 109 degrees of freedom					
Multiple R-squared: 0.1424,			*Adjusted R-squared: 0.1345*		
F-statistic: 18.1 on 1 and 109 DF,			*p-value: 4.456e-05*		

On the significance of the placement of dormouse nests on the topography, nesting on higher ground was not found for the dormouse a pattern of relevance, on chapter3, page47.

Table B.3: One sample t-tests on deviation from mean altitude. There is no significant evidence within the values of a cluster or community, that values deviate significantly from nest mean elevation. Therefore, I must uphold the null hypothesis that the mean elevation of clusters, and to a lesser extent that of communities, is not significantly different than that of dormice nests (60 m).

> t.test(co$CoAvgElev, alternative="less", mu=60)		
t = -1.6589,	*df = 6,*	*p-value = 0.07411*
alternative hypothesis: true mean is less than 60		
95 percent confidence interval: -Inf 60.9685		
sample estimates: mean of x = 54.34935		
> t.test(ClAvgElev, alternative="less", mu=60)		
t = -1.4514,	*df = 16,*	*p-value = 0.083*
alternative hypothesis: true mean is less than 60		
95 percent confidence interval: -Inf 60.68129		
sample estimates: mean of x = 56.6426		

Table B.4: Analysis of Variance on habitat membership results between training and validation sets.

> anova(lm(mbshpStgth~dataset))	*Df*	*Sum Sq*	*Mean Sq*	*F value*	*Pr(>F)*
dataset	*1*	*0.00268*	*0.0026754*	*0.665*	*0.4164*
Residuals	*126*	*0.50694*			

Table B.5: Regression between occupied nests and their average distance to 5 closest buildings.

> summary(lm(ClNrNests~avgDstHs))					
Residuals:					
	Min	1Q	Median	3Q	Max
	-9.5166	-2.2528	0.4724	4.3218	5.6496
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.589428	0.913446	10.498	< 2e-16	***
avgDstHs	0.008293	0.002903	2.856	0.00513	*
Significance codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '
Residual standard error: 4.2 on 109 degrees of freedom					
Multiple R-squared: 0.06964,		Adjusted R-squared: 0.0611			
F-statistic: 8.158 on 1 and 109 DF,		p-value: 0.005134			

Table B.6: Regression between occupied nests and their min. distance to parcel boundaries.

> summary(lm(ClNrNests~DistPB))					
Residuals:					
	Min	1Q	Median	3Q	Max
	-10.9349	-2.2418	0.1786	2.8239	4.8836
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	13.588820	0.447014	30.399	< 2e-16	***
DistPB	-0.020747	0.003388	-6.125	1.47e-08	***
Significance codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '
Residual standard error: 3.756 on 109 degrees of freedom					
Multiple R-squared: 0.256,		Adjusted R-squared: 0.2492			
F-statistic: 37.51 on 1 and 109 DF,		p-value: 1.47e-08			

Table B.7: Regression between occupied nests and their average distance to paved roads.

> summary(lm(ClNrNests~DistRd))					
Residuals:					
	Min	1Q	Median	3Q	Max
	-9.0992	-2.1142	-0.4916	2.7618	7.4116
Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	13.921256	0.810557	17.175	< 2e-16	***
DistRd	-0.006941	0.002468	-2.813	0.00583	**
Significance codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '
Residual standard error: 4.204 on 109 degrees of freedom					
Multiple R-squared: 0.06766,		Adjusted R-squared: 0.05911			
F-statistic: 7.911 on 1 and 109 DF,		p-value: 0.005831			

Table B.8: Regression between occupied nests and their average distance to unpaved roads.

```
> summary(lm(ClNrNests~DistDRd))
Residuals:
                           Min          1Q       Median       3Q        Max
                      -10.6293     -3.1707      -0.8352   4.0584     5.7871
Coefficients:
                      Estimate  Std. Error      t value  Pr(>|t|)
(Intercept)          12.534315    0.674040       18.596    <2e-16   ***
DistDRd              -0.009124    0.008160       -1.118     0.266
Significance codes:     0 '***'   0.001 '**'     0.01 '*'  0.05 '.'      0.1 ' '
Residual standard error: 4.329 on 109 degrees of freedom
Multiple R-squared:  0.01134,             Adjusted R-squared:  0.002269
F-statistic:  1.25 on 1 and 109 DF,        p-value: 0.266
```

Table B.9: Regression between occupied nests and their average distance to urban areas.

```
> summary(lm(ClNrNests~DistUr))
Residuals:
                           Min          1Q       Median       3Q        Max
                      -12.3121     -3.1158      -0.3005   4.5054     4.8633
Coefficients:
                      Estimate  Std. Error      t value  Pr(>|t|)
(Intercept)          7.1109614   1.3628274        5.218  8.71e-07   ***
DistUr               0.0010512   0.0002845        3.695  0.000345   ***
Significance codes:     0 '***'   0.001 '**'     0.01 '*'  0.05 '.'      0.1 ' '
Residual standard error: 4.105 on 109 degrees of freedom
Multiple R-squared:  0.1113,              Adjusted R-squared:  0.1032
F-statistic: 13.66 on 1 and 109 DF,        p-value: 0.0003453
```

B.0.1 Figures

When designing a proper way to weight the data acquired from different clusters, I came to realise that, both compactness ratio (based on an isoperimetrical measurement) and the number of nests, had an inverse correlation (i.e. -0.8369 see subsection2.2.3). Here,you can find the plotted lineal regression that fits the data, in figureB.1.

Such figure is the result of rolling back a logarithmic transformation that was necessary to straighten the skewed residuals. Such transformation expressed a probability, which had to be transformed back, by applying a logistic function. In this way, now we are plotting both original response (y) and explanatory (x) variables back.

$$p = \exp(y_i)/(1+\exp(y_i)); \quad \textit{which is} \quad p = \exp(z)/(1+\exp(z)) \quad \textit{in R}.$$

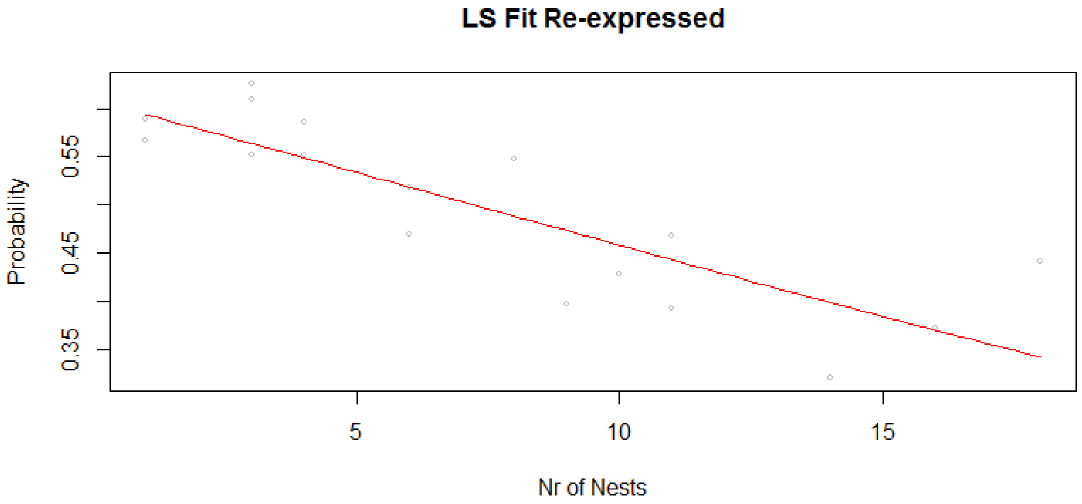


Figure B.1: Lineal model that best fits the ratio of compactness, explained by the number of nests, after putting the data through a log transformation. The inverse correlation exposes the adaptive behaviour of the common dormouse to landscape features (i.e. a dense understory).

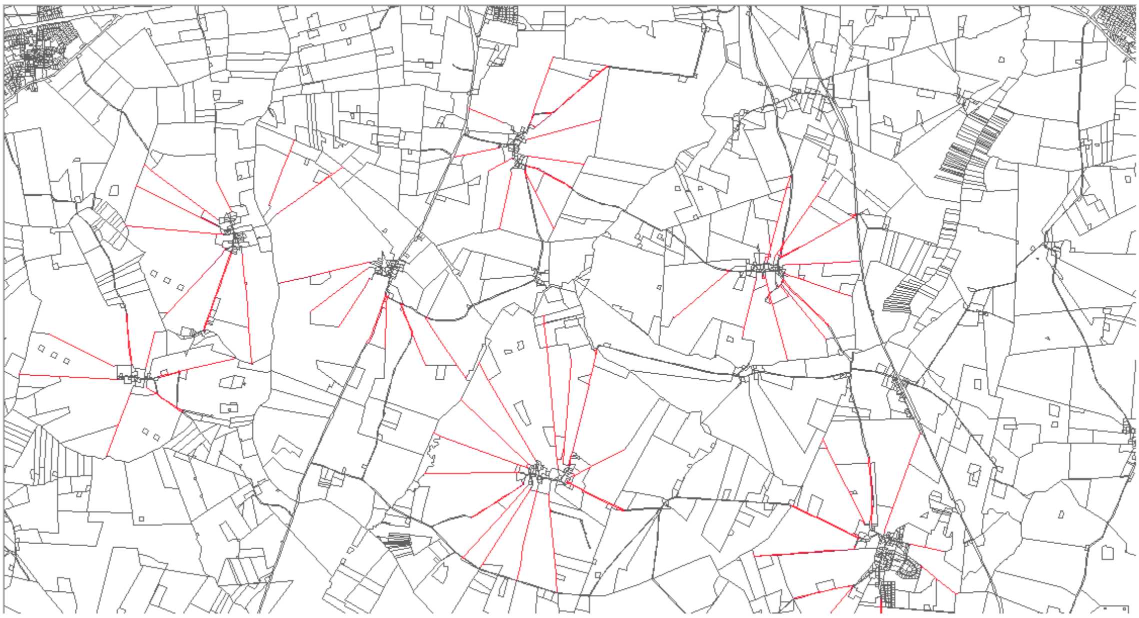


Figure B.2: This image depicts agricultural parcels south-west from the dormouse forest in Slagelse. Their concentric pattern, irradiating from inhabited nuclei, indicate that they are long time remnants. It presents evidence that the parcelling, and perhaps even the land use distribution of the area, has not undergone profound changes since the industrial revolution.

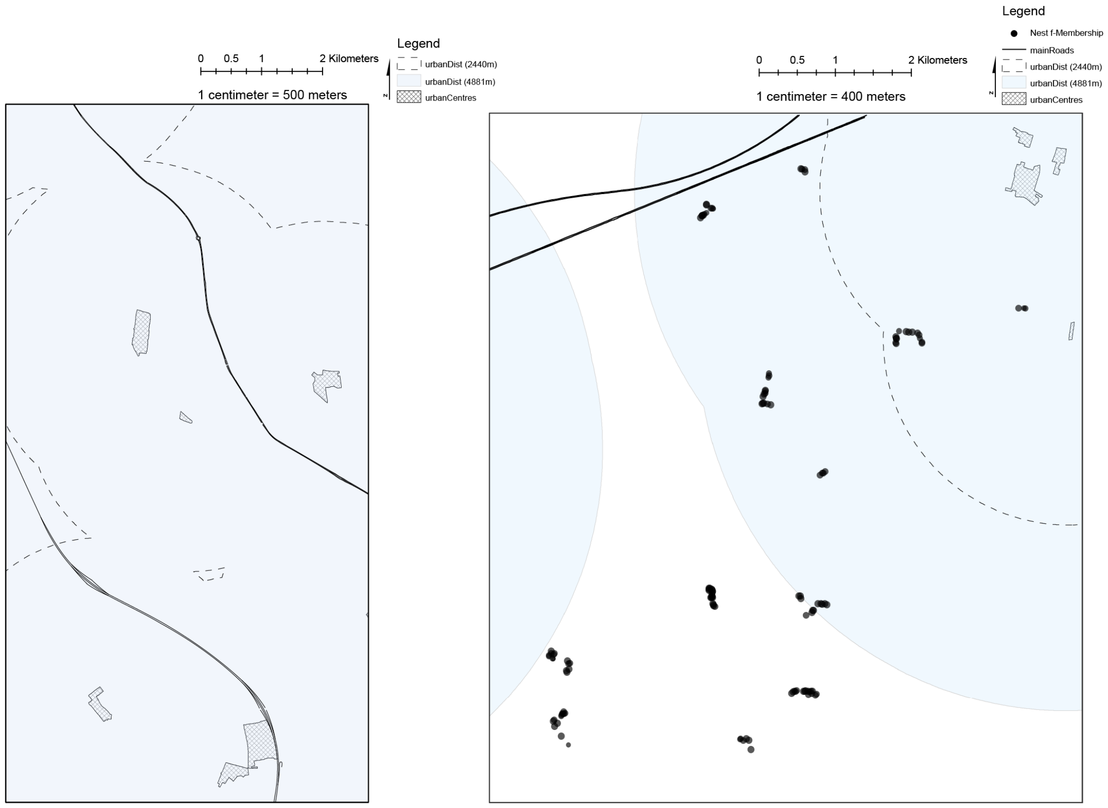


Figure B.3: Distance to urban centres according to the proxy variable *DistUr*. On the left it is depicted the area of north-east Sjælland, which does not have any space left due to the distribution of urban centres. On the right it is depicted the area of south-west Sjælland and the location of nests. Extra information on the nest location is provided by the size of the point, which is in relation to the strength of membership of the nest. Large points are strong nests, small points are weak nests. As it is possible to see, this euclidean measure does not forbid the presence of nests, and, what is more, the strengths of their membership results almost unscathed. I ran an OLS regression on the strength of fuzzy value explained by distance to urban centres and, although it resulted significant (p-value: 0.06138), the sensitiveness of the variable is close to null ($\beta : 7.924e - 06$), and only a two percent of the variability on strength of membership to an occupied nest was explained by the variance of the distance to urban centres (R^2: 0.0275). This makes of such variable, a very fuzzy variable in value.

B.0.2 Key variables

Table B.10: List of key variables, ordered by their significance. The top seven variables are very significant, which means that their values are the least random in relation to the presence of occupied nests. The five bottom variables have different reasons for not being significant. The vegetation index, for instance, does not have great variability, since this is an exclusively nemoral species. Other variables like the ratio of suburbanity and number of addresses per hectare, do not have either much variability, for the training area is rural. This means that such variables would correlate much easier if the training area included, even partly, a metropolitan area. The average tree height within the community does not correlate well because of a problem of scale. Observe the very same variable at a nest scale (*surTreeHgh*) does present significance, even though dormice tend to select specific location with a large variety of vegetation heights.

> summary of individual OLS regression results per variable

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
DistMB	-0.0031810	0.0003792	-8.388	1.96e-13	***
hSumRdKmHa	-16.1307	1.9516	-8.266	3.68e-13	***
DistPB	-0.020747	0.003388	-6.125	1.47e-08	***
nestElev	0.21023	0.03879	5.420	3.6e-07	***
nDrVarRt	7.965	1.774	4.490	1.78e-05	***
DistUr	0.0010512	0.0002845	3.695	0.000345	***
wAvgTArAut	-25.833	6.991	-3.695	0.000346	***
avgDistHs	0.008293	0.002903	2.856	0.00513	*
surTreeHgh	-0.14874	0.05276	-2.819	0.00559	*
DistRd	-0.006941	0.002468	-2.813	0.00583	*
hNghSbUrRt	-38.9252	26.2299	-1.484	0.141	
hNghAddSHa	-0.9643	0.7163	-1.346	0.181	
coAvgTrHgh	0.5269	0.4525	1.165	0.246	
DistDRd	-0.009124	0.008160	-1.118	0.266	
nAvgNDVI	-2.312	7.768	-0.298	0.7666	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '*

Table B.11: The calculation of the fuzzy membership function was carried out statistically, by using the Median and IQR. This methodology is explained in chapter2.2.7.

> summary of variable filter values, ordered by individual OLS regression estimate

Variables:		*Estimate*	*1st. Quantil*	*Median*	*3rd. Quantil*	Filter *(Median ± (1.5×IQR))*
hNghSbUrRt	---	*mask*	*0.000000*	*0.000000*	*0.013570*	*x =1 ≤ (0 - 0.020355) < x=0*
wAvgTArAut	*(°C)*	*-52.12*	*11.70*	*11.75*	*11.80*	*x =1 < (11.60 - 11.90) < x=0*
hSumRdKmHa	*(km/ha)*	*-13.06*	*0.1991*	*0.3215*	*0.4360*	*x =1 (0.03385 - 0.67685) < x=0*
nDrVarRt	---	*-1.527*	*1.853*	*2.015*	*2.140*	*x=0 < (1.5845 - 2.4455) < x=1*
hNghAddSHa	*(add/ha)*	*mask*	*0.0000*	*0.0000**	*0.0000*	*x =1 < (0 - 2) < x=0*
nAvgNDVI	---	*mask*	*0.7469*	*0.7887*	*0.8090*	*x=0 < (0.6955 - 0.8818) < x=1*
nestElev	*m (a.s.l)*	*0.04616*	*54.18*	*57.54*	*66.70*	*x=0 < (38.76-76.32) < x=1*
nTrHghVrRt	*m*	*-0.03111*	*8.154*	*10.680*	*13.950*	*x=0 < (1.986-8.154) < x=1* *x=1 < (13.950-19.374) < x=0*
avgDistHs	*m*	*0.02244*	*149.10*	*274.10*	*412.80*	*x=0 < (0-149.10) < x=1* *x=1 < (412.8-669.65) < x=0*
DistPB	*m*	*-0.01920*	*9.3800*	*33.2300*	*100.6000*	*x=1 < (0-170.06) < x=0*
DistRd	*m*	*-0.01255*	*159.80*	*274.60*	*383.30*	*x=0 < (0-159.80) < x=1* *x=1 < (383.30-609.85) < x=0*
DistDRd	*m*	*-0.01150*	*26.69*	*48.61*	*90.60*	*x=1 < (0-144.475) < x=0*
DistMB	*m*	*-0.002143*	*202.1000*	*897.4000*	*1875.0000*	*x=1 < (0-3406.75) < x=0*
DistUr	*m*	*mask*	*3977.0*	*5020.0*	*5741.0*	*x=0 < (2374-7666) < x=1*

** Mean value = 0.2793, due to the skewness of the data in a rural area, containing a town. I will take the values ≤2 add/ha, common throughout the rual area, as good.*