Using soil samples and footwear to improve forensic palynology

A study about where a soil sample could best be taken along a transect from parking space to dumping site when investigating a crime scene

Abstract

This study investigates the transfer of soil to shoe soles when a perpetrator disposes of a corpse in a forested area. During such an action soil is transferred to the shoe soles, but it is unknown which point(s) along the transect transfers the most soil to the shoe soles. This information is needed for taking a soil sample at a crime scene to make a comparison between the crime scene and the shoes which are thought to be linked with the crime. Previous research has mainly been focused on case studies while experimental data is lacking. This study will therefore focus on collecting experimental data by simulating the dumping of a corpse at five different locations. The pollen underneath the boots and the pollen along the transect are compared with each other to see where the highest similarity between the boots and the transect could be found. The similarity indicates where the transfer of soil to the shoe soles is highest and thus where a soil sample during investigation of a crime scene should be taken. The point with the highest similarity to the boots varies between the locations. This indicates that the transfer of soil to shoe soles does not happen at a set point along the transect, but depends on different factors, like moisture, ground coverage and compacted soil. The locations Heemskerk, IJsselstein and Bosrand showed the highest similarity with the points which were the muddiest, had the lowest amount of compacted soil or had limited ground coverage. The locations Schaapsallee and Utrechtse Heuvelrug showed the highest similarity with the average of the total transect because the moisture, ground coverage and compacted soil were comparable throughout the complete transect. In addition, the total transect showed a high similarity with the boots at every location. However, when the amounts of moisture, ground coverage and compacted soil varied along a transect, there was always one point along that transect with a higher similarity to the boots than they had with the total transect. In conclusion, the point(s) where soil is transferred to shoes are the muddiest part, the points with limited ground coverage or the least compacted soil. However, if the whole transect shows similarities in these factors, it is best to take multiple samples along the transect from which the average could be used.

Keywords – Forensic palynology, footwear, pollen analysis, soil transfer, DCA

Kelly van Leeuwen BSc - 11157690
MSc Earth Sciences – Future Planet Ecosystem Sciences
MSc Thesis (42EC)
Research performed: June 2021 – February 2022
Date: 17 February 2022

University of Amsterdam (IBED)
Netherlands Forensic Institute (NFI)
Supervisor: dr. S.C.A. Uitdehaag (NFI)
Examiner: dr. W.D. Gosling (UvA)
Assessor: dr. C.N.H. McMichael (UvA)

Contents

1. Introduction

In 2020 alone, 2675 people have been murdered in the Netherlands (Politie, 2021). To find out what has happened, evidence is collected. In court, this evidence against a suspect is evaluated to conclude whether or not it is legally and convincingly proven that the suspect committed the crime (Strafrechtadvocaten Netwerk, 2021). This means that a judge does not have reasonable doubt about the guilt of a suspect and that there is no plausible room for alternative scenarios. This concept is based upon the idea that society wants to minimize the number of innocent suspects that are found guilty and the number of guilty suspects that remain free. Forensic scientists help a judge's deciding process whether the suspect is guilty by investigating the traces from the crime scene (Bell et al., 2018). The investigation of traces is the domain of forensic science, where scientific techniques are studied and used to provide evidence for investigations (Bell et al., 2018; Tilstone et al., 2006).

Since the late nineteenth and early twentieth centuries, forensic science has improved significantly (Bell et al., 2018; Tilstone et al., 2006). Fingerprint recognition, blood-stain pattern analysis and ABO blood group recognition were extra forms of evidence that helped judges draw their conclusions about the guilt of a suspect (Tilstone et al., 2006). One of the most important improvements of forensic science was the invention of DNA typing, which allows experts to identify suspects by comparing the DNA found on an item or the victim with the DNA of possible suspects (Butler, 2011). DNA typing is based on the exchange of DNA between victims, suspects and items and allows experts to find similarities between these. Another improvement of forensic science is the field of forensic palynology which is based on the exchange of pollen grains and spores from an area to items (Mildenhall et al., 2006). Forensic palynology can be used to find similarities between items and an area, for example between footwear and the crime scene.

Forensic palynology entails the use of pollen and spores to solve crimes (Mildenhall et al., 2006). This field has been around since the 1950s , but is still underutilized (S. C. A. Uitdehaag, 2021). This is mainly caused by some challenges which afflict the field of forensic palynology. One of these challenges is the lack of specialists to identify pollen (Alotaibi et al., 2020; Mildenhall et al., 2006; Tiemens, 2009; Walsh & Horrocks, 2008; Wiltshire, 2009). Another challenge is that the determination of pollen types and the comparison between the pollen assemblages on an item and at the crime scene is time-consuming (Alotaibi et al., 2020; Mildenhall et al., 2006). However, the biggest challenge in this field is the low amount of research that has been done (Mildenhall et al., 2006; S. C. A. Uitdehaag, 2021). Information about this field can often be found in anecdotes or grey literature, but there is a lack of experimental designs (S. C. A. Uitdehaag, 2021; Walsh & Horrocks, 2008). Although experimental research is lacking, forensic palynology is used in the Netherlands since the early 1990s. An example of this is a murder case where the similarity between the pollen assemblage of a shovel and the crime scene strengthened evidence against the suspect (de Leeuwe, 2014; S. Uitdehaag et al., 2014). This anecdote shows how forensic palynology can be valuable evidence against a suspect.

The value of forensic palynology can also be found in the fact that pollen and spores leave invisible traces and are therefore useful signs to find similarities between items and a crime scene (S. Uitdehaag et al., 2014). The pollen assemblages on items are compared to the pollen assemblages at the location of a crime scene to find these similarities. However, to be able to find these similarities it is necessary to know how the pollen assemblages in the environment look and which pollen assemblages will be transferred to the shoe soles while walking in this environment (Mildenhall et al., 2006). This is especially important when a location has varying pollen assemblages (see figure 1), because it is possible that the pollen assemblage under the shoe sole of the perpetrator shows a mixture of the pollen assemblages throughout the transect (S. C. A. Uitdehaag, 2021). It is also possible that the shoe soles only present one dominant pollen assemblage depending from

which point a shoe sole picks up a soil trace. It is thus unclear which pollen assemblage will be found under a shoe sole and where a soil sample should be taken to be able to compare the shoes with the crime scene.

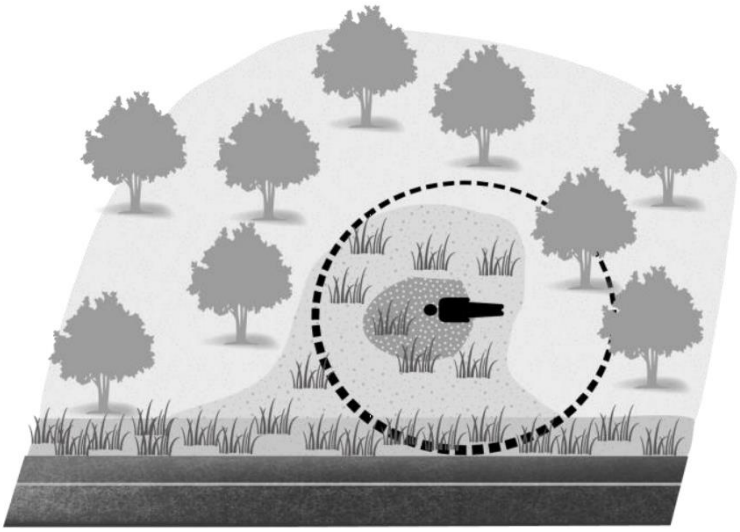

Figure 1. This figure shows an example of a murder case with a dashed line representing the crime scene (S. C. A. Uitdehaag, 2021). The different tints of grey resemble vegetation zones with different pollen assemblages. If a perpetrator walked from the road towards the location of the dead person to dump a body, this perpetrator can transfer soil with different pollen assemblages to his shoe soles.

1.1 Research aim and expected results

This research aims to figure out where a soil sample should be taken along a transect from parking space to dumping site. Realistic crime scenes with some degree of heterogeneity in plants along the transect were chosen to test this. Therefore the following question will be answered: "At which point along a walked transect from a parking space to a dumping site is it best to take a soil sample when investigating a crime scene?"

To answer this question, this study investigates the pollen assemblages along the walked transect, the pollen assemblage underneath a shoe sole and the similarity between those pollen assemblages. This leads to knowledge about which pollen assemblage(s) are transferred to a shoe sole. The point along the transect which transfers most of the soil to a shoe sole will show the highest similarity to the shoes with regards to the pollen assemblage. This point should therefore be used to investigate the similarities between the soil traces underneath a shoe sole and the area of the crime scene. To validate the results a detrended correspondence analysis (DCA) will be performed. This will show if the counts were consistent throughout the study.

It is expected that the results will show that the point with the highest similarity with the shoes has less ground coverage of plants and plant material than the rest of the transect, or is a point near the body where the perpetrator has spent the most time (van der Wal, 2021). The hypothesis that will be tested is therefore: the point along the transect from a parking lot to the dumping site shows the highest similarity with the shoes when this point has less ground coverage than the rest of the transect and/or is near to the corpse.

1.2 Theoretical framework

1.2.1 What is forensic palynology?

The principle of Locard states that if two objects come together they will exchange material (Saferstein, 2015). This holds for items, objects and people that have been at a crime scene and were in contact with plants or the soil. The pollen in the soil and on the plants can be transferred to the shoes and clothes of the people that

have walked over there. Consequently, pollen from plants and the soil are exchanged with items at the crime scene. This exchange of pollen is used to find similarities between items and a crime scene by examining the pollen and spores found on the items and the crime scene (Mildenhall et al., 2006; Saferstein, 2015). This field is called forensic palynology and is used to provide evidence to the court.

The variation in distribution of pollen and spores is the essential factor in forensic palynology (Bryant et al., 1990; Mildenhall et al., 2006; Walsh & Horrocks, 2008). This variation in distribution, also called the pollen rain, is among other things a consequence of different ways of pollen dispersal (Bryant et al., 1990; Walsh & Horrocks, 2008). The four ways of dispersal are self-pollination and dispersal by wind, water and insects. The ways of dispersal contribute to the distance pollen can travel and thus where those pollen end up. However, there are more factors influencing how far pollen can travel, like the density of a forest, the rainfall and the size of pollen (Jacobson & Bradshaw, 1981). All these factors together regulate the pollen rain and therefore also the pollen assemblage. The variation of pollen assemblages between different locations in combination with the rarity of some pollen types at a location are useful as associative evidence: evidence that can link an item to a crime scene (Horrocks & Walsh, 1998; Walsh & Horrocks, 2008; Wiltshire, 2009).

1.2.2 Forensic palynology and footwear

The challenges of forensic palynology also apply to footwear. Most of the published papers in forensic palynology and footwear are based upon case studies (e.g. Bull et al., 2006; Morgan et al., 2009; Wiltshire, 2016). This means that the results are derived from truly happened murders. Consequently, in these kinds of papers the results are based on one situation and therefore lack experimental and statistical power. Conclusions drawn from case studies are therefore subject to a certain extent of uncertainty. There are however also some studies performed with controlled experiments (Adams-Groom, 2018; Pereira et al., 2019; Riding et al., 2007).

The studies based upon case studies showed that soil from different areas can be transferred to shoe soles (Bull et al., 2006; Morgan et al., 2009; Wiltshire, 2016). The different case studies showed that underneath a shoe sole of a perpetrator pollen from the crime scene can be found as well as pollen from locations the perpetrator walked before or after visiting the crime scene. Morgan et al. (2009) and Wiltshire (2016) see this mixing of different pollen under a shoe sole as a problem for finding similarities between the shoes and the area of the crime scene. However, if it is known in which areas a perpetrator has been, this knowledge can also be used to see how similar this pollen assemblage is with the combination of pollen assemblages of the visited areas. Nevertheless, most people walk on pavements which means that combinations of pollen assemblages from different areas are not always a problem.

The studies based upon controlled experiments show more uncertainties in the conclusions they draw about the origin of soil underneath a shoe (Adams-Groom, 2018; Pereira et al., 2019; Riding et al., 2007). Riding et al. (2019) investigated for example the mixing of soil traces under shoes by visiting 6 sites with either pristine boots and boots that were worn before. They showed that when mixing occurs, the pollen assemblages most of the time had the highest similarity with the last visited site. However, when the boots were worn before, the boots and the site became less similar which could affect real forensic investigations. Adams-Groom (2018) performed a study where experts were asked to figure out if people walked at a designated place. People were allowed to walk at certain places with new or old shoes. Adams-Groom (2018) showed that experts could figure out whether or not a shoe has been at a designated area, but that they had doubts. This is important information because this was a closed set comparison which means that it is not a realistic set up (Adams-Groom, 2018). For real forensic investigations this comparison will thus be even more difficult. Pereira et al. (2007) investigated the importance of the timing in the year when soil samples are taken. They point out the importance of the timing in the year when soil samples have been taken to compare

with the shoe soles. Pollen assemblages vary through the year due to the flowering season with many pollen and the degradation of pollen in the soil. If there is a delay between visiting the area with the boots and collecting the soil samples in that area, there is a possibility that the pollen assemblages have changed which makes it harder to see similarities between the boots and the area.

The difference in results between studies based upon case studies and studies based upon controlled experiments shows that the latter one needs more research to draw more realistic conclusions about which soil traces are transferred to the soles of shoes. This research will therefore use an experiment to draw conclusions about the transfer of soil to shoes.

2. Methodology

2.1 Research design

2.1.1 Field work

To answer the research question, the dumping of a dead body is simulated at five different locations (see section 2.2 for more information about the locations). At every location a 50 metre transect was laid out along which a person (ca. 75 kg) walked with an extra weight (ca. 60 kg) to simulate the carrying of a corpse (see figure 2). This transect was based on the most likely route that a perpetrator would walk from the parking lot to a possible dumping location whilst passing through heterogeneous vegetation. The vegetation is an important factor for this study as it gives an indication of the pollen in the soil. Without a certain extent of heterogeneity in pollen assemblages it will become more complicated to find (significant) differences in the similarity between boots and points along the transect. The heterogeneity was even so crucial that variation in the vegetation at one location, IJsselstein, was found more important than how realistic the transect to the dumping location was. This will be explained in more detail in section 2.2.

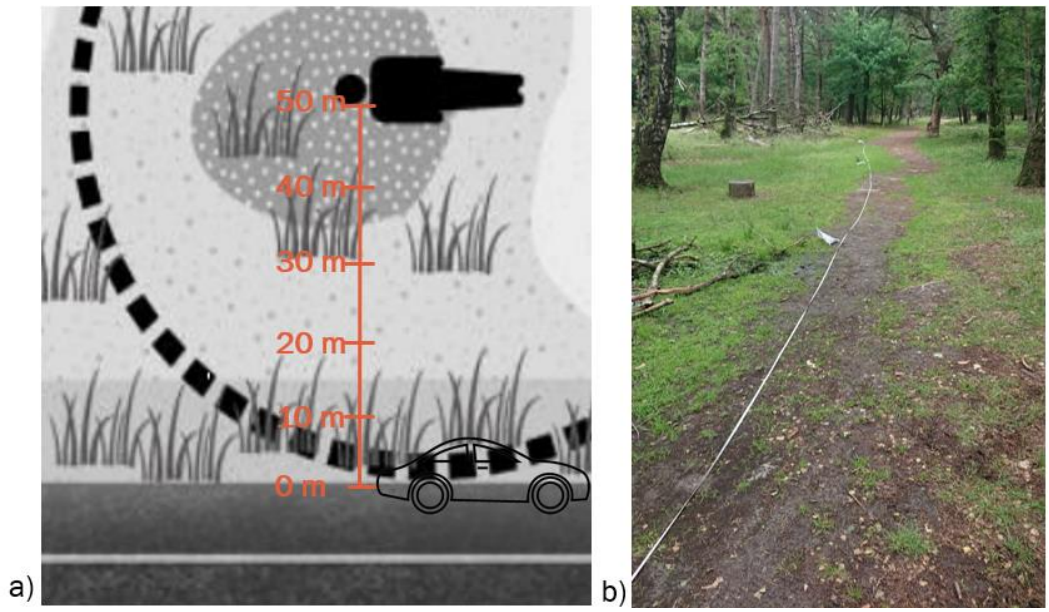

Figure 2. Schematic (a) and realistic (b) overview of the fieldwork set-up.

At the start of the transect a person of about 75 kilograms, the 'perpetrator', put on rubber rain boots (size 45, no brand). Another person of about 60 kilograms, the 'corpse', was taken on the back of the perpetrator as extra weight, simulating the carrying of a corpse. The perpetrator walked along the transect until the 50 metre point where the person on the back would be put down, simulating the dumping of a corpse. The perpetrator walked back along the transect to the 0 metre point next to the car. There the boots were changed for normal shoes and the boots were put in a clean trash bag. This act has been repeated five times with five different pairs of boots.

Hereafter, soil samples were taken every 10 metres along the transect (0m, 10m, 20m, 30m, 40m and 50m). The first was to remove any ground coverage, e.g. leaves, grasses, etc., from the top of the soil if present. With a small shovel a square of roughly 10 by 10 centimetres was carved out. The top layer of 1 centimetre of this square was put in a plastic sample bag and taken to the laboratory. The shovel was cleaned between the sample points by wiping off the soil with a bare hand. This was always done at the point where the sample was taken to prevent cross contamination between the different sample points.

The boots and the soil samples were taken home to the researcher's house where the soil samples were kept in the refrigerator (± 5 °C) or freezer (± -18 °C) depending on how long it would take to bring the samples to the laboratory. If it would take less than seven days after sampling the samples would be kept in the refrigerator. If it would be longer than seven days the samples were kept in the freezer. The exact dates of freezing and unfreezing can be found in table 1.

The boots were stored at room temperature for one day. The soil was then taken from the soles of the boots by scraping the soil with a scalpel. This was always done on a clean board. With a spoon and a dough scraper the soil per pair of boots was put in a sample bag. The samples were put in the refrigerator or freezer depending on how long it would take to bring the samples to the laboratory. This was done for every pair of boots. After putting the soil in a sample bag, all the materials used were cleaned with dish soap and hot water before starting with the next boots. The boots were also cleaned with dish soap and hot water (± 60 °C) before going to a new location.

Table 1. Shows the dates of sampling, freezing and unfreezing per location. The – sign indicates that the samples were not frozen, but kept in the fridge.

Location	Sampling date	Date of freezing	Date of unfreezing
Bosrand	22-06-2021	23-06-2021	30-07-2021
Utrechtse Heuvelrug	30-06-2021	01-07-2021	30-07-2021
IJsselstein	06-07-2021	07-07-2021	30-07-2021
Schaapsallee	10-08-2021	-	-
Heemskerk	07-11-2021	-	-

2.1.2 Lab work

Before the samples were taken to the laboratory, all the samples which were in the freezer, were taken out and put in the refrigerator for one day. In the laboratory the samples were prepared by mixing the samples while in the sample bag by kneading the soil. The mixing was done to homogenise the sample. With a density ring 1 cubic centimetre of soil was taken out of every sample to be able to calculate the concentration of pollen in each sample. Rocks and plant material were not transferred to a new sample bag, because they are not transferred to the boots. The density ring was cleaned with cold water and a brush before moving on to the next sample. A lab technician from the University of Amsterdam, A.L. (Annemarie) Philip, subsequently extracted the pollen from the soil following the procedure shown in Appendix A. She also added 18407

Lycopodium sp. spores (Lundt University, batch: 050220211, sd: 592) to use as a reference for calculating the concentration of pollen and spores in the samples.

The prepared samples were used to count a minimum of 300 pollen in the samples (microscope: Zeiss Axioscope 5). This was done for statistical power. *Lycopodium sp.* spores are not included in this total count. The count sheets used can be found in Appendix B.

2.2 Locations

The locations visited for this study were based upon how suitable they are for dumping a dead body and on the rate of variation in vegetation (see figure 3 and table 2 for more information about the locations). The factors which therefore influenced the decision for a location were: the reachability by car, the coverage from roads and houses and how heterogenous the vegetation was along the walked transect. Besides these factors, there were two factors which affected whether or not enough soil traces were picked up along the transect by the boots: time since the last rainfall and the amount of ground coverage (van der Wal, 2021). The locations were therefore visited within a maximum of 1 day after the last rainfall and locations with a high level of ground coverage were avoided.

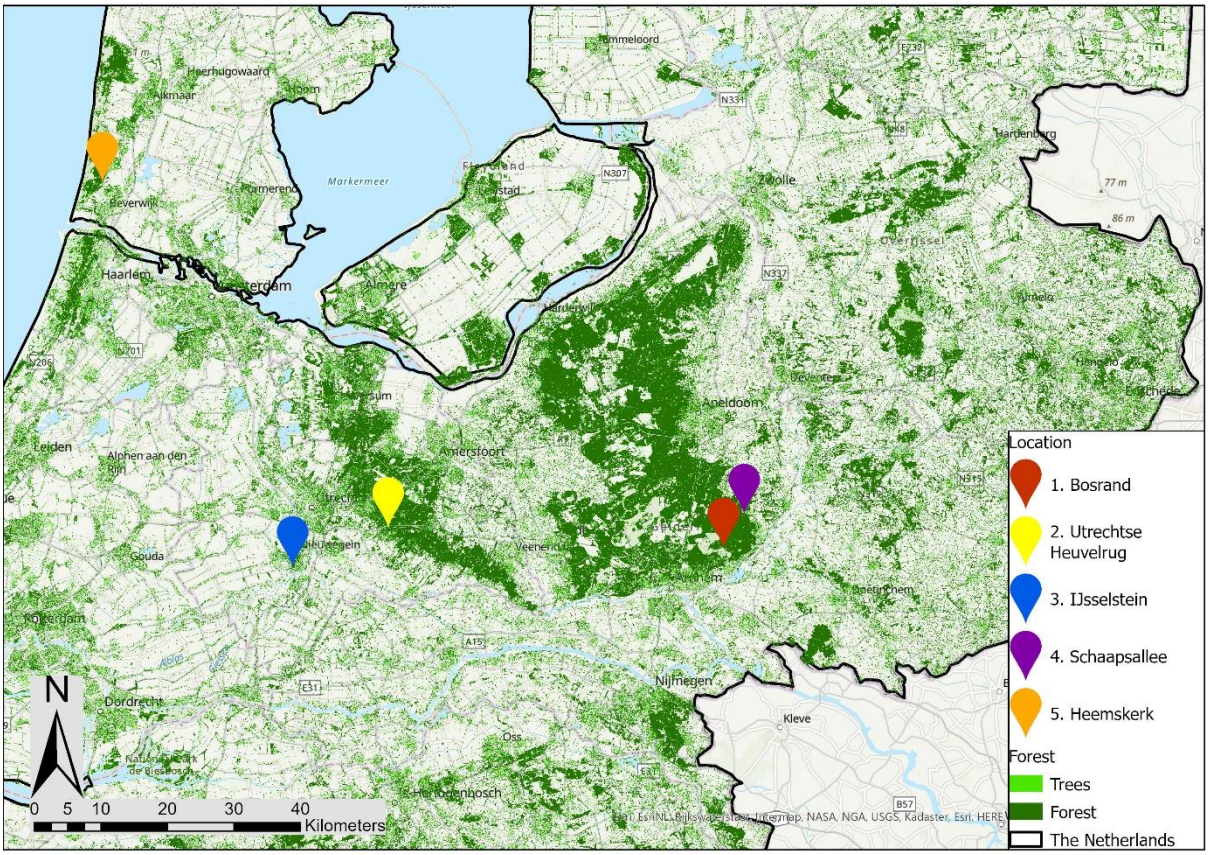

Figure 3. Map of the visited sites.

Table 2. Information about the visited sites.

Location	Coordinates	Sampling date	Time since last rainfall
Bosrand	52,033646; 6,018407	22-06-2021	0 days
Utrechtse Heuvelrug	52,062835; 5,283443	30-06-2021	0 days
IJsselstein	52,009914; 5,074248	06-07-2021	1 day
Schaapsallee	52,07913; 6,064438	10-08-2021	0 days
Heemskerk	52,527083; 4,649399	07-11-2021	0 days

2.2.1 Location 1: Bosrand

Bosrand is a location in Nationaal Park Veluwezoom in the province of Gelderland. This location has been visited on the 22nd of June 2021. The first part of the transect was on the normal trail and consisted therefore of compacted soil (see figure 4 for an overview of the transect). After 30 metres the transect continued on grass. The soil became looser after that point. The amount of ground coverage became higher as well. On the left side of the transect *Betula sp.* was most dominant and on the right side *Quercus sp.* was most prominent. This location was near a heathland which contained mainly *Calluna vulgaris*.

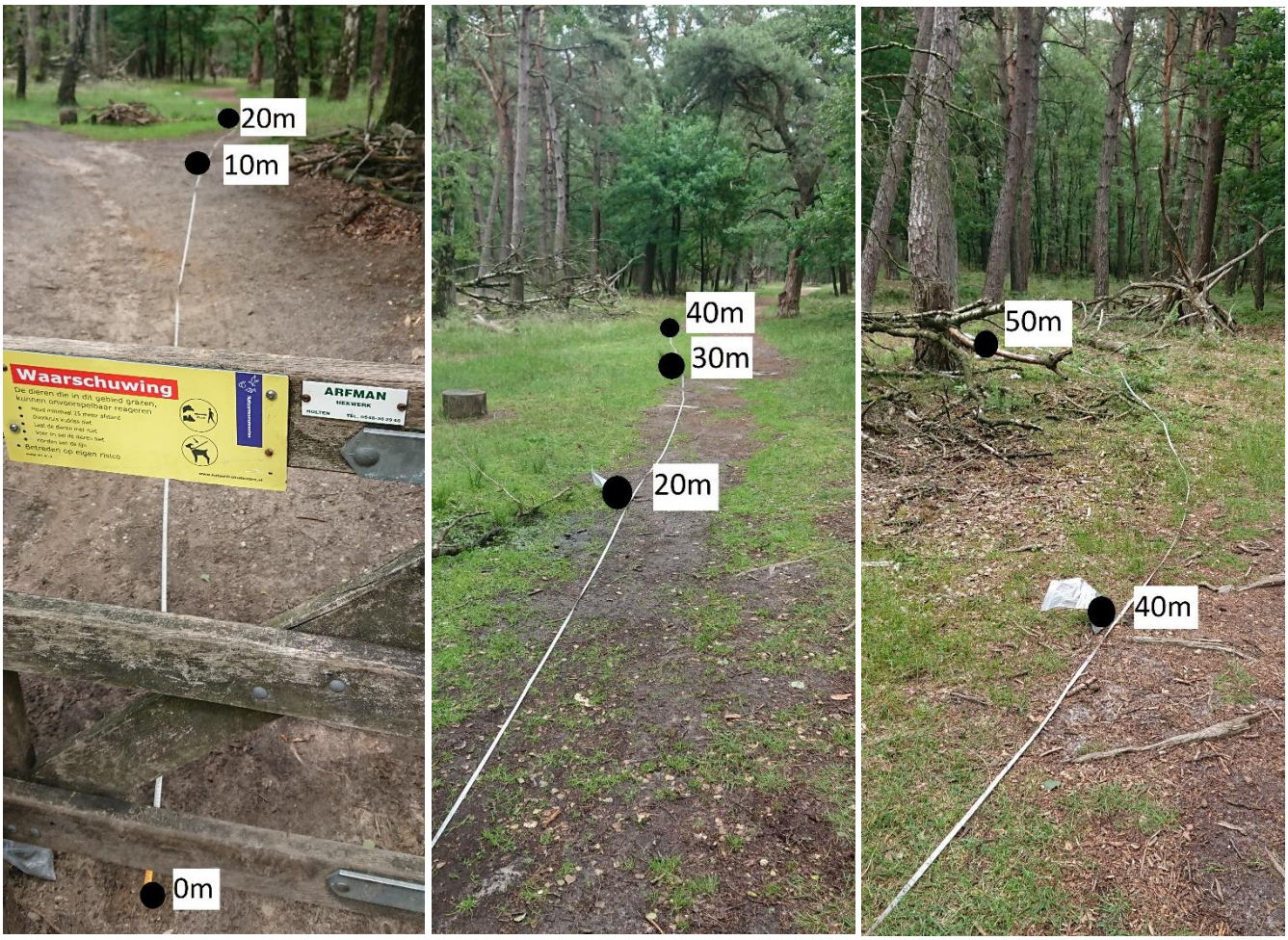

Figure 4. Overview of the walked transect at the location Bosrand.

2.2.2 Location 2: Utrechtse Heuvelrug

The Utrechtse Heuvelrug is located at the East side of Utrecht. This location has been visited on the 30th of June 2021. The Utrechtse Heuvelrug was visually blocked by trees from the parking lot and the road, but after this line of trees the location was very open. The first 15 metres of the transect consisted of bare and compacted soil. The next 20 metre was partly covered with grass and the last 15 metre consisted again of more bare soil (see figure 5 for an overview of the transect). Along the transect *Urtica sp.* and *Quercus sp.* were dominant. The last 20 metres were close to a ditch where *Pteropsida sp.* and *Typha sp.* were prominent.

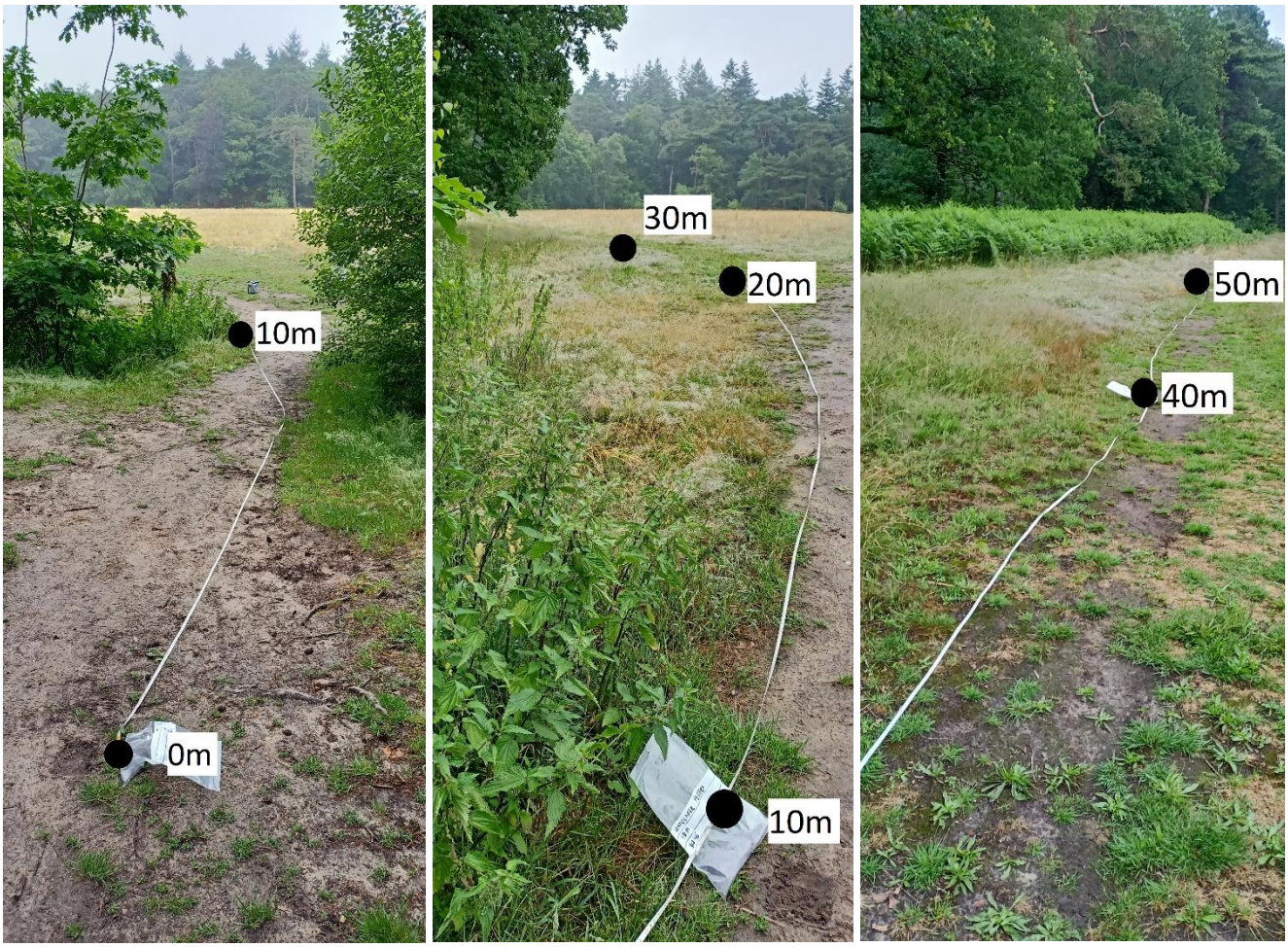

Figure 5. Overview of the walked transect at the location Utrechtse Heuvelrug.

2.2.3 Location 3: IJsselstein

IJsselstein is a village located on the South-west side of Utrecht. The location visited is a forest located in the South-East side of IJsselstein and is called IJsselbos. This location was visited on the 6th of July 2021. It is important to notice that the start of this transect was not directly next to a parking lot due to the homogenous vegetation at this point. The start was therefore about 100 metre further away from the parking lot. The first 10 metres of the transect were covered by grass. The rest was all bare soil. An important feature of this location was a muddy part around the 30 metre point (see figure 6 for an overview of the transect). The beginning of the transect consisted mainly of *Urtica sp.* and *Plantago sp*. Thereafter, the forest became closer and *Quercus sp.* became more dominant while *Urtica sp.* and *Plantago sp.* did not change in their dominancy.

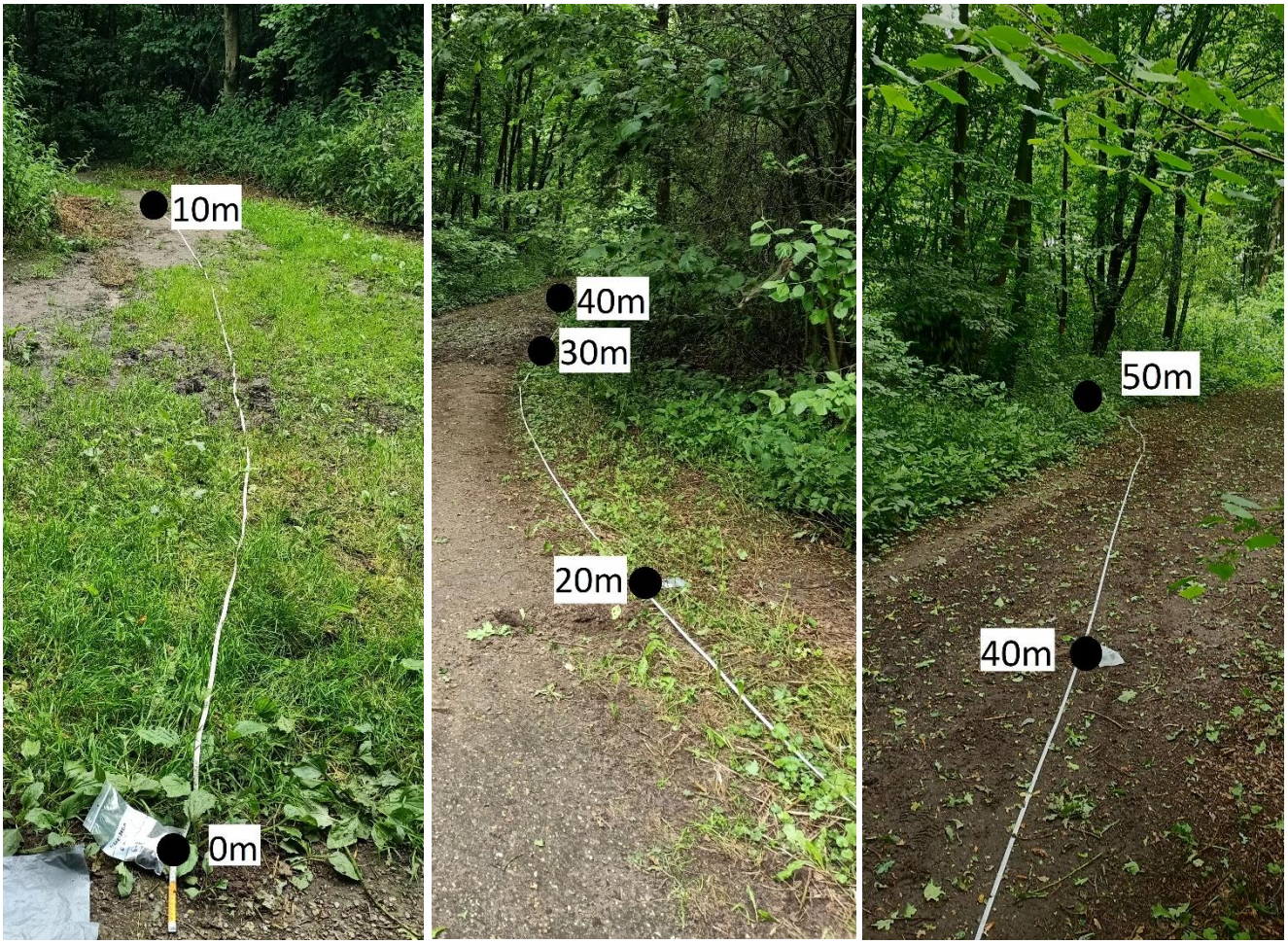

Figure 6. Overview of the walked transect at the location IJsselstein.

2.2.4 Location 4: Schaapsallee

Schaapsallee is located in the Nationaal Park Veluwezoom in the province of Gelderland, the same area as Bosrand. Bosrand was however located more in the middle of the area while Schaapsallee is located at the North-East of this area. This location was visited on the 10th of August 2021. The first 10 metres were slightly muddy and were covered by some patches of grass. The rest of the transect was muddy but less than the first 10 metres. From the 10 metre point onwards till the last 5 metres the transect showed no coverage from leaves or grass. The last 5 metres were completely covered with needles and pine cones, but this layer of ground coverage was so thin that stepping onto this resulted in pushing the needles and pinecones into the soil (see figure 7 for an overview of the transect). Along the transect *Pinus sp.* was most prominent. The first 10 metres were more open, thereafter *Betula sp.* and *Sorbus sp.* occurred together with the *Pinus sp.* trees.

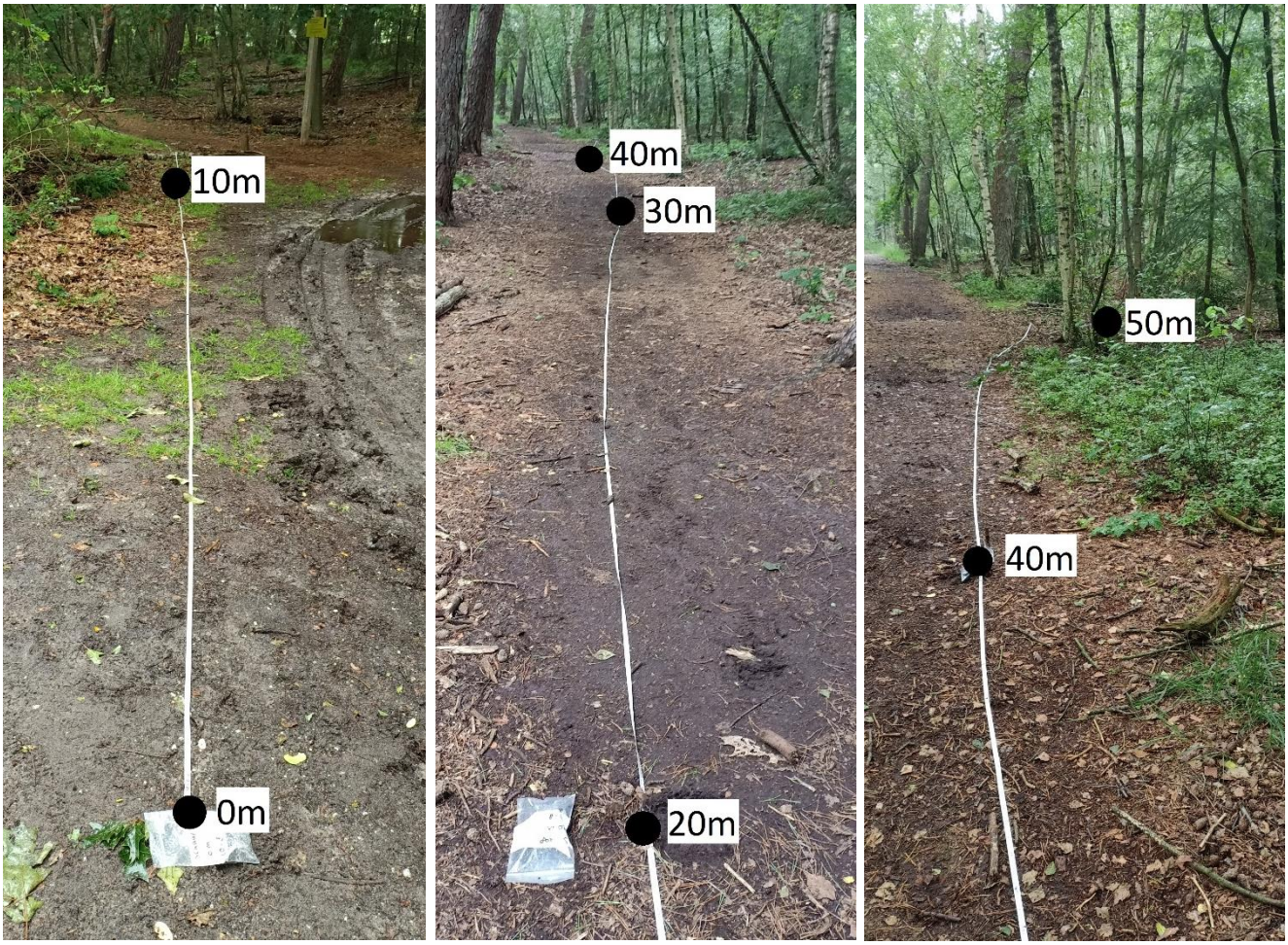

Figure 7. Overview of the walked transect at the location Schaapsallee.

2.2.5 Location 5: Heemskerk

Heemskerk is located at the coast of the Netherlands and is a dune forest. The location was visited on the 7th of November 2021. The first couple of metres were covered with grass. This was followed by a path of bare soil which existed of compacted soil. Only the last 5 metres were again covered, this time with dead leaves and some herbs. The 30 metre point was muddy. The first couple of metres were open, just like the path between the 20 and 40 metre point (see figure 8 for an overview of the transect). Near the 10 and 20 metre point *Quercus sp.* and some tall grasses were prominent. At the 50 metre point *Quercus sp.* and herbs were dominant.

Figure 8. Overview of the walked transect at the location Heemskerk.

2.3 Statistical analysis

2.3.1 Pollen counts

Once the samples were prepared by A.L. Philip, they were used to count a total of 300 pollen per sample. However, in some samples the amount of pollen from one species covered more than 50 per cent of the total count. This could result in count numbers of other species which are too low to be a good representation of the real amount of pollen of those species. For one of those samples (the 0 metre point of the location Bosrand) with one species covering 73.3 per cent of total counts, it was counted to a total of 200 pollen excluding the most dominant species. A Wilcoxon signed rank test has been performed to see if the extra counts make a significant difference in the percentage data of the counted species. If this result is not significant, the total count of 300 pollen has been retained.

2.3.2 Analysis per location

To be able to do a statistical analysis over the differences in similarity between the boots and the points along the transect, it was necessary to add an extra soil sample to the dataset of each location. This extra sample represented a soil mixture of the total transect to be able to compare the boots to a mixture of the transect. This was calculated by adding the count data of all the soil samples per species together and using this to calculate the relative data per species (see figure 9 for the formula). It is important to notice that there has

not been any corrections for differences in concentrations between the points along the transect. Variation in pollen concentrations along the transect will give the impression that more soil has been picked up from points with higher concentrations. To prevent any misinterpretations, it is best to correct for variations in concentrations when calculating relative data for a mixture of the entire transect. However, this is not done because the standard deviations in the concentration data were too big to be used (see Appendix D for de SD's).

$$A_j = \frac{x_{j0m} + x_{j10m} + x_{j20m} + x_{j30m} + x_{j40m} + x_{j50m}}{x_{n0m} + x_{n10m} + x_{n20m} + x_{n30m} + x_{n40m} + x_{n50m}}$$

Figure 9. Formula used for the calculations of the relative occurrence of a specific species along the transect (A_j). The count data per species was added together (x_{j0m}, x_{j10m}, x_{j20m}, x_{j30m}, x_{j40m}, x_{j50m}) and divided by total counts of all the soil samples (x_{n0m}, x_{n10m}, x_{n20m}, x_{n30m}, x_{n40m}, x_{n50m}). This was done for every species found in the soil samples. This resulted in numbers representing the relative occurrence of each species along the transect.

The relative data was also calculated for each separate soil and boot sample by dividing the pollen count per species by the total count of that sample. This was done at every location. The relative data has been used to calculate the squared chord index (see figure 10 for the formula). The squared chord index is a measure of the distance between two different samples. An index score of zero means that both samples are exactly the same and the bigger the number gets the less similar the two samples are. Uitdehaag (2021) showed that this measure is the best way to define the similarity between two pollen assemblages.

$$D\,(x_1, x_2) = \sum_j^n (\sqrt{x_{1j}} - \sqrt{x_{2j}})^2$$

Figure 10. Formula used for the squared chord method. In this case it is used to calculate the distance (D) of dissimilarity between pollen assemblage of the soil and shoe samples. For each pollen type (j) the square root of the relative amounts of the shoe sample (x_1) and one soil sample (x_2) are subtracted from each other. That number is squared and these squared numbers for all the pollen types (n) are summed. This gives the dissimilarity. Null indicates that the samples are completely similar and the higher the number the more dissimilar the samples are.

For this study, the squared chord index was calculated per location for the combination of every boot with every point along the transect including the mixture of the total transect. An ANOVA (or the non-parametric version, the Kruskal Wallis test) was performed to see if index scores differ significantly between the points along the transect. Another ANOVA (or the non-parametric version, the Kruskal Wallis test) was performed to see if the index scores between the boots differed significantly. If one of the boots differed significantly and was an extreme outlier, the first ANOVA performed, was run again without this boot to see if this influenced the results.

The results of the first ANOVA performed would then be compared to pictures which were taken in the field to see what has caused the difference in similarity. The best significant fit of all the boots with a specific point along the transect would be the best sampling location at a crime scene. To test whether this result is consistent throughout the locations, another ANOVA was performed with all the squared chord indexes of every location.

The above mentioned calculations have also been done for the similarity between a point along the transect of a location and the other points along the transect. A one-way RM-ANOVA (or the non-parametric version, the Friedman test) was performed to see if the mean index score differed significantly from each other (samples are the same). This was done to see how different the points along the transect were and thus to see

if the location was heterogeneous enough to find differences between soil traces on a boot and the points along the transect.

2.3.3 Validation of the results

To validate the results of the ANOVA's performed, a detrended correspondence analysis (DCA) was performed. This test shows the variation in the samples. It is expected that, if the counts were consistent throughout the samples, the samples from one location will cluster together in the DCA.

There are four different DCA's performed: a DCA with only the soil samples (1), a DCA with the soil and boot samples (2), a DCA with the soil samples from this study and from Van der Wal's (2021) study (3) and a DCA with the boot and soil samples from this study and from Van der Wal's (2021) study (4). Van der Wal's data is included to see if there are differences in counts between two people. Only the data of location 1, 3, 5 and 6 are included because Van der Wal (2021) did not have data from location 2 and 4. For DCA 1 and 3 the boot samples are left out because boots can show mixtures of the total transect which could make them look different from the specific soil samples along the transect. However, if the DCA shows that the boots are very different from the soil samples, this can show that the counts were not consistent.

The DCA's were performed with the relative data excluding the data with the mixture of the complete transect. All the species that have at least one sample with a relative abundance of 5 per cent were included except for spores (see Appendix C for the species which are included per DCA). The DCA's including Van der Wal's (2021) data contained adjusted data, because Van der Wal (2021) sometimes used higher taxonomic ranks (see table 3 for the adjustments). Besides this, Van der Wal (2021) counted a specific morphology as *Sphagnum sp.* while I counted it as *Pteropsida sp.* (trilete). We both defined the same morphology under another group. Both groups were taken together under the name *Sphagnum sp.*

Table 3. This table shows the adjustments that have been made with the data from this study and Van der Wal's (2021) data before the DCA was performed.

Van der Wal's (2021) data	Data from this study	Notes
Sphagnum sp.	*Pteropsida sp.* (trilete)	Both *Sphagnum sp.* and *Pteropsida sp.* (trilete) are spores and are therefore deleted from the DCA.
Ericaceae sp.	*Calluna sp.*	All the data was grouped together under *Ericaceae sp.*
Plantago undif.	*Plantago lanceolata, Plantago major* and *Plantago undif.*	All the data was grouped together under *Plantago undif.*
Asteraceae undif.	*Asteraceae liguliflorae* and *Asteraceae undif.*	All the data was grouped together under *Asteraceae undif.*
Typha latifolia	*Typha latifolia* and *Typha angustifolia*	All the data was grouped together under *Typha undif.*

2.3.4 Data storage

The data collected for the study is stored at https://zenodo.org/record/6107133.

3. Results

3.1 Pollen counts

The 0 metre point of the location Bosrand consisted of 73.3 per cent *Betula sp.* by a total count of 307 pollen. To figure out if the result in count numbers of other species than *Betula sp.* were misrepresented due to the low counting numbers of those species, there has been counted to a total of 206 pollen excluding *Betula sp.*. This resulted in a total count of 704 of which 498 *Betula sp.* pollen (70.7%). The Wilcoxon signed rank test performed to see if the percentage data differed between the sample with and without extra counting's was insignificant (V = 80, p = 0.5511). This means that a total count of 300 pollen per sample gives sufficient statistical power.

3.2 Results per locations

3.2.1 Location 1: Bosrand

An overview of the percentage and concentration data of location Bosrand can be found in Appendix D resp. Appendix E. The percentage data shows that this location has relatively high amounts of *Betula sp.* in the area (on average 67.1 per cent of the total counts). This means in absolute numbers that the concentration of *Betula sp.* at this location is on average 301875 pollen/cm^3 (lower limit = 208277 pollen/cm^3, upper limit = 437534 pollen/cm^3) in comparison to a total average concentration of 449622 pollen/cm^3 (lower limit = 311203 pollen/cm^3, upper limit = 649607 pollen/cm^3).

The results of the Kruskal-Wallis test of the differences in the squared chord index between the points along the transect can be found in figure 11 ($X^2 = 29.177$, $p = 5.6e^{-5}$). This figure shows that there is a significant difference for the 40 and 50 metre points. The 50 metre point has the lowest similarity with the boots and the 40 metre point has the highest similarity with the boots. An interesting, but not significant result, is that the total transect has the second highest similarity with the boots.

No boots were left out of the analysis because the Kruskal-Wallis test that was performed to see if there were any differences between the boots was insignificant ($X^2 = 2.8653$, $p = 0.5806$).

The Friedman test performed to see if the points along the transect were different from each other showed insignificant results ($X^2 = 8.67$, $p = 0.123$). This means that none of the points along the transect were significantly different from each other.

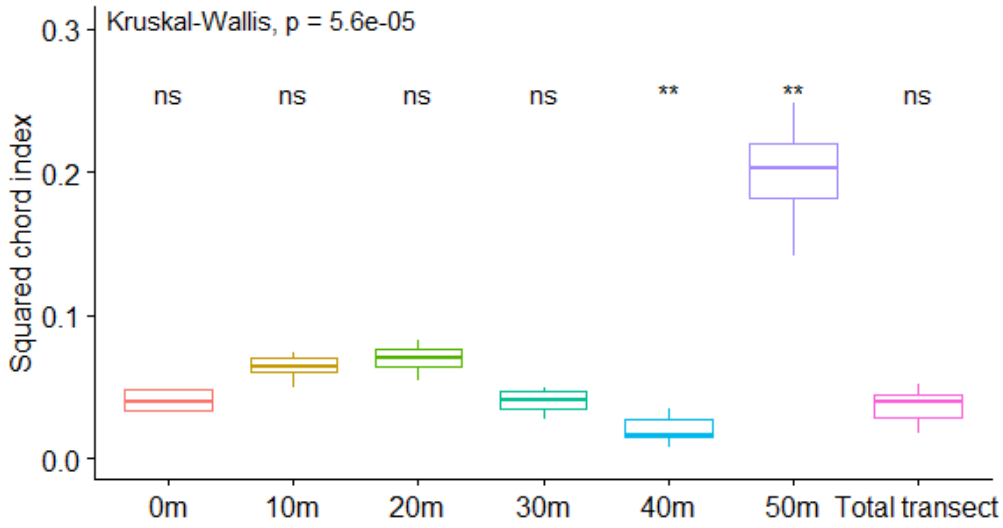

*Figure 11. Boxplot that represents the similarity of the boots with the points along the transect at the Bosrand. The x-axis shows the different points along the transect. The y-axis shows the squared chord index, a measure of the distance between two samples. The lower this number the more similar two samples are. The two samples are in this case a boot with a point along the transect. Each colour would therefore represent the similarity between the five boots and one point on the transect. The asterisks represent the significance (ns: P>0.05, *: P≤0.05, **: P≤0.01, ***: P≤0.001, ****: P≤0.0001).*

3.2.2 Location 2: Utrechtse Heuvelrug

An overview of the percentage and concentration data of location Utrechtse Heuvelrug can be found in Appendix D resp. Appendix E. The percentage data shows that this location has relatively high amounts of *Pinus sp.*, *Quercus sp.* and *Poaceae sp.* in the area (together 65.5 per cent of the total counts). This means in absolute numbers that the concentration of *Pinus sp.*, *Quercus sp.* and *Poaceae sp.* is together 167068 pollen/cm^3 in comparison to a total concentration of 255028 pollen/cm^3 (lower limit = 190604 pollen/cm^3, upper limit = 341229 pollen/cm^3).

The results of the Kruskal-Wallis test of the difference in the squared chord index between the points along the transect can be found in figure 12 (X^2 = 14.381, p = 0.02566). This figure shows that there is a significant difference for the 30 metre points. This point has the lowest similarity with the boots. However, it was decided to leave boot 1 out of the analysis, because the Kruskal-Wallis test that was performed to see if there were any differences between the boots was significant for this boot (X^2 = 10.656, p = 0.03072). The results of the ANOVA of the difference in the squared chord index between the points along the transect excluding boot 1 can be found in figure 13 (F = 10.52, p = 2.02e^{-5}). These results show that the pattern of how similar the boots are to a point on the transect stays roughly the same with or without boot 1 in the analysis. However, the results become more often significant. The ANOVA shows that the results are significant for the 20 and 30 metre points and for the total transect. The 30 metre point still has the lowest similarity with the boots. The 20 metre point and the total transect have the highest similarity with the boots.

The one-way RM-ANOVA performed to see if the points along the transect were different from each other showed insignificant results (F = 0.379, p = 0.858). This means that none of the points along the transect were significantly different from each other.

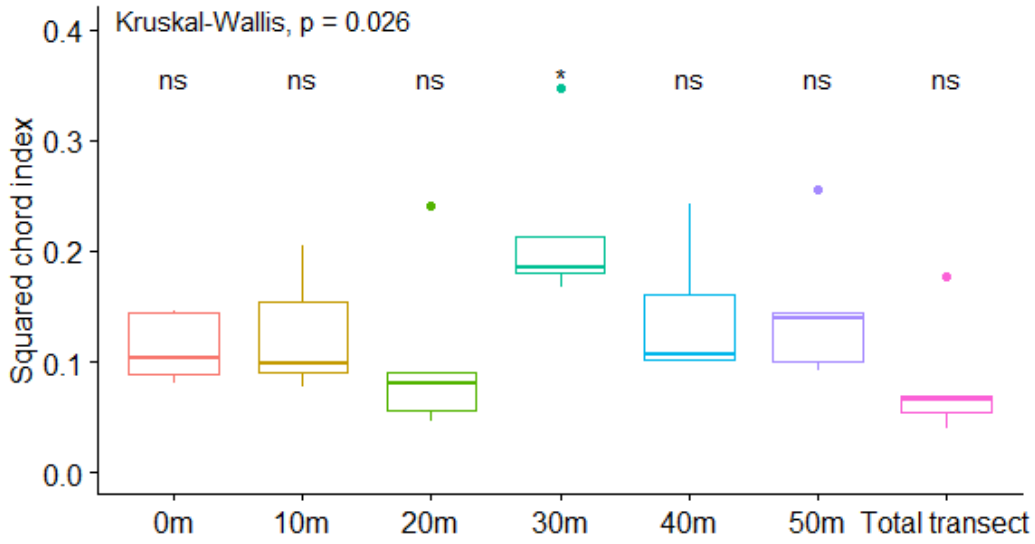

*Figure 12. Boxplot that represents the similarity of the boots with the points along the transect at the Utrechtse Heuvelrug. This figure includes boot 1. The x-axis shows the different points along the transect. The y-axis shows the squared chord index, a measure of the distance between two samples. The lower this number the more similar two samples are. The two samples are in this case a boot with a point along the transect. Each colour would therefore represent the similarity between the five boots and one point on the transect. The asterisks represent the significance (ns: P>0.05, *: P≤0.05, **: P≤0.01, ***: P≤0.001, ****: P≤0.0001).*

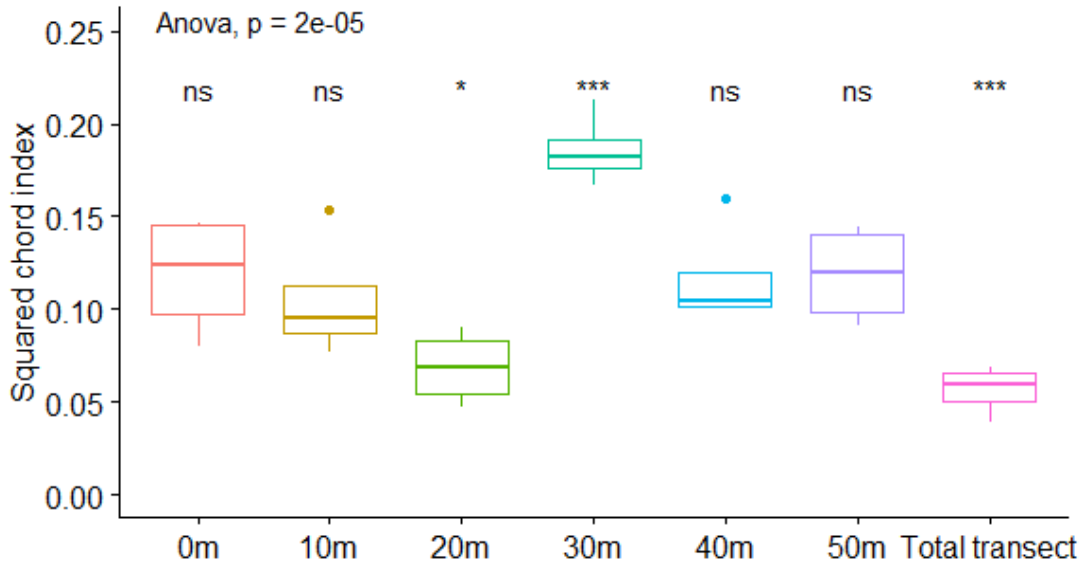

*Figure 13. Boxplot that represents the similarity of the boots with the points along the transect at the Utrechtse Heuvelrug. This figure excludes boot 1. The x-axis shows the different points along the transect. The y-axis shows the squared chord index, a measure of the distance between two samples. The lower this number the more similar two samples are. The two samples are in this case a boot with a point along the transect. Each colour would therefore represent the similarity between the five boots and one point on the transect. The asterisks represent the significance (ns: P>0.05, *: P≤0.05, **: P≤0.01, ***: P≤0.001, ****: P≤0.0001).*

3.2.3 Location 3: IJsselstein

An overview of the percentage and concentration data of location IJsselstein can be found in Appendix D resp. Appendix E. The percentage data shows that this location has relatively high amounts of *Quercus sp.* in the area (on average 46.1 per cent of the total counts). This means in absolute numbers that the concentration of *Quercus sp.* is on average 54375 pollen/cm^3 (lower limit = 43078 pollen/cm^3, upper limit = 68636 pollen/cm^3)

in comparison to a total concentration of 118052 pollen/cm^3 (lower limit = 94699 pollen/cm^3, upper limit = 147164 pollen/cm^3).

The results of the ANOVA test of the difference in the squared chord index between the points along the transect can be found in figure 14 (F = 10.03, p = 6.3e^{-6}). This figure shows that there is a significant difference for the 0 and 30 metre points and for the total transect. The 0 metre point has the lowest similarity with the boots and the 30 metre point and the total transect have the highest similarities with the boots.

No boots were left out of the analysis because the ANOVA test that was performed to see if there were any differences between the boots was insignificant (F = 1.392, p = 0.26). However, the boots showed a relatively high variation between the boots. These are the outliers shown in figure 14.

The Friedman test performed to see if the points along the transect were different from each other showed insignificant results (X^2 = 10.4, p = 0.0651). This means that none of the points along the transect were significantly different from each other.

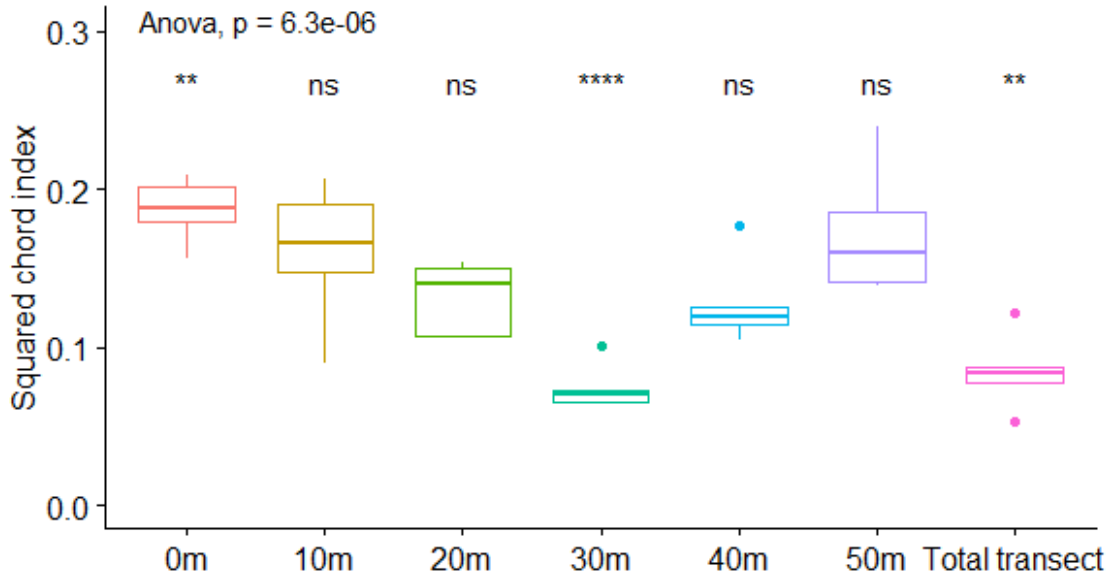

*Figure 14. Boxplot that represents the similarity of the boots with the points along the transect at the IJsselstein. The x-axis shows the different points along the transect. The y-axis shows the squared chord index, a measure of the distance between two samples. The lower this number the more similar two samples are. The two samples are in this case a boot with a point along the transect. Each colour would therefore represent the similarity between the five boots and one point on the transect. The asterisks represent the significance (ns: P>0.05, *: P≤0.05, **: P≤0.01, ***: P≤0.001, ****: P≤0.0001).*

3.2.4 Location 4: Schaapsallee

An overview of the percentage and concentration data of location Schaapsallee can be found in Appendix D resp. Appendix E. The percentage data shows that this location has relatively high amounts of *Betula sp.* and *Pinus sp.* in the area (together 73.4 per cent of the total counts). This means in absolute numbers that the concentration of *Betula sp.* and *Pinus sp.* is together 378618 pollen/cm^3 in comparison to a total concentration of 515679 pollen/cm^3 (lower limit = 311210 pollen/cm^3, upper limit = 1258551 pollen/cm^3). However, low amounts of *Lycopodium sp.* have been counted which means that there is relatively big uncertainty in the calculated concentrations (see Appendix E).

The results of the Kruskal-Wallis test of the difference in the squared chord index between the points along the transect can be found in figure 15 (X^2 = 18.724, p = 0.0047). This figure shows that there is a significant difference for the 0 metre point and for the total transect. The 0 metre point has the lowest similarity with the boots and the total transect has the highest similarity with the boots.

No boots were left out of the analysis because the Kruskal-Wallis test that was performed to see if there were any differences between the boots was insignificant (X^2 = 7.5837, p = 0.1081). However, the boots showed a relatively high variation between the boots at the 0 and 10 metre points (see figure 15).

The one-way RM-ANOVA performed to see if the points along the transect were different from each other showed insignificant results (F = 1.397, p = 0.259). This means that none of the points along the transect were significantly different from each other.

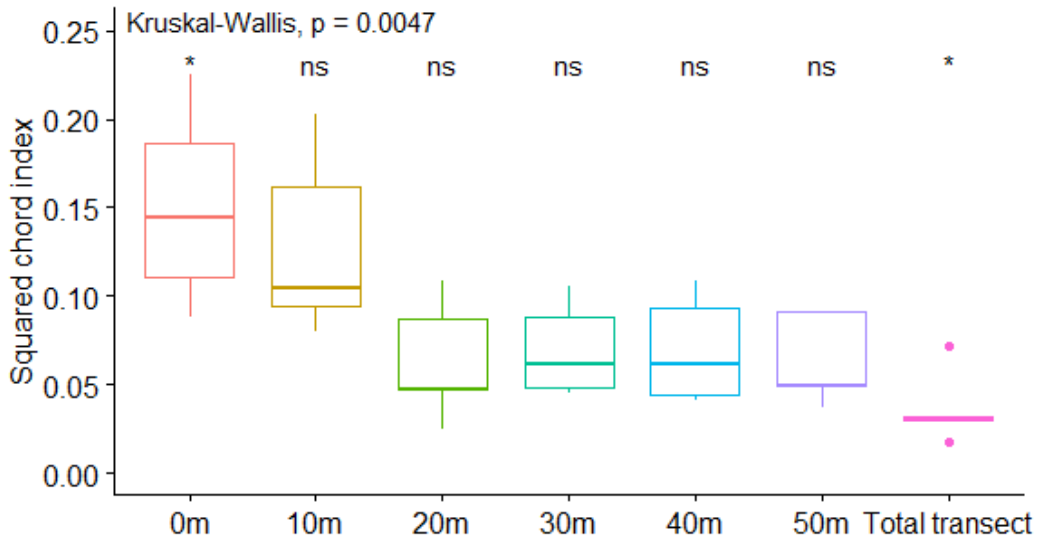

*Figure 15. Boxplot that represents the similarity of the boots with the points along the transect at the Schaapsallee. The x-axis shows the different points along the transect. The y-axis shows the squared chord index, a measure of the distance between two samples. The lower this number the more similar two samples are. The two samples are in this case a boot with a point along the transect. Each colour would therefore represent the similarity between the five boots and one point on the transect. The asterisks represent the significance (ns: P>0.05, *: P≤0.05, **: P≤0.01, ***: P≤0.001, ****: P≤0.0001).*

3.2.5 Location 5: Heemskerk

An overview of the percentage and concentration data of location Heemskerk can be found in Appendix D resp. Appendix E. The percentage data shows that this locations has relatively high amounts of *Quercus sp.*, *Poaceae sp.* and *Pinus sp.* in the area (together 59.2 per cent of the total counts). This means in absolute numbers that the concentration of *Quercus sp.*, *Poaceae sp.* and *Pinus sp.* is together 27068 pollen/cm^3 in comparison to a total concentration of 45727.1 pollen/cm^3 (lower limit = 38501 pollen/cm^3, upper limit = 54310 pollen/cm^3).

The results of the Kruskal-Wallis test of the difference in the squared chord index between the points along the transect can be found in figure 16 (X^2 = 28.64, p = 7.1e^{-5}). This figure shows that there is a significant difference for the 0, 30 and 50 metre points. The 0 and 50 metre points have the lowest similarity with the boots and the 30 metre point has the highest similarity with the boots. An interesting, but not significant result, is that the total transect has the second highest similarity with the boots.

No boots were left out of the analysis because the Kruskal-Wallis test that was performed to see if there were any differences between the boots was insignificant (X^2 = 3.3878, p = 0.4951).

The one-way RM-ANOVA performed to see if the points along the transect were different from each other showed insignificant results (F = 2.057, p = 0.203). This means that none of the points along the transect were significantly different from each other.

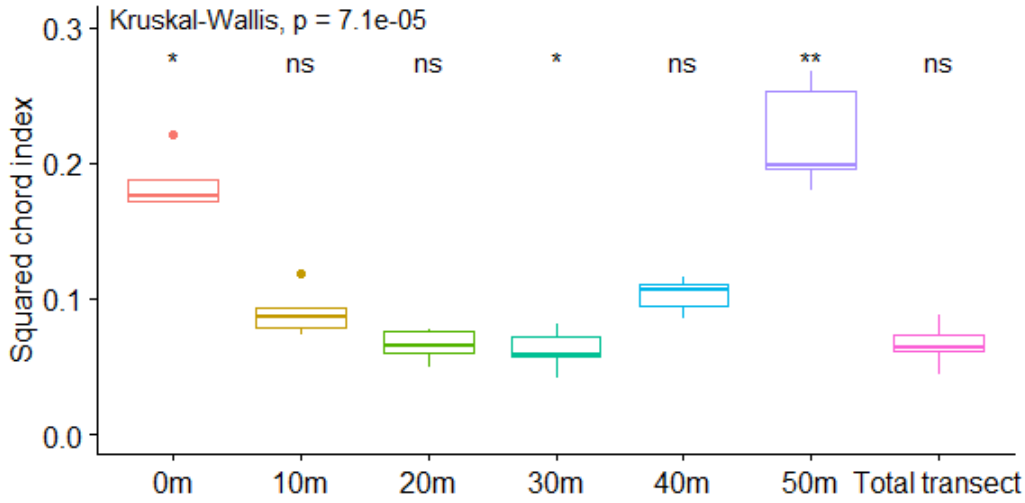

*Figure 16. Boxplot that represents the similarity of the boots with the points along the transect at the Heemskerk. The x-axis shows the different points along the transect. The y-axis shows the squared chord index, a measure of the distance between two samples. The lower this number the more similar two samples are. The two samples are in this case a boot with a point along the transect. Each colour would therefore represent the similarity between the five boots and one point on the transect. The asterisks represent the significance (ns: P>0.05, *: P≤0.05, **: P≤0.01, ***: P≤0.001, ****: P≤0.0001).*

3.2.6 All locations combined

To see if the results of the separate location are consistent throughout all the locations, a Kruskal-Wallis test has been performed with the index scores of all the locations. The results of this test can be found in figure 17 (X^2 = 42.319, p = 1.59e^{-7}). This figure shows that there is a significant difference for the 0 and 50 metre points and for the total transect. The 50 metre point has the lowest similarity with the boots. The 0 metre point and the total transect have the highest similarity with the boots. Interestingly, there is relatively much variation between the boots at every point along a transect.

No boots were left out of the analysis, because the Kruskal-Wallis test that was performed to see if there were any differences between the boots was insignificant (X^2 = 3.6112, p = 0.4612).

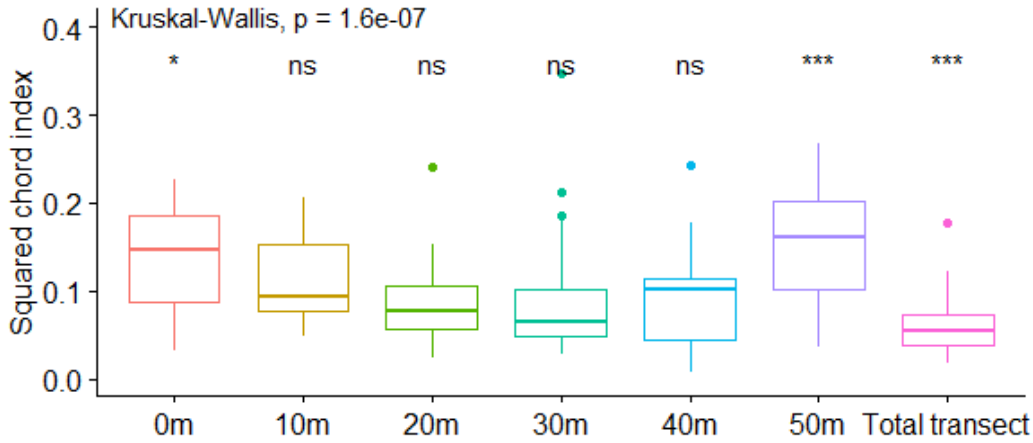

*Figure 17. Boxplot that represents the similarity of the boots with the points along the transect for all locations combined. The x-axis shows the different points along the transect. The y-axis shows the squared chord index, a measure of the distance between two samples. The lower this number the more similar two samples are. The two samples are in this case a boot with a point along a transect. Each colour would therefore represent the similarity between 25 boots and one point on a transect. The asterisks represent the significance (ns: P>0.05, *: P≤0.05, **: P≤0.01, ***: P≤0.001, ****: P≤0.0001).*

3.3 Validation of the results

In figure 18 the results of the DCA between the samples of the different locations are shown. This figure only includes the soil samples for this study and not the boot samples. This figure shows the DCA scores for the soil samples per location. The locations show different combinations of DCA1 and DCA2 scores which makes it possible to separate the locations from each other. The DCA1 is mainly driven by *Quercus sp.* (negative), *Ulmus sp.* (negative) and *Betula sp.* (positive). The DCA2 is mainly driven by *Hedera helix* (negative) and *Plantago lanceolata* (positive). An overview of all the pollen species which drive the variation in DCA1 and DCA2 are shown in Appendix F.

As shown in figure 18, there are some similarities between the locations. Bosrand and Schaapsallee have comparable values for DCA2 but vary in their DCA1 values. This also applies to IJsselstein and Heemskerk. For the Utrechtse Heuvelrug and Heemskerk the opposite accounts: those locations differ in the DCA2 scores and have comparable DCA1 scores.

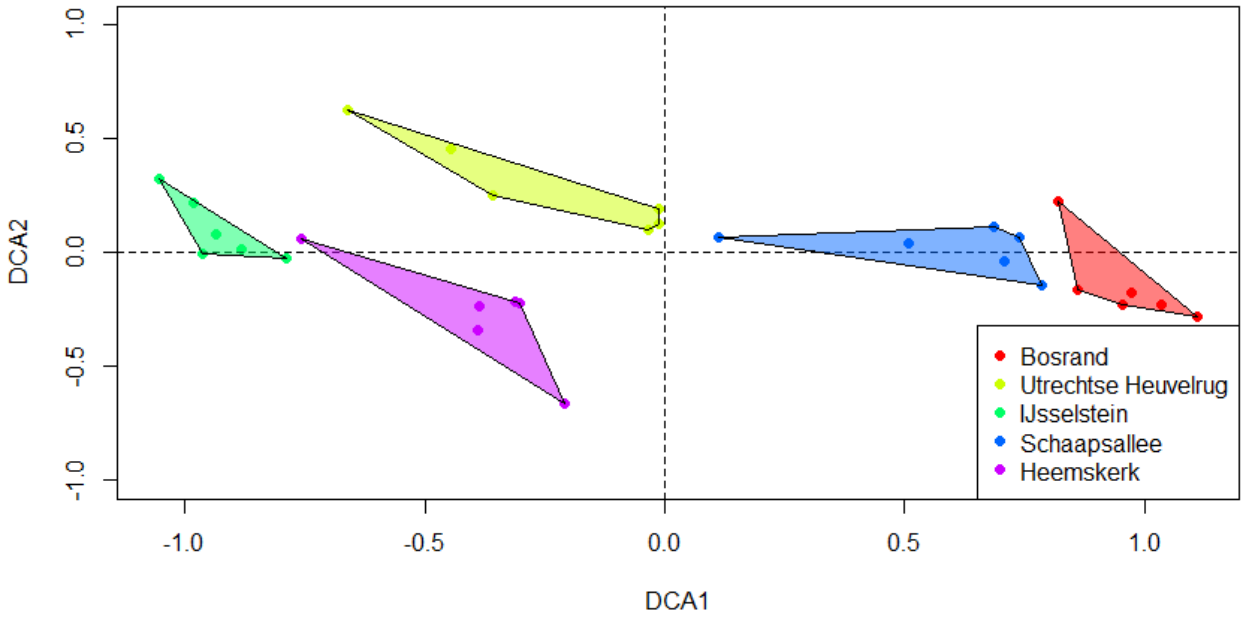

Figure 18. Scatterplot that represents the results of the DCA between all the soil samples collected for this study. Each colour shows a different location. Each location has six different points which represent the places of where the soil samples were taken (0m, 10m, 20m, 30m, 40m and 50m).

In figure 19 the results of the DCA between the samples of the different locations are shown. This figure includes the soil and boot samples collected for this study. This figure shows a comparable, but mirrored pattern as figure 18, which only included the soil samples. The DCA1 is mainly driven by *Quercus sp.* (positive), *Ulmus sp.* (positive), *Asteraceae* undif. (positive) and *Betula sp.* (negative). The DCA2 is mainly driven by *Hedera helix* (negative) and *Plantago lanceolata* (positive). An overview of all the pollen species which drive the variation in DCA1 and DCA2 are shown in Appendix F.

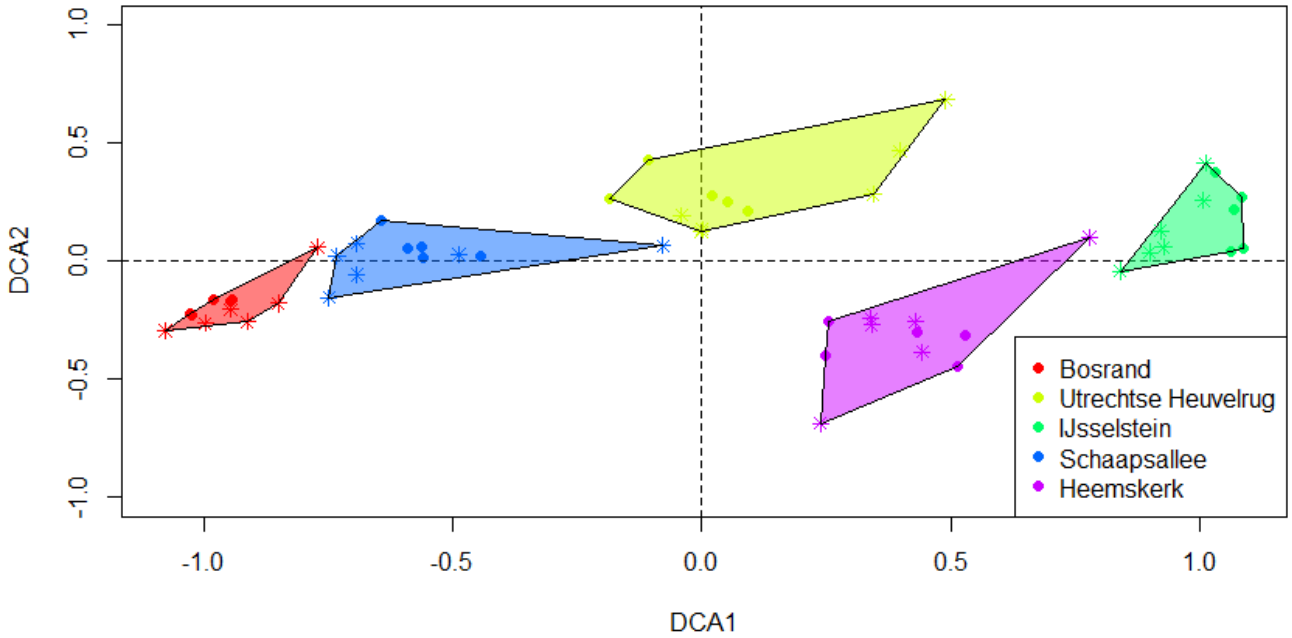

Figure 19. Scatterplot that represents the results of the DCA between all the soil (*) and boot (•) samples collected for this study. Each colour shows a different location. Each location has eleven different points which represent the places of where the soil samples were taken (0m, 10m, 20m, 30m, 40m and 50m) and the different boots that were used (boot 1, 2, 3, 4 and 5).

Figure 20 shows the results of the DCA including the soil samples of the locations that Van der Wal (2021) visited. Locations 5 and 6 from Van der Wal (2021) are comparable to each other for the DCA1 and DCA2 scores. Locations 1 and 3 differ from the rest of his locations in the DCA2 scores. Interestingly, the locations from Van der Wal (2021) show comparable results in DCA2 scores with Bosrand, but not for the rest of the locations visited for this study. IJsselstein on the contrary shows comparable DCA1 scores as three locations of Van der Wal's (2021) locations. The rest of the scores vary between Van der Wal's data and the data from this study. Besides this, by adding Van der Wal's (2021) data to the DCA the DCA scores of the locations visited for this study also changed. There is for example an overlap between Heemskerk and the Utrechtse Heuvelrug while the DCA without Van der Wal's (2021) data did not show this overlap.

The DCA1 and DCA2 scores are partly driven by other species than the scores of the DCA excluding Van der Wal's (2021) data. The DCA1 is mainly driven by *Betula sp.* (positive), *Pinus sp.* (positive) and *Asteraceae* undif. (negative). The DCA2 is mainly driven by *Fagus sp.* (positive), *Quercus sp.* (positive), *Apiaceae* undif. (positive) and *Ericaceae sp.* (negative). An overview of all the pollen species which drive the variation in DCA1 and DCA2 are shown in Appendix F.

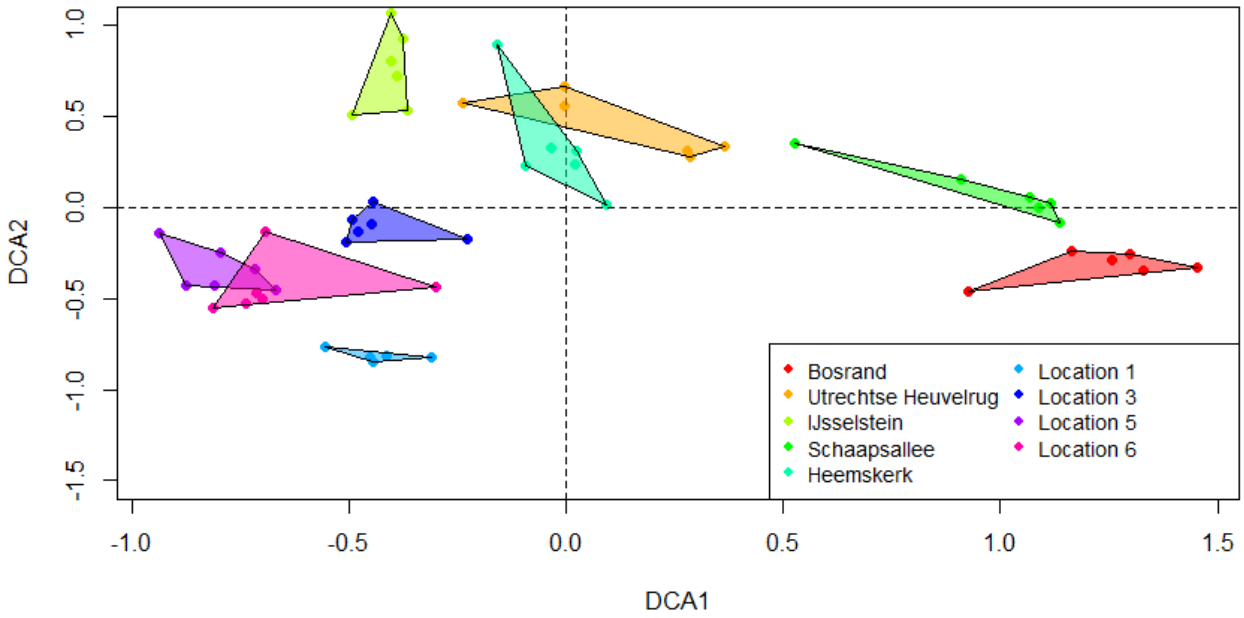

Figure 20. Scatterplot that represents the results of the DCA between all the soil samples collected for this study and for Van der Wal's (2021) study. Each colour shows a different location. Each location has six different points which represent the places of where the soil samples were taken (0m, 10m, 20m, 30m, 40m and 50m).

In figure 21 the results of the DCA between the samples of the different locations are shown. This figure includes the soil and boot samples collected for this study and for Van der Wal's study. This figure shows a comparable, but mirrored pattern as figure 20, which only included the soil samples. The DCA1 is mainly driven by *Betula sp.* (negative), *Pinus sp.* (negative) and *Asteraceae* undif. (positive). The DCA2 is mainly driven by *Fagus sp.* (positive), *Quercus sp.* (positive), *Apiaceae* undif. (positive) and *Ericaceae sp.* (negative). An overview of all the pollen species which drive the variation in DCA1 and DCA2 are shown in Appendix F.

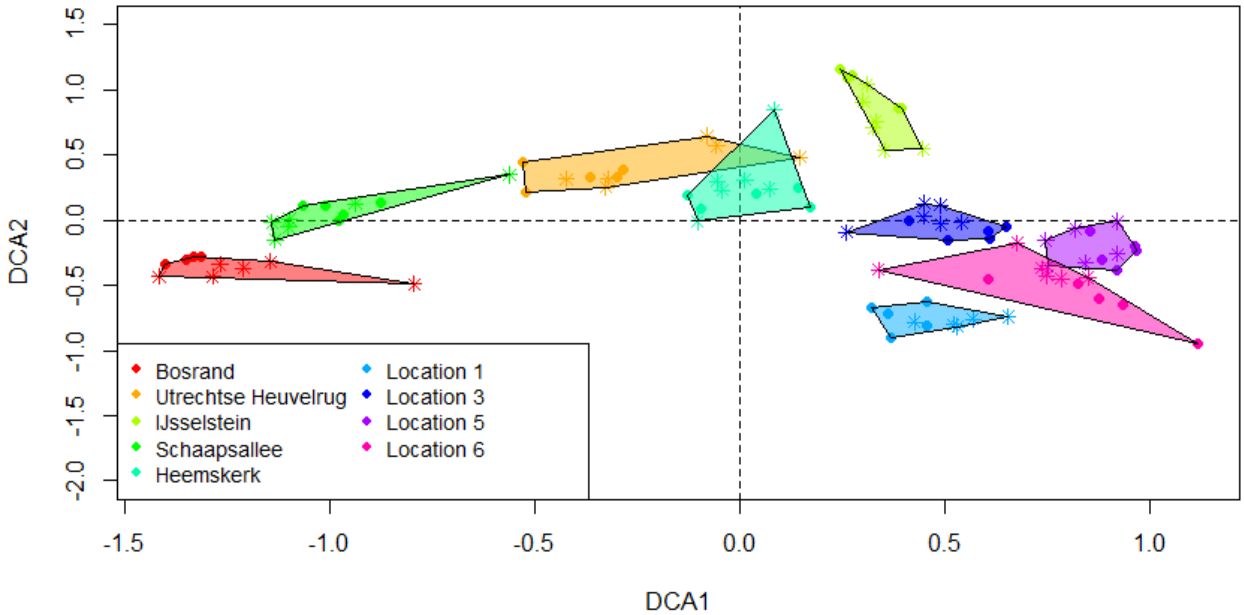

Figure 21. Scatterplot that represents the results of the DCA between all the soil (*) and boot (•) samples collected for this study and for Van der Wal's (2021) study. Each colour shows a different location. Each location has eleven different points which represent the places of where the soil samples were taken (0m, 10m, 20m, 30m, 40m and 50m) and the different boots that were used (boot 1, 2, 3, 4 and 5).

4. Discussion

4.1 Discussion per location

4.1.1 Location 1: Bosrand

Bosrand was a location with high amounts of *Betula sp.* pollen in the samples. This is not surprising because the location has been visited during the flowering period of *Betula sp*. The high amounts of pollen have caused large standard deviations in the concentration data. The percentage data shows lower standard deviations and could thus be used for further analysis.

The Kruskal-Wallis test showed that the boots had the highest similarity with the 40 metre point and the lowest with the 50 metre point. The 50 metre point was completely covered with dead leaves and needles. This might have caused that little to no soil had been transferred to the boots. However, the 40 metre point was also partly covered with dead leaves and needles, but the difference with the 50 metre point was that the 40 metre point showed some patches of bare soil. Nevertheless, there were more points along the transect which showed some degree of coverage and bare patches of soil. This leads to the question why the 40 metre point shows more similarity with the boots than the other points. The most likely answer to this question is that the soil at the 40 metre point was more compacted than the rest of the transect. This concludes for this location that bare soil which is not completely compacted is the best sampling location to compare the soil with the boots.

However, It should be said that the pollen assemblages along the transect did not vary significantly from each other and this makes the location less suitable for this study. Nevertheless, the sampling points were only 10 metres apart from each other, so it was to be expected that these points were not significantly different from each other. The 50 metre point had, however, relatively high squared chord index scores compared to the other points along the transect. This could have been caused by the thick leaf layer on top of the soil which preserves the moisture and organic content. This improves the preservation of pollen in the soil which can result in higher pollen concentrations and preserved pollen from previous years (Havinga, 1984).

4.1.2 Location 2: Utrechtse Heuvelrug

The Utrechtse Heuvelrug was a more open location compared to Bosrand. The main pollen found in the soil were *Pinus sp., Quercus sp.* and *Poaceae sp*. The same could be said about all the sampled boots, except for boot 1. Boot 1 showed a similar concentration of *Pinus sp.* as the other boots, but had lower amounts of *Quercus sp., Poaceae sp.* and *Betula sp*. The total concentration of pollen underneath this boot was also lower. This resulted in percentage data that was different from the rest of the boots. Boot 1 was most comparable to the 0m point concerning the concentration. This indicates that the boots stopped collecting soil after the first point of the transect or that they did not pick up any soil until the last point. Due to those differences in concentration and species underneath the boot, the pollen assemblage was significantly different from the other boots. For this reason boot 1 was deleted from the rest of the analysis.

The analysis excluding boot 1 showed that the 20 metre point and the average of the total transect had the second highest and the highest similarity with the boots and that the 30 metre point had the lowest similarity. The 30 metre point was the point on the transect with the highest ground coverage of (withered) grass. The soil was covered with such an amount of grass that it was almost impossible to walk on bare soil. However, the point with the second highest similarity, the 20 metre point, was the second most covered part of the transect. In contrast to the 30 metre point, this point showed some patches of bare soil so that it was possible to transfer soil to the boots. The rest of the transect however had less ground coverage. It was also

indicated that moisture level and the degree of how much it was compacted was comparable throughout the transect. This lack of variation in these factors might be the reason that the total transect showed the highest similarity with the boots. The boots probably picked up about the same amount of soil along the transect. The total transect showed a relatively high similarity with the 20 metre point (index = 0.058) which could have caused that the 20 metre point also showed a high similarity with the boots.

4.1.3 Location 3: IJsselstein

IJsselstein was the most human-regulated location visited. There was mainly a monoculture of *Quercus sp.*, but the shrubs and herbs changed along the transect which made it a suitable location for this study. The percentage and concentration data did not show any outliers which need explanation.

At this location the boots showed the highest and the second highest similarity with the 30 metre point and the total transect and the lowest similarity with the 0 metre point. The 0 metre point was the only point along the transect with a relatively high percentage of ground coverage with *Plantago* and grasses. It was therefore difficult to transfer any soil to the boots. This is probably the reason that the boots showed the lowest similarity with this point. The point with the highest similarity, the 30 metre point, on the other hand did not have any ground coverage. This point was also next to a puddle which made the moisture level of this point relatively high. The 40 metre point also showed some degree of moisture, but it was less muddy than the 30 metre point. The level of moisture seems to be an accurate predictor of the soil which will be transferred to the boots. However, the average of the total transect at this location showed the second highest similarity with the boots. This could be a coincidence because the average pollen assemblage of the total is comparable to the 30 metre point (index = 0.042).

4.1.4 Location 4: Schaapsallee

The Schaapsallee was located near the Bosrand and showed some similarities with this location in their pollen assemblage. This location for example was also dominated by *Betula sp.* and *Pinus sp*. This location however showed high standard deviations in the concentration data, mainly for the 10 metre point. This was due to low amounts of *Lycopodium sp.* spores that have been counted in the samples. However, the walked transect at the 10 metre point deviated from where the soil sample was taken due to a branch that was lying in the way. The high concentrations of pollen in the soil at the 10 metre point were therefore not transferred to the boots. Besides this, the boots showed a relatively high variation in how similar they were to the points along the transect. This could have been caused by horse manure lying between the 40 and 50 metre point in which the 'perpetrator' walked for three times out of the five. This manure could still contain pollen from the location where the horse has been eating (Arguelles et al., 2015).

The data has been used for further analysis, despite the fact that the concentration data had high standard deviations, despite the set route was not completely walked and despite the boots showed some degree of variation. This has been done, because in reality these situations will also occur. The analysis showed that the 0 metre point had the lowest similarity with the boots and that the total transect had the highest similarity. The 0 metre point was however not any different from the rest of the transect concerning moisture, ground coverage and the compactness. The fact that the walked transect did not show a lot of differences in these factors might be the reason that the total transect had the highest similarity with the boots. It could also be that the total transect was averaging out for the horse manure between the 40 and 50 metre point. Although some boots walked through the horse manure and others did not, the averaging out of this manure by calculating the pollen assemblage of the total transect could have caused that all the boots had the highest similarity with the total transect.

4.1.5 Location 5: Heemskerk

Heemskerk was the only location in the dunes and was also the only location visited in the autumn. The concentration data showed therefore lower amounts of pollen than the rest of the locations. This location was dominated by *Quercus sp.*, *Poaceae sp.* and *Pinus sp.*. The percentage and concentration data did not show any outliers which need explanation.

The analysis for this location showed that the boots had the lowest and the second lowest similarity with the 50 and 0 metre points. The 30 metre point had the highest similarity with the boots. Both the 0 and 50 metre points did not show any bare soil to be transferred to the boots. The 0 metre point was covered with grass and the 50 metre point was covered with different types of herbs. On the contrary, the point with the highest similarity did not have any ground coverage. The soil at this point was muddy and not compacted. These factors were the biggest differences with the rest of the transect. The degree of moisture was therefore the best predictor of which part of the transect would be mostly transferred to the boots.

4.1.6 All locations combined

To examine whether the point along the transect that transfers most of the soil to the boots is the same throughout the locations, the results of the similarity between the boots and the points along the transect were combined for all the locations. The results showed that the 50 and the 0 metre point had the lowest and the second lowest similarity and the total transect had the highest similarity. This indicates that the 50 and 0 metre points are throughout the locations transfer the least soil to the boots. This result was to be expected for the 50 metre point, because this point was covered with leaves, herbs or grass at every location. As discussed above, points with complete ground coverage transfer relatively little soil to the boots. Interestingly, the 0 metre point showed the second lowest similarity while this point varied in ground coverage, moisture and compactness throughout the locations. An explanation for why this point still has an overall low similarity with the boots might be that three locations (IJsselstein, Schaapsallee and Heemskerk) had a low similarity due to ground coverage. However, there is relatively much variation within the data which could result in a false positive. The only point with less variation is the total transect, which has the highest similarity with the boots. This is no surprise as the total transect had the highest or the second highest similarity with the boots at every location. The results of the combinations of all the locations thus show that the total transect is a good place for taking soil samples when investigating a crime scene. However, places with complete ground coverage can better be avoided.

4.2 Validation of the results

The DCA for the different locations including only the soil sample collected for this study showed a clustering of the samples from the same locations. This means that it is possible to differentiate between one location from another on the basis of the differences and similarities between the pollen counts. This indicates that the pollen counts were consistent throughout the samples. In addition, adding the boot samples collected for this study to the DCA still showed the same clustering of locations. This strengthens the conclusion that the pollen counts were consistent throughout the samples.

When adding the pollen data from Van der Wal (2021) to the DCA a separation between the locations visited for his study and for this study arose. This could have had multiple causes. The first possible cause is that Van der Wal (2021) and I counted differently. However, as mentioned in the methods, the data has been corrected as much as possible for this. The second reason could be that Van der Wal (2021) did his research during the late autumn while I did my field work mostly during the summer. The positive DCA1 scores were

mainly driven by differences in *Betula sp.* and *Pinus sp.* between the samples. This research was done just after the flowering season of *Betula sp.*, except for the location Heemskerk, while Van der Wal's (2021) research was done long after the flowering season (Flora van Nederland, 2019; Hájková et al., 2020). However, if this is true it is expected that the pollen data from Heemskerk is more similar to the data from Van der Wal (2021). This is not true which leads to the third reason for the deviation between his and my pollen data. The third reason could be that the locations, where the research has been performed, differed. As mentioned, the DCA1 scores are mostly driven by differences in *Betula sp.* and *Pinus sp*. The locations that Van der Wal (2021) visited had less *Betula sp.* and *Pinus sp.* than the locations I visited except for IJsselstein (Flora van Nederland, 2019). It is therefore expected that the pollen data from IJsselstein is more similar to the data from Van der Wal (2021) which is true. In consequence of this, it is likely that Van der Wal (2021) and I counted in the same and in a consistent manner and that our data varied because of differences in sampling areas.

4.3 Comparison with previous studies

This study shows that the soil that is transferred to the soles of shoes is not always from one specific place along the transect, but that it could be a mixture as well. This depends on factors along the transect, like moisture, ground coverage and compactness of the soil.

These results are interesting in comparison to the findings of Riding et al. (2019). Riding et al. (2019) investigated which soil traces were transferred to the shoe soles of people walking in different areas. They concluded that most of the time the most dominant soil trace came from the last visited site. The last visited site in the case of this study can be compared to the 0m point next to the car from this study. This was however not always the most dominant type of pollen underneath the shoe soles. The variation between those two studies might be a result of the different ways of doing the field work, i.e. visiting multiple sites versus visiting different points at one site. Visiting different points at one site causes that the transfer of soil to the shoes depends more on the ground coverage: if the 0 metre point along the transect is covered, no soil will be transferred to the shoes. However, when visiting multiple sites, the last site will probably not be completely covered so that the shoes will still pick up some soil.

The findings of this study support some of the studies which are based upon case studies. Bull et al. (2006), Morgan et al. (2009) and Wiltshire (2016) showed that underneath a shoe sole of a perpetrator pollen from the crime scene can be found, but also pollen from locations that the perpetrator walked on before or after visiting the crime scene. In comparison to those studies, this study also shows that the soil that is transferred to a shoe sole could be from more than just one point along the transect. It could thus be a mixture of multiple points. However, this study also shows that there is not always a mixture. Sometimes one point along the transect showed the highest similarity with the boots and is the most dominant part of the transect that has been transferred to the shoe soles. This depends on the factors along the transect, like moisture, ground coverage and the compactness of the soil. These factors were not taken into account by Bull et al. (2006), Morgan et al. (2009) and Wiltshire (2016).

In contrast, Van der Wal (2021) took these factors into account in his study. He concluded that the points along a transect where most of the soil is transferred to the shoes, are the points with the least coverage and the point near the corpse where the perpetrator spends most time. This study supports the first conclusion of Van der Wal (2021). However, the second conclusion is not supported by this study. This could be a consequence of the locations visited. For this study the visited locations were either covered with grass, leaves or herbs at the point near the corpse or the transect showed similarities in the factors so that the average of the transect was the best predictor of the soil that was transferred to the shoes.

4.4 Future research

This study gave new insights into the transfer of soil to shoes which can hopefully advance the field of forensic palynology. However, new questions about the application of palynology in forensic studies were also raised: (I) What is the influence of the experimental design on the outcome of this study concerning the statistical power and the way of calculating the pollen assemblage of the total transect? (II) Should the points with total ground coverage be left out when calculating the pollen assemblage of the total transect? (III) Would measuring factors like moisture and ground coverage instead of estimating these factors give the same results? (IV) Will using previously worn boots affect the results? (V) Will it influence the results if the sampling of the soil will be done later than the boot sampling? (VI) And what will the influence be on the results if the body would be buried?

The first question (I) regarding the experimental design is two-sided. The first part of the question is about statistical power. For each location only five pairs of boots were used which means that each group only consisted of five measurements. It is often said that statistical power is reached from thirty or more measurements per group. It would therefore improve this study when the same transect is walked thirty times to reach statistical power and have more reliable results. The second part of this question is about the way of calculating the pollen assemblage for the total transect. In this study there has been chosen for adding up the total counts of each point along the transect per species. These numbers were used to calculate new percentage data. However, no corrections have been made for differences in concentrations between the points along the transect. Higher concentrations of pollen in the soil at one point along the transect will give the impression that relatively more soil has been picked from this place. However, not more soil was picked up, there are just more pollen at that place. Correction for differences in concentration data could have influenced the pollen assemblage of the total transect of the locations Bosrand and Schaapsallee, because these locations showed relatively high variation in concentration data between the points along the transect.

The second question (II) about leaving out points with total ground coverage is related to the previous question. When calculating the pollen assemblage of the total transect, each point along the transect was taken into account. However, this study and Van der Wal's (2021) study showed that points with complete ground coverage did not transfer soil to the shoe soles. It would therefore be good to see if the pollen assemblage of the total transect would show a higher similarity with the boots when those points with complete ground coverage are left out in calculating this assemblage. This would be advantageous because less samples along the transect should then be taken while the matches between the pollen assemblages of the soil and the boots become better.

The third question (III) regards the measuring of the factors moisture, ground coverage and the compactness of the soil. Those factors were only estimated and compared by sight between the different points along a transect. To be able to draw firmer conclusions about the factors that influence the transfer of soil to a shoe sole, it would be useful to measure these factors. These factors can then be included in the statistical analysis to find out if the conclusions drawn in this study are significant.

The fourth question (IV) was about the influence of used shoes on the results of this study. In reality, the chances that totally clean or new shoes will be used for disposing of a corpse are rather small. Previous studies took this into account and performed the field work with used shoes (Adams-Groom, 2018; Pereira et al., 2019; Riding et al., 2007). However, they researched soil transfer to boots at different locations instead of at multiple points at one location. It would therefore be good to combine these previous studies with this research to get a more complete overview of soil transfer to shoes.

The fifth question (V) that was raised was about the timing of sampling. In reality, there is a chance that a suspect or an item connected with the crime scene is not immediately found. This could result in a delay

between the visiting of the area for the dumping and the sampling of the soil. The pollen assemblages can have changed in the time between visiting and sampling (Pereira et al., 2019). It would be useful to investigate if, how and how much this delay results in higher dissimilarities between the boots and the soil.

The last question (VI) regarding the actual burying of a corpse relates to the finding of this study that less compacted soil transfers better to shoes than compacted soil. By burying a corpse it is necessary to dig a grave and thus to loosen the soil. Based on the finding that less compacted soil transfers better to shoes, it is to be expected that the soil dug up for making a grave will be transferred to the shoes. This might give another pattern of which point along the transect transfers most soil. Soil samples for the investigation of a crime scene should then also be taken at a different place: next to the grave.

5. Conclusion

This study investigated the question: "At which point along a walked transect from a parking space to a dumping site is it best to take a soil sample when investigating a crime scene?" To answer this question the dumping of a corpse has been simulated at five different locations. The pollen assemblages along the walked transect were compared to the pollen assemblage underneath the boots to find out at which point most soil was transferred to the boots. A soil sample could best be taken at this point while investigating a crime scene.

The study showed that the point along the walked transect where most soil is transferred to the boots depended on different factors along the transect, like muddiness, ground coverage and compactness of the soil. If there is one point along the transect with more mud than the rest of the transect, this is the point with the highest similarity to the boots. However, points with relatively high amounts of ground coverage with (dead) leaves, grasses or herbs, show the lowest similarities to the boots. Low similarities also arise when the soil is compacted. However, if the total transect is comparable in these factors, the pollen assemblage of the total transect shows the highest similarity with the boots. The total transect at every location had a relatively high similarity with the boots even though the factors along the transect varied. However, when the factors varied there was always one point along the transect with a higher similarity to the boots. This point never had complete ground coverage, was not fully compacted and was muddier than the rest.

In conclusion, it would be best to take a soil sample along the walked transect at a crime scene at the point which is muddier than the rest, shows no complete ground coverage and is not completely compacted. If this point does not exist, it would be best to take multiple samples along the transect from which the pollen assemblage of the total transect can be calculated.

6. Acknowledgements

I would like to thank William Gosling for his supervision during my thesis. He helped me with pollen determination, with the statistical analysis and he supported me during the entire process. I would also like to acknowledged Stefan Uitdehaag from the Netherlands Forensic Institute for his substantial feedback. In addition, I would like to thank Crystal McMichael for being the second assessor for this thesis.

Furthermore, I would like to thank Annemarie Philip for preparing my samples as soon as she could. I would also like to thank Majoi de Novaes Nascimento and Quinten Matthijs for the many fun moments in the lab. At last, I want to thank Jasmijn Theel for her feedback on my thesis and Tijl de Bruijn for all his help and mental support during the entire journey.

7. References

Adams-Groom, B. (2018). Assessment of pollen assemblages on footwear for evidence of pollen deriving from a mock crime scene: A contribution to forensic palynology. *Grana*, *57*(3), 223–234. https://doi.org/10.1080/00173134.2017.1310293

Alotaibi, S. S., Sayed, S. M., Alosaimi, M., Alharthi, R., Banjar, A., Abdulqader, N., & Alhamed, R. (2020). Pollen molecular biology: Applications in the forensic palynology and future prospects: A review. *Saudi Journal of Biological Sciences*, *27*(5), 1185–1190. https://doi.org/10.1016/j.sjbs.2020.02.019

Arguelles, P., Reinhard, K., & Shin, D. H. (2015). Forensic Palynological Analysis of Intestinal Contents of a Korean Mummy. *The Anatomical Record*, *298*(6), 1182–1190. https://doi.org/10.1002/ar.23141

Bell, S., Sah, S., Albright, T. D., Gates, S. J., Denton, M. B., & Casadevall, A. (2018). A call for more science in forensic science. *Proceedings of the National Academy of Sciences*, *115*(18), 4541–4544. https://doi.org/10.1073/pnas.1712161115

Bryant, V. M., Jones, J. G., & Mildenhall, D. C. (1990). Forensic palynology in the United States of America. *Palynology*, *14*(1), 193–208. https://doi.org/10.1080/01916122.1990.9989380

Bull, P. A., Parker, A., & Morgan, R. M. (2006). The forensic analysis of soils and sediment taken from the cast of a footprint. *Forensic Science International*, *162*(1), 6–12. https://doi.org/10.1016/j.forsciint.2006.06.075

Butler, J. M. (2011). *Advanced topics in forensic DNA typing: Methodology*. Academic press.

de Leeuwe, R. (2014, September 27). *Een graf zegt meer dan een lichaam*.

Flora van Nederland. (2019, March 24). *Ruwe berk—Betula pendula*. https://www.floravannederland.nl/planten/ruwe_berk

Hájková, L., Kožnarová, V., Možný, M., & Bartošová, L. (2020). Influence of climate change on flowering season of birch in the Czech Republic. *International Journal of Biometeorology*, *64*(5), 791–801. https://doi.org/10.1007/s00484-020-01869-1

Havinga, A. J. (1984). A 20-year experimental investigation into the differential corrosion susceptibility of pollen and spores in various soil types. *Pollen et Spores*, *26*, 541–558.

Horrocks, M., & Walsh, K. A. J. (1998). Forensic palynology: Assessing the value of the evidence. *Review of Palaeobotany and Palynology*, *103*(1), 69–74. https://doi.org/10.1016/S0034-6667(98)00027-X

Jacobson, G. L., & Bradshaw, R. H. W. (1981). The Selection of Sites for Paleovegetational Studies. *Quaternary Research*, *16*(1), 80–96. https://doi.org/10.1016/0033-5894(81)90129-0

Mildenhall, D. C., Wiltshire, P. E. J., & Bryant, V. M. (2006). Forensic palynology: Why do it and how it works. *Forensic Science International*, *163*(3), 163–172. https://doi.org/10.1016/j.forsciint.2006.07.012

Morgan, R. M., Freudiger-Bonzon, J., Nichols, K. H., Jellis, T., Dunkerley, S., Zelazowski, P., & Bull, P. A. (2009). The Forensic Analysis of Sediments Recovered from Footwear. In K. Ritz, L. Dawson, & D. Miller (Eds.), *Criminal and Environmental Soil Forensics* (pp. 253–269). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9204-6_16

Pereira, J. S., Ribeiro, H., & Abreu, I. (2019). Spatial and temporal environmental pollen analysis of footwear worn in the area of Barcelos, North-West Portugal, in a forensic context. *Aerobiologia*, 1–6.

Politie. (2021, April 15). *data.politie.nl—Geregistreerde misdrijven; soort misdrijf, wijk, buurt, jaarcijfers*. https://data.politie.nl/#/Politie/nl/dataset/47018NED/table?ts=1620143946675

Riding, J., Rawlins, B., & Coley, K. (2007). Changes in soil pollen assemblages on footwear worn at different sites. *Palynology*, *31*, 135–151. https://doi.org/10.2113/gspalynol.31.1.135

Saferstein, R. (2015). *Criminalistics: An Introduction to Forensic Science* (11th ed.). Pearson.

Strafrechtadvocaten Netwerk. (2021). Uitleg 'wettig en overtuigend' bewezen. *Bewijs in Strafzaken*. https://bewijs-in-strafzaken.nl/uitleg-wettig-en-overtuigend-bewezen/

Tiemens, D. (2009). *De toepassing van grondonderzoek binnen de opsporing*. https://www.politieacademie.nl/kennisenonderzoek/kennis/mediatheek/PDF/79150.PDF

Tilstone, W. J., Savage, K. A., & Clark, L. A. (2006). *Forensic science: An encyclopedia of history, methods, and techniques*. ABC-CLIO.

Uitdehaag, S. C. A. (2021). *Forensic soil comparison: Towards objective methods for a more robust evidential value*. University of Utrecht.

Uitdehaag, S., Donders, T. H., & Kuiper, I. (2014). *Palynology in forensic casework; possibilities and pitfalls*. https://doi.org/10.13140/RG.2.2.30618.44488

van der Wal, S. (2021). *Envrionmental Forensics: Origin and build-up rate of forest soil on the shoes of culprits*. University of Amsterdam.

Walsh, K. A. J., & Horrocks, M. (2008). Palynology: Its Position in the Field of Forensic Science. *Journal of Forensic Sciences*, *53*(5), 1053–1060. https://doi.org/10.1111/j.1556-4029.2008.00802.x

Wiltshire, P. E. J. (2009). Forensic Ecology, Botany, and Palynology: Some Aspects of Their Role in Criminal Investigation. In K. Ritz, L. Dawson, & D. Miller (Eds.), *Criminal and Environmental Soil Forensics* (pp. 129–149). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9204-6_9

Wiltshire, P. E. J. (2016). Protocols for forensic palynology. *Palynology*, *40*(1), 4–24. https://doi.org/10.1080/01916122.2015.1091138

Appendix A. Preparation of (microfossil) pollen

These are the steps Ms. Philip will follow to prepare the samples for microscopy.

- Boil the samples in 10% KOH.
- Sieve the samples (212 m meshes) and pour into centrifuge tubes.
- Centrifuge (after each step in the treatment the samples are centrifuged at ca. 4500 rpm)
- Wash 2 times (or more to clean the sample) with water until after centrifuging the water is clear.
- Wash 1 time with Acetic acid 96%, to remove the water from the sample.
- Acetolysis: use mixture of 9 parts acetic anhydride and 1 part of H2SO4.
    - Be careful with mixing the liquids: slowly add the H2SO4 to the acetic anhydride! Stirring and, at the same time, cooling in a water bath is essential for getting the right acetolysis mixture. No contact between the mixture and water!
- Heat the samples in the acetolysis mixture to 100 degrees C for ca. 10 minutes.
- Wash 4 times with water.
- Wash 3 times with ethanol 96%.
- In case the samples contain minerogenic material: use Bromoform/ ethanol mixture (specific gravity 2.0)
- Bromoform-treatment:
    - Use Bromoform s.g. 2 and centrifuge during 10 minutes at 1500 rpm. In the liquid the pollen (and other organic matter) will be separated from the minerogenic material, which is transported to the bottom of the tube. Microfossils are in the collar of the centrifuge tube. Important: do not use the break of the centrifuge in this case!
    - Pour the collar into a tube filled for 1/3 with ethanol 96%. Centrifuge (4500 rpm)
    - Pour the residue into small tubes of ca 5 ml with ethanol 96%. Centrifuge (4500 rpm).
    - Put a drop of glycerin on the residue and put it for one night in an oven at ca.40-50 degrees C.

Appendix B. Count sheet lay-out

Pollen Count Data Sheet ***V. 1-1*** (WDG)	Site:	Level:
Slide:	Analyst:	Date:

Tree Pollen		**Herb Pollen**	
Betula:		Poaceae [Gramineae][6]:	
Pinus:			
Ulmus:		Cereal[6]:	
Quercus:		Cyperaceae:	
Tilia cord. / plat.:		*Calluna:*	
Alnus:		*Empetrum:*	
Fagus:		Ericaceae undif.:	
Carpinus:		Asteraceae (Asteroideae/Cardueae) undif. [Compositae (tub.)][7]:	
Fraxinus:			
Acer:		*Artemisia* type (Asteroideae):	
Populus:		*Solidago v.* [*Aster*] type (Asteroideae)[8]:	
Abies:		*Achillea* [*Anthemis*] type (Asteroideae)[8]:	
Picea:		*Arctium* [*Serratula*] type (Cardueae)[8]:	
Taxus:		*Cirsium* type (Cardueae):	
		Centaurea (Cardueae)*:*	
		Asteraceae (Lactucae) undif. [Compositae (Lig.)][9]:	
TOTAL TREE:			
Shrub Pollen		*Armeria m.* type [A/B]*:*	
Corylus avellana type (±*Myrica* [ ]):		*Caltha* type*:*	
		Caryophyllaceae:	
Salix:		*Campanula* type:	
Buxus:		Chenopodiaceae[10]:	
Juniperus:		Brassicaceae [Cruciferae]:	
Hippophae:		*Epilobium* type:	
Hedera helix:		*Filipendula:*	
Ilex:		*Helianthemum:*	
Ephedra[1]*:*		Lamiaceae [Labiatae][11]:	
		Fabaceae [Leguminosae][12]:	
		Oxyria type[13]	
TOTAL SHRUB		*Plantago* undif.*:*	
Spore		*Plantago lanceolata* type[14]*:*	
Botrychium:		*Polygonum*[15]*:*	
Equisetum:		*Potentilla* type*:*	
Lycopodium[2]*:*		*Ranunculus* type[16]:	
Ophioglossum:		Rosaceae undif.[17]:	
Osmunda:		Rubiaceae undif.[18]:	
Polypodium:		*Rumex acetosa* type*:*	
Selaginella:		*Rumex*[19]*:*	
Pteridium aquilinum:		Saxifragaceae undif.[20]:	
Pteropsida [*Filicales*] (monolete) undif.[3]:		*Succisa:*	

		Thalictrum:	
Pteropsida (trilete) undif.[4]:		*Urtica*[21]*:*	
		Apiaceae [Umbelliferae] undif.[22]:	
		Valeriana:	
TOTAL SPORE:			
Obligate aquatic spores[5]			
Sphagnum:			
Isoetes:		TOTAL HERB:	

Aquatic Pollen		*Potamogeton*[24]*:*		**Indeterminate**	
Alisma type*:*		*Sparganium erectum:*		Broken:	
Hydrocotyle vulgaris:		*Typha angustifolia* type*:*		Concealed:	
Menyanthes trifoliata:		*Typha latifolia* type*:*		Corroded:	
Myriophyllum[23]*:*				Crumpled:	
Nuphar:				Degraded:	
Nymphaea alba type*:*		TOTAL AQUATIC:		TOTAL INDETERMINATE:	

Pre-Quaternary microfossils:		Dinoflagellate cysts:	
Exotic:		TOTAL DRY LAND / MAIN SUM:	

NOTE:

1 *Ephedra* can be further sub-divided into *Ephedra distachya* and *E. fragilis.*

2 *Lycopodium* is often added as an exotic marker, however differentiation between added and fossil pollen should be possible. This trilete spore may be subdivided *Lycopodium annotium* type, *Lycopodium clavatum.* Degraded grains may key out as *Pteropsida* (trilete) undif.[3].

3 This category could include any of the above spores due to the loss of outer coat making differentiation impossible. Polypodeace. Other monocolpate spores that can be determined include *Thelyptris palustris, Dryopteris dilatata, Dryopteris filix-mas,* and *Dryopteris cristata* type, MWC 1991.

4 Other trilete spores may key out as *Adiantum capillus-veneris, Anogramma leptophylla, Anthoceros punctatus* type*, Botrychium lunaria* type, *Cryptogramma crispa, Diaphasiastrum* type, *Huperzia, selago, Hymenophyllum, Lycopodiella inundata, Lycopodium annotium* type, *Lycopodium clavatum*[1] *Ophiglossum vulgatum* type A/B, *Osmunda regalis, Phaeoceros laevis, Pilularia globulifera* microspores, *Pilularia* microspores, *Riccia* type, *Selaginella selagnoidies,* and *Trichomanes speciosum,* MWC 1991.

5 These species are dependant upon waterlogged conditions and can occur is such abundance that they swamp samples and therefore must be calculated outside the main sum.

6 Grasses are defined as below; Poaceae (wild grass group) = Mean annulus diameter < 8μm, mean grain size < 37μm, surface scabrate or verrucate; Cereal undif. can be further sub-divided into *Hordeum* group = Mean annulus diameter 8-10μm, mean grain size 32-45μm, surface scabrate; *Glyceria* = water grass easily mistaken for the *Hordeum* group; *Avena-Triticum* group = Mean annulus diameter >10μm, mean pollen grain size > 40μm, surface verrucate; and *Secale cereale* = Mean annulus diameter 8-10μm oblong grain outline (high pollen index), surface scabrate, MWC 1991.

7 Asteraceae can be divided into Asteroidea and the Lactucoidea. The Lactucoidea can be further sub-divided into the Cardueae and Lactucae. The old division of the Compositae based upon pollen morphology known as the Compositae Liguliflorae can now be referred to as Asteraceae (Lactucae) undif. however the Compositae Tubuliflorae covers a wider taxonomic grouping and is now referred to as Asteraceae (Asteroidea/Cardueae) undif. BWE 1994.

8 Divisions as suggested by BWE 1994.

9 Asteraceae (Lactucae) undif. are part of the Asteraceae Lactucoidea, S 1991

10 Includes Amaranthaceae, MWC 1991.

11 Lamiaceae includes species which key to *Marrubium vulgare, Mentha* type, *Prunella* type, *Stachys sylvatica* type, *Stutellaria* type and *Teucrium,* MWC 1991.

12 Fabaceae includes species which key to *Astragalus danicus* type, *Coronilla varia*, *Galega officinalis* type, *Hippocrepis comosa*, *Lotus* type, *Medicago sativa*, *Onobrychis* type, *Ononis* type, *Ornithopus perpusillus*, *Robinia pseudoacacia*, *Trifolium* type, *Trifolium spadiceum*, *Ulex* type, *Vicia* type and *Vicia cracca* type, MWC 1991.

13 *Oxyria* type includes *Oxyria digyna, Rumex cripus, R. conglomeratus, R. sanguineus, R. pulcher, R. maritimus. Oxyria* and *R. acetosella* are <26μm, *Oxyria* also has a more clearly circular porus which is ringed in phase contrast, MWC 1991.

14 Varieties of *Plantago* can be keyed out to *P. maritima* type, *P. major, P. media,* and *P. coronopus*, MWC 1991..

15 Varieties of *Polygonum* can be keyed out to *P. amphibium, P. aviculare* type, *P. bistorta* type, and *P. persicaria* type.

16 Within *Ranunculus* type it may be possible to identify *Eranthis hyemalis* and *Pulsatilla vulgaris* MWC 1991

17 Rosaceae undif. can be sub-divided into *Sanguisorba minor* ssp. *minor, Agrimonia eupatoria, Crataegus, Dryas octopetala, Malus sylvestris, Mespilus germanica, Potentilla* -type, *Prunus, Pyrus pyraster, Rosa, Rubus, Sorbus,* MWC 1991.

18 *Galium* type is the only definable member of the Rubiaceae and includes *Galium, Asperula, Rubia* and *Sherardia*, MWC 1991.

19 Certain varieties of *Rumex* also key out to *Oxyria* type or *Rumex obtusifolius* type, MWC 1991.

20 Saxifragaceae undif. includes *Saxifraga androsacea, S. cernua* type, *S. granulata* -type, *S. hirsuta* type, *S. oppsitifolia* type and *S. stellaris* type, MWC 1991

21 *Urtica* can be sub-divided into *Urtica dioica, U. pilulifera* and *U. urens*, MWC 1991.

22 Apiaceae undif. can be further sub-divided into groups with similar morphological characteristics, MWC 1991

23 *Myriophyllum* can be sub-divided in to *M. alterniflorum, M. spicatum* and *M. verticillatum*, MWC 1991.

24 *Potamogeton* can be divided in to two subgenera *Potamogeton* subgenus *Potamogeton* type and *Potamogeton* subgenus *Coleogeton,* MWC 1991.

References

BWE 1994 Bennet, K.D., Whittingto, G. & Edwards, K.J. 1994 Recent plant nomenclatural changes and pollen morphology in the British Isles *Quaternary Newsletter* **73:**1-6

MWC 1991 Moore, P.D., Webb, J.A. & Collinson, M.E. 1991 *Pollen Analysis* (2nd ed.) Blackwell Oxford

S 1991 Stace, C. 1991 *New Flora of the British Isles* Cambridge University Press Cambridge

Appendix C. Species included in DCA's

Table 4. Overview of the species included in the DCA's.

DCA	Description	Species
1	Includes soil samples collected for this study.	*Betula sp., Pinus sp., Ulmus sp., Quercus sp., Alnus sp., Fagus sp., Corylus avellana, Hedera helix, Ilex sp., Poaceae sp., Calluna sp., Asteraceae* undif., *Plantago lanceolata, Urtica sp.* and *Apiaceae* undif.
2	Includes soil and boot samples collected for this study.	*Betula sp., Pinus sp., Ulmus sp., Quercus sp., Alnus sp., Fagus sp., Corylus avellana, Salix sp., Hedera helix, Ilex sp., Poaceae sp., Calluna sp., Asteraceae* undif., *Plantago lanceolata, Urtica sp.* and *Apiaceae* undif.
3	Includes soil samples collected for this study and Van der Wal's (2021) study.	*Betula sp., Pinus sp., Ulmus sp., Quercus sp., Alnus sp., Fagus sp., Corylus avellana, Salix sp., Hedera helix, Ilex sp., Poaceae sp., Cyperaceae sp., Ericaceae sp., Asteraceae* undif., *Plantago* undif.*, Saxifragaceae sp., Urtica sp., Apiaceae* undif., and *Sagitarria sp.*
4	Includes soil and boot samples collected for this study and Van der Wal's (2021) study.	*Betula sp., Pinus sp., Ulmus sp., Quercus sp., Alnus sp., Fagus sp., Corylus avellana, Salix sp., Hedera helix, Ilex sp., Poaceae sp., Cyperaceae sp., Ericaceae sp., Asteraceae* undif., *Plantago* undif.*, Saxifragaceae sp., Urtica sp., Apiaceae* undif., and *Sagitarria sp.*

Appendix D. Percentage data per location

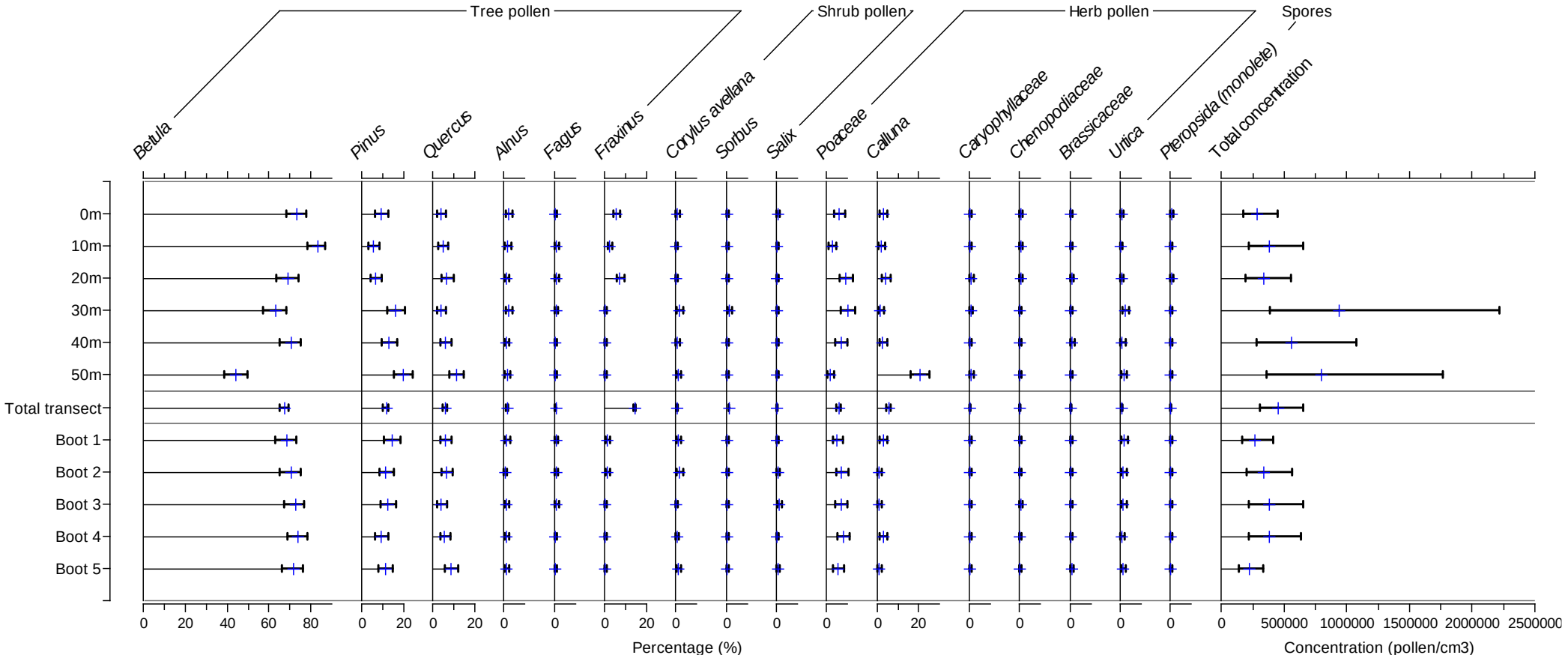

Figure 22. Overview of the relative data for each sample at the location Bosrand. The total concentration of pollen per cm^3 are also mentioned.

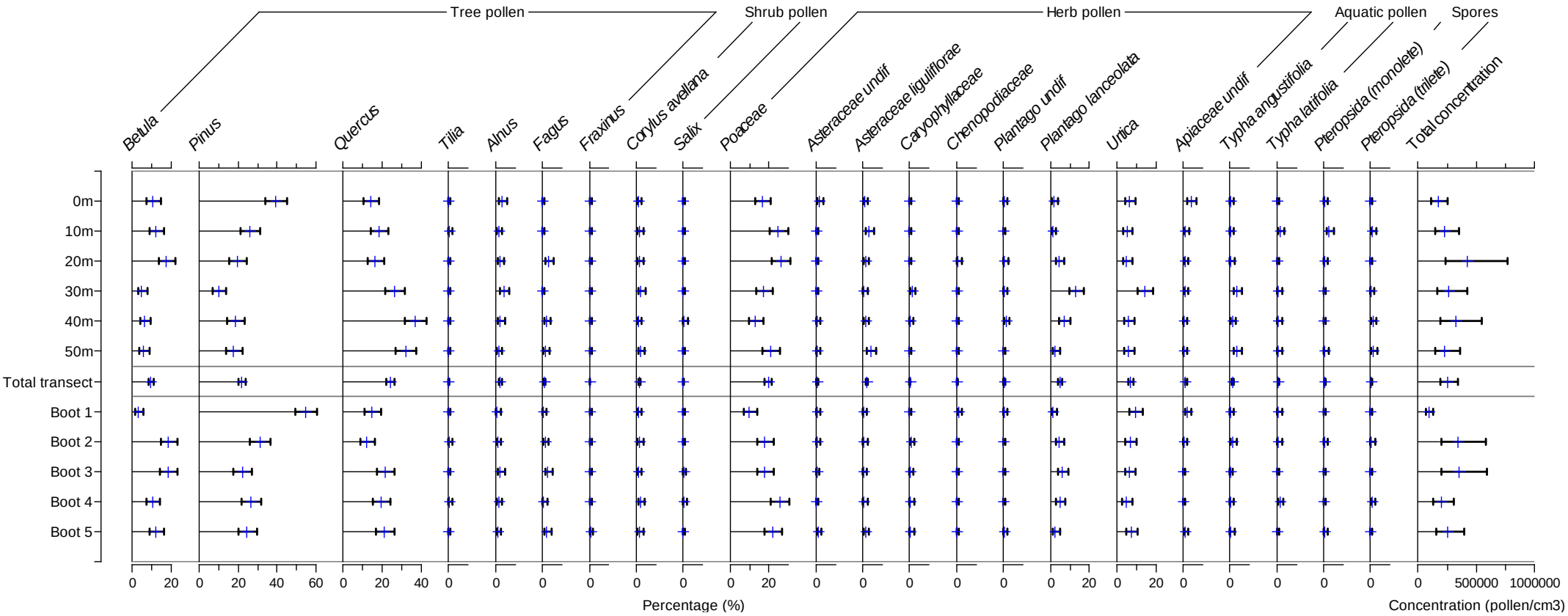

Figure 23. Overview of the relative data for each sample at the location Utrechtse Heuvelrug. The total concentration of pollen per cm^3 are also mentioned.

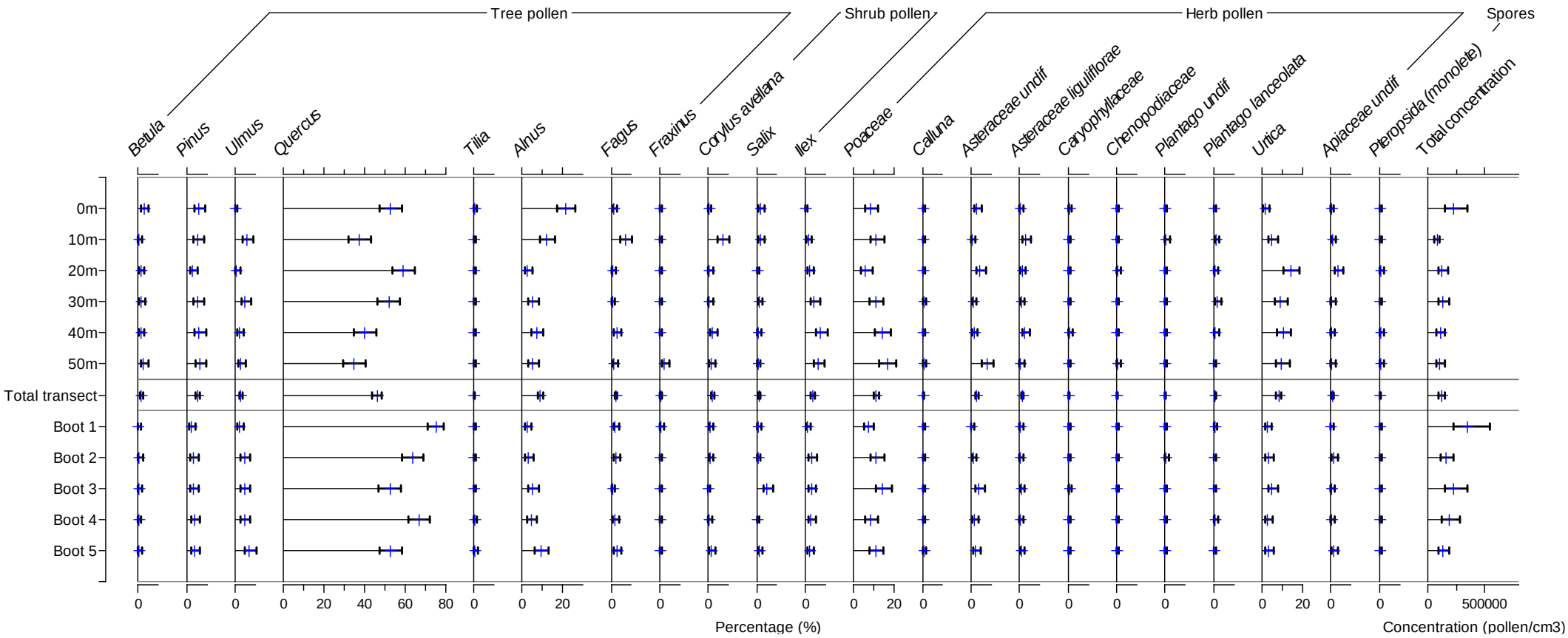

Figure 24. Overview of the relative data for each sample at the location IJsselstein. The total concentration of pollen per cm^3 are also mentioned.

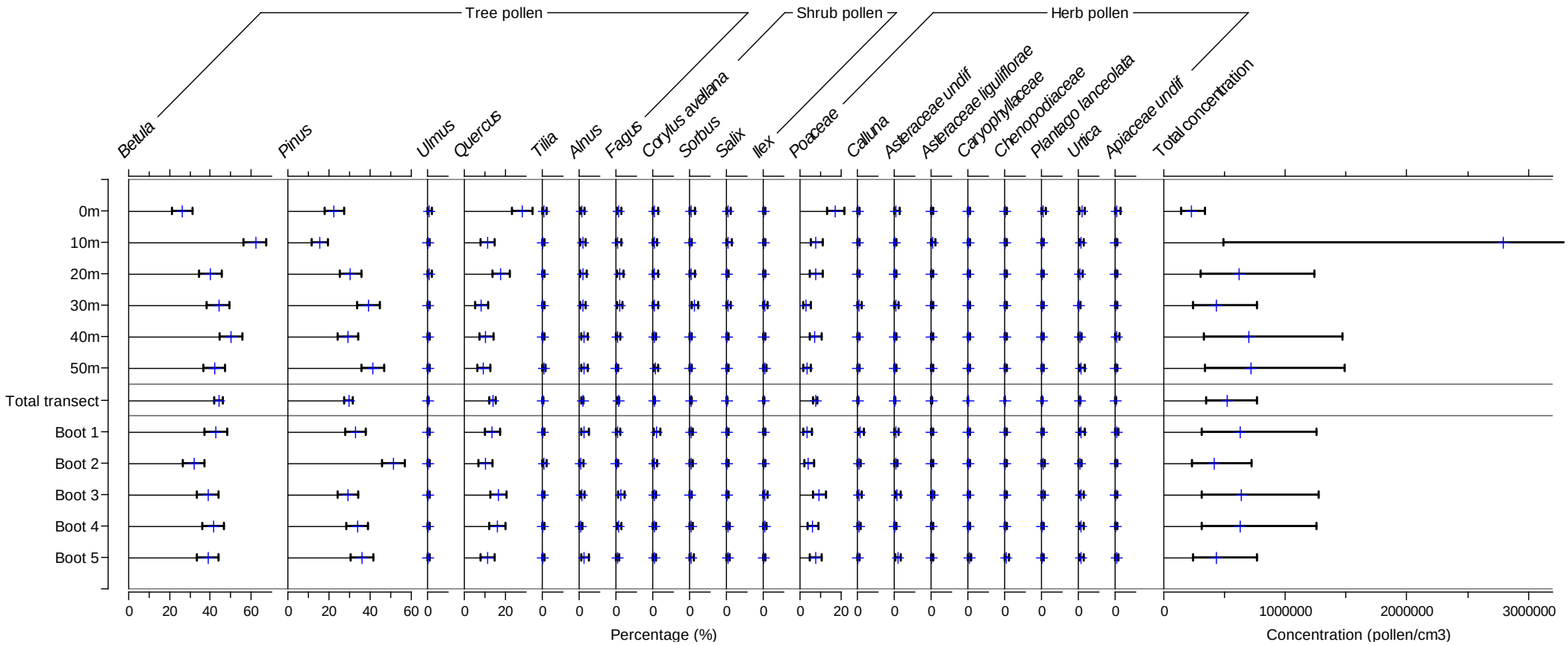

Figure 25. Overview of the relative data for each sample at the location Schaapsallee. The total concentration of pollen per cm^3 are also mentioned.

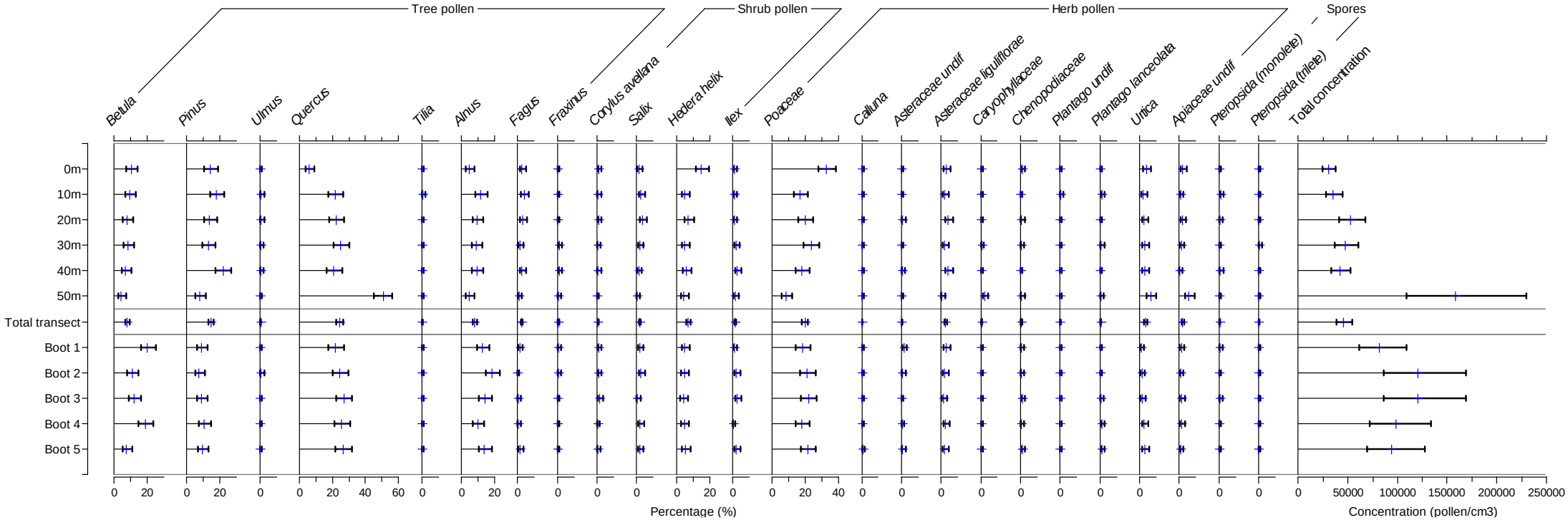

Figure 26. Overview of the relative data for each sample at the location Heemskerk. The total concentration of pollen per cm^3 are also mentioned.

Appendix E. Concentration data per location

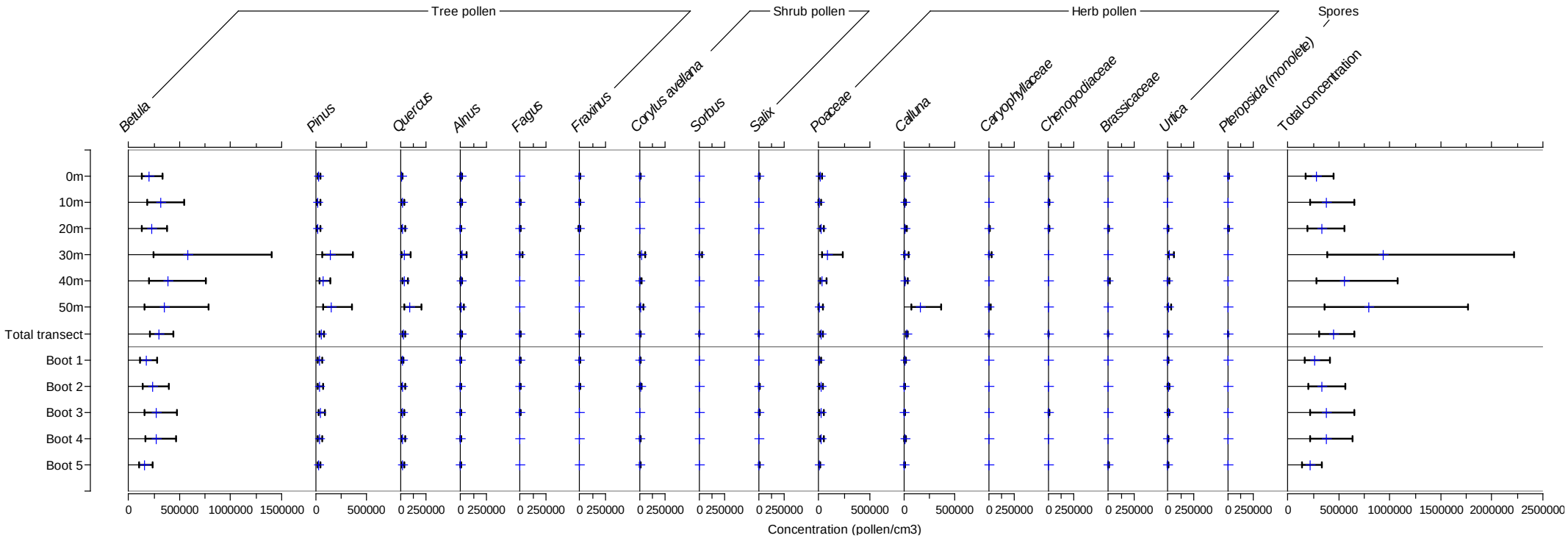

Figure 27. Overview of the concentration data (pollen/cm^3) for each sample at the location Bosrand.

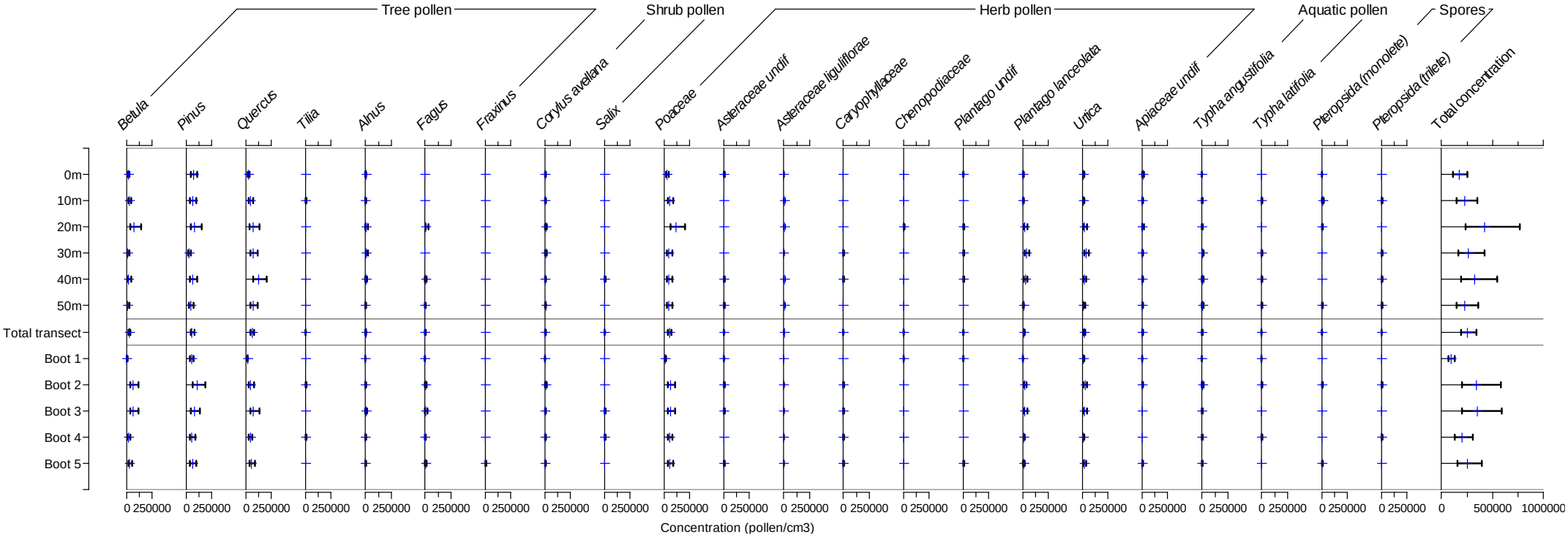

Figure 28. Overview of the concentration data (pollen/cm^3) for each sample at the location Utrechtse Heuvelrug.

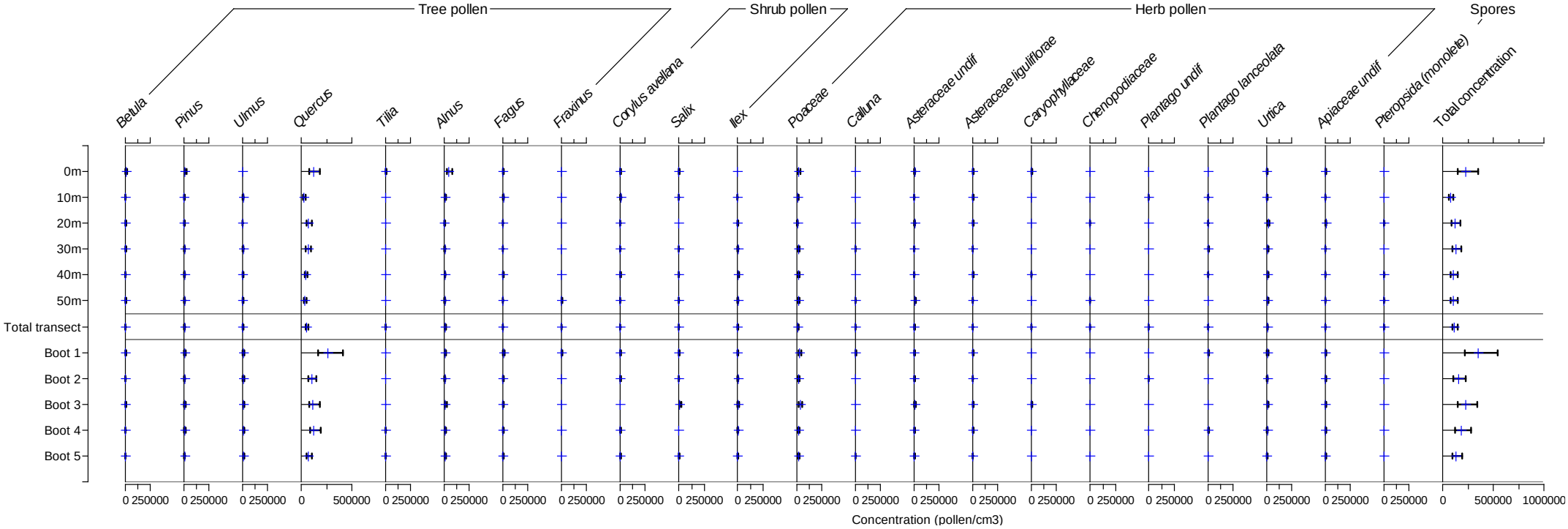

Figure 29. Overview of the concentration data (pollen/cm^3) for each sample at the location IJsselstein.

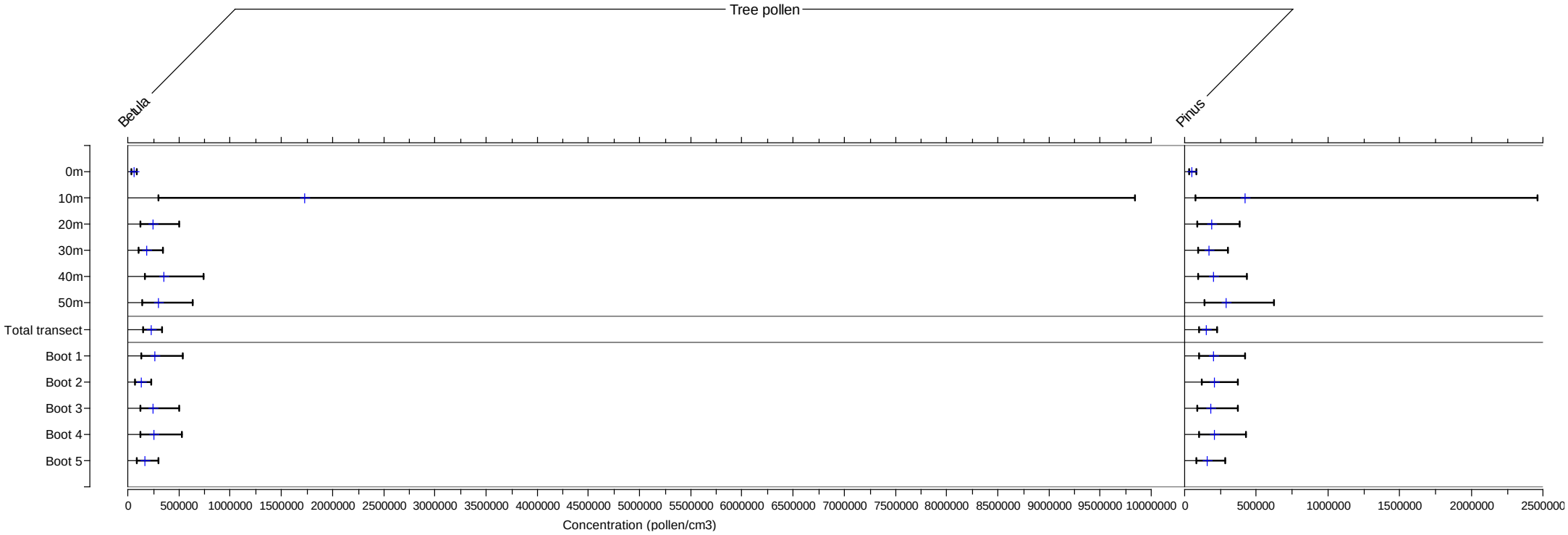

Figure 30a. Overview of the concentration data (pollen/cm^3) for each sample at the location Schaapsallee.

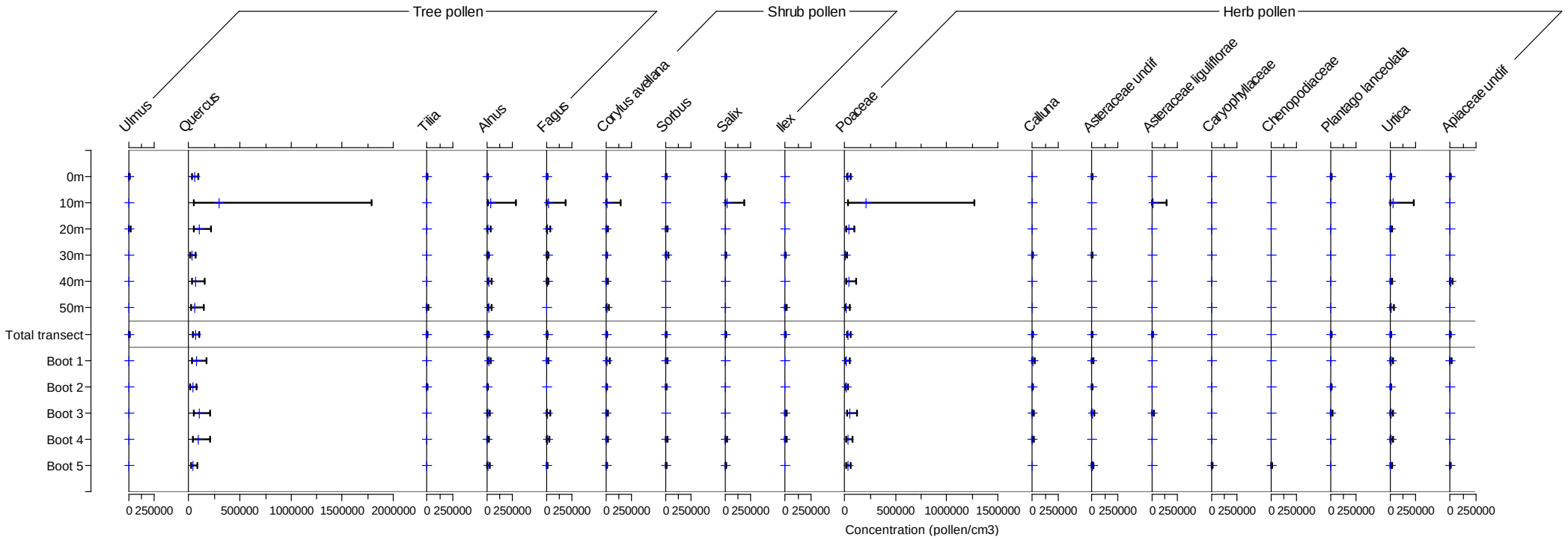

Figure 30b. Overview of the concentration data (pollen/cm^3) for each sample at the location Schaapsallee.

Figure 30c. Overview of the concentration data (pollen/cm^3) for each sample at the location Schaapsallee.

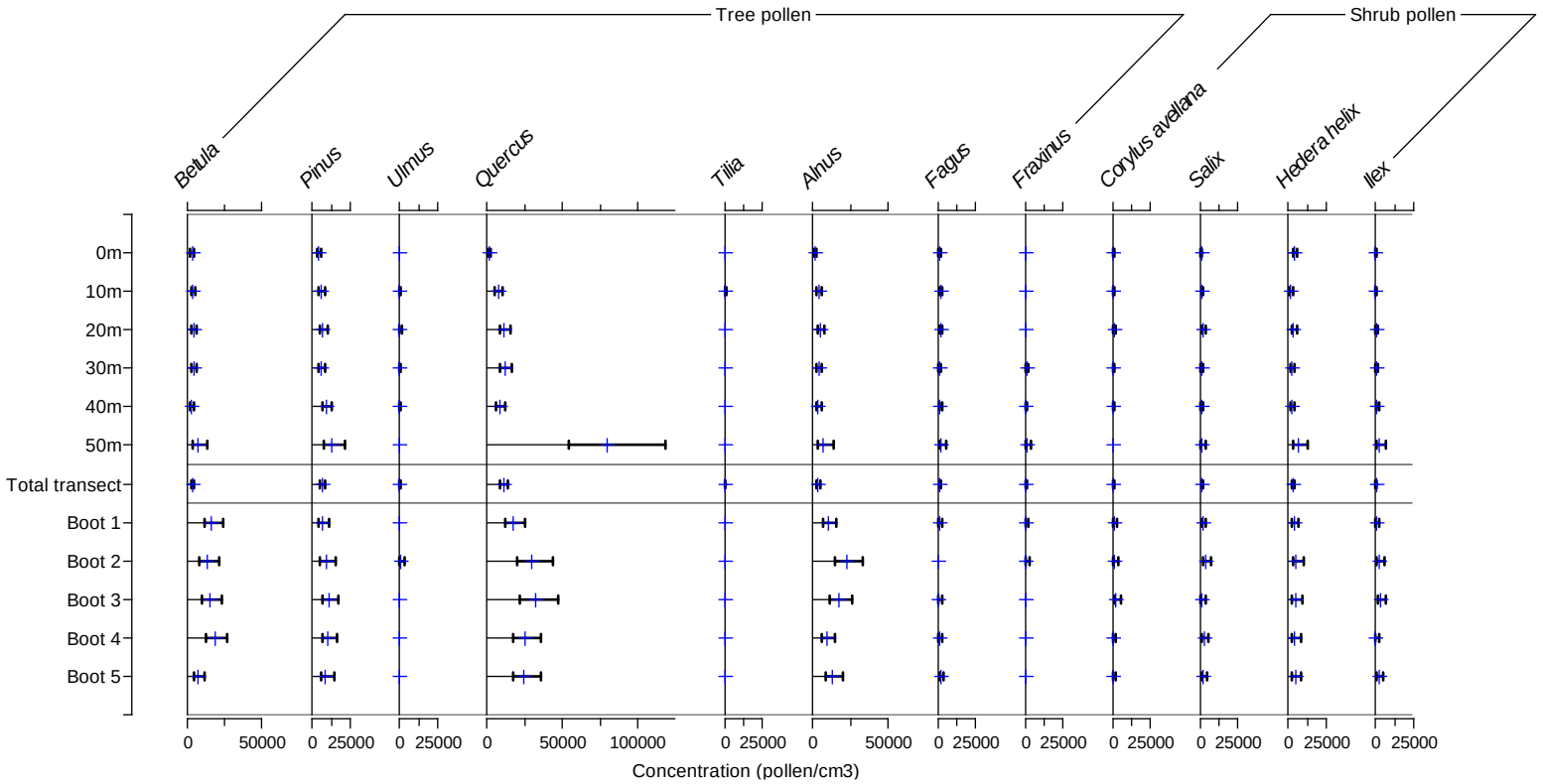

Figure 31a. Overview of the concentration data (pollen/cm^3) for each sample at the location Heemskerk.

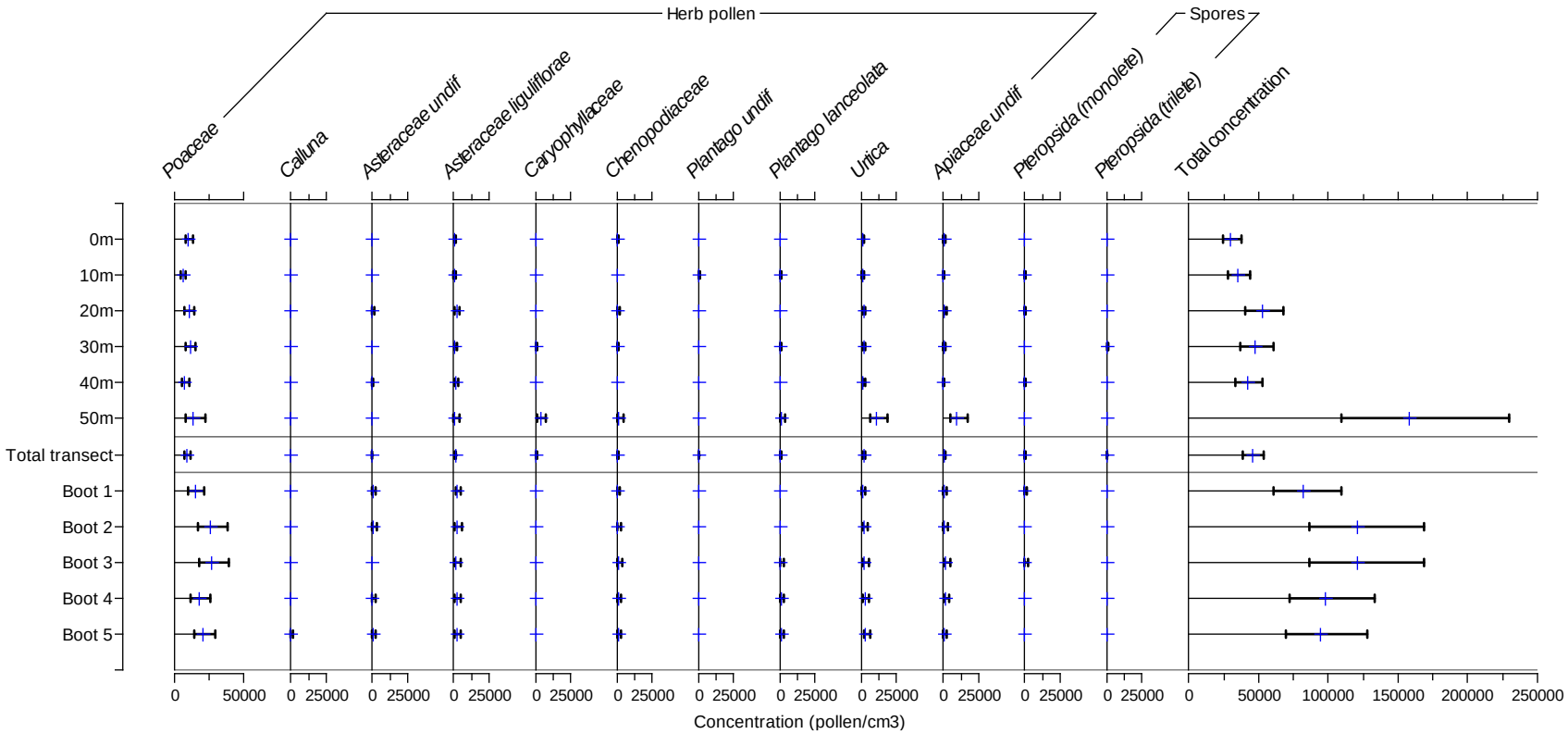

Figure 31b. Overview of the concentration data (pollen/cm^3) for each sample at the location Heemskerk.

Appendix F. DCA's with intrinsic variables

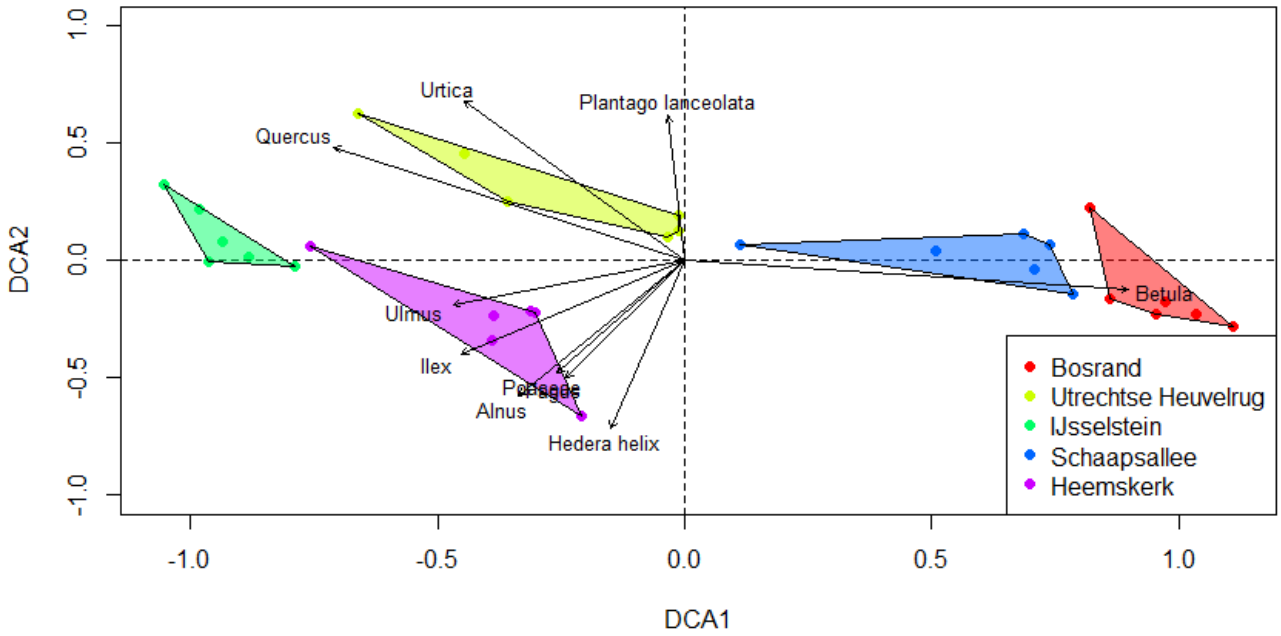

Figure 32. Scatterplot that represents the results of the DCA between the different soil samples. This figure only includes the data from the soil samples collected for this study. Each colour shows a different location. Each location has six different points which represent the places of where the soil samples were taken (0m, 10m, 20m, 30m, 40m and 50m). The intrinsic variables which have driven this distribution pattern are shown with arrows. Only the species which drive the pattern with a significance of p≤0.01 are included in this figure.

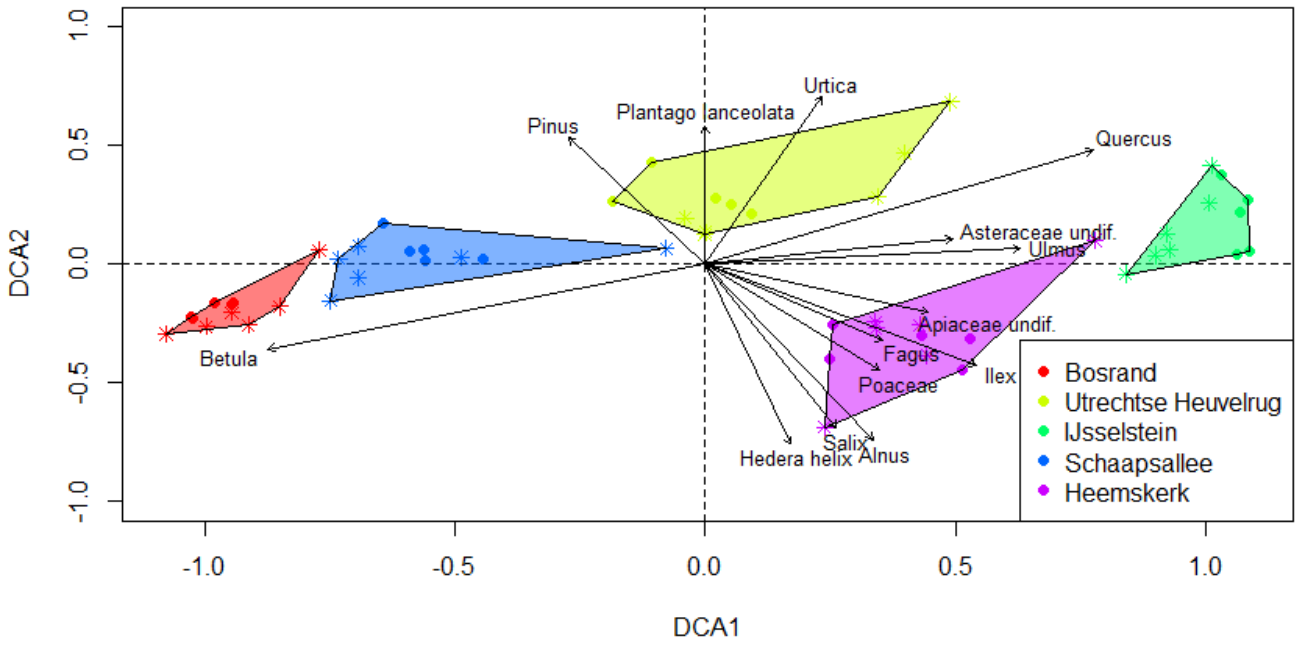

Figure 33. Scatterplot that represents the results of the DCA between all the soil (*) and boot (•) samples. This figure only includes the data from the soil and boot samples collected for this study. Each colour shows a different location. Each location has eleven different points which represent the places of where the soil samples were taken (0m, 10m, 20m, 30m, 40m and 50m) and the boots that were used (boot 1, 2, 3, 4 and 5). The intrinsic variables which have driven this distribution pattern are shown with arrows. Only the species which drive the pattern with a significance of p≤0.01 are included in this figure.

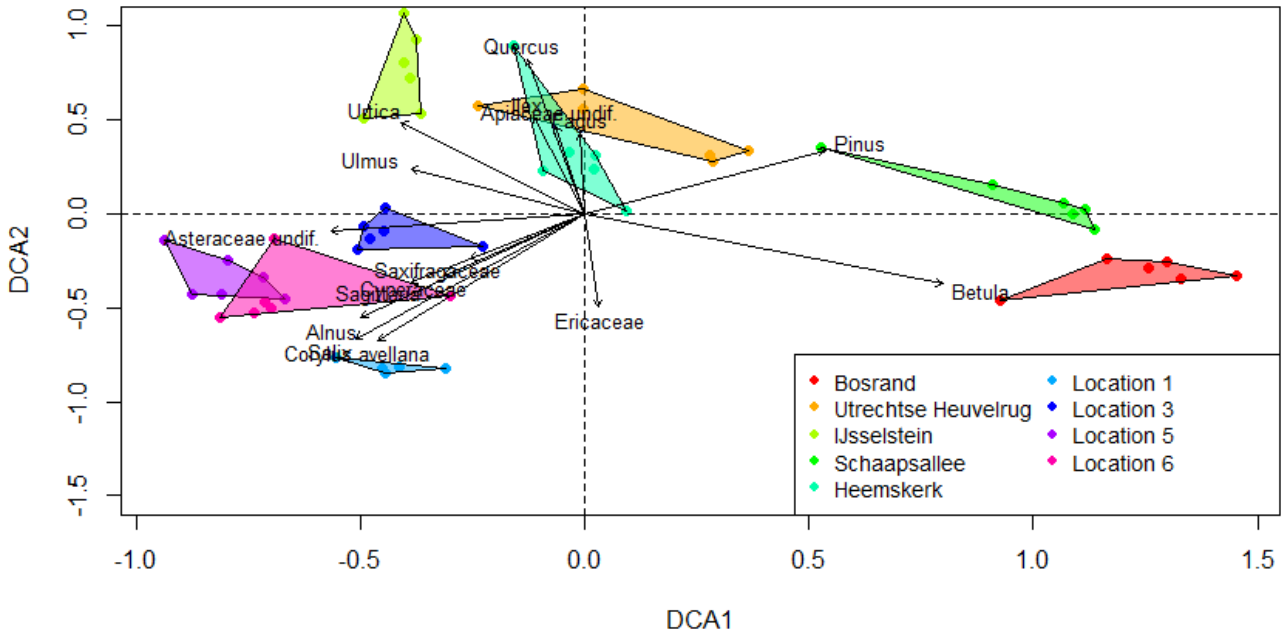

Figure 34. Scatterplot that represents the results of the DCA between the different soil samples. This figure includes the data from the soil samples collected for this study and for Van der Wal's study. Each colour shows a different location. Each location has six different points which represent the places of where the soil samples were taken (0m, 10m, 20m, 30m, 40m and 50m). The intrinsic variables which have driven this distribution pattern are shown with arrows. Only the species which drive the pattern with a significance of p≤0.01 are included in this figure.

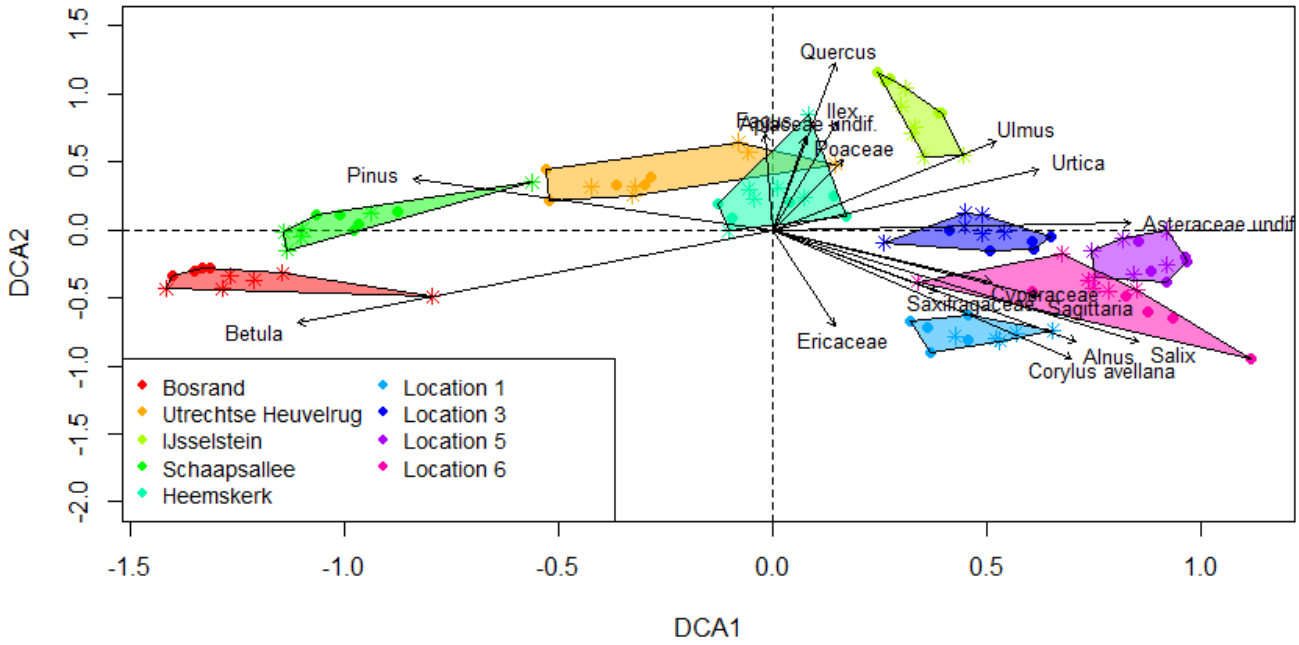

Figure 35. Scatterplot that represents the results of the DCA between all the soil (*) and boot (•) samples. This figure includes the data from the samples collected for this study and for Van der Wal's study. Each colour shows a different location. Each location has eleven different points which represent the places of where the soil samples were taken (0m, 10m, 20m, 30m, 40m and 50m) and the boots that were used (boot 1, 2, 3, 4 and 5). The intrinsic variables which have driven this distribution pattern are shown with arrows. Only the species which drive the pattern with a significance of p≤0.01 are included in this figure.