Utrecht University

The Palynology of Urban Environments

A new method for the classification of palynological distinct urban environments for forensic purposes

Author:	J.W.K. van Vliet
Student number:	3838366
Master Program:	Earth, Life and Climate
Supervisor:	dr. T.H. Donders
	S. Uitdehaag MSc

Date:	15-10-2018

Abstract

Pollen are used by both Earth scientists and forensic investigators. Both frequently cope with small quantities of pollen with an unknown source. Provenancing of samples to the source location can be a delicate business, especially in urban environments. Few data is available and the variation of pollen assemblages within an urban area is not yet readily evident. This research presents the data of pollen from various urban environments and provides the tools for the appropriate classification of samples with an anthropogenic background. This is done by deriving characteristics such as dominance and diversity from the pollen data using ecologic and statistic methods usually applied in climate studies. The characteristics are combined in a provenancing framework, which has both illustrative and quantitative components. Validation and improvement of the provenancing model occurs through testing with additional data from arbitrary locations provided by the Netherlands Forensic Institute.

The derived characteristics show differentiation between multiple palynological environments within urban areas and together they provide a first classification. When the pollen source is unknown, a multivariable approach shows the capability of confining the search area by excluding improbable urban environments. In some scenarios it is possible the indicate the most probable environment of the pollen source. An iterative approach by further extension of the database and enhancing of the model is recommended for more reliable classification and provenancing.

Keywords: Earth Science, Forensics, palynology, pollen, urban, classification, provenancing

1. Introduction

Forensic science recognizes palynology as a field with potential. While many successful case studies using palynology are readily available, the scientific literature lags behind with the discussion of the underlying scientific principles (Walsh & Horrocks 2008). Much experimentation and modelling is needed before forensic palynology can be used to its full efficiency. Multidisciplinary cooperation is essential (Mildenhall et al. 2006) just as enhancing the information flow on forensic palynology from scientists to the general public and crime fighters (Mathewes 2006). This thesis considers the pollen from anthropogenic and urban locations, still a 'terra incognita' in palynology. Presented are pollen assemblages along with a first attempt to the new classification criteria and a framework for the subdivision of urban areas in distinct palynological environments. This will provide guidance in the tracing of the pollen provenance in order to make palynology better applicable in forensic investigations. Forensic studies encompass multiple disciplines and will attract scientists with various backgrounds. Therefore, a brief introduction on pollen and forensic palynology is required.

1.1 Palynology

Palynology is the study of palynomorphs, which include pollen, spores and other organic microscopic particles. It is an interdisciplinary scientific field which collaborates with mainly biology, earth sciences and archaeology. Several successful applications of palynology are found in: taxonomy (Punt et al., 1976-2009), (paleo)botany, climate reconstruction (Davis et al. 2003; Sadori et al. 2016), diet reconstruction (Kuijper & Turner 1992), stratigraphy (Donders et al. 2007) and forensics (Mildenhall 2004).

Pollen are the male reproductive cells of seed plants, developed in the stamen. Through time and evolution all species of seed plants have developed their own unique pollen morphology. The pollen morphology is important for the process of pollination, the transport of pollen to the stigma, which eventually leads to fertilization and the production of a seed. In adaptation to their habitat, seed plants have developed several mechanisms of pollination since they can utilize water, wind or animals or have specialized in self-pollination (Faegri & Pijl 1979). In palynological research the variety in morphologies is considered an important aid in identification of species.

Abiotic pollination by water and wind is non-directional and requires large quantities of pollen for a chance of successful fertilization. Wind-driven pollination (e.g. used by grasses and many gymnosperms) is sometimes recognized by the presence of sacci or vesicles on the pollen (figure 1.1), which increases the buoyancy for further dispersal (Schwendemann et al. 2007). Plants that have specialized in pollination through insects, birds or mammals have a lower production of pollen as this mechanism is more efficient. For enhanced transport, these biotic pollinated seed plants developed flowers and have adapted their pollen morphology. The exine wall is often echinate (figure 1.2), which facilitates the pollen to attach to animals. Plants specialized in self-pollination require a minimum amount of pollen as the transportation distance of pollen to the pistil is much smaller.

The uniqueness of pollen morphology makes it an efficient tool in taxonomy and evolutionary research. An additional necessary property that makes pollen applicable for research is its rate of preservation. The sporopollenin exine walls of pollen are chemically very inert under anoxic conditions, which contributes to the high preservation level of pollen (Ariizumi & Toriyama 2011). Pollen have been traced back to Paleozoic sediments (Eyles et al. 2002). The oldest found palynomorphs date back even further, to the Precambrian; acritarchs have been found aging over

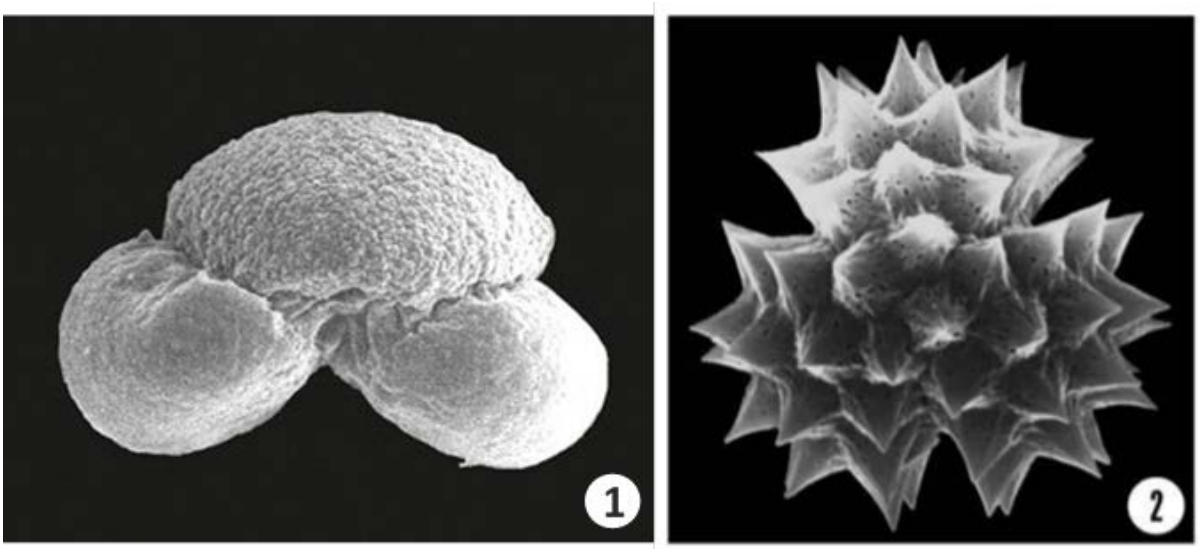


Figure 1: SEM images of 1) Pinus strobus, with clearly visible sacci and 2) the echinate pollen Helichrysum arenarium. From Punt and Hoen 2009; Schwendemann et al. 2007

1500 Myr (Yoon et al. 2004).

1.2 Forensic palynology

A rather underexploited application is the palynology related to crime investigations. Pollen are abundant, can be found on nearly all locations and are easily preserved. Therefore, forensic palynologists use pollen grains and spores to assist the fact finder in the court room. This can be achieved in several ways:

a) (dis)associate an object or person to a crime scene or victim
b) disprove an alibi
c) narrow down a search area
d) retrieve the place of origin (e.g. drugs, contraband, weapons, money,)
e) estimate exposure time or time of death

(Mildenhall 1990; Mildenhall et al. 2006; Adams-Groom 2012)

The use of palynology in forensic research started in the latter part of the 20th century. The first European documented use of palynology in forensics is a murder case in Vienna in 1959 (Erdtman 1969). One of the earliest cases in the USA is the investigation on honey fraud in the 70's. The pollen assemblages in honey provided evidence for the origin of the honey. Honey labelled and sold in the USA as 'domestic produced honey' turned out to be cheap imported honey from Central America (Bryant & Jones 2006).

A significant part of the available scientific literature on forensic palynology are reviews on the current status of forensic palynology per country (Mildenhall 1990; Bryant & Jones 2006; Mathewes 2006). Secondly the literature consists of case studies where pollen are used as conclusive or

supportive evidence. Examples include: A rape case where palynology gave information on the location of the crime scene (Wiltshire et al. 2014); the identification of 32 skeletons as soviet soldiers (Szibor et al. 1998); war crime investigations in NE Bosnia by Brown (2006).

1.3 Problem and research question

Forensic palynology occupies a very small niche within the forensic sciences. While palynology proved to have a high potential in criminal investigations it is not commonly used and the awareness is low. This underutilization can be attributed to several gaps in the palynological knowledge. A major problem is the lack of proper databases and reference collections (Mildenhall et al. 2006). The available pollendata insufficiently reflect modern anthropogenic created environments e.g. parks, road verges, vacant lots and gardens; where crime scenes often occur. This is because the palynological datasets originate from Quaternary research with native pollen of natural environments. This is in contrast to forensic palynology which has to cope with urban environments, numerous non-native pollen (Mildenhall 1990), mixed sources and disturbed environments (Wiltshire 2009).

The data collection on urban pollen and its availability in scientific literature needs to be increased to make palynology better applicable in forensic science. A dataset which includes the classification of different types of urban environments will contribute to this. This reference material, accompanied with proper distinctive features, can potentially be incorporated in an enhanced provenancing method. When the pollen source is unknown it is important to have a standardized method for the recovery of the source location.

The Research Questions for this thesis project are:

I. **Can palynology be used to distinguish between anthropogenic urban environments for forensic purposes, and which aspects of the data are hereby most informative?**
II. **Is it possible to construct a framework for the provenancing of pollen samples with an urban background?**

The general research approach to resolve these questions is the spatial sampling of multiple urban environments, followed by a pollen analysis.

There only is a single prevailing climate in the Netherlands, which is temperate maritime, Cfb according to the Köppen climate classification. Therefore, there is only a small gradient in natural conditions. This in contrast to the artificial gradient caused by human influences. An urban area is a heavily anthropogenic manipulated area and is all but homogenous. While the primary vegetation of a city may be focussed in parks and gardens, much urban nature may go unnoticed. Patches of green can be found on seemingly arbitrary locations such as road verges, vacant lots or construction sites. Flora in the Netherlands however is carefully managed and for every location different vegetation is maintained. This leads to the hypothesis that within an urban area multiple urban environments can be distinguished with pollen. To test this, the sampling sites have been chosen based on their presumed palynological distinctness. It is thought to be distinct if the level of anthropogenic influence is different. Five of such typical urban locations are sampled: an old park, a new park, several gardens, a vacant lot and roadsides verges. The exact number of palynological distinct urban environments is unknown. It is however possible to think of more possibly distinct areas such as a construction site or a meadow. The scope of this research limits the number of tested environments. The five sampled environments are selected on basis of their accessibility and their occurrences in

crime investigations. In addition, a natural reference site (a forest) is sampled, which is expected to differ from the urban environments. All sampling sites are located in the same region (Utrecht) to avoid background noise caused by regional differences. The region of Utrecht was selected as sampling area for logistic reasons and for its well-documented tree population that can be consulted via the *Bomenkaart* (City of Utrecht). It is hypothesized that all locations are palynological distinct and that the combination of palynological characteristics are indicative for a specific urban environment. It is necessary to collect multiple samples at all six locations. This ensures an average composition of a location plus the associated ranges of pollen abundances. Many crime scenes are typically restricted to only a few square meters. Palynology in forensic research can only be a suitable tool if the palynological composition within the crime scene is homogeneous (Horrocks et al. 1998). The homogeneity of a sampling site will be tested with clustered sampling in the Old Park.

The first objective for the pollen analysis is the determination of the pollen assemblages. These assemblages are obtained through palynological processing in the lab and microscopic analysis of the collected samples. In addition, nine multivariable methods and characteristics are analysed: the pollen percentage diagram, Principal Component Analysis (PCA), cluster analysis, environmental-specific species, diversity, AP/NAP-ratio, spore percentage, concentrations and dominance. A combination of these features allows the palynological classification of urban environments.

The classification will be used for a new provenancing method: a classification framework including a scoring mechanism based on the computed variables. To review the effectiveness of the framework the scoring mechanism is applied on the original data and furthermore on 10 independent samples. The framework can then be improved wherever necessary. The working theory is that accuracy of the provenancing shall increase with the extension of the database and adjustments in the scoring mechanism until the database reflects the true characteristics of urban environments. This research should be seen as a first start for the construction of an urban provenancing database.

2. Material and Methods

2.1 Sampling strategy

In order to address the research questions it is necessary to take into account the forensic aspect of the sample material. A recurring sample type is soil preserved in a shoe profile. With this in mind surface samples were collected (Horrocks et al. 1998). Another possibility would have been a subsurface, which could represent cases associated with digging, e.g. graves.

The sampling of a location was achieved by scraping of the surface layer of an approximately 10x10cm patch until about 10 grams of soil was collected. The grid was carefully chosen as it must be a representative part of the surroundings and be consistent with the other sampling grids. For forensic purposes soil with much botanical remains were avoided because they limit the possible transport of soil by a shoe profile. A bare soil surface was thus preferred. All sampling locations were saved in a GPS device and furthermore a vegetation survey was conducted in the field. This survey primarily contains the record of all nearby trees within a ~30m radius. As the samples were collected during winter few herbs are included in the vegetation survey (some ornamental bulbous plants were present). The vegetation record was further completed with the data available via the 'Bomenkaart' (City of Utrecht). This database contains the record of the approximate 160,000 trees present in the area of Utrecht.

Urban environment	Sampling Location	Abbreviation	Colour code
Old Park	Wilhelminapark	Wil	Blue
Young Park	Maximapark	Max	Purple/Pink
Gardens	Various households	Tui	Red
Vacant lot	Minnaert, Uithof	Min	Brown
Road verge	Weg tot de Wetenschap	WtdW	Yellow
Forest	Amelisweerd	Ame	Green

Table 1: Overview of the urban environments with their corresponding associations

2.2 Locations

Between February and March 2016 from 6 locations a total of 36 samples were collected and 28 were eventually microscopically analysed. The aim of this research is to make palynological statements on the differences between urban environments. Table 1 gives an overview of the selected locations and presumed environments. For convenience, the corresponding abbreviation and colour code were added. These features reoccur in the plots and figures. The sampling locations can also be found on the map shown in figure 2. In addition, the vegetation record is added as an Appendix A.

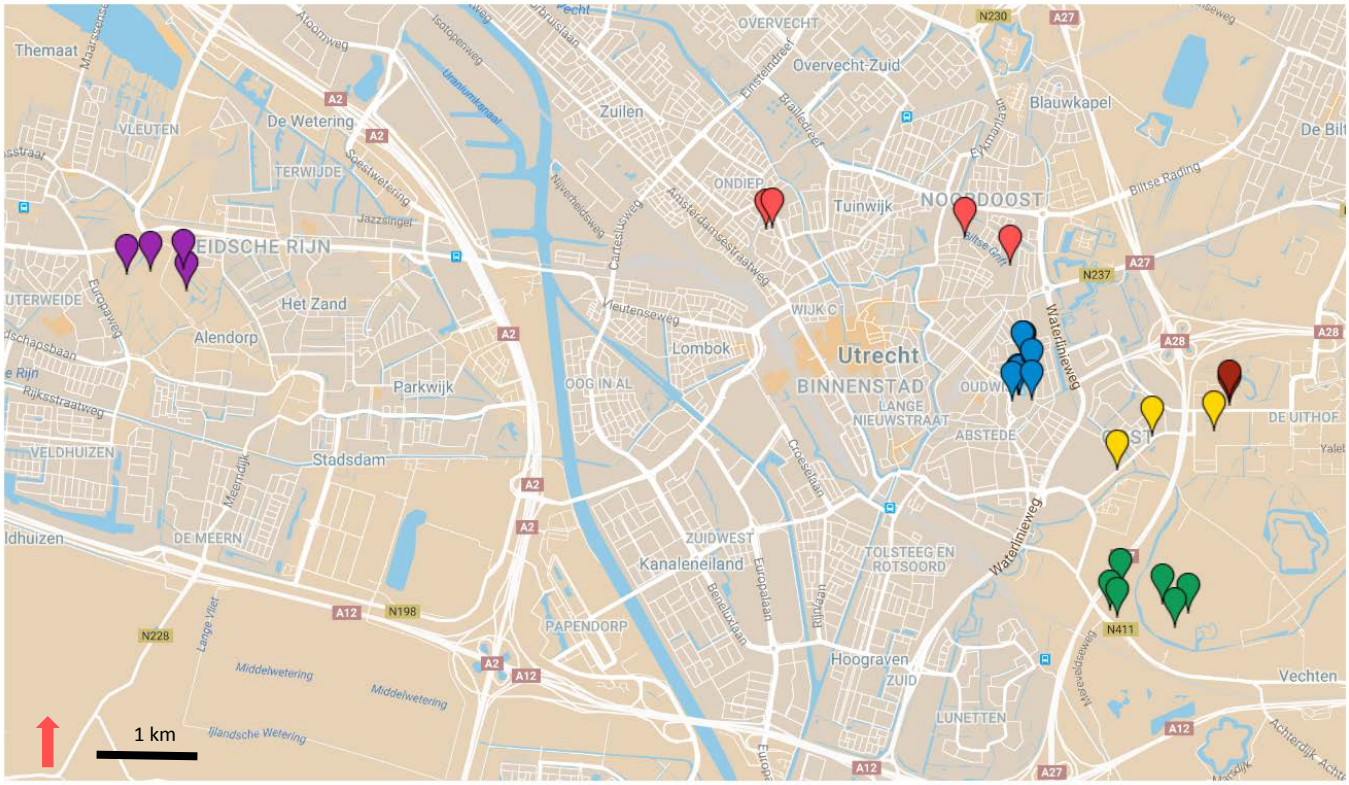


Figure 2: Map of Utrecht containing the 28 sampling sites. The colour indicates the urban environment which is in accordance with table 1. (blue: old park, purple: young park, red: gardens, brown: vacant lot, yellow: road verge, green: forest) Original image from Google maps

2.2.1 Old Park

The Wilhelminapark originates from the late 19th century and was designed as an English landscape garden by H. Copijn (Schackmann & van Rossum 2005). It became a national heritage site in 2001 (*Rijksmonumentenregister*). The vegetation is very well-documented and the flora management policy is described in a recurring report (van Berghem et al. 2009). This policy includes an annual check-up of every individual tree. There is high diversity of trees in the Wilhelminapark, both native and exotic. A majority of the trees is at least several decennia old. Due to strict maintenance, there are few possibilities for the growth of younger trees. The term old park therefore directs to both the age of the park and the maturity of the trees within. The available vegetation records, the maturity of the vegetation, the age of the park and the central position in the city make the Wilhelminapark a good sampling location for this research.

Figure 3: Map of the Wilhelminapark. In green, the main park is indicated. Furthermore, indicated are the 9 sampling sites. The sampling order for the clusters from west to east: A2, A1, A3 & B1, B2, B3. Original image from Google Earth

In February 2016, 9 samples were collected for microscopic pollen analysis (visible in red in figure 3). They were collected in two clusters (A & B) and three separated sites (C, D & E). Both clusters consist of three

samples collected within a radius of 10 m. Cluster A is situated in centre of the park, while cluster B is situated in the very north of the park at an approximate distance of 300 m from A. The remaining samples C, D & E are taken in the directions SW, NE and E of centre cluster A. The chosen clustering was performed to analyse the internal variation of the Wilhelminapark and to compare this variation with the external variation with the other presumed urban environments. It is expected to find very few palynological differences within a single cluster as the sampling sites are within a ~10m radius. Between the sampling sites A-E more differences in the pollen assemblages are expected as the vegetational diversity in the park is high with at approximately 100 heterogenous distributed arboreal species according to the Bomenplan (van Berghem et al. 2009). However, the nine derived characteristics and exploration techniques are predicted to show significant coherence between the sampling sites within the park; especially in contrast to the characteristics of sampling sites outside the park. The external variation is expected to be high for both the pollen assemblages and the derived characteristics.

2.2.2 Young Park

The Maximapark is a very recent park; it was officially opened in 2013, though some parts are still under construction. It is located in the western part of Utrecht, in the residential area Leidsche Rijn (figure 2 & 4). All samples are collected in the *Binnenhof*, the centre part of the park. It is designed to become a classical urban park. In contrast to the Wilhelminapark the diversity of trees is lower and less exotic with much *Platanus, Populus, Tilia* and *Fagus* present. With an average age of <15 years (most trees were planted between 2003 and 2006) the trees are not mature, in particular in comparison with the Wilhelminapark. It is expected to take at least 20 years before the park reaches a more mature state (City of Utrecht 2014).

Figure 4: Map of the Maximapark. The park is shown in dark green. The Binnenhof, the sampling area, is shown in light green. The sites of the four analyzed samples are shown in red. Original image from Google Earth.

Ten samples were collected, though only four (3, 5, 6 & 10) were eventually selected for microscopic analysis. The samples are not clustered, but rather evenly distributed across the park. This was done in order to capture the average vegetation composition of the Maximapark, which is needed for the broader comparison between urban environments. Max 3 is the most varied site with much *Salix* and a combination of *Fagus, Populus, Prunus, Quercus and Tilia.* The site Max 5, next to a small pond, is close to multiple

Platanus and *Populus* trees, while Max 6 is surrounded by young *Fagus* trees. The trees in proximity to the site of Max 10 are mainly *Tilia*.

2.2.3 Gardens

Gardens are a major part, though very diverse aspect, of the urban nature in a city as it is less influenced by management policies. Five arbitrarily chosen gardens, with varied vegetation coverage were selected for sampling (figure 2). All gardens were approximately between $10m^2$ and $30m^2$. Wherever possible the samples were taken at the centre of a garden. Permission was granted before sampling the gardens. Four of the garden samples were eventually analysed. Sample 1 and 2 originate from the same street (Cornetstraat) but are approximately 50 m apart. The first, Tui 1, has a rather high vegetation density due to two large trees (one is *Larix*). Furthermore, some space is reserved for the growth of some vegetables. The second garden is tiled and further completed with some beds of flowers, plants and shrubs. The other two samples were collected at respectively the Griftkade and the F.C. Dondersstraat. The garden at the Griftkade is situated near a small waterway accompanied with *Platanus xhispanica* (also: *xacerifolia*) from 1940 (*Bomenkaart*). The garden has some common exotic plants like *Hydrangea*, *Rhododendron*, *Buxus*, *Vinca* and *Narcissus*. The fourth garden has *Ranunculus ficaria* and *Hibiscus syriacus*. Few trees can be found in the adjacent street: *Tilia, Platanus* and a single *Gingko biloba.*

2.2.4 Vacant lot

Vacant lots can be recognized by either ruderal vegetation or neglected patches of green. These habitats are typically represented by vegetation such as *Chenopodium, Plantago, Rumex, Cirsium, Poaceae* and *Polygonum*. Two samples have been taken at the university campus representing an area of unutilized wasteland with no vegetation. In early 2016 the many *Pterocarya* trees were chopped down. Currently only several *Tilia* adjacent to the Padualaan remained and four other trees (not determined). It must be mentioned that during the summer of 2016 the area transformed into a much more vegetated area. Further important notice is that the site is in close proximity (~150m) to the botanical gardens of the university.

2.2.5 Road verge

Road verges can have a varied vegetation composition based on their setting, e.g. next to a grand avenue or a neglected, abandoned road. Four samples have been collected from the verge of the main road towards the university, the *Weg tot de Wetenschap*. The road was newly reconstructed in 2015. The first sample, closest to the campus, has no nearby trees and the only vegetation consist of grasses. Sample 2 and 3 are collected more towards the city centre and are in close proximity to several backyards. The tree composition alongside the road consists of *Fraxinus, Tilia* and *Ulmus.*

2.2.6 Forest

Nieuw-Amelisweerd is a large estate accompanied with gardens, a forest and farmlands. The numerous monumental trees from the sampled part of the Amelisweerd originate from 1765-1900 (Maes et al. 2009). The forest has a much higher vegetation density than the other sampled locations. The forest is mainly composed of *Quercus* and *Fagus* further completed with *Fraxinus* and some *Ulmus* and *Acer*. There is however some variation in the ratio between the mentioned trees at the different sampling sites (figure 5). Sampling site Ame 1 and Ame 3 have an even distribution, Ame 2 has more *Quercus* and *Fagus.* This in contrast to site Ame 4 which has mainly *Quercus* and *Fraxinus,* while site *5* is dominated by *Fraxinus*. *Fagus* is dominant at sampling site Ame 6.

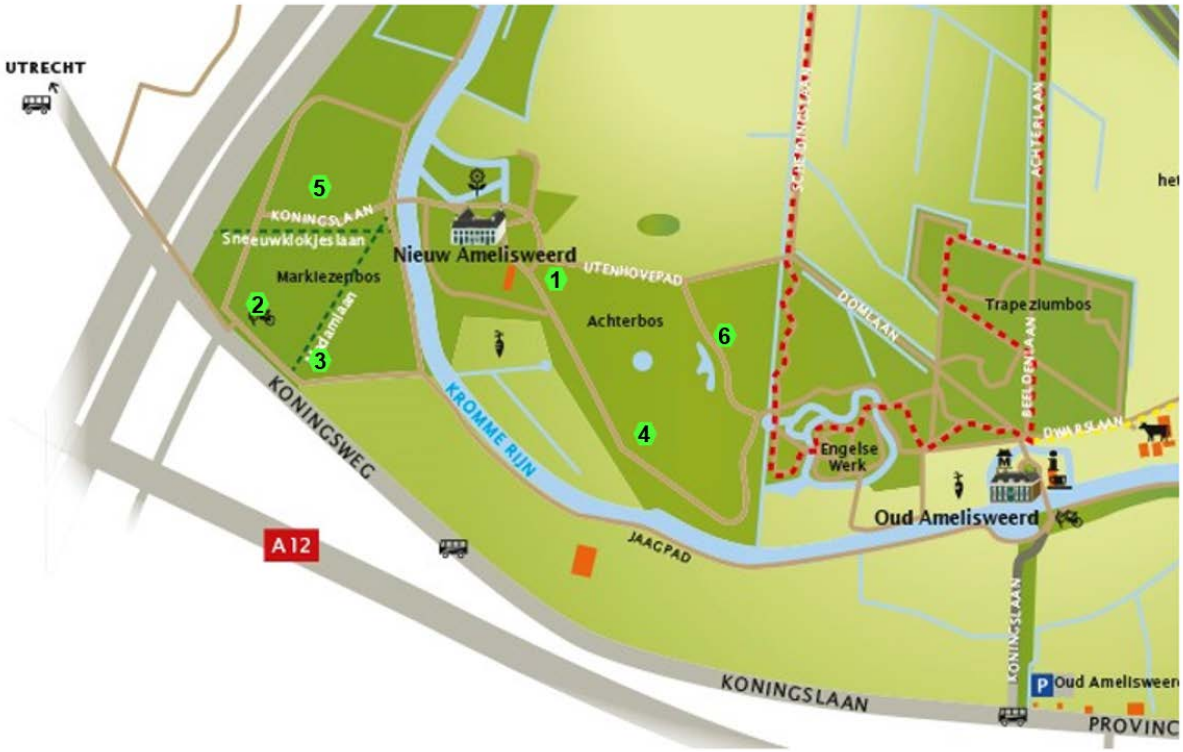


Figure 5: Sampling sites in the forest of Amelisweerd. Original map from the website of the City of Utrecht

2.3 Palynological processing

For the determination of the pollen assemblage of a sample the pollen must be separated from the bulk soil. This is achieved using the characteristic properties of pollen. Such characteristics are composition, size and density. The detailed protocol can be found in Appendix B. The protocol is similar to the protocol used at the Netherlands Forensic Institute, with some minor adjustments to conform to the guidelines of the method applied at the University Utrecht. Both protocols involve decalcifying, the addition of *Lycopodium* spores, acetolysis and heavy liquid separation. Differences are found in the used volumes, centrifuge rotation speed and the sequence of decalcifying, addition of *Lycopodium* spores and the sieving process.

After weighing the samples (~1.4g; also, Appendix C) the first step is decalcifying. 8 ml *Lycopodium*-solution with known number of spores is added in order to allow calculations on concentration (Stockmarr 1971). After removing the carbonates, KOH (5%) is added and the sample is heated at 70° C to remove the humic acids. The particles <7 micron and >250 micron are then excluded by sieving. The next step is acetolysis, a method to remove organic matter (e.g. cellulose, lipids). A beneficial consequence is the darkening of the pollen, which promotes the identification (Erdtman 1960). The acetolysis is followed by heavy liquid separation. Sodium-polytungsten with a density of 2.1 g/cm^3 is used to sink most minerals. The remaining material now resembles the main characteristics of pollen as it has a similar size, density and resistance to chemicals. Between all proceedings (multiple) centrifugation steps at 1700 rpm is applied. As a final step the microscopic slides are made with a basis of glycerol instead of paraffin, which improves pollen determination due to increased mobility.

2.4 Microscopic analysis

The microscopic slides were analysed using a light microscope (Leitz Diaplan) with magnification of 200x-1000x. On each slide, at least 300 pollen were determined. Fern and moss spores (no fungi) were determined but are not included in this counting as they are excluded in the percentage sum of the pollen diagram. Counting continued if one pollen type was anomalously high until the 300 was reached without including the abundant pollen-type. For reliable concentrations, an additional condition was taken into account: the number of counted *Lycopodium* spores. Counting continued if the number of lycopodium spores was too low (<20). A counting >100 *Lycopodium* spores was preferred for more reliable concentration calculations. Frequently consulted literature for pollen determination includes: Leitfaden der Pollenbestimmung (Beug 2004), Pollen Analysis (Moore et al. 1997) and The Northern European Pollen Flora I-IX (Punt et al.). The pollen data was collected in a spreadsheet which is supplemented as Appendix D. For convenience both scientific names and common English names are shown.

3. Data treatment

The data acquired in the spreadsheets can be analysed in various ways. In order to be able to distinguish between the presumed environments, nine multivariable methods and characteristics have been derived from the pollen data.

- Pollen percentage diagram
- Cluster Analysis
- PCA
- Environmental specific species
- Diversity
- AP/NAP-ratio
- Spore percentage
- Concentration
- Dominance

3.1 Pollen diagram

The pollen data is commonly visualized in a pollen percentage diagram. Tilia 1.7 provides a suitable tool for the graphing of the palynological data. The sorting is based on the type of vegetation: trees/shrubs, herbs/bushes, aquatics and spores. The totals of these groups are supplemented next to the diagram in combination with the pollen concentration. All species which make up at least 1% of the pollen percentage sum are individually shown in the diagram. The remaining pollen species are stacked in 'other' groups. Furthermore, some pollen species are stacked in 'undifferentiated' groups. Reasons for this grouping are a high degree of morphologic resemblance or a close genetic relationship. The pollen sum includes all pollen. The plant spores are not included in the pollen sum, but their quantities are instead shown as percentage of the pollen sum. This explains values >100% in the ratio diagram.

3.2 Cluster Analysis

A cluster analysis creates a dendrogram which link samples, based on their percentage sum, according to their distance indices. There are numerous varieties of cluster analyses. The algorithm used for the dendrogram is UPGMA (Unweighted Pair Group Method with Arithmetic mean). In this research, a Bray-Curtis distance metric is used. This similarity index is both asymmetric and quantitative, which is appropriate for pollen data. Symmetric measures include the evaluation of absent species, this in contrast to asymmetric measures which ignores absent species. The latter is preferred for pollen data as the absence of a pollen type in samples does not necessarily indicate a common origin. A quantitative analysis has the preference over a binary analysis as some pollen are frequently present though in significant different abundances. The analysis is conducted using PAST, which uses the equation by Bray and Curtis (1957).

3.3 Principal Component Analysis

In this research with 28 samples a total of 84 pollen types, 6 spore types and an undetermined group are distinguished. This generates a 91-dimensional dataset. Using an ordination technique, it is possible to summarize and highlight trends of such multivariate dataset. There are multiple

ordination techniques available for palynological data such as principal component analysis (PCA), correspondence analysis (CA) and detrended correspondence analysis (DCA)(Shi 1993).

For this research, a Principal Component Analysis (PCA) is selected as there is sufficient overlap between the samples and the number of samples is limited. For this technique, the data is transformed to percentages. The PCA reduces the multivariate data to principal components, which are hypothetical variables covering a portion of variance within the data. If the data can be reduced to two variables, it will allow the visualization of the data in a PCA-plot. It is important to notice that the most important components correlate to their underlying variables. To do so it focuses on the pollen types which cause the largest differences between the pollen assemblages. Pollen types which have a high abundance in some samples and an absence in other samples will have a high variance. This in contrast to those species which occur evenly in all samples and cause few variations between the samples. The pollen types with more variation will have more influence or loading on the principal components. There is a component for every dimension (pollen type), but the first few principle components take up most of the variance.

It is possible to plot the sample assemblages in a PCA plot in which the x- and y-axis are composed of two principal components. The sample assemblages can be plotted using their principal component score which is based on the loading and the abundance of each species i present in the sample:

$$Sample\ PCscore = \sum_{i} loading_i \times abundance\ percentage_i$$

The PCA plot results in a representative visualization of the data with a minimum loss of information. It produces a data point for every sample, those with a high degree of similarity are clustered together. The principal component analysis was conducted using the software PAST.

3.4 AP/NAP-ratio

The degree of vegetation cover in a landscape can be very variable. A major aspect involving the openness of a landscape is vegetation type, the ratio between trees + shrubs and herbs + shrubs in particular. Calculating the ratio between arboreal pollen grains and nonarboreal pollen grains is a commonly used method to distinguish between some environments. Since the introduction of the AP/NAP-ratio (Faegri & Iversen 1964) its meaning as a paleoclimatic index and as a vegetation cover index has been intensively discussed. In general, the trees are overrepresented in the pollen assemblage. The patchiness and the size of the trap have influence on the sediment accumulation. Following Favre et al. (2008) it is possible to neglect variation in pollen production. Under the same climatic conditions, the pollen assemblages should be influenced similarly by the heterogeneity in pollen productivity. In this research, no normalization is needed as the AP/NAP-ratio does not have to reflect the 'true distribution' of the vegetation. It is merely a method to distinguish between urban environments based on the pollen assemblages. Favre et al. (2008) furthermore states that slight variation in the ratio do not make much sense in respect to vegetation cover. Larger changes however do indicate differences in vegetation cover.

3.5 Pollen concentration

The concentration of pollen is a variable that could give information on the density of the vegetation or give an indication of the local pollen production. The absolute quantity of pollen and spores in a sample is therefore another variable that could distinguish between the environment within an urban area. The pollen concentration of a specie i (C_i) is calculated using the following equation:

$$C_i = \frac{n_{pollen}}{n_{Lyc}} \times \frac{C_{Lyc} \times V_{Lyc}}{m_{sample}}$$

This equation consists of the counted number of palynomorphs of specie i (n_{pollen}), the counted number of *Lycopodium* in the sample (n_{Lyc}), the added concentration and added volume *Lycopodium* spores (C_{Lyc} & V_{Lyc}) and the mass of the sample (m_{sample}).

3.6 Dominance

The distribution of species in a community can be measured using the dominance (1-Simpson index), which gives an indication for the dominancy of species (Simpson 1949). The software of PAST v3 (Hammer et al. 2001) uses the following equation:

$$D = \sum_i (\frac{n_i}{n})^2$$

Where n is the pollen sum including the spores. n_i represents the number of pollen or spores for species i.

This index results in a value between $1/n$ and 1. A value of $1/n$ means that all species are equally present within the sample. The value of 1 means that the sample consist of a single pollen/spore type. Important to notice is that this analysis is done on the pollen and that no correction has been applied on the pollen production of a species. This dominance therefore does not represent the real vegetation distribution.

3.7 Diversity

The diversity of the vegetation is one of the characterising features of an environment. The most direct method to calculate the diversity is the species richness. This represents the number of species found in a sample. This method for reconstructing the (paleo)biodiversity is not without controversy. There are multiple underlying biases (Odgaard 2007):

- Dominance: higher dominance causes the loss of less abundant species and result in a lower richness
- Resistance to decay: a relative low content of sporopollinin, causes faster decay of pollen and therefore a lower richness
- Sample size: Increasing the sample size results in a higher species richness and vice versa. This bias can be solved using rarefaction
- Taphonomy: the preservation of pollen depends on the taphonomic conditions such as soil type, soil moisture and climatic conditions
- Taxonomic precision: The level of taxonomic determination differs among pollen. Some can only be identified on family level, possibly lowering the species richness

The biases cause difficulties in reconstruction the true (paleo)biodiversity. The true diversity is therefore only partially reflected in the pollen composition. Still, this characteristic can be used for classification and provenancing as the mutual comparison of samples and the database is much more important than its value as diversity proxy.

In statistics, there are multiple types of diversity available. In this research, the diversity of a sampling site and the diversity of a sampling location are calculated. The number of species found in a single sample reflects the sample-diversity and represents the local diversity. Unfortunately, this is affected by the size of the count. Using individual rarefaction, it is possible to estimate how many species would have been found if the pollen sum was smaller, for example the size of the smallest pollen sum (C1 = 252). The software PAST for paleontological statistics uses the rarefaction algorithm from Krebs (1989).

$$E(S_n) = \sum_{i=1}^{s}\left[1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}}\right]$$

In this equation $E(S_n)$ is the expected number species and *s* reflects the total number of species. N is the total number of individuals and Ni the number of individuals for species i. The sample size is given with n. The algorithm consists of a ratio of two binomial coefficients, which reflects the probability of a specie to occur in the counting. The probability of each specie is then summed to result in an expected number of species at a specific sample size.

There is also a location-diversity. This reflects the total number of species found within the samples of a single location, therefore representing a broader diversity than the sample-diversity. Unfortunately the number of samples differs per location. Therefore sample-based rarefaction is applied towards the smallest sample number, which is 2 (the vacant lot). The analysis available in PAST, 'Mao's Tau' sample rarefaction, uses Equation 5 from Colwell et al. (2004) and is based on the presence-absence of species in a sample.

$$\tilde{\tau}(h) = S_{obs} - \sum_{j=1}^{H} \alpha_{jh} s_j \qquad \alpha = \begin{cases} \frac{(H-h)!(H-j)!}{(H-h-j)!H!} & for\ j+h \leq H \\ 0 & for\ j+h > H \end{cases}$$

This equation on the estimated richness, $\tilde{\tau}(h)$, involves the total number of observed species S_{obs} and a sum of presence probabilities. This summed probability is composed of a combinatorial coefficient α_{jh}, the total number of species present only in j samples (S_j), the number of samples (H) and the number of samples (h) for which the richness should be estimated. As $\tilde{\tau}(H) = S_{obs}$, the combinatorial coefficient α should be 0 for $h = H$. The derivation and a more detailed explanation is available in Colwell et al. (2004)

3.8 Environmental-specific species

In forensics environmental-specific and rare species are often a key factor, the presence of a single species can potentially be the breakthrough in an investigation. This because a rare species can have a large, case-specific, potential by directly referring to a site, object or person. In a murder case in Wales for example where *Juglans* pollen could be linked to a location where 80 years earlier the only known specimen of a Walnut tree in that area was cut down. (Mildenhall et al. 2006). In this research, environmental-specific species are those species that are either exclusively found in one

single sample or are found in multiple samples retrieved but from a single environment. The hypothesis is that some species require such specific growth conditions that they are limited to specific (urban) environments. Furthermore, it is thought that some exotic species only occur in the most anthropogenic environments, e.g. more exotic species are expected in the garden than at a road-side verge.

3.9 Spores

Spores contain valuable information for the distinction between samples. It is necessary to again mention that the definition of spores for this research only include the spores from ferns and mosses. No determination is done on the spores from fungi. To avoid absolute abundances the spore data should be transformed to true percentages. This differs from the percentages used in the pollen diagram, which focuses on pollen. Spores were therefore excluded from the pollen sum; they are the reproductive units of cryptogams and fungi instead of seed plants (spermatophyte).

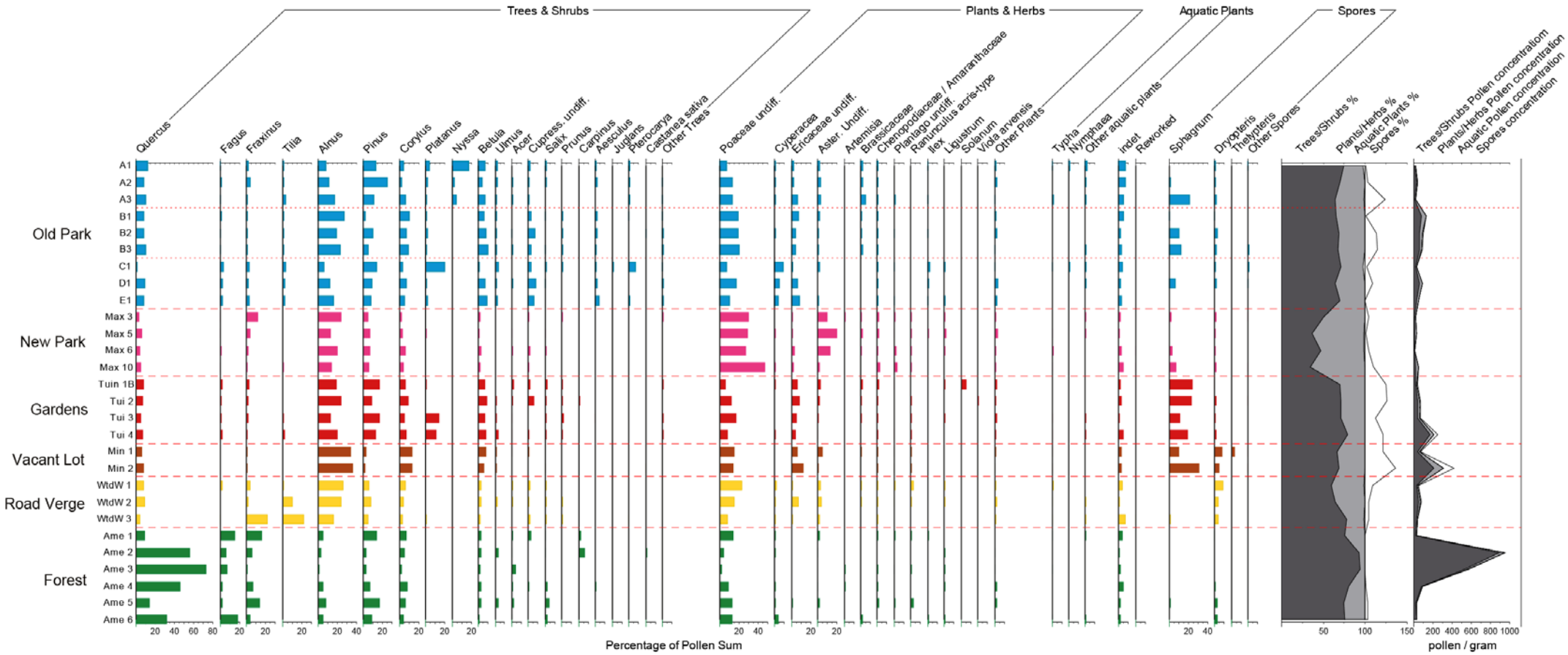


Figure 6: Urban environments of Utrecht, Pollen percentage diagram of selected taxa (>1%) per location; vegetation ratio and concentration. The different locations are grouped by colour in accordance with table 1 and are sorted on the Y-axis. A dotted line separates the clusters made in the Wilhelminapark. In horizontal direction the identified pollen types are shown.

4. Results

During the lab process it was expected to find only low concentrations of carbonates in the samples. The samples of the Maximapark however reacted strong with the HCl, just as Tui 2. The carbonates in the Maximapark originates from the soil used for the construction of the park.

4.1 Pollen diagram

The visualization of the pollen data in a pollen diagram can be seen in figure 6. Based on the pollen diagram the following remarks can be made:

- The Wilhelminapark contains many arboreal species; *Aesculus, Pterocarya, Nyssa* and 'other trees' are mainly limited to the old park. In appendix D the other arboreal species present in the Wilhelminapark can be found.
- Within the clusters A and B in the Wilhelminapark the S*phagnum* percentage is fluctuating.
- The Maximapark contains a wide range of plants and herbs; grasses and *Asteraceae* have high abundances in the young park.
- In contrast to the urban locations the forest of Amelisweerd has large quantities of *Quercus, Fagus* and *Fraxinus pollen* while being low on *Alnus.*
- The vacant lot is the only location where no pollen from aquatic species are found. A vacant lot is in general rather 'vacant'.
- The road verge and the forest contrast the other locations for their absence in spores of *Sphagnum*.
- Pollen from vegetable plants can be found in gardens.

Details on the latter columns containing ratios, spores and concentration are pointed out at their respective paragraphs

4.2 Cluster Analysis

The dendrogram in figure 7 shows Bray Curtis sample distances using a UPGMA algorithm. From these distances, the point of within-group variation is set at a similarity of 0.725. The value was computed by making a histogram using the distribution of the distances. At 0.725 a minimum splits the plot in two groups (Appendix E). Above this value the internal variation is lower than the external variation, meaning that the linkages should come from samples of one location. Below this value the linkages do not necessarily represent environmental matches, based on these samples.

The forest samples split into two clusters separating Ame 1 and Ame 5 from the other samples. The same is true for the road verge for which WtdW 1 and WtdW 2 are clustered but separated from WtdW 3. Most of the samples from the Old Park show internal resemblance. B1 and B2 are the samples with the highest degree of similarity. The main cluster consists of the B-group together with D and E. The others (A1, A2, A3 and C) are outliers though A1 and A2 seem to be closely related. Sample A3 is included in the garden cluster while C is a very isolated sample. The vacant lot and the young park both form their own cluster in the dendrogram with consistent internal resemblance.

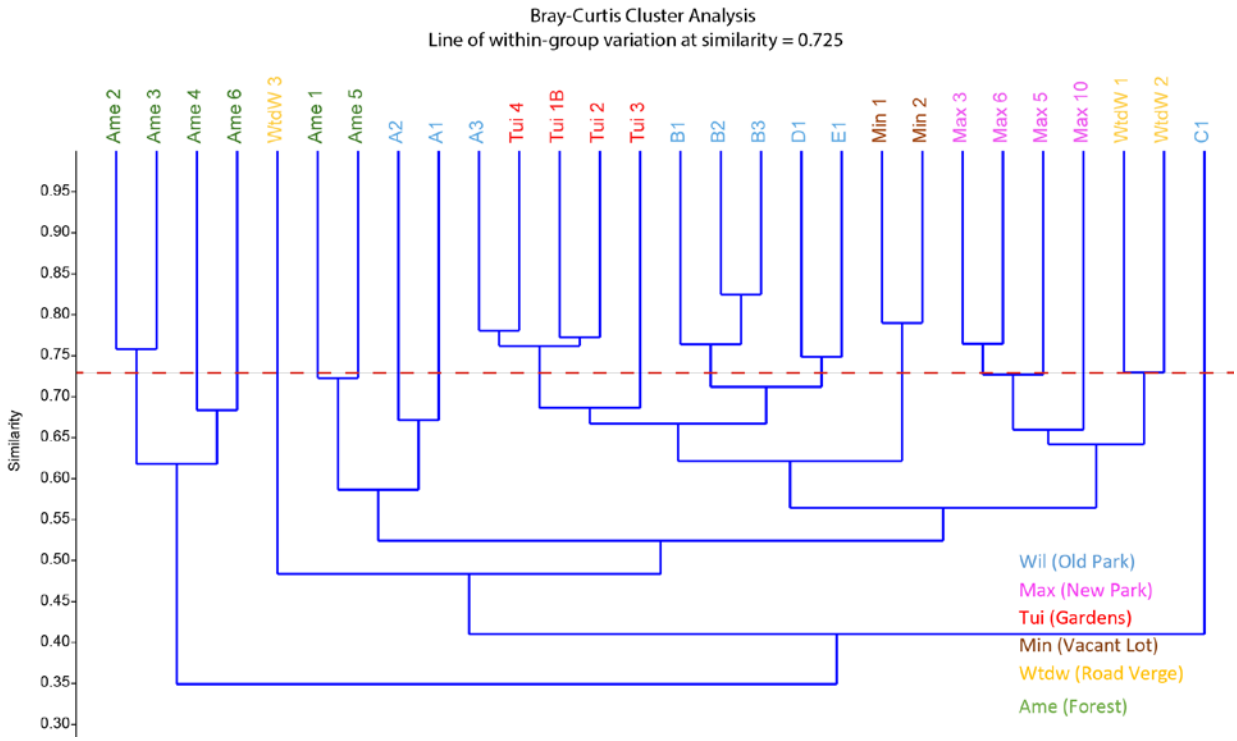


Figure 7: Dendrogram computed from the Bray-Curtis distance indices.

4.3 PCA

The major part of the variance is covered by the first 3 components. PC1 consists of 53% of the variance while PC2 and PC3 take up 13.2% and 11.7% of the variance. The main positive loading of the first component is *Quercus*; the negative loadings are primarily *Alnus, Poaceae* and *Sphagnum*. PC2 is mainly positively influenced by *Poaceae* and *Asteraceae Liguliflorae* and negatively by *Sphagnum.* PC 3 has strong positive loadings for *Pinus, Fraxinus* and *Platanus* while it has negative loadings for *Alnus, Quercus* and *Sphagnum.* The PCA-plot in figure 8 consist of PC1 and PC2. This is preferred over a combination with PC1 and PC3 because the latter causes much overlap of most of the sampling locations. The combinations of PC2 with PC3, which would eliminate the possible overrepresentation of *Quercus,* is not preferred because that plot would only show 24.9% of the total variance.

The PCA-plot of PC1 and PC2 (figure 8) shows a clear separation of the locations. The two principal components show 66% of the total variance, with component one consisting of 53% of the variance and component 2 only 13%. The most significant difference between the locations is therefore visible on the PC1-axis (the x-axis in figure 8). This results in the division between the urban locations and the natural reference site. The forest is located on the positive side of the x-axis. This is the direction which is mainly composed of *Fagus* and *Quercus*.

Another clear distinction is the isolation of the new park from the other urban environments, primarily caused by the second component. The new park situated in the top left corner of the plot, in the direction mainly influenced by *Poaceae* and *Asteraceae*. The samples in the bottom left corner resemble samples which are rich in *Sphagnum* spores and many arboreal pollen. These remaining data points are separated in the vacant lot, gardens and the verge/old park group. The verge and the old park are the only locations that are indistinguishable based on this PCA-plot. The overlap of the

verge with sample B1 is partly because of the lack of *Sphagnum* in B1. The sample of WtdW3 from the road verge lies within the old park cluster. This is because the sample contains much more arboreal pollen than the other samples along the road verge.

For verification of the correct ordination method, other ordination techniques have been run. Both CA and DCA strongly resemble the pattern as seen in the PCA-plot. The gradient of the DCA is

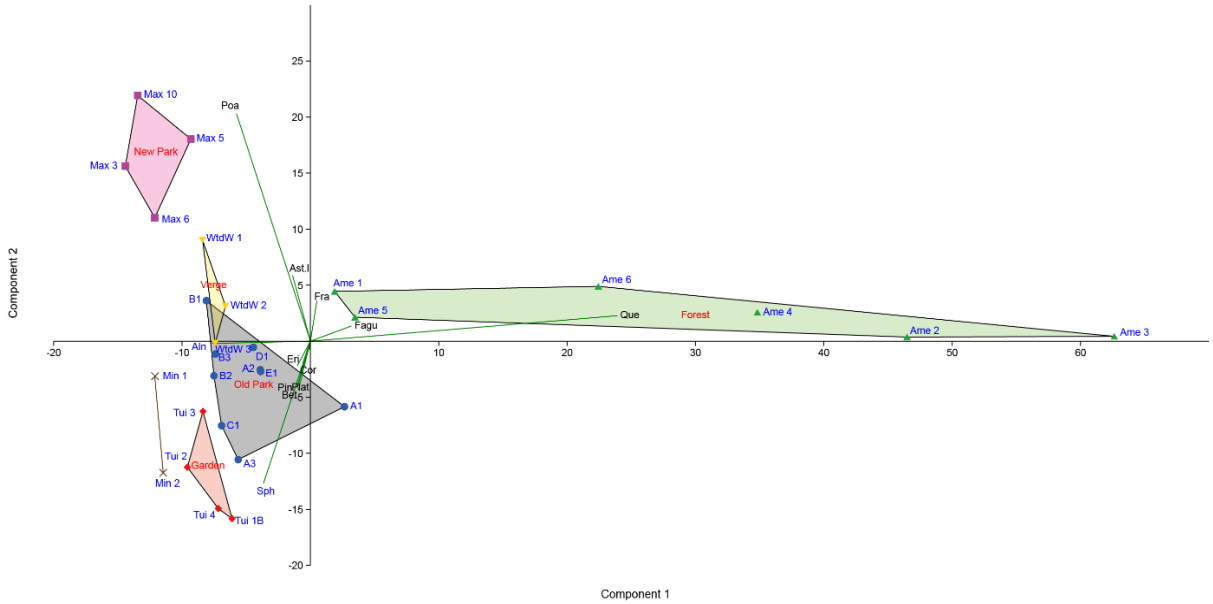


Figure 8: The PCA plot with 53% of the variation given by PC 1 and 13% by PC2. Shown are all analysed samples and the pollen types which cause the most variation.

2.24.

4.4 AP/NAP-ratio

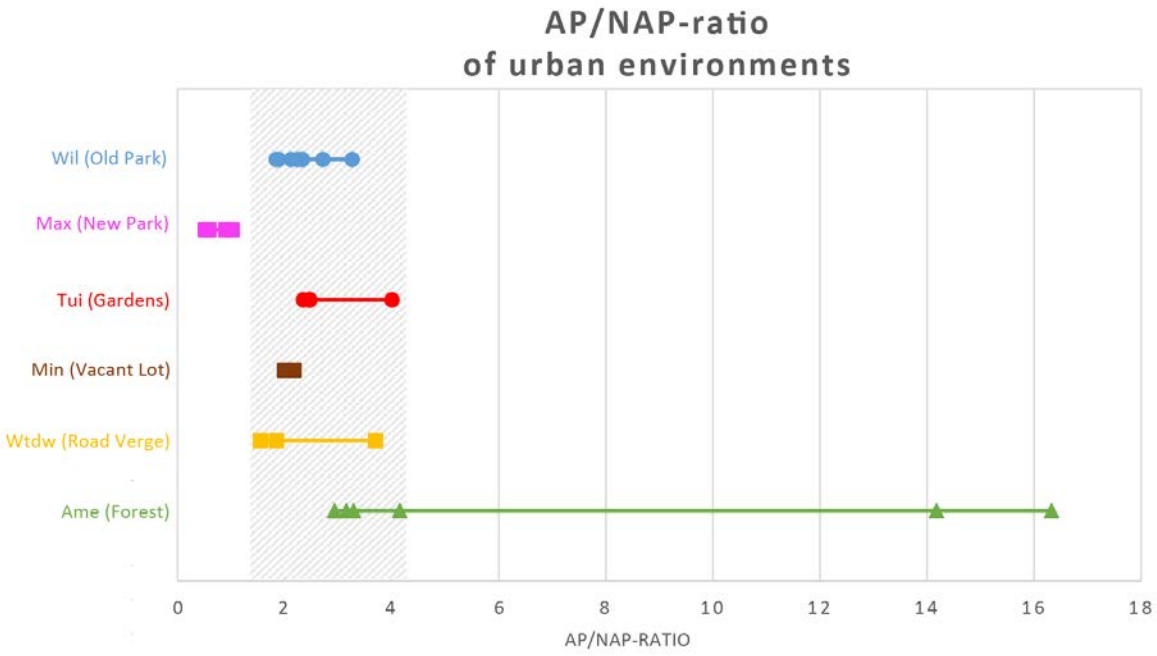


Figure 9: AP/NAP-ratios per urban environment

The results from the AP/NAP-ratios are visible in figure 9. It becomes clear that the AP/NAP-ratio is rather constant for all but two locations: The Maximapark and a part of the Amelisweerd. The young Maximapark deviates by having a ratio <1, visible in the pollen diagram as low percentages of arboreal pollen and higher percentages of plants and herbs. The other distinct deviation is the high arboreal percentage in some forest samples. The Amelisweerd has two samples which have a significant higher AP/NAP-ratio than all other samples: 14.2 and 16.3. This can be explained by the very high quantities of *Quercus* pollen found in sample Ame 2 and Ame 3. The other samples, all with a positive ratio, show significant overlap, which is visible in the figure as the dashed area.

4.5 Pollen concentration

The pollen concentration was calculated using *Lycopodium* spores. A very low number of *Lycopodium* causes a much larger uncertainty in the calculated concentration of those samples. Three samples might be affected by a low *Lycopodium* counting. In sample Tui 4 at a total pollen sum (incl. spores) of 443 only 21 *Lycopodium* spores were counted. Only 11 *Lycopodium* spores were counted in both sample Ame 2 and Ame 3 at a pollen sum of respectively 881 and 519.

The majority of the samples have a pollen concentration <100,000 pollen/gram. However, the overall average concentration of the analysed samples is 120,000 pollen/gram. Figure 10 shows that two forest samples have a much higher concentration relative to the other samples, 950,000 and 565,000 pollen/gram. This coincides with the high (*Quercus)* dominance found in the same samples, Ame 2 and Ame 3. Next to the forest some above average values can be found at the gardens and the vacant lot. All sampled locations show significant overlap in the concentration range <100,000 pollen/gram.

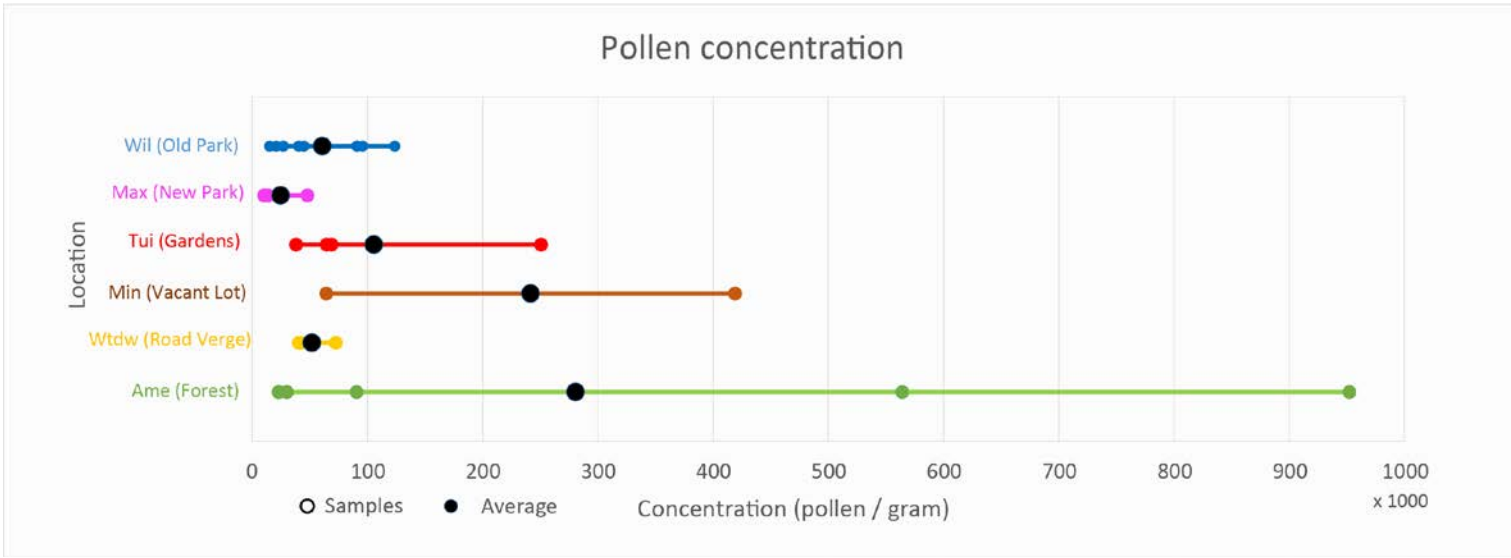


Figure 10: Ranges of pollen concentration per urban environment with indicated average.

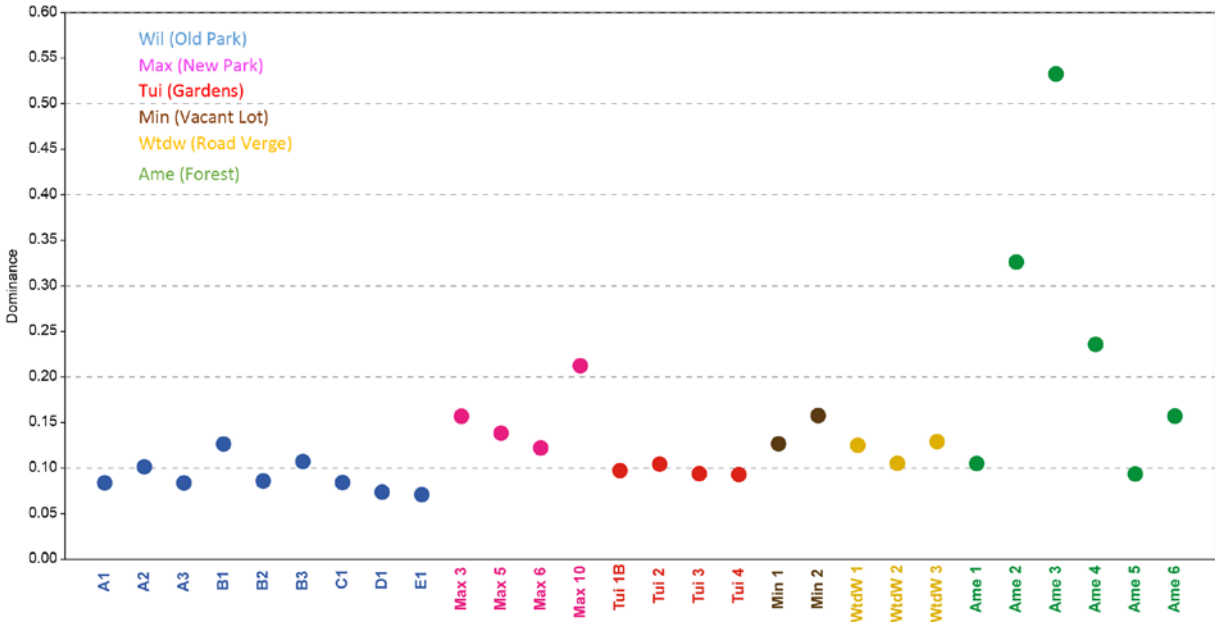


Figure 11: Dominance plot per urban environment

4.6 Dominance

The distribution of individual pollen taxa among the species is visualized as the dominance in figure 11. Most of the analysed samples have a dominance between 0.07 and 0.18. A rather low value which means that the individual pollen taxa are rather evenly distributed amongst the determined species. Multiple outliers in dominance can be found in the forest, with values of 0.33 and 0.53. As a consequence, the internal variation in Amelisweerd is high; between 0.09 and 0.53. The gardens have the least internal variations with values of approximately 0.10. The lowest dominance, therefore the most equally distributed composition, is found in the old park. This is closely followed by the samples from the gardens and the road verge.

4.7 Diversity

The highest diversity in a sample is determined in the samples C1 (old park) and Tui 1B (garden), as visible in figure 12. On average, the highest sample diversities are found in the old park, although the young park, gardens and the road verge do not have much different values for this sample diversity. Approximately 30 species have been determined per sample. The vacant lot and the forest have a lower diversity. The location-diversity shows a similar trend as the sample-diversity. It can be noticed that the highest average sample-diversity is found at the old park, while the highest location-diversity is found at the New park.

Both (average) sample-diversity and location-diversity split the locations in two environmental groups. The low diversity group consist of the vacant area and the forest; having a sample-diversity of ~22 and a location-diversity of ~33. These locations also have the largest internal variation in sample-diversity. The high diversity group consists of the parks, gardens and road verges; have a sample-diversity of approximately 30 species while the location-diversity is ~45. Although two major environmental groups can be distinguished the individual sample-diversity shows some overlap between the two groups.

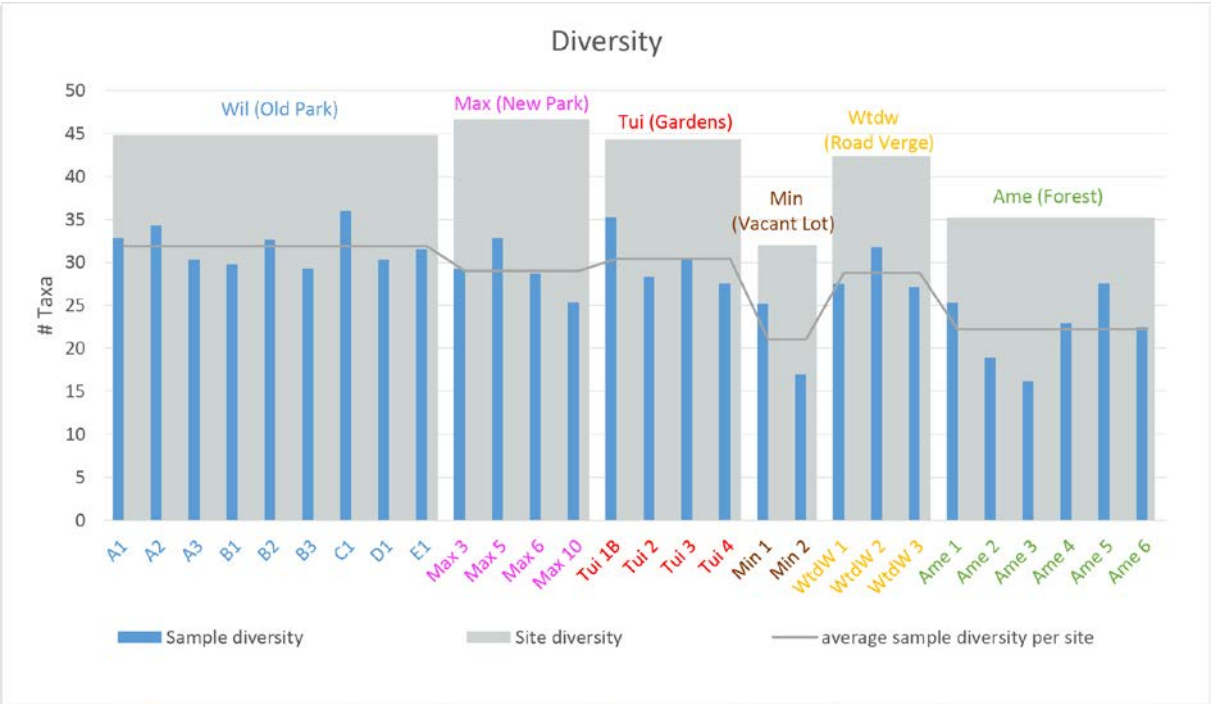


Figure 12: Diversity plot showing both α- and β-diversity; also shown is the average α-diversity.

4.8 Environmental-specific species

In the analysed samples, there are several species which are limited to a single environment (figure 13 & 14). Most of these species are restricted to only one sample, while 6 species are found in multiple samples of one location. Scanning a sample for the species of figure 14, could give an indication for the presumed environment.

There are 7 species which have been found in all samples: *Alnus, Betula, Corylus, Fraxinus, Pinus, Poaceae, Quercus*. Except for the P*oaceae* they are all arboreal pollen. Other frequently occurring species are *Asteraceae liguliflorae, Chenopodiaceae, Dryopteris, Fagus, Tilia* and *Ulmus*. Only the combination of multiple of these species and their abundance or percentage could provide indications for a (mis)match with an environment, comparison should then occur via the pollen diagram.

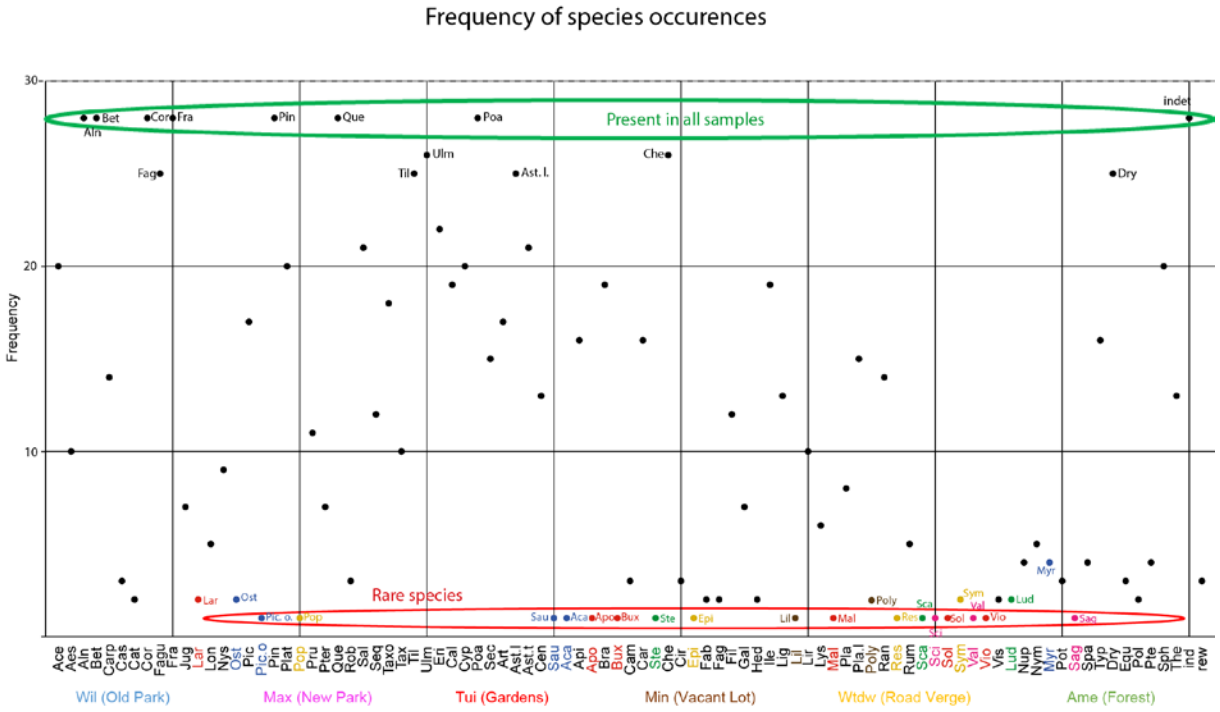


Figure 14: Plot showing the frequency of pollen occurrences. Some species are common in all samples while other are environment-specific or are only present in a single sample. Colours are indicative for their environment.

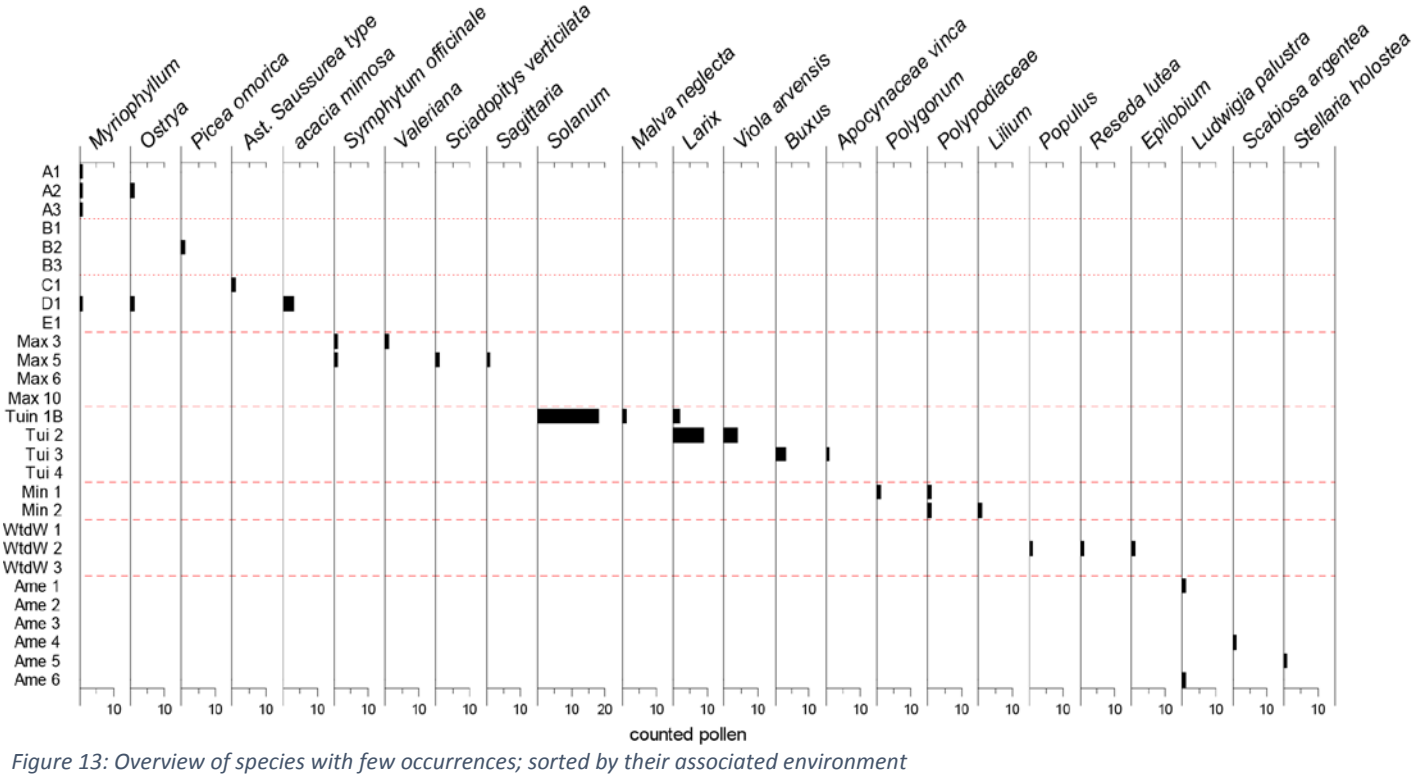


Figure 13: Overview of species with few occurrences; sorted by their associated environment

4.9 Spores

The spores identified in the samples are mainly *Sphagnum* and *Dryopteris* (peat moss and wood fern). Observing the spore percentages (figure 15), it is possible to distinguish between two groups. The group with relative high percentages (~12% - ~37%) consists of the gardens and the vacant lot. Lower percentages (<10%) are found in the group composed of the new park, the road verge and the forest. The old park has a very large range of possible spore percentages and fits within both groups. Two locations, the old park and the forest, have samples were no spores were found. The highest value of spores is found in sample Min 2 with 36.85%.

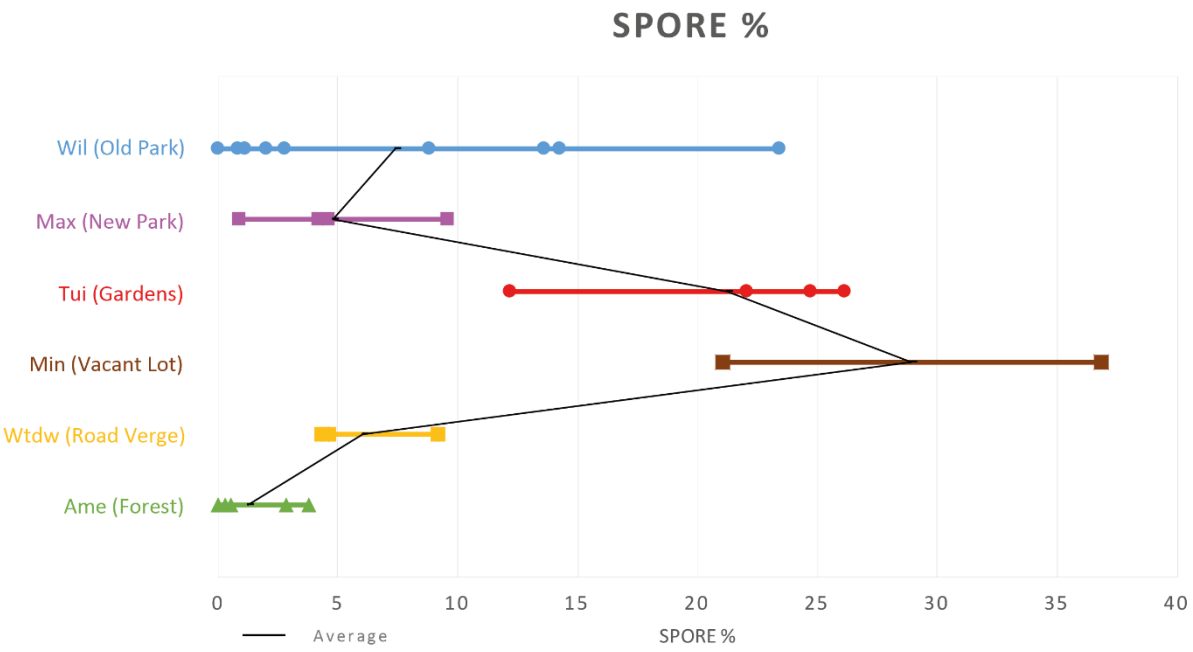


Figure 15: Spore percentage, based on the ratio of spores and the total sum

5. Discussion

5.1 Research question I: Distinguishing and classifying urban environments

5.1.1 Pollen Diagram

A pollen diagram is readily used in ordinary direct palynological comparison. The pollen diagram in this research shows some differences between the urban environments and would indeed be suitable for direct comparison. The focus should lay on differences in arboreal diversity and herb diversity and the quantities of *Quercus, Fagus* and *Fraxinus.*

It should be questioned how the pollen diagram can be implemented in a provenancing framework. The pollen diagram does not readily translate into classification conditions. Automatic scoring by a model may be difficult. When used for provenancing it can best be utilized for reconnaissance or final confirmation of the classification. The pollen diagram enables the primary comparison of samples, encompassing the linking but also the elimination of possible source locations; mainly due to high visualizing aspects.

5.1.2. Cluster analysis

The cluster analysis can be used for provenancing if the samples of a single location are clustered together in the cluster diagram. This is primarily true for the vacant lot and the young park. Provenancing could still be possible if the samples of a single location are grouped in multiple clusters and/or isolated samples. This is true for the forest (Ame 2-4 vs Ame 1 and Ame 5) and the road verge (WtdW 1-2 vs WtdW 3). It might be necessary to find additional explanation for the isolated samples. C1 is the sample with the highest diversity, possibly explaining the isolated place in the diagram. WtdW 3 has the highest percentage of *Fraxinus* compared to all other samples.

Provenancing becomes limited when samples of multiple locations are grouped in a single cluster. This is true for the old park and the garden where sample A3 seems to be related with the garden samples. These locations could therefore be interchanged during provenancing based on this type of cluster analysis. It might be possible to resolve this issue in various ways. A possible cause for the clustering of A3 with the garden samples is the high *Sphagnum* concentration which can be a very local (<m scale) signal. This is supported by looking at the within variation in the A- and B-cluster, in which the *Sphagnum* percentage fluctuates. It might be better for future provenancing to exclude spores from the cluster analysis. Another solution might be to find another, better suitable, type of distance computation or clustering.

5.1.3 PCA

The ordination method shows good distinction between the environments. With extension of the database is might be better to switch to the DCA as ordination method as the gradient length will become larger. At the current sample size however both PCA and DCA function well by separating all environments in the plot. During provenancing the PCA will be a good method to find the source

location. To avoid large changes in the plot it is advisable to make the provenancing sample a passive sample.

5.1.4 AP/NAP-ratio

The AP/NAP-ratio is an effective method for the identification of a young park. The low AP/NAP-ratio turns out to be caused by a lower arboreal concentration rather than a flourishing plants and herbs community. This is probably caused by the numerous immature trees which have a lower pollen production. This effect is demonstrated at the site of Max 6 which is situated in a very *Fagus* dominated part of the park. This is however not visible within the pollen data as very few *Fagus* pollen were counted.

Most samples in the database however show significant overlap, based solely on the AP/NAP-ratio it is impossible to appoint the source location to samples from other urban environments. Only the exclusion of some environments would be possible. Caution should be used as small differences in AP/NAP-ratio do not necessarily represent true changes in vegetation cover. Further extension of the database might reduce the overlap. Until then combining the AP/NAP-ratio with other characteristics in a scoring mechanism for the provenancing framework can be an aid to extract more information on the source.

5.1.5 Pollen concentration

The environments show much overlap based on pollen concentration. Based solely on pollen concentration it is impossible to identify the source location except for highly concentrated forest samples. The high pollen density of the forest is partly due to the dense vegetation of many mature *Quercus* trees which have a high pollen production. The high density of pollen found in the garden Tui 4 deviates from the other garden samples. Potting soil can sometimes cause such deviations. But this is improbable for this sample as no anomalies or indications for exogenic pollen are found in the pollen diagram. While directly identifying the source remains largely impossible, it is possible to eliminate environments as a possible source in some cases. Furthermore, if the pollen concentration is combined with other characteristics it indeed has the potential to identify the pollen source.

The pollen concentration is the end product of production (vegetation density and production per individual), persistence (preservation conditions, species and soil type) and recovery (sampling and the lab process). These conditions should be considered when applying the pollen concentration as a characteristic in provenancing. A source with a very small dispersal distance or a source with a very large pollen production could alter the signal on a very local scale. Lateral difference on even meter scale could hamper the provenancing. Increasing the database would certainly give further insight on whether the data ranges are based on local anomalies or true differences in (urban) environmental conditions.

5.1.6 Dominance

Provenancing using only the dominance of a sample is very difficult. The differences are in fact so small that the possible error makes it impossible to make a proper distinction based on dominance. Only the forest could possible identified and only when it has values over 0.24, which is the upper error for the highest non-forest sample. The high dominance in Ame 2 (0.33) and Ame 3 (0.53) is

caused by high abundances of *Quercus.* Not only are Ame 2 and Ame 3 the samples with the highest percentage of *Quercus* they furthermore have the highest pollen concentration. In this case there is a clear correlation between the dominance and pollen concentration in sample Ame 2 and Ame 3. This is not necessarily true for other samples. For example, there can be a high pollen concentration but with all pollen types evenly present in the sample, causing a low dominance.

Until further expansion of the data set it is recommended, whenever a high dominance is found, to find the source for the high dominance using for example the pollen diagram. This is recommended because currently the only high dominance in this dataset is a forest. There are more ('natural') environments, not included in this study, which could possibly cause a high dominance. One could think of environments such as wetlands with large quantities of reed that could result in a similar dominance.

5.1.7. Diversity

The determined richness of a sample is influenced by the dominance of pollen taxa in a sample. It is therefore interesting to look at this dependency (figure 16). A higher dominance causes the estimated diversity to decrease. This relation should continue towards a dominance of 1 equalling a diversity of 1, as the pollen assemblage would then be composed of a single species. The trend line found in this data set $y = -39.816x + 33.66$ approaches this value with a dominance of 0.82 for a diversity of 1.

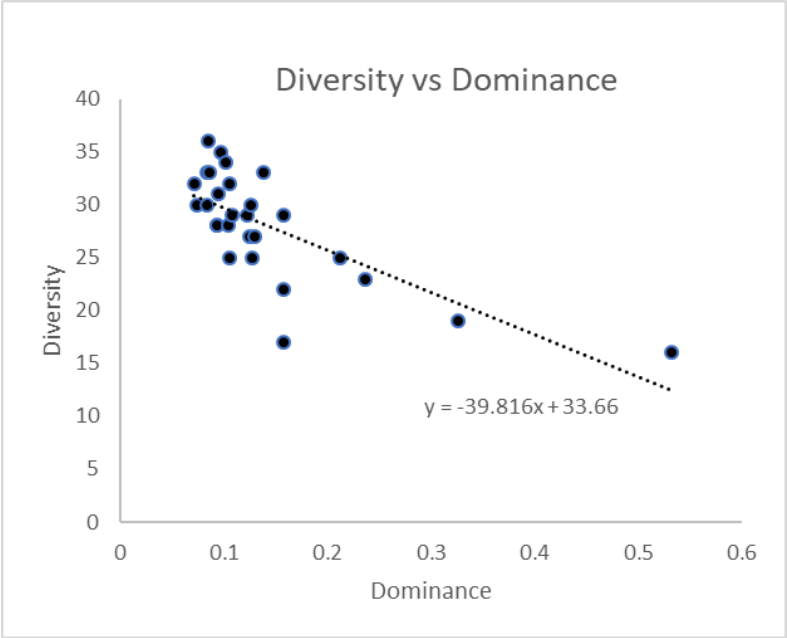


Figure 16: Diagram showing the influence of dominance on the determined diversity.

The number of samples needed to reach a stable diversity follows from the sample rarefaction curves. Only the Old park with nine samples reaches a stable diversity. The site-diversity shown in figure 12 however represents the Mao's Tau sample rarefaction diversity found at two samples, as the vacant lot has only two samples. As provenancing is associated with a limited number of samples, a stable site-diversity is then not reached using sample rarefaction. The diversity characteristic can therefore best be utilized using only the sample-diversity using individual rarefaction.

5.1.8. Environmental-specific species

Several pollen types were determined in a single or only a few samples. However, many of them are not suitable as environmental-specific species. Some plants are common in multiple environments and their pollen are expected to be found in the pollen assemblages with extension of the dataset. Examples are: *Symphytum officinale, Valeriana officinale, Malva neglecta, Viola arvensis, Polygonum and Epilobium.* Some other unsuited species are: *Myriophyllum, Saggitaria, Populus and*

Ludwigia palustre. The first two are aquatic species and indicate a body of water rather than a park. *Populus* is infrequently determined due to its vulnerability to decay and *Ludwigia palustre* is very rare in the Netherlands and its determination accuracy should be treated with caution.

A few pollen-types can be considered as environmental-specific species. In the Old Park *Ostrya* and *Picea omorica* were found which are uncommon arboreal species in the Netherlands. It should be noticed *Picea Omorica* is also used as Christmas tree, which could lead to occurrences in gardens and houses. In the Young Park *Sciadopitis verticalata* was determined, a rare species in the Netherlands, originally from Japan. The occurrence could possibly be explained by the Japanese Garden in the Maximapark. It can be used as an indicator species, though caution should be used for its occurrences as garden tree. In the garden the occurrence of tomato pollen can be explained by a former kitchen garden. *Buxus* is a common hedge plant and *Larix* and *Apocynaceae vinca* are ornamental species often found in gardens. *Reseda lutea* is typical for road verges but has the potential to be found in vacant areas as well.

There are much more species that can potentially be added to this list and therefore the database should be extended with more samples. The strength of environmental-specific species as characteristic increases if the assemblage of those species in a sample increases. Furthermore, with extension of the database, pollen occurrences caused by stochastic processes can be filtered out. Only then environmental-specific species can be used as a characteristic for the classification and provenancing of urban environments.

5.1.9. Spores

Possible environmental factors contributing to the differences in spore content are vegetation cover and humidity. Spore percentage is sometimes considered to have a strong local signal and therefore strong lateral differences. This is demonstrated in the data from the clusterized samples A and B of the Wilhelminapark. However, the ranges are similar to several pollen species. The spores used as characteristic are calculated as percentage of the total sum. A consequence of this method is the possible influence of the pollen concentration on the spore percentage. However, no such influence is found in this dataset (R^2<0.1).

There are considerable differences in spore percentages between the sampled locations. The differences are large enough to differentiate between two environmental groups and the characteristic can therefore be used for classification and provenancing purposes.

	Old park		New Park		Garden		Vacant lot		Road verge		Forest	
Dominance	0.06	0.15	0.11	0.25	0.08	0.12	0.1	0.17	0.09	0.15	0.08	0.59
Pollen conc.	15000	124000	10000	49000	37000	251000	64000	420000	40000	73000	22000	953000
AP/NAP-ratio	1.8	3.3	0.5	1.1	2.3	4.1	1.9	2.2	1.5	3.7	2.1	16.4
Diversity	29	34	25	33	28	35	17	25	27	32	16	25
Spores (%)	0	23.4	0.8	9.6	12.1	26.1	21	36.8	4.3	9.2	0	3.8
Environmental-specific	Ost, Pic. O, Sau, Aca, Myr, (Cat, Lon)		Sci, Val, Sag		Lar, Apo, Bux, Mal, Sol, Vio, (Cat, Lon)		Lil, Poly		Pop, Epi, Res, Sym		Ste, Sca, Lud	
Pollen Diagram	Ran	Plat, Nys, Aes, Jug, Pte, Lon, Pru, Sec		Poa, Ast, Sec, Plan		Plat, Pru	Pin	The	Spa	Til	Aln, Eri	Que, Fag, Fra, Car
PCA												
Cluster												
	lowest	highest	lowest	highest	lowest	highest	lowest	highest	lowest	highest	lowest	highest

Table 2: Overview of the urban environments with associated variable values. For dominance, pollen concentration, AP/NAP-ratio, diversity and spore percentage the minimum and maximum values are shown. For rare species the presence of certain rare species are given. For the pollen diagram species with low and high abundancies are shown. The full name of the pollen types can be found in Appendix D. The PCA and cluster analysis cannot be summarized by minimum and maximum values.

Table 3: A visualization of the data ranges in Table 2. It shows how a single characteristic indicates for a environment or a group of environments. A fully red row would mean total overlap of all the data. All different colours in one row would mean total distinction between all environments. Two colours within one cell means there is partial overlap with another group or environment.

5.1.10 Summary

The results of the data exploration techniques are summarized in table 2. For every location the range of the data points is given by the minimum and maximum values of each characteristic. This table should provide a good first tool in the provenancing of a sample. The PCA and the cluster analysis should be analysed visually. In order to show how the different characteristics classify the environments in different groups, the table is translated into table 3. It shows that PCA, cluster analysis, environmental-specific species and the pollen diagram are competent in classifying different environments. The dominance, pollen concentration, AP/NAP-ratio, diversity and spore percentage have less classifying abilities; as they have much overlap and can often only distinct between two groups. All together however, they form the first criteria for the division between multiple (mainly urban) environments.

It should be discussed how valid this classification is. Is it truly possible to distinguish between multiple palynological environments in urban areas? At the start of this research five urban areas where selected and sampled of which was theorised they could be palynological distinct. The following proposition should be discussed:

- **Some urban environments are not included in this research.**
  It is most probable more palynological distinct urban environments exist than those analysed in this research. In the interest of further research in urban palynology it is recommended to include additional environments whenever the database is extended. Several suggestions for further analysis are: construction sites, meadows and watersides.
- **Some proposed urban environments are non-existent.**
  It could be argued whether gardens are a single environment as the heterogeneity between gardens can be very high. Dividing this environment in multiple environments however would arise new problems concerning definition, database and fragmentation. It is probably more appropriate to accept the broad range of the characteristics associated with the garden environment.
- **Some proposed urban environments should be merged.**
  Evident differences (e.g. AP/NAP-ratio, PCA, pollen diagram) between old and new park justify the proposed division of a park in two separated environments. Resemblance of the old park with other environments is not caused by comparison within a single larger existing environment but merely by comparison of the broad-ranged, average valued old park with other smaller, more specific environments. Merging with the old park would lead to the loss of differentiation between those smaller environments and is therefore not preferred.
- **The sample size of the database is too small.**
  A consequence of the small sample size, the environmental limits are too small in comparison with the true ranges. While this remark is certainly valid for this research, it was a presumption made at the start of this research. This pilot study on palynology in urban environments was designed in such way that it can and should be extended if promising results are shown.

It can be concluded that the characteristics show enough differences to construct an initial classification, summarized in Table 2 and 3. However, the classification based on the nine characteristics has it flaws, mainly caused by the limited sample size of the database and possibly by

non-incorporated palynological environments. This can be resolved with the extension of the database.

5.2 Research question II: Framework for a new provenancing method

An appropriate and reliable provenancing should not be based on a single characteristic, but on as many characteristics as possible. However, the characteristics must truly mean something. As seen in the first part of the discussion not all characteristics can be specified to a single environment. Rather than discard the less informative characteristics of this study it is possible to combine them, in order to minimize the loss of data and optimize their provenancing potential.

The summarizing tables 2 and 3 provide the classification conditions of the selected urban environments discussed in part 5.2. The conditions are either given by minimum and maximum values, the presence of certain species or their place in visualizing plots. Using these conditions, it is possible to introduce a provenancing method for urban pollen from soils. A framework will be introduced in which sample data is compared with the database. This should result in the most probable pollen source, or at least exclude some environments as possible pollen source. This approach can be summarized as followed:

1) Sample data → 2) Comparison with database → 3) Presumed pollen source

The framework in combination with the sample data should provide the provenancing solution. The question arises in what way the comparison should be set up and how scoring should proceed. We propose to split the provenancing framework into two separate parts. One focusses on visualization and the second is a scoring mechanism. The framework is elaborated in more detail in the flowchart of figure 17.

Part 1, the visualization, consists of: the pollen diagram, rare species, PCA and cluster analysis. They are best analysed by plotting the sample values passively into the plots and diagrams of the database values. For the pollen diagram this means the addition of an additional horizontal row, while in PCA this is a single point.

Part 2, which consist of the other characteristics (dominance, diversity, pollen concentration, AP/NAP-ratio and spore percentage), can best be merged in a scorings mechanism. The scorings mechanism works through comparison of the sample values with the characteristics outer values of each urban environment (table 2). It is possible to assign one point, assuming the characteristics should be equally weighed, for every sample characteristic matching with the environmental values. If we assign a maximum of one point for each of the five characteristics the environment with the highest score, with maximum of 5, should most resemble the urban environment associated with the origin of the sample. It is however possible that a sample receives a high score for another location as well, which can be considered as a false positive. In both scenarios the visualized characteristics, part 1 of the framework, should be used for confirmation.

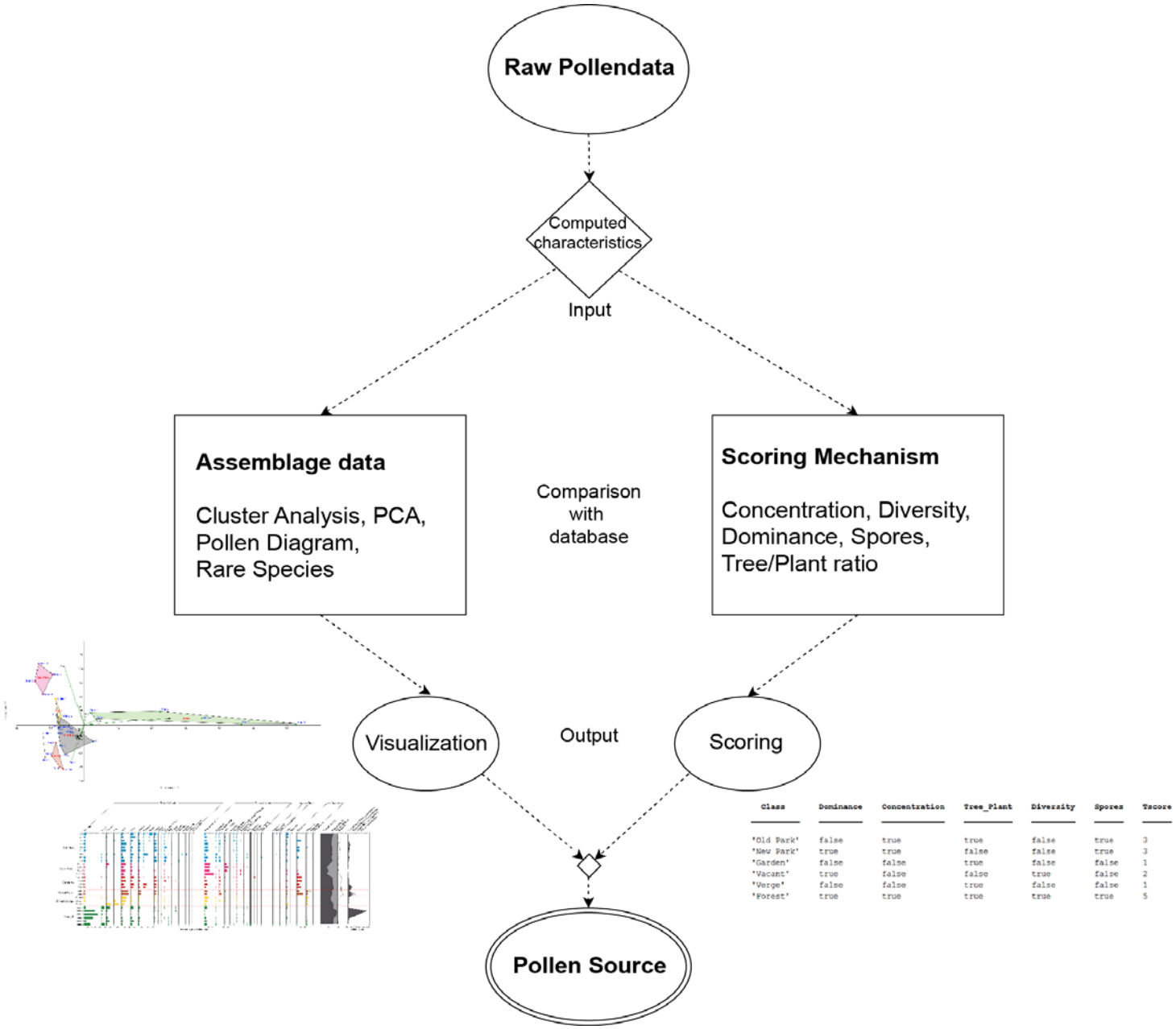


Figure 17: Flowchart for the proposed provenancing method including a scoring mechanism and a visual comparison of assemblage data. Rare Species represents the environmental specific species and the Tree/Plant ratio represents the AP/NAP-ratio.

Class	Dominance	Concentration	Tree_Plant	Diversity	Spores	Tscore
'Old Park'	false	true	true	false	true	3
'New Park'	true	true	false	false	true	3
'Garden'	false	false	true	false	false	1
'Vacant'	true	false	false	true	false	2
'Verge'	false	false	true	false	false	1
'Forest'	true	true	true	true	true	5

Table 3: Scoring result for sample Ame 6. TScore (total score) is the result based on the number of returned 'true' values. This value reflects the similarity with an urban environment and ranges between 1-5. Tree_Plant represents the AP/NAP-ratio.

It is even possible to automate the scoring mechanism. For efficiency a script (Appendix F) has been written, which can speed up the comparison between a sample and the pollen data collected in this research. The script gives either a 0 ('false') or 1 ('true') for each characteristic. This results in a maximum score of 5 if the sample matches an environment on basis of concentration, diversity, dominance, spore percentage and AP/NAP-ratio. The script gives the result in the form of table 4. The shown table shows the result of the forest sample Ame 6. As Ame 6 is incorporated in the database the logical result is a maximum score of 5 for the forest environment.

To test the model, it has been run not only for Ame 6 but for all the samples in the database. The results can be found in table 5. As expected all samples have the maximum score for their source location, indicated in dark green. More interesting are the false positives, high scores for the 'wrong' environments. Most of the high values are scores from the verge-samples (WtdW 1-3). WtdW 1 is similar to a New park, while WtdW 2 could be mistaken for a sample from an Old Park. WtdW 3 scores high for the New Park and the Forest. Three more false positives are found in Tui 1 and Tui 3 and Ame 4 which score high for the Old Park environment. The Old Park environment in general has above average scores. This is caused by the broad range of this dataset. This in contrast to the forest-samples (Ame 1-6) which have more extreme values and therefore low similarities with other environments. The low similarities with the vacant-environment extends to most of the other samples. This can be explained by the very small vacant lot data set and therefore very small data range of the vacant-environment.

This scoring mechanism was made as a replacement for five less informative characteristics. At first these characteristics were only able to distinct between two or three groups (recall table 3), while using the scoring mechanism there is much more differentiation between the environments. Still some overlap exists, especially with the Old Park environment, but compared to the initial characteristics the overlap is reduced.

	A1	A2	A3	B1	B2	B3	C1	D1	E1	Max1	Max5	Max6	Max10	Tui1	Tui2	Tui3	Tui4	Min1	Min2	WtdW1	WtdW2	WtdW3	Ame1	Ame2	Ame3	Ame4	Ame5	Ame6
Old Park	5	5	5	5	5	5	5	5	5	2	3	3	2	4	3	5	2	3	1	3	5	2	3	1	1	2	4	3
New Park	2	2	2	2	0	1	2	2	2	5	5	5	5	2	1	1	0	1	1	4	2	4	1	0	0	0	3	3
Garden	1	2	2	2	2	3	1	2	2	1	0	1	2	5	5	5	5	2	0	2	1	3	1	0	0	1	3	1
Vacant	0	0	1	1	1	2	0	1	0	1	1	0	0	1	2	1	2	5	5	0	1	1	1	1	0	2	0	2
Verge	1	1	2	3	1	3	1	3	3	2	0	2	1	2	3	3	0	3	1	5	5	5	1	0	0	0	2	1
Forest	2	3	0	3	1	2	2	1	2	1	3	1	3	2	3	2	3	3	3	3	2	4	5	5	5	5	5	5

Table 4: Score for all sample incorporated in the database. Expected scores of 5 are shown in dark green. Other high scores are shown in orange (5) and pink (4). Very dissimilar values are shown in light green.

5.3 Validation of the framework

To test the validity of the provenancing method ten additional samples from arbitrary (urban) locations are provenanced by the model. Table 6 shows the results from the scoring mechanism. Appendix G and H provide the pollen data and derived characteristics.

	1 meadow	2 vacant	3 wildered	4 wildered	5 wildered	6 wildered	7 old park	8 old park	9 old park	10 old park
Old Park	2	0	2	3	2	2	1	1	2	3
New Park	0	2	0	0	1	1	1	2	2	1
Garden	0	1	1	1	1	1	0	1	1	2
Vacant	1	2	1	1	2	0	2	0	1	0
Verge	1	1	1	2	1	2	0	0	2	1
Forest	2	2	3	3	4	3	4	4	3	3

Table 5: Tscore for additional samples. Red=improbable, light green=plausible, Green=probable

The meadow sample is very dissimilar from the new park, gardens, vacant lot and the roadside verge. This is explained as the sample originates from a meadow, which is probably an independent environment not included in this research. The dissimilarity is also visible in the pollen composition, the cluster analysis and the PCA-plot (figure 18A), where the meadow sample is isolated from the other groups. In an enhanced provenancing framework the meadow should be implemented as a new additional environment.

The vacant lot has low scores in the scoring mechanism and is dissimilar with the old park, gardens and the verge. From the scores it is not evident that the sample originates from a vacant lot, neither this is clear in the cluster analysis and the pollen diagram. Apparently, the vacant data retrieved from the Minnaert, Uithof, is not representative or at least insufficient for such environment. The vacant sample in the PCA-plot (figure 18B) is also rather isolated but a dotted line is given to show how the vacant environment is extended if the new sample is incorporated in the database.

The wildered patch (samples 3-6) is dissimilar from both the new park and the gardens. Instead, based on the scoring mechanism, the samples can be interpreted as forest samples, though sample 4 suggest old park influences. Recall that this is based on a combination of dominance, concentration, spore percentage and diversity. The pollen diagram shows a resemblance with an old park and garden. According to the PCA, as shown in figure 18C, the wildered lot significantly overlaps the old park environment. The cluster analysis groups the wildered patch as an independent environment. It can be hypothesized that such a wildered patch translates in an urban environment similar to a mix of an old park and a forest.

The old park samples (7-10) would be provenanced to a forest based on the characteristics processed in the scoring mechanism. This false positive from the scoring mechanism is partly compensated by the illustrative comparison. Based on the pollen diagram samples 7-10 have most similarities with the old park. In the PCA-plot of figure 18D the samples are shown in orange. It is an isolated group between the forest group and the old park group. When the database is further extended using this new data a dotted line can be drawn to extend the old park group.

The provenancing model and the database in the current state should be operated with caution. The 28 samples plus the additional samples show that a confident provenancing based on the selected characteristics is not yet possible. However, during provenancing differentiation between urban environments does occur which allows segregation between improbable and probable source environments.

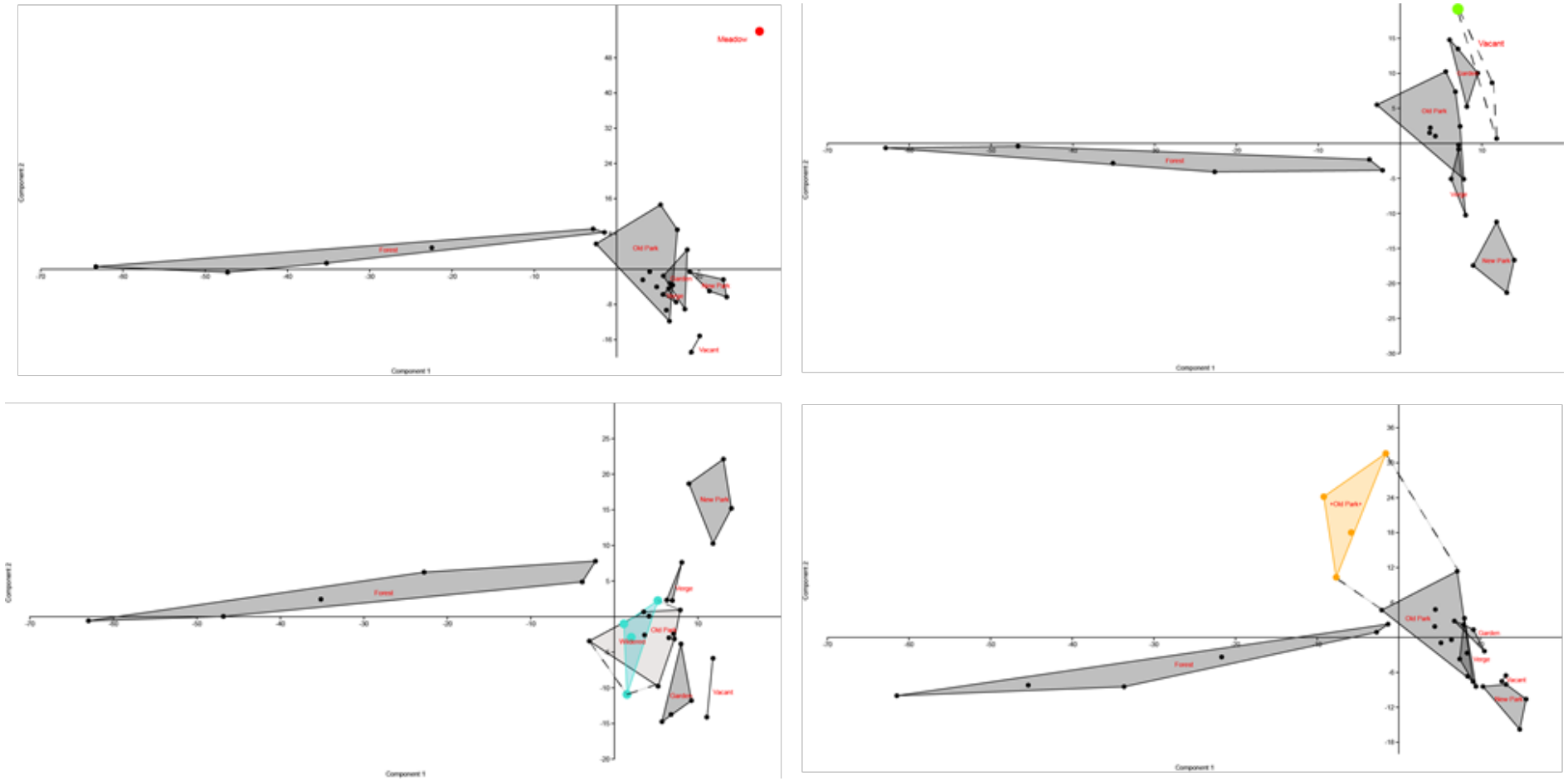


Figure 18: PCA plots with additional samples. Red) meadow; Green) vacant; Blue) wildered patch; Orange) old park.

5.4 Final results of the enhanced model

The best continuation of this research should be an iterative approach. Samples should be added to the database until the classification and the resulting scores do no longer change. A first step for this process is the addition of the extra old park and vacant lot samples to their associated database. Furthermore, new distinctive environments should be added to the database. The meadow for example, which has shown to be significant different from the selected environments.

To show this iterative approach the internal check can be rerun and consequently table 5 can be updated. The internal check results in a slightly altered score for both the old park and the vacant lot (Table 6). The very low scores found at the initial vacant environment are resolved. Interesting results are found at the old park environment. In general, all samples score relatively high values for an old park. It must be noticed however that most of these samples have at least one other high scoring environment, except for the samples from the old park, which only scores high for the old park environment.

	A1	A2	A3	B1	B2	B3	C1	D1	E1	Max1	Max5	Max6	Max10	Tui1	Tui2	Tui3	Tui4	Min1	Min2	WtdW1	WtdW2	WtdW3	Ame1	Ame2	Ame3	Ame4	Ame5	Ame6
Old Park	5	5	5	5	5	5	5	5	5	4	4	4	2	4	3	5	3	4	2	3	5	4	4	1	1	3	4	4
New Park	2	2	2	2	0	1	2	2	2	5	5	5	5	2	1	1	0	1	1	4	2	4	1	0	0	0	3	3
Garden	1	2	2	2	2	3	1	2	2	1	0	1	2	5	5	5	5	2	0	2	1	3	1	0	0	1	3	1
Vacant	1	1	3	2	1	2	1	2	1	3	3	2	2	2	2	1	2	5	5	2	2	2	2	1	0	2	1	3
Verge	1	1	2	3	1	3	1	3	3	2	0	2	1	2	3	3	0	3	1	5	5	5	1	0	0	0	2	1
Forest	2	3	0	3	1	2	2	1	2	1	3	1	3	2	3	2	3	3	3	3	2	4	5	5	5	5	5	5

Table 6: Sample scores based on the enhanced model. In red are shown the altered results caused by extension of the database

With the iterative approach the database should not only be extended but the provenancing method could be improved as well. Several suggestions for the improvement of the model:

1. Extend the database by increasing the number of samples per environment. Extension should continue until the classification or the provenancing results are minimal.
2. Extend the database by including more possible urban environments such as construction sites and watersides.
3. Revision of the characteristics weights in the scoring mechanism. In the current model equal weights were used. However, some characteristics may have a higher indicative value during provenancing than others. How the weights should be altered is dependent on how the data spread reacts to extension of the database.
4. Computation of the characteristics should be incorporated in the model. This can be achieved by using the raw pollen data as input for the scoring mechanism, instead of the derived variables. This avoids the use of multiple software applications (Excel, Matlab, Past and Tilia).
5. Alteration of the scoring method for the scoring mechanism. Instead of using the minimum and maximum values, other methods could result in more appropriate results. Suggestions include the use of the standard deviation or the exclusion of outliers or extreme values.

6. Conclusions

This research with its multidisciplinary approach between forensics and earth sciences explored the palynology of urban environments. The analysis of the urban palynology has shown differentiation between multiple environments. It is possible to classify palynological samples on a higher resolution than 'urban'. Nine characteristics have been derived from the pollen data, all of them were able to distinguish at least two palynological groups. With combined effort the data characteristics are able to classify between six environments. However, the proposed classification in its current state should be used with caution. More palynological data from urban areas is necessary to reduce the existing overlap in the proposed classification. The palynological database should be extended until the change in the characteristics ranges are in equilibrium, thereby getting a better insight into the true internal variation of an environment and the variation between environments. The clusters in the old park show that internal variation is reduced if the sampling area is smaller. A suggested research subject for future studies would be the relationship between palynological composition and distance from the centre of an urban environment.

The proposed classification can be applied during (forensic) provenancing. The five characteristics, which classification that could only differentiate between two palynological groups, (dominance, concentration, pollen percentage, diversity and AP/NAP-ratio) are combined in a scoring mechanism. The other characteristics should be analysed visually. Based on the proposed classification in its current state the provenancing is limited to eliminating the improbable source locations. The research should continue using an iterative approach. With extension of the database, both in sample size and the number of environments, the provenancing should improve. With these improvements the scoring mechanism can possibly be further enhanced using both changes in computation and altered variable boundaries.

It should be stressed that this proposed method is still in its infancy. The presented pollen data along with the classification and the provenancing framework poses a satisfactory solution to the aims of this research and pave the way for better exploitation of (forensic) palynology in urban areas.

7. Acknowledgements

I would like to thank Timme Donders and Stefan Uitdehaag for their supervision, support and especially their unlimited patience. In addition, many thanks go to the entire 5th floor students room for their everlasting help, fun and distraction. At last I would like to thank all other colleagues, friends and family; anyone who contributed to this thesis by either grammar checks, visual enhancements, food supplies or any kind of mental support.

References

-, Rijksmonumentenregister. Available at: https://monumentenregister.cultureelerfgoed.nl/monuments/514409?MonumentId=514404.

Adams-Groom, B., 2012. *Forensic Ecology Handbook* N. Márquez-Grant & J. Roberts, eds., Chichester, UK: John Wiley & Sons, Ltd.

Ariizumi, T. & Toriyama, K., 2011. Genetic Regulation of Sporopollenin Synthesis and Pollen Exine Development. *Annual Review of Plant Biology*, 62, pp.437–460.

van Berghem, G., Tepp, I. & Kipp, F., 2009. *Hand aan de Schop, Bomenplan voor het Wilhelminapark en omgeving*,

Beug, H.-J., 2004. *Leitfaden der Pollenbestimmung*, Munich: Dr. Friedrich Pfeil.

Bray, J.R. & Curtis, J.T., 1957. An Ordination of the upland forest community of southern Wisconsin.pdf. *Ecology Monographs*, 27(4), pp.325–349.

Brown, A.G., 2006. The use of forensic botany and geology in war crimes investigations in NE Bosnia. *Forensic Science International*, 163, pp.204–210.

Bryant, V.M. & Jones, G.D., 2006. Forensic palynology: Current status of a rarely used technique in the United States of America. *Forensic Science International*, 163, pp.183–197.

City of Utrecht, 2014. *Beheerplan Maximapark*,

City of Utrecht, Bomenkaart Gemeente Utrecht. Available at: http://utrecht.gemgids.nl/bomenkaart/.

Colwell, R.K., Mao, C.X. & Chang, J., 2004. INTERPOLATING, EXTRAPOLATING, AND COMPARING INCIDENCE-BASED SPECIES ACCUMULATION CURVES. *Ecology*, 85(10), pp.2717–2727.

Colwell, R.K., Xuan Mao, C. & Chang, J., 2004. Interpolating, Extrapolating, and Comparing Incidence-Based Species Accumulation Curves. *Ecology*, 85(10), pp.2717–2727.

Davis, B.A.S. et al., 2003. The temperature of Europe during the Holocene reconstructed from pollen data. *Quaternary Science Reviews*, 22, pp.1701–1716.

Donders, T.H. et al., 2007. Late Neogene continental stages in NW Europe revisited. *Earth-Science Reviews*, 85, pp.161–186.

Erdtman, G., 1969. *Handbook of palynology: morphology, taxonomy, ecology. An Introduction to the Study of Pollen Grains and Spores*, Copenhagen: Munksgaard.

Erdtman, G., 1960. The acetolysis method. *Svensk Botanisk Tidskrift*, 54, pp.561–564.

Eyles, N., Mory, A.J. & Backhouse, J., 2002. Carboniferous–Permian palynostratigraphy of west Australian marine rift basins: resolving tectonic and eustatic controls during Gondwanan glaciations. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 184(3–4), pp.305–319.

Faegri, K. & Iversen, J., 1964. *Textbook of pollenanalysis*, Copenhagen: Munskgaard.

Faegri, K. & Pijl, L. van der (Leendert), 1979. *The principles of pollination ecology* 3rd rev. e., Pergamon Press.

Favre, E. et al., 2008. A contribution to deciphering the meaning of AP/NAP with respect to vegetation cover. *Review of Palaeobotany and Palynology*, 148(1), pp.13–35.

Hammer, Ø., Harper, D.A. & Ryan, P.D., 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1), p.9pp.

Horrocks, M., Coulson, S.A. & Walsh, K.A.J., 1998. Forensic Palynology : Variation in the Pollen Content of Soil on Shoes and in Shoeprints in Soil. *Journal of Forensic Sciences*, 43(2), pp.119–122.

Krebs, C.J., 1989. *Ecological Methodology* Second Edi., New York: Harper & Row.

Kuijper, W.J. & Turner, H., 1992. Diet of a Roman centurion at Alphen aan den Rijn, The Netherlands, in the first century AD. *Review of Palaeobotany and Palynology*, 73, pp.187–204.

Maes, B. et al., 2009. *Ecologische waarden Landgoederen Oud en Nieuw Amelisweerd en Rhijnauwen*, Utrecht.

Mathewes, R.W., 2006. Forensic palynology in Canada: An overview with emphasis on archaeology and anthropology. *Forensic Science International*, 163, pp.198–203.

Mildenhall, D.C., 2004. An Example of the Use of Forensic Palynology in Assessing an Alibi. *Journal of Forensic Sciences*, 49(2), pp.312–316.

Mildenhall, D.C., 1990. Forensic palynology in New Zealand. *Review of Palaeobotany and Palynology*, 64, pp.227–234.

Mildenhall, D.C., Wiltshire, P.E.J. & Bryant, V.M., 2006. Forensic palynology: why do it and how it works. *Forensic Science International*, 163, pp.163–72.

Moore, P.D., Webb, J.A. & Collinson, M.E., 1997. *Pollen Analysis* Second Edi., Oxford: Blackwell Science.

Odgaard, B.V., 2007. POLLEN METHODS AND STUDIES | Reconstructing Past Biodiversity Development. In *Encyclopedia of Quaternary Science*. Elsevier, pp. 2508–2514.

Punt, W. et al., *The Northern European Pollen Flora I-IX*, Elsevier.

Punt, W. & Hoen, P.P., 2009. The Northwest European Pollen Flora, 70: Asteraceae — Asteroideae. *Review of Palaeobotany and Palynology*, 157(1), pp.22–183.

Sadori, L. et al., 2016. Pollen-based paleoenvironmental and paleoclimatic change at Lake Ohrid (south-eastern Europe) during the past 500 ka. *Biogeosciences*, 13, pp.1423–1437.

Schackmann, M. & van Rossum, M., 2005. *De aanleg van het wilhelminapark e.o. in Utrecht*. University Utrecht.

Schwendemann, A.B. et al., 2007. Aerodynamics of saccate pollen and its implications for wind pollination. *American journal of botany*, 94(8), pp.1371–81.

Shi, G.R., 1993. Multivariate data analysis in Palaeoecology and palaeobiogeography - a review. *Paiaeogeography, Palaeoclimatology, Palaeoecology*, 105, pp.199–234.

Simpson, E.H., 1949. Measurement of diversity. *Nature*, 163(4148), p.688.

Stockmarr, J., 1971. Tablets with spores usedin absolute pollen analysis. *Pollen and Spores*, 13(4 (January 1971)), pp.615–621.

Szibor, R. et al., 1998. Pollen analysis reveals murder season. *Nature*, 395(6701), pp.449–50.

Walsh, K. a. J. & Horrocks, M., 2008. Palynology: Its Position in the Field of Forensic Science. *Journal of Forensic Sciences*, 53(5), pp.1053–1060.

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*. Dordrecht: Springer Netherlands, pp. 129–149.

Wiltshire, P.E.J. et al., 2014. Palynology and mycology provide separate classes of probative evidence from the same forensic samples: A rape case from southern England. *Forensic Science International*, 244, pp.186–195.

Yoon, H.S. et al., 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular biology and evolution*, 21(5), pp.809–18.

Used software:

Adobe Illustrator CC -Adobe

Matlab -The Mathworks

Microsoft Office -Microsoft

Past v3 -(Hammer et al. 2001)

Tilia v1.7 -TiliaIT

Appendices

Appendix A: Vegetation record

Vegetation record of the sampling sites with a ~30m radius. Between brackets the number of individuals is shown.

Species	Old Park	New Park	Gardens	Vacant lot	Road verge	Forest
Acer campestre	B, D1, E1	Max5(2x)				Ame2
Acer platanoides	A, E1(3x)					
Acer pseudoplatanus	A, B, D1, E1(2x)					Ame1(2x), Ame2, Ame3(7x), Ame4, Ame5, Ame6(4x)
Acer saccharinum	C1					
Acer undiff.			Tui1			
Aesculus hippocastanum	A, E1(2x)					Ame1(6x), Ame6
Alnus glutinosa	C1, E1					
Betula ermanii	A, B(3x)					
Betula pendula	B					
Betula undiff.			Tui3			
Carpinus betulus		Max3(2x)				Ame1(2x), Ame2(5x), Ame3(3x), Ame6(3x)
Fagus sylvatica	A, D1	Max3, Max6(36x), Max10(2x)				Ame1(15x), Ame2(17x), Ame3(9x), Ame4(7x), Ame6(18x)
Fraxinus augustifolia		Max3(2x)				
Fraxinus excelsior	A, D1				WtdW3(4x)	Ame1(10x), Ame2(13x), Ame3(6x), Ame4(50x), Ame5(>100x), Ame6(11x)
Ginkgo biloba			Tui4			
Gymnocladus dioica	A					
Larix			Tui1			
Malus cultivars	C1					
Malus hybride	C1					
Metasequoia glyptostroboides	A					
Ostrya carpinifolia	B					
Paulownia tomentosa	A					
Picea abies						Ame6
Pinus nigra	B(3x)					
Pinus wallichiana	A					
Platanus xhispanica	C1(2x)	Max5(19x)	Tui3(10x), Tui4			
Populus xcanadensis		Max3(5x), Max5(17x)				Ame2(2x), Ame5(2x)
Populus nigra		Max10(7x)				
Prunus avium						Ame1
Prunus padus	C1					

(Continued)

Pterocarya fraxinifolia	C1			Min (chopped)		
Quercus macranthera	A, D1					
Quercus rubra	B(3x)					
Quercus robur	D1(3x), E1					Ame1(x10), Ame2(21x), Ame3(6x), Ame4(49x), Ame6(11x)
Robinia pseudoacacia	A, D1(4x), E1					
Salix Sepulcralis	B, C1					
Salix undiff.		Max3(>20x)				
Taxodium distichum	C1(2x), D1(2x)	Max6(2x)				
Taxus baccata	B, E1(2x)					Ame6(6x)
Tilia cordata		Max3				
Tilia americana	E1	Max10(31x)				
Tilia platyphyllos	C1			Min1(4x), Min2(6x)		
Tilia xeuropaea	A, C1(3x), D1, E1(2x)	Max10(5x)	Tui4(2x)		WtdW2(5x), WtdW3(12x)	
Ulmus glabra	B, C1, D1				WtdW3(3x)	
Ulmus undiff.						Ame1(4x), Ame2(9x), Ame3(15x), Ame5(5x)
Ulmus xhollandica	D1				WtdW3	
Composition of: Prunus, Tilia, Fagus, quercus		Max3(>10), Max6(>10)				
Amelanchier			Tui3			
Buxus			Tui2, Tui3			
Cytisus			Tui2			
Forsythia			Tui3			
Galanthus nivalis			Tui3(>1)			
Hedera helix					WtdW2(>1)	
Hibiscus syriacus			Tui4			
Hydrangea macrophylla			Tui2, Tui3(4x)			
Liriodendron tulipifera			Tui2			
Narcissus		Max3(>10)	Tui3(>4)			
Ranunculus ficaria			Tui4			
Rhododendron			Tui2, Tui3			
Rosa			Tui1			
Sambucus			Tui1			
Taraxacum officinale			Tui1			
Tulipa			Tui1(>1)			
Vinca			Tui3			

Appendix B: Sample protocol

Decalcifying & sieving

1) Weigh some sample material and transfer <1.5 g of sample material to a 15 ml tube
2) Add several drops of 5% HCl
3) Add 8 ml lycopodium-HCl solution with HCl. Homogenise the sample after 4 ml and 8 ml.
4) Centrifuge for 5 min at 1700 rpm
5) Decant and homogenise
6) Add an excess of KOH and homogenise
7) Heat for 60 min at 70° C
8) Centrifuge for 4 min at 1700 rpm
9) Decant, dilute with aqua dest. and homogenise
10) Centrifuge 4 min at 1700 rpm
11) Decant and add 1ml aqua dest.
12) Sieve the sample over a 250 micron sieve with abundant water and collect the filtrate
13) Sieve the filtrate using a 7 micron sieve with abundant use of water and collect the residue from the sieve.
14) Transfer material back to 15 ml tube. Use aqua dest to fill 15 ml tube.
15) Centrifuge for 5 min at 1700 rpm
16) Decant and fill tube with demineralised water

Acetolysis

17) Centrifuge for 4 min at 1700 rpm
18) Decant water
19) Add 4 ml glacial acetic acid and vortex tube
20) Centrifuge for 5 min 1700 rpm
21) Decant and add again 4 ml glacial acetic acid. Vortex thereafter
22) Centrifuge for 5 min 1700 rpm. Decant afterwards
23) Mix acetolysis mixture (9 parts acetic anhydride with 1 part sulphuric acid)
24) Add 4 ml acetolysis mixture and homgenise
25) Heat for 10 min at 100° C in a water bath. Vortex after 5 min and afterwards
26) Centrifuge for 5 min at 1700 rpm
27) Decant and add aqua dest.
28) Centrifuge for 5 min at 1700 rpm
29) Decant and add aqua dest.
30) Centrifuge for 5 min at 1700 rpm

(Continued)

Heavy liquid separation

31) Vortex and centrifuge 3 min at 1700 rpm. Decant water
32) Add 4 ml sodium-polytungstate solution (ρ=2.1)
33) Vortex and centrifuge for 15 min at 1700 rpm
34) Transfer float to new tube and add aqua dest.
35) Add 4 ml sodium-polytungstate (ρ=2.1) to tube containing the sink
36) Vortex all tubes (float + sink) and centrifuge for 15 min at 1700 rpm
37) Decant tube containing the float into recovery vial for sodium-polytungstate
38) Transfer the second float into 50 ml tube and add first float and demineralised water
39) Clean the tube with remaining sink and decant all material into recovery vial for sodium-polytungstate.
40) Centrifuge 50ml tubes for 5 min at 1700 rpm
41) Decant into recovery vial and add demineralised water
42) Centrifuge for 5 min at 1700 rpm
43) Decant into recovery vial and transfer material into 15 ml tube and add water
44) Centrifuge for 3 min at 1700 rpm
45) Decant into recovery vial

Slide preparation

46) Transfer material to Eppendorf cup using ethanol (99%)
47) Centrifuge 2 min at 1700 rpm with open eppendorf cup
48) Decant and add glycerol (same volume as sample)
49) Heat at 70° C for one night (with open eppedorf cups)
50) Add several drops of glycerol and homogenise
51) Put one drop of material on slide and add one drop of glycerol
52) Mix and distribute using a wooden toothpick
53) Cover with cover slip
54) Use cover slip lacquer to seal slide
55) Let the slide rest for a couple hours

Appendix C: Sample weights

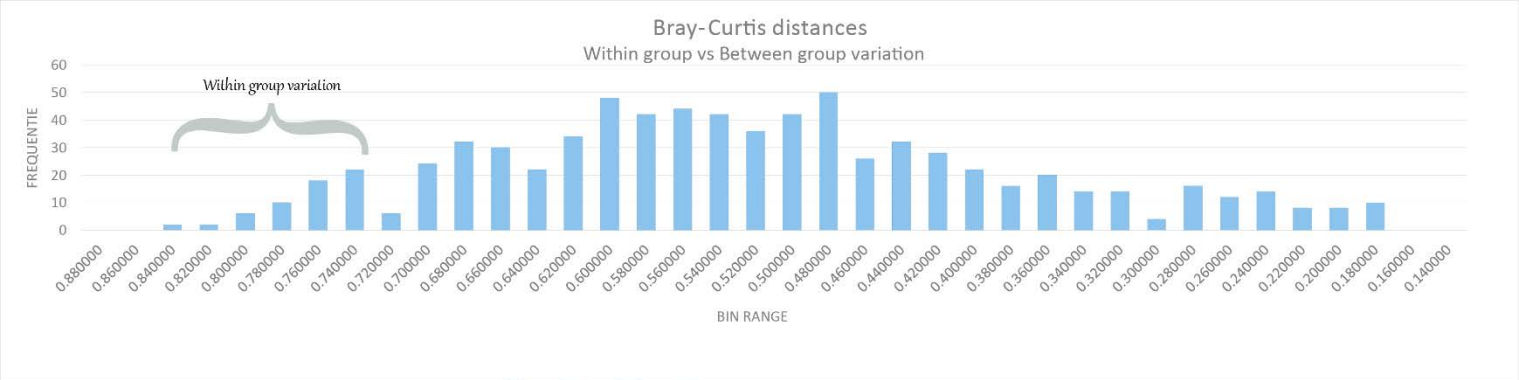
Weights are in grams; samples in bold were palynological analysed in this research

Old Park		**New Park**		**Gardens**		**Vacant Lot**		**Road Verge**		**Forest**	
Wilhelminapark		**Maximapark**		**Several households**		**Minnaert, Uithof**		**Weg tot de Wetenschap**		**Nieuw Amelisweerd**	
A1	**1.361**	Max 1	1.403	Tui 1	1.403	**Minn 1**	**1.401**	**WtdW 1**	**1.396**	**Ame 1**	**1.397**
A2	**1.411**	Max 2	1.396	**Tui 1b**	**1.405**	**Minn 2**	**1.405**	**WtdW 2**	**1.395**	**Ame 2**	**1.403**
A3	**1.470**	**Max 3**	**1.393**	**Tui 2**	**1.400**			**WtdW 3**	**1.401**	**Ame 3**	**1.395**
B1	**1.409**	Max 4	1.403	**Tui 3**	**1.404**			WtdW 4	1.402	**Ame 4**	**1.392**
B2	**1.455**	**Max 5**	**1.401**	**Tui 4**	**1.403**					**Ame 5**	**1.403**
B3	**1.440**	**Max 6**	**1.404**	Tui 5	1.403					**Ame 6**	**1.399**
C1	**1.393**	Max 7	1.403								
D1	**1.458**	Max 8	1.404								
E1	**1.401**	Max 9	1.398								
		Max 10	**1.395**								

Appendix D: Pollen data – absolute countings

Name	*Full name*	*English name*	*A1*	*A2*	*A3*	*B1*	*B2*	*B3*	*C1*	*D1*	*E1*	*Max 3*	*Max 5*	*Max 6*	*Max 10*	*Tui 1B*	*Tui 2*	*Tui 3*	*Tui 4*	*Min 1*	*Min 2*	*WtdW 1*	*WtdW 2*	*WtdW 3*	*Ame 1*	*Ame 2*	*Ame 3*	*Ame 4*	*Ame 5*	*Ame 6*
Lyc	Lycopodium	wolf's-foot clubmoss	296	120	255	34	44	51	114	45	85	418	199	510	176	128	81	83	21	84	28	148	64	99	161	11	11	50	199	156
Ace	Acer	Maple	0	6	2	0	1	2	0	4	0	0	2	3	0	7	4	2	1	1	0	4	4	0	2	4	21	2	6	2
Aes	Aesculus	Horse Chestnut	2	6	0	5	5	3	3	6	13	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Aln	Alnus	Alder	30	45	65	94	63	81	16	41	52	111	59	87	92	62	90	54	72	127	263	124	88	51	17	23	4	18	29	18
Bet	Betula	Birch	25	15	26	22	26	34	3	22	30	7	6	14	10	24	35	28	30	29	44	14	11	7	9	24	11	10	11	3
Carp	Carpinus	Hornbeam	0	2	0	0	0	1	1	2	2	2	0	0	0	1	3	0	0	1	1	2	0	0	7	54	1	0	0	0
Cas	Castanea sativa	Sweet chestnut	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	9	0	0	0	0
Cat	Catalpa	Catalpa	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Cor	Corylus	Hazel	16	10	23	35	23	33	7	25	16	8	13	26	27	21	36	21	28	50	97	31	15	9	24	42	12	30	22	17
Fagu	Fagus	Common beech	2	4	2	4	2	2	9	7	8	2	3	5	2	6	5	3	8	0	0	12	0	2	48	57	40	8	10	71
Fra	Fraxinus	Common ash	7	15	5	3	4	4	7	5	8	56	19	8	7	9	7	9	4	4	7	19	5	70	50	55	6	26	52	14
Jug	Juglans	Walnut tree	0	1	0	1	1	1	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Lar	Larix	Larix	0	0	0	0	0	0	0	0	0	0	0	0	0	2	9	0	0	0	0	0	0	0	0	0	0	0	0	0
Lon	Lonicera	Honeysuckle	0	0	1	1	0	1	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nys	Nyssa	Tupelo	63	9	15	1	1	1	0	0	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
Ost	Ostrya	Hop-hornbeam	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pic	Picea	Spruce	0	1	1	0	2	2	0	1	0	3	1	0	0	1	1	1	1	1	0	1	0	0	1	1	0	0	1	1
Pic.o	Picea omorica	Serbian spruce	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pin	Pinus	Pine	46	100	40	8	32	17	36	30	24	24	29	27	35	54	31	68	47	10	9	26	29	17	42	23	16	25	62	34
Plat	Platanus	Plane	12	8	4	4	6	4	49	1	6	0	3	0	1	2	0	55	41	0	0	1	1	1	0	2	1	0	1	0
Pop	Populus	poplar	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Pru	Prunus	Fruit trees	0	0	0	1	1	1	2	0	0	1	0	0	0	1	2	5	2	0	0	0	1	1	1	0	0	0	0	0
Pter	Pterocarya	Wingnut	4	2	5	1	0	0	17	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Que	Quercus	Oak	42	28	35	28	24	33	1	27	24	11	24	16	31	26	23	20	25	19	58	38	32	10	28	488	375	172	50	121
Rob	Robinia	black locust	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Sal	Salix	Willow	4	5	0	2	0	1	2	1	0	0	1	3	0	6	3	3	3	0	1	4	4	4	1	5	0	8	15	8
Seq	Sequoia	dawn redwood	1	0	0	1	7	1	1	1	0	1	0	0	0	0	0	0	0	1	0	6	1	1	1	0	0	0	0	0
Taxo	Taxodium	Baldcypress	4	3	2	7	13	7	4	23	12	1	0	6	0	2	2	0	2	0	0	0	3	4	7	0	0	1	0	0
Tax	Taxus	Yew	1	0	0	1	0	0	2	0	7	1	0	0	1	2	8	0	0	0	0	0	0	1	1	0	0	0	0	0
Til	Tilia	Lime	4	7	13	6	6	4	8	8	8	0	2	2	7	2	2	7	10	2	3	5	38	73	1	3	0	1	3	0
Ulm	Ulmus	Elm	8	10	6	6	5	5	8	6	8	1	3	3	4	2	4	4	14	5	10	1	8	0	1	33	0	2	14	2
Eri	Ericales	Erica	7	8	24	17	14	10	0	16	16	6	1	13	5	20	21	10	16	3	0	5	5	4	0	0	0	0	2	1
Cal	Calluna	Common heather	0	1	1	10	0	5	3	4	9	0	4	0	12	2	11	12	0	21	89	0	24	1	0	1	0	1	1	0
Cyp	Cyperacea	Sedges	4	1	2	0	1	0	23	17	13	1	2	0	4	1	0	0	2	0	0	6	4	1	1	3	0	1	4	14
Poa	Poaceae	Grasses	23	48	44	61	54	69	17	56	31	133	130	110	301	18	45	65	26	56	100	112	58	27	43	33	12	33	49	50
Sec	Secale	Rye	2	6	7	6	9	3	0	0	0	3	1	5	6	1	0	3	2	1	1	0	0	0	0	0	0	0	0	0
Art	Artemisia	Artemisia	0	1	1	1	1	0	0	1	0	6	2	1	0	0	0	1	2	0	0	0	2	0	1	2	5	7	3	3
Ast.l	Ast. Liguliflorae	Composite family	5	6	4	1	2	3	2	0	2	38	76	43	11	7	5	2	0	9	1	8	12	3	6	1	0	1	3	1
Ast.t	Ast. Tubuliflorae	Composite family	2	2	4	4	2	1	1	0	1	6	14	10	2	1	0	0	0	9	4	5	1	2	1	1	0	0	1	0
Cen	Ast. Centaurea	Centaury	0	2	4	1	0	0	0	0	0	1	0	3	1	1	1	0	1	1	3	0	0	2	0	0	0	0	1	0
Sau	Ast. Saussurea type	Saussurea	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
aca	acacia mimosa	Mimosa	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Api	Apiacea	Parsley family	1	3	0	0	3	0	0	0	0	1	2	0	4	1	0	4	0	2	0	2	1	2	0	2	0	2	1	1
Apo	Apocynaceae vinca	Periwinkle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Bra	Brassicaceae	Crucifers	6	7	18	2	2	7	0	0	0	6	9	5	1	8	2	3	1	4	1	3	4	0	0	0	0	0	0	9
Bux	Buxus	box	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Cam	Campanulaceae	Bellflower family	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
Car	Caryophyllaceae	Pink family	1	0	0	1	1	0	1	0	2	0	3	1	2	0	1	0	0	2	0	3	1	0	0	1	1	0	1	1
Ste	Stellaria holostea	Addersmeat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Che	Chenopodiaceae	Goosefoot family	2	2	1	1	1	1	1	1	5	7	4	15	4	2	1	1	3	4	4	3	1	1	0	1	2	4	0	
Cir	Circaea lutetiana	Enchanter's-nightshade	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
Epi	Epilobium	Rosebay willowherb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Fab	Fabaceae	Legume family	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fag	Fagopyrum	Buckwheat	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Fil	Filipendula	Filipendula	1	1	0	0	1	1	0	1	0	0	2	1	0	2	0	2	0	0	0	0	1	0	0	0	0	1	0	1
Gal	Galium	Galium	0	2	0	1	0	0	0	0	2	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1	0
Hed	Hedera	Ivy	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ile	Ilex	Holly	2	0	1	0	1	0	5	3	2	1	6	0	2	1	1	1	0	1	0	0	1	0	3	2	1	0	1	3
Lig	Ligustrum	Privet	0	0	0	0	0	0	1	0	2	1	9	1	0	1	0	1	0	0	0	1	1	0	0	2	1	4	0	3
Lil	Lilium	lily	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Lir	Liriodendron	Tuliptree	0	1	0	1	0	0	1	3	2	0	0	3	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0
Lys	Lysimachia vulgaris	Garden loosestrife	0	2	1	0	0	0	0	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Mal	Malva neglecta	Common Mallow	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pla	Plantago	Plantain	0	0	2	0	1	0	0	0	0	2	0	3	4	0	0	0	0	0	0	3	1	0	0	0	1	0	0	0
Pla.l	Plantago lanc.	Ribwort Plantain	0	0	1	0	0	0	1	0	1	3	1	4	19	1	0	0	1	1	1	3	0	0	3	0	1	0	3	0
Poly	Polygonum	Knotweed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Ran	Ranunculus acris-type	Buttercup family	0	0	0	0	0	0	0	0	0	1	1	2	3	0	1	2	1	2	2	11	3	0	1	0	1	0	8	0
Res	Reseda lutea	Yellow mignonette	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Rum	Rumex	Dock & Sorrel	0	0	0	0	1	1	0	0	0	1	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Sca	Scabiosa argentea	Scabiosa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Sci	Sciadopitys verticilata	Japanese umbrella-pine	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sol	Solanum	Tomato	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sym	Symphytum officinale	Comfrey	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Val	Valeriana	Valerian	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vio	Viola arvensis	Viola	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
Vis	Viscum Album	Mistletoe	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
Lud	Ludwigia palustra	Marsh seedbox	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Nup	Nuphar	Water-lily	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
Nym	Nymphaea	Water-lily	3	0	0	0	0	0	3	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Myr	Myriophyllum	Watermilfoil	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pot	Potamogeton	Pondweed	1	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sag	Sagittaria	Arrowhead	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Spa	Sparganium	Bur-reed	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0
Typ	Typha	Bulrush / cattail	1	0	4	0	0	0	1	0	0	0	1	7	2	0	0	2	1	1	2	5	1	3	1	0	1	0	0	1
Dry	Dryopteris	Woodfern	3	4	9	1	9	5	2	5	0	10	3	3	14	1	8	2	8	28	38	41	14	12	1	0	0	2	9	10
Equ	Equisetum	Horsetail	0	0	0	0	0	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pol	Polypodiaceae	Polypodiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
Pte	Pteridium aquilinum	Bracken	1	0	0	0	0	2	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Sph	Sphagnum	Peat moss	0	6	80	2	34	42	0	21	0	10	1	14	46	80	88	47	72	37	224	2	1	3	0	0	0	0	3	0
The	Thelypteris	Maiden ferns	0	1	0	0	1	1	1	2	0	0	0	1	1	0	0	0	0	13	3	1	1	0	0	0	0	0	2	1
ind	indet		24	25	16	15	6	11	9	7	9	7	5	10	29	7	10	4	16	9	20	19	4	22	13	10	6	17	7	5
rew	Reworked		0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	Pollen sum		362	397	381	353	324	351	252	329	320	457	452	428	646	328	372	402	363	376	723	479	373	324	315	881	519	377	370	385
	Total sum		366	408	470	356	368	401	257	358	320	478	456	446	708	409	469	451	443	455	989	523	389	339	316	881	519	379	384	396

Appendix E: Histogram of the Bray-Curtis distances



Appendix F: Script for provenancing

In this script Tree/Plant ratio and TP is equal to AP/NAP-ratio

```
% Code for provenancing of your sample
% Running the code it will ask to provide the values of the sample variable:
% Dominance, Concentration(pollen/g), Diversity (at #252), Tree/Plant ratio and Spore
percentage

Sdom = input('Enter sample dominance: ');
Scon = input('Enter sample concentration: ');
Sdiv = input('Enter sample diversity: ');
STP = input('Enter sample Tree/Plant ratio: ');
Sspo = input('Enter sample spore percentage: ');
Spollensum = input('Enter pollensum exl spores between 141-252: ');

% This results in comparison of your sample with the database values.
% The location with the highest score has the best resemblence with the
% variable of the analyzed sample.

load Locationdata;
load rarefactiondata;

domminvalues = [min(OldPark(:,1)) min(NewPark(:,1)) min(Garden(:,1)) min(Vacant(:,1))
min(Verge(:,1)) min(Forest(:,1))];
dommaxvalues = [max(OldPark(:,1)) max(NewPark(:,1)) max(Garden(:,1)) max(Vacant(:,1))
max(Verge(:,1)) max(Forest(:,1))];
domscore = (domminvalues <= Sdom) == (dommaxvalues >= Sdom);
Dominance = transpose(domscore);

conminvalues = [min(OldPark(:,2)) min(NewPark(:,2)) min(Garden(:,2)) min(Vacant(:,2))
min(Verge(:,2)) min(Forest(:,2))];
conmaxvalues = [max(OldPark(:,2)) max(NewPark(:,2)) max(Garden(:,2)) max(Vacant(:,2))
max(Verge(:,2)) max(Forest(:,2))];
conscore = (conminvalues <= Scon) == (conmaxvalues >= Scon);
Concentration = transpose(conscore);

TPminvalues = [min(OldPark(:,3)) min(NewPark(:,3)) min(Garden(:,3)) min(Vacant(:,3))
min(Verge(:,3)) min(Forest(:,3))];
TPmaxvalues = [max(OldPark(:,3)) max(NewPark(:,3)) max(Garden(:,3)) max(Vacant(:,3))
max(Verge(:,3)) max(Forest(:,3))];
tpscore = (TPminvalues <= STP) == (TPmaxvalues >= STP);
Tree_Plant = transpose(tpscore);

rOldPark = rarefactionOldPark(Spollensum - 140,2:end);
rNewPark = rarefactionNewPark(Spollensum - 140,2:end);
rGarden = rarefactionGarden(Spollensum - 140,2:end);
rVacant = rarefactionVacant(Spollensum - 140,2:end);
rVerge = rarefactionVerge(Spollensum - 140,2:end);
rForest = rarefactionForest(Spollensum -140,2:end);
divminvalues = [min(rOldPark) min(rNewPark) min(rGarden) min(rVacant) min(rVerge)
min(rForest)];
divmaxvalues = [max(rOldPark) max(rNewPark) max(rGarden) max(rVacant) max(rVerge)
max(rForest)];
divscore = (divminvalues <= Sdiv) == (divmaxvalues >= Sdiv);
Diversity = transpose(divscore);

spominvalues = [min(OldPark(:,5)) min(NewPark(:,5)) min(Garden(:,5)) min(Vacant(:,5))
min(Verge(:,5)) min(Forest(:,5))];
spomaxvalues = [max(OldPark(:,5)) max(NewPark(:,5)) max(Garden(:,5)) max(Vacant(:,5))
max(Verge(:,5)) max(Forest(:,5))];
sposcore = (spominvalues <= Sspo) == (spomaxvalues >= Sspo);
Spores = transpose(sposcore);

score = [domscore;conscore;tpscore;divscore;sposcore];

Tscore = transpose(sum(score));

%output

Endscore = table(Class,Dominance,Concentration,Tree_Plant,Diversity,Spores,Tscore);

output = Endscore
```

Appendix G: Absolute countings of additional samples

Coloured are the pollen types which are not present in the original dataset

Name	Full name	English name	1 meadow	2 vacant	3 wildered	4 wildered	5 wildered	6 wildered	7 old park	8 old park	9 old park	10 old park
Lyc	Lycopodium	Wolf's-foot clubmoss	-	238	97	117	102	80	143	137	45	32
Abi	Abies	Fir	0	3	0	1	0	0	0	1	0	0
Ace	Acer	Maple	0	0	1	0	0	3	47	54	6	2
Aes	Aesculus		0	0	0	0	0	0	3	7	67	43
Aln	Alnus	Alder	0	0	23	28	75	36	15	16	10	16
Bet	Betula	Birch	6	7	14	10	17	19	7	5	8	15
Carp	Carpinus	Hornbeam	0	1	0	1	3	1	2	0	1	1
Cas	Castanea sativa	Sweet chestnut	0	0	0	0	0	0	0	0	0	0
Cat	Catalpa	Catalpa	0	0	0	0	0	0	0	0	0	0
Ced	Cedrus	Cedar	0	0	0	0	0	0	32	53	8	6
Cor	Corylus	Hazel	0	30	25	14	55	41	4	1	6	2
Fagu	Fagus	Common beech	0	5	4	4	3	3	4	3	25	33
Fra	Fraxinus	Common ash	0	0	1	3	0	1	1	0	3	1
Jug	Juglans	Walnut tree	0	0	0	0	0	0	0	0	1	1
Jun	Juniperus	Juniper	0	0	2	0	0	0	0	0	1	0
Lar	Larix		0	0	0	0	0	0	0	0	0	0
Lon	Lonicera	Honeysuckle	0	1	1	0	0	0	0	0	0	0
Nys	Nyssa	Tupelo	0	0	0	0	0	0	0	0	0	0
Ost	Ostrya	Hop-hornbeam	0	0	0	0	0	0	0	0	0	0
Pic	Picea	Spruce	1	1	4	2	1	1	1	2	3	1
Pic.o	Picea omorica	Serbian spruce	0	0	0	0	0	0	0	0	0	0
Pin	Pinus	Pine	143	0	21	18	16	15	11	13	20	12
Plat	Platanus	Plane	0	0	0	0	0	0	0	1	0	1
Pop	Populus	poplar	0	0	0	0	0	0	0	0	0	0
Pru	Prunus	Fruit trees	0	1	0	0	0	0	3	0	0	0
Pter	Pterocarya	Wingnut	0	0	0	0	0	0	0	3	13	18
Que	Quercus	Oak	0	5	16	17	35	19	28	13	24	33
Rob	Robinia	Black locust	0	0	0	0	0	0	0	0	0	0
Sal	Salix	Willow	1	1	0	1	1	0	1	1	2	1
Seq	Sequoia	Dawn redwood	0	0	0	0	0	0	0	0	0	0
Taxo	Taxodium	Baldcypress	0	0	0	0	0	0	0	0	0	0
Tax	Taxus	Yew	0	0	0	0	0	0	0	0	0	0
Til	Tilia	Lime	0	2	0	1	1	1	10	4	5	4
Ulm	Ulmus	Elm	0	1	0	0	2	4	1	0	1	4
Eri	Ericales	Erica	13	87	16	7	33	23	1	0	2	0
Cal	Calluna	Common heather	0	0	0	0	0	0	0	0	0	0
Cyp	Cyperacea	Sedges	0	9	18	8	14	8	0	0	1	0
Poa	Poaceae	Grasses	55	20	19	18	13	46	4	7	11	20
Sec	Secale	Rye	3	17	3	1	3	4	0	1	1	2
Art	Artemisia	Artemisia	0	1	0	0	1	1	0	0	1	0
Ast.l	Ast. Liguliflorae	Composite family	1	7	5	1	0	3	3	5	1	1
Ast.t	Ast. Tubuliflorae	Composite family	0	5	1	0	0	1	3	1	2	2
Cen	Ast. Centaurea	Centaury	0	0	0	0	0	0	0	0	0	0
Sau	Ast. Saussurea type	Saussurea	0	0	0	0	0	0	0	0	0	0
aca	acacia mimosa	Mimosa	0	0	0	0	0	0	0	0	0	0
Api	Apiacea	Parsley family	0	0	0	0	0	0	0	0	0	0
Apo	Apocynaceae vinca	Periwinkle	0	0	0	0	0	0	0	0	0	0
Bra	Brassicaceae	Crucifers	3	1	1	0	0	2	2	1	0	2
Bux	Buxus	Box	0	0	0	0	0	0	0	0	0	0
Cam	Campanulaceae	Bellflower family	0	0	0	0	0	0	0	0	0	0
Car	Caryophyllaceae	Pink family	0	4	0	0	3	1	0	1	0	0
Ste	Stellaria holostea	Addersmeat	0	0	0	0	0	0	0	0	0	0

Name	Full name	English name	1 meadow	2 vacant	3 wildered	4 wildered	5 wildered	6 wildered	7 old park	8 old park	9 old park	10 old park
Che	Chenopodiaceae	Goosefoot family	0	3	0	2	1	7	0	1	1	1
Cir	Circaea lutetiana	Enchanter's-nightshade	0	0	0	0	0	0	0	0	0	0
Epi	Epilobium	Rosebay willowherb	0	0	0	0	0	0	0	0	0	0
Fab	Fabaceae	Legume family	0	0	0	0	0	0	0	0	0	1
Fag	Fagopyrum	Buckwheat	0	0	0	0	0	0	0	0	0	0
Fil	Filipendula	Filipendula	0	0	0	0	0	0	0	0	0	0
Fra	Frangula alnus	Alder buckthorn	0	0	0	0	0	0	0	0	0	1
Gal	Galium	Galium	0	0	0	0	0	0	0	0	0	0
Hed	Hedera		0	0	0	0	0	1	0	0	0	0
Hip	Hippophae rhamnoide	Common sea buckthorn	0	0	0	0	0	0	0	0	1	2
Ile	Ilex	Holly	0	0	0	0	0	0	0	0	0	0
Lig	Ligustrum	Privet	0	0	0	0	0	0	0	1	1	1
Lil	Lilium	Lily	0	0	0	0	0	0	0	0	0	0
Lir	Liriodendron	Tuliptree	0	0	0	0	0	0	0	0	0	0
Lys	Lysimachia vulgaris	Garden loosestrife	0	0	0	0	0	1	0	0	0	0
Mal	Malva neglecta		0	0	0	0	0	0	0	0	0	0
men	Mentha-typ	Mint family	0	1	0	0	0	0	0	0	0	0
Myr	Myrica gale	Bog-myrtle	0	5	1	2	2	2	0	0	0	0
Pla	Plantago	Plantain	0	0	0	0	0	0	0	0	0	0
Pla.l	Plantago lanc.	Ribwort Plantain	0	0	0	0	0	0	0	1	2	0
Poly	Polygonum	Knotweed	0	0	0	0	0	0	0	0	0	0
Pot	Potentilla	Cinquefoils	0	0	0	1	0	0	0	0	1	1
Ran	Ranunculus acris-type	Buttercup family	0	1	2	0	0	1	0	1	0	2
Res	Reseda lutea	Yellow mignonette	0	0	0	0	0	0	0	0	0	0
Rum	Rumex		0	0	0	0	1	0	0	0	1	1
Sam	Sambucus nigra	Elder	0	0	1	1	0	6	0	0	0	0
Sca	Scabiosa argentea	Scabiosa	0	0	0	0	0	0	0	0	0	0
Sci	Sciadopitys verticilata	Japanese umbrella-pine	0	0	0	0	0	0	0	0	0	0
Sol	Solanum		0	0	0	0	0	0	0	0	0	0
Sym	Symphytum officinale	Comfrey	0	0	0	0	0	0	0	0	0	0
Val	Valeriana		0	0	0	0	0	0	0	0	0	0
Vib	Viburnum	Viburnum	0	0	0	0	0	0	0	1	0	0
Vio	Viola arvensis		0	0	0	0	0	0	0	0	0	0
Vis	Viscum Album	Mistletoe	0	0	0	0	0	0	0	0	0	0
Lud	Ludwigia palustra	Marsh seedbox	0	0	0	0	0	0	0	0	0	0
Nup	Nuphar	Water-lily	0	1	0	0	0	0	0	0	0	0
Nym	Nymphaea	Water-lily	0	0	0	0	1	0	0	0	0	0
Myr	Myriophyllum	Watermilfoil	0	0	0	0	0	0	0	0	0	0
Pot	Potamogeton	Pondweed	0	1	0	0	0	1	0	0	0	0
Sag	Sagittaria	Arrowhead	0	0	0	0	0	0	0	0	0	0
Spa	Sparganium	Bur-reed	0	5	0	0	0	1	0	0	0	0
Typ	Typha	Bulrush / cattail	0	0	1	0	0	0	0	0	0	0
Dry	Dryopteris		0	29	0	0	0	0	0	0	0	0
Equ	Equisetum	Horsetail	0	0	0	0	0	0	0	0	0	0
Pol	Polypodiaceae	Polypodiaceae	0	0	0	0	0	0	0	0	0	0
Pte	Pteridium aquilinum	Bracken	0	1	0	0	0	0	0	0	0	0
Sph	Sphagnum	Peat moss	0	32	0	0	0	0	0	0	0	0
The	Thelypteris	Maiden ferns	0	0	0	0	0	0	0	0	0	0
ind	indet		0	0	0	0	0	0	0	0	0	0
rew	Reworked		0	0	0	0	0	0	0	0	0	0
	Pollen sum		226	226	180	141	281	253	183	198	230	231
	Total sum			288								

Appendix H: Variable values of additional samples

	1 meadow	*2 vacant*	*3 wildered*	*4 wildered*	*5 wildered*	*6 wildered*	*7 old park*	*8 old park*	*9 old park*	*10 old park*
Dominance	0.464	0.137	0.095	0.109	0.151	0.106	0.138	0.166	0.128	0.103
Concentration	-	-	-	-	-	-	-	-	-	-
AP/NAP-Ratio	2.013	0.345	1.647	2.439	2.903	1.321	13.077	8.429	7.846	5.243
Diversity (pollen sum)	9 (226)	29.36 (252)	22 (180)	21 (141)	20.25 (252)	28.95 (252)	21 (183)	26(198)	31 (230)	31 (231)
Spores %	0*	27.43	0*	0*	0*	0*	0*	0*	0*	0*

*Not counted; though possibly present in the assemblage