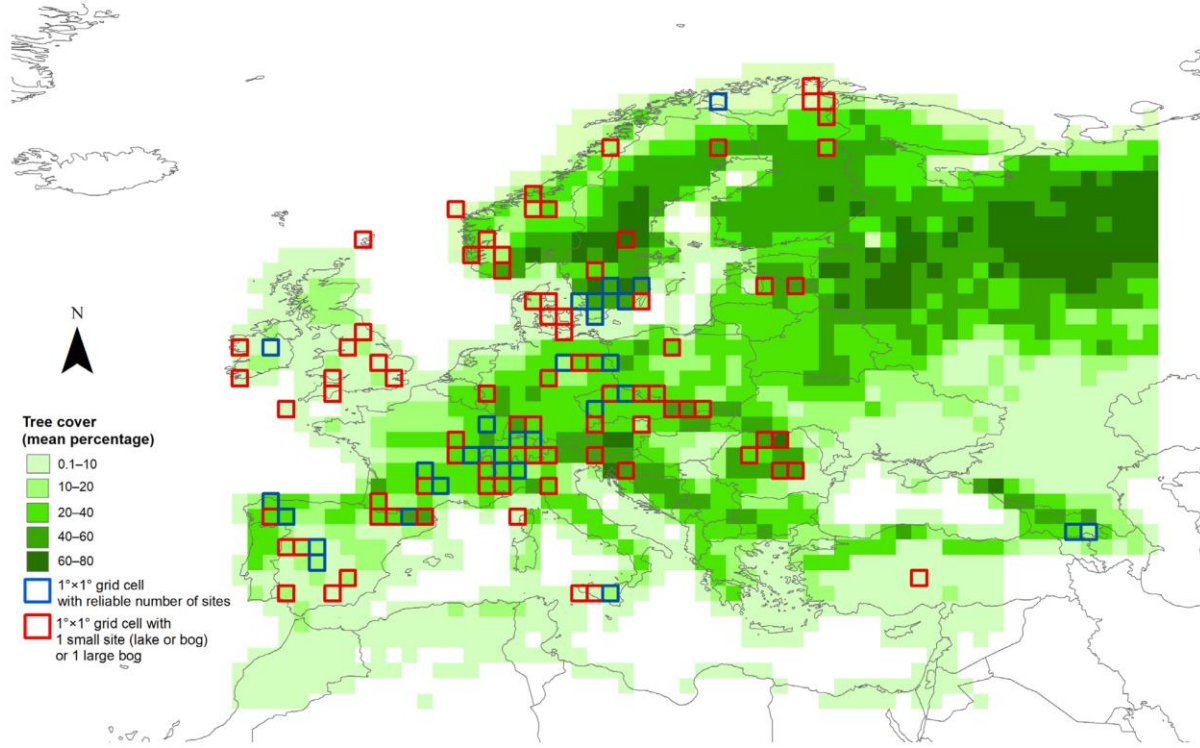


A review of pollen-based quantitative vegetation reconstruction models



Grid cells with sites (bogs and lakes) used for validation and tree cover at 2000 AD (Serge et al., 2023)

T.E.P. (Tobias) Vervaart

6543324

MSc Environmental Biology (Paleoecology)

Writing assignment

7.5 EC

22-11-2024

Examiner: Thomas Giesecke

Second reviewer: Derek Karsseberg

Abstract

Quantitative vegetation reconstruction aims to provide information about past vegetation cover. One method is using pollen diagrams that are supposed to reflect the relative abundance of taxa in the surrounding vegetation. However, the relationship between pollen and vegetation is affected by production and dispersal bias. Therefore, a need for methods arose that could correct for these biases. To quantify the effect of production bias, estimates for pollen productivity (PPEs) were developed. Secondly, to account for dispersal bias, accurate pollen dispersal models that described the movement pattern of pollen grains with different physical characteristics were needed.

This review first presents an overview of the development of dispersal and deposition models from the basic idea of the R-value model by Margaret Davis to the Extended R-value model by Colin Prentice that was further developed by Shinya Sugita. It continues with a reconstruction of the development of PPEs and dispersal models that should correct for production and dispersal bias is given. Lastly, three current quantitative vegetation reconstruction approaches (the Landscape Reconstruction Algorithm (LRA), the Multiple Scenario Approach (MSA) and the Extended Downscaling Approach (EDA)) are compared in the context of a theoretical intermediate scale landscape reconstruction in the Netherlands for the Lateglacial and the Holocene.

While the LRA has been often applied on regional-continental scales and sporadically on local scales, the MSA and EDA have been only applied on local scale reconstructions. The LRA reconstructs regional landscapes based on the pollen assemblages of large lakes (REVEALS) and local landscapes on the pollen assemblages of small lakes (LOVE). The MSA and EDA reconstruct landscapes based on the input of abiotic landscape parameters and the simulation of theoretical pollen assemblages that are then compared to the empirical assemblage. The MSA can yield multiple likely landscape scenarios, whereas the EDA only reconstructs one scenario using an optimization key.

In the light of a potential landscape reconstruction of the Dutch Lateglacial and Holocene, application of the LRA could be impractical due to the low availability of large lakes (100-500ha). The availability of high-quality abiotic landscape information makes the application of the MSA, and the EDA seem quite suitable, but both methods have not yet been applied on intermediate scales and scaling up may be bothered by high simulation times, especially for the MSA.

Layman summary

Quantitative vegetation reconstruction is a field within palaeoecology that aims to obtain information about plant abundances for time periods in the past and quantify this information, which may be used in maps of past vegetation cover. The main method for this is the analysis of pollen grains deposited in lake or bog sediments. Pollen grains can be identified under a microscope into different taxonomic groups of plants. Counting the grains to a fixed number and tallying the counts of the separate groups yields proportions for each pollen group and therewith, a general estimation of the vegetation cover can be made by constructing a pollen diagram, in which the relative abundances of all pollen groups are displayed. However, the pollen proportions do not equal vegetation proportions, because plant species produce pollen in different rates and disperse them with varying effectiveness. To reconstruct whole landscapes, we need models that correct for these differences. For that, we need measurements of pollen production for different species, and we need models that are able to describe the pollen dispersal pattern. We also need information about the area for which the pollen found in the sediment are representative.

In the first section of this review, the development of correction models for these pollen percentage models is outlined. The second section more closely reviews methods to deal with differential pollen production and dispersal. The last section then reviews three current models that make quantitative reconstructions of past vegetation.

One of the models (the Landscape Reconstruction Algorithm (LRA)) reconstructs the vegetation based on the pollen assemblages of both large basins (coarse scale) and small basins (fine scale). The other two models (the Multiple Scenario Approach (MSA) and the Extended Downscaling Approach (EDA)) first reconstruct one (EDA) or more (MSA) possible landscape(s) and simulate theoretical pollen compositions for these theoretical landscapes. Then they compare their simulated pollen compositions with the actual pollen composition for that time and select landscapes based on the highest similarity.

For a landscape reconstruction in the Netherlands, the LRA may not be the best choice, because there are few large basins available for the first step of this approach. On the other hand, there is a lot of environmental data available that can be used as input for both the MSA and EDA. The main bottleneck for these methods is that they have only been applied on smaller scales, and upscaling may be hindered by computational performance of the simulations.

Contents

Abstract.....	2
Layman summary.....	3
Introduction	5
Chapter 1: The development of the ERV-models	7
1.1 The R-value model	7
1.1.1 Theory	7
1.1.2 Application	9
1.2 The Extended R-value model	9
1.3 Pollen source area and basin size	12
1.4 Pollen source area and basin type	13
1.5 Relevant source area and vegetation heterogeneity	15
Chapter 2: Production and dispersal bias	17
2.1 Production bias	17
2.2 Dispersal bias	18
2.2.1 Dispersal models	19
Chapter 3: Comparing the reconstruction models	21
3.1 Landscape Reconstruction Algorithm	21
3.1.1 REVEALS	21
3.1.2 LOVE	23
3.2 Multiple Scenario Approach	25
3.3 The Extended Downscaling Approach	27
Discussion.....	30
Production & dispersal bias	30
Intermediate scale landscape reconstruction	31
Conclusion.....	32
Literature	33

Introduction

Modern pollen analysis was developed approximately 100 years ago by Lennart von Post. Originally a geologist, through contact with botanist Rutger Sernander von Post became interested in peat stratigraphy and Holocene climate development (Birks & Berglund, 2017). Inspired by pioneering work in quantitative palynology (Weber, 1893, in: Birks & Berglund, 2017) and pollen observations in peat (Lagerheim, 1902, in: Birks & Berglund, 2017), Von Post decided to use fossilized pollen material preserved in peat as a tool for synchronizing the local stratigraphies of peatlands in Sweden by calculating pollen percentages over the stratigraphic series (Birks & Berglund, 2017). He concluded that the changes in fossil pollen percentages must reflect changes of the regional vegetation over time (Davis, 1963).

The first pollen diagrams were presented during his lecture at the Scandinavian Meeting of Natural Scientists (von Post, 1916). He visualised the stratigraphic developments as 'closed' percentage data, meaning that the pollen percentages are calculated as the pollen counts from each taxon¹ divided by the pollen sum - the total of counts - and then plotted against the sample depths. The interpretation of the diagrams led to a discussion. von Post argued that the changes in pollen percentages over a stratigraphic series corresponded with vegetational changes, both in direction as in magnitude. Henrik Hesselman, a forester attending the meeting, countered that this linearity could not be assumed, because the pollen percentages are relative representations of abundance; any changes in them could be caused by **interdependence** (Hesselman, 1916, in: Davis, 1963). In other words, a change in the abundance of taxon B also affects the abundance of taxon A in a percentage diagram, even though the absolute abundance of taxon A does not necessarily have to be changed (Davis, 1963).

While the problem of interdependence is only inherent to pollen percentage data, two other biases are fundamental to pollen analysis in general; **production bias** and **dispersal bias** (Prentice, 1985). Production bias is caused by differential pollen production between plant species related to i.e. the pollination syndrome. Plants that disperse their pollen primarily through wind often produce large quantities of pollen, whereas plants who use animals as a vector for pollination have a lower pollen production (e.g. Rempe, 1937). But even within species, pollen production can differ substantially depending on environmental conditions; for instance free standing trees often produce more pollen than trees in closed stand (e.g. Iversen, 1941). Dispersal bias is associated with differences in pollen grain morphology. Factors such as grain shape, size and weight influence the fall speed and thus the distance a pollen grain can travel before being deposited (Tauber, 1965). As a result, production and dispersal bias can cause over- and underrepresentation of certain pollen taxa in a pollen diagram in comparison to the actual vegetation cover. To maximize the representativity of pollen taxa in a pollen diagram to the (past) vegetational abundance, production, and dispersal bias need to be accounted for (Tauber, 1965).

One possibility for this is through pollen deposition models that aim to correct for biases in pollen deposition (Birks & Berglund, 2017). The R-value model by (Davis, 1963) incorporated both production and dispersal bias into a single representativity factor, the R-value. The Extended R-value model (ERV; Prentice, 1985), enabled production and dispersal biases to be accounted for separately. Production and dispersal bias can be accounted for in different ways. Production bias is often measured by the calculation of pollen productivity estimates (PPE's), but the process is laborious, and the estimates are not easily generalized (e.g. Fang, 2019). There are various ways to

¹ Pollen identification is based on morphological types that are often only identifiable on the genus or family level, therefore the term 'taxon' or 'pollen type' is most appropriate.

measure dispersal bias, for instance with dispersal models such as the Gaussian Plume model (Prentice, 1985; Sutton, 1947, in *inter alia*: Tauber, 1965) or the Langrangian Stochastic model (e.g. Kuperinen et al., 2007), but also other methods such as following gene flow through paternity assignment are possible (Butcher et al. (2020).

Present day, multiple models and techniques are available for quantitative vegetation reconstruction. However, their application so far has been limited, especially on intermediate landscape scales such as small countries like the Netherlands. This paper aims to review three approaches: (1) *the Landscape Reconstruction Algorithm (LRA)* (Sugita, 2007a, b), (2) *the Multiple Scenario Approach (MSA)* (Bunting & Middleton, 2009) and *the Extended Downscaling Approach (EDA)* (Theuerkauf & Couwenberg, 2017). First, an overview of the development of pollen deposition models based on the R-value model of Davis (1963) will be presented. Secondly, the most important biases causing differential pollen deposition, production, and dispersal bias, will be highlighted. Lastly, the (dis)advantages in terms of functionality and applicability of the abovementioned models will be discussed in the light of an (intermediate scale) potential landscape reconstruction of the Netherlands.

Chapter 1: The development of the ERV-models

1.1 The R-value model

1.1.1 Theory

During the early days of pollen analysis (ca. 1920-1950), it became clear that the relationship between pollen percentages and vegetation abundance was often not linear and that biases were involved. To derive relevant interpretations on vegetation development from pollen diagrams, it was therefore necessary to identify and correct for these biases (Fagerlind, 1952). This ultimately led to the development of **representation factors** or **R-values**. Mathematically, an R-value reflects the ratio of pollen proportion to vegetation cover for every counted taxon in a pollen sample (Davis, 1963):

$$R_a = \frac{\text{Species } a \text{ pollen percentage}}{\text{Species } a \text{ vegetational percentage}} \quad (1)$$

One of the assumptions is that pollen percentages depend on (1) the pollen production and (2) the dispersal effectiveness of each individual species (Davis, 1963). However, in pollen percentage diagrams, the relative abundances are not only dependent on the species' own production and dispersal effectiveness, but also on those of the other species (Prentice & Webb, 1986).

This leads to interesting considerations regarding under- and overrepresentation. The contemporary thought was that species with high pollen production and/or high dispersal effectiveness were always overrepresented in pollen diagrams (Davis, 1963). However, when regarding example 1 and 2 in [Table 1](#), species *c* remains present at the same vegetational abundance, but it is proportionally more abundant in the pollen record in example 2, because the abundance of the massive pollen producer species *a* is halved. Secondly, when looking at species *b*, it is equally represented in example 1, overrepresented in example 2 and underrepresented in example 3, which nuances the idea of general under- and overrepresentation. Although taxa with high pollen productivity may appear in higher abundances in a pollen diagram than their vegetation abundance suggests, their eventual representation is also dependent on the other pollen contributors in the sample (Davis, 1963).

Table 1: Hypothetical comparisons of vegetational percentages and the percentages. From Davis (1963)

	1 Vegetational percentage	2 Pollen percentage	3 R (ratio of pollen percentage to vegetational percentage)	4 $R_a:R_b:R_c$
Example 1.				
Species <i>a</i>	40%	80%	2	10
Species <i>b</i>	10%	10%	1	5
Species <i>c</i>	50%	10%	0.2	1
	<hr/> 100%	<hr/> 100%		
Example 2.				
Species <i>a</i>	20%	50%	2.5	10
Species <i>b</i>	30%	37.5%	1.25	5
Species <i>c</i>	50%	12.5%	0.25	1
	<hr/> 100%	<hr/> 100%		
Example 3.				
Species <i>a</i>	68%	85%	1.25	10
Species <i>b</i>	22%	13.75%	0.625	5
Species <i>c</i>	10%	1.25%	0.125	1
	<hr/> 100%	<hr/> 100%		

Because of this interdependence the individual R-values could not be easily generalized for multiple sites. However, since it is also assumed that the number of pollen grains produced by each species and dispersed from the plants is directly proportional to the number of plants of that species, it follows that if the plants *a*, *b* & *c* occurred at an equal ratio, the ratio of their respective pollen types would be a constant ratio $R_a:R_b:R_c$ (Table 1). The ratio $R_a:R_b:R_c$ can be calculated by measuring modern forest cover and comparing the vegetation percentages with pollen percentages from surface samples. The ratio between the individual R-values then determines the ratio $R_a:R_b:R_c$ (Table 1). With the R-value ratio calculated, one now can assign vegetation cover percentages from fossil pollen material. By dividing the pollen counts (or pollen percentages) by the R-value ratio, one yields a 'corrected percentage' which corresponds after conversion to percentages of the vegetation proportions (Table 2). These **relative R-values** (Rrel) then are more useful for between-site comparisons (Parsons & Prentice, 1981).

Table 2: Hypothetical example of objective interpretation of fossil pollen percentages by correction for differences in representation. From Davis (1963)

	1 Fossil pollen percentage (or number counted)	2 $R_a:R_b:R_c$	3 Corrected relative number $\left(\frac{\text{pollen number}}{R}\right)$	4 Corrected percentage= vegetational percentage
Species <i>a</i>	64	10	6.4	32%
Species <i>b</i>	28	5	5.6	28%
Species <i>c</i>	8	1	8	40%
			<hr/> Total	<hr/> 100%
			20.0	

1.1.2 Application

The R-value correction method was applied to a pollen diagram from Brownington Pond, Vermont, describing the North American 'Pine Zone' and a modern surface sample from the same area (Davis, 1963). The correction method yielded vastly different conclusions on the abundance of some taxa than previous research. For instance, even though pine often contributed more than 50 percent of all tree and shrub pollen, their vegetational abundance using Davis's R-value method was calculated to be only a few percent. Conversely, the abundances of larch, fir, maple, and poplar appeared to be strongly underestimated by previous researchers based only on their pollen percentages.

However, the calculation of R-values for either high pollen contributors (e.g. pine, oak) and low pollen contributors (e.g. larch, fir, maple) was statistically problematic, because they yielded extreme R-values with high standard errors that affected the reliability of the relative R-values of all taxa (Davis, 1963; Parsons et al., 1983). Davis manually adjusted the extreme R-values, but this made the R-values also more subjective. One solution for low pollen contributors (low pollen proportion compared to actual regional vegetation abundance) would be to omit them from the analysis. This significantly improved all R-value estimates in Brownington Pond (Parsons et al., 1983). The unusually high R-values (high pollen proportion compared to actual regional vegetation abundance) from e.g. pine and oak in Brownington Pond may have been caused by pollen coming from outside the area in which the vegetation survey was conducted. This way the surveyed vegetation does not match the vegetation that contributed to most of the pollen in the sample. This could be alleviated by calculating a source area that most accurately reflects the average regional vegetation (Parsons & Prentice, 1981).

1.2 The Extended R-value model

The Extended R-value (ERV) model builds on the theoretical R-value framework as proposed by Davis (1963). Rapid developments in computer science in the 1980s opened new possibilities for vegetation reconstructions using simulations and model building. Before the development of the ERV models, the 'modern analogue' technique was often used. Here, fossil pollen spectra are compared to modern pollen spectra to find a best fit (Parsons & Prentice, 1981). Some researchers preferred this method over the R-value model because they saw too many insuperable problems (e.g. Faegri, 1966). However, others identified the modern analogue method as rather qualitative, despite the potential application of simulations and numerical techniques (Parsons & Prentice, 1981). The latter authors therefore advocated the further development of correction methods, pioneered by Davis (1963) and Andersen (1970). The basal formula for the R-value correction model is

$$y_{ik} = \alpha_i x_{ik} \quad (2)$$

where y_{ik} is the pollen deposition rate of taxon i at site k , α_i a representation factor for taxon i and x_{ik} the quantity of taxon i in the vegetation at site k (Parsons & Prentice, 1981). However, the parameters in eq. 2 yield absolute values that are only applicable to pollen influx data (absolute pollen analysis). In percentage modelling, these parameters are relativised to

$$R_{ik} = p_{ik} / u_{ik} \quad (3)$$

consisting of the representation factor R_{ik} as the proportional value of α_i . y_{ik} is displayed as p_{ik} , representing the pollen proportion of taxon i at site k , whereas x_{ik} is displayed as u_{ik} , being the vegetation proportion of taxon i at site k .

The Extended R-value model was designed to account for three inherent problems with R-value modelling (Parsons & Prentice, 1981):

1. Estimating R-values from more than one site

When comparing R values from many sites, one can choose to simply average the R values. However, while this preserves the R-value ratio, the average R-value is prone to be disproportionately influenced by a few sites with abnormally low vegetation percentages. On the other hand, *a priori* averaging the pollen and vegetation spectra is less sensitive for inaccuracy but fails to preserve the ratios of R-values (Parsons & Prentice, 1981).

Maximum Likelihood (ML) estimates of R-values could alleviate this problem. The ML estimate is based on a multinomial probability distribution and gives an estimate of the 'underlying' value that has most likely produced the observed values. In this sense, it can be considered a 'weighted average R-value' because it preserves the R-value ratios, while the sites that make the most statistically reliable contributions are weighed the heaviest. This approach is based on the basic R-value model ($y_{ik} = \alpha_i x_{ik}$) from Davis (1963) but slightly adapted. The estimates of α that yield the maximum of the ML function can be extracted and represent the R-values, while through this method also the standard deviations of the estimates can be calculated (Parsons & Prentice, 1981).

2. Taking account of the 'background'

Vegetation is often not homogeneously distributed around a sampling site, which means that the pollen spectra reflect the 'averaged vegetation' around a site. It is therefore important that the source area around the sampling site is appropriate. Long-distance transport can seriously inflate the R-value of pollen taxa that are not common in the regional vegetation composition, so the source area should not be too small (Tauber, 1965; Faegri, 1966). But a source area that is too large does not account for the less well dispersed taxa nearer to the sampling site (Parsons & Prentice, 1981)

The solution in the ERV model is incorporating a background component w_i into a slightly more flexible pollen-vegetation model based on the model from Andersen (1970) (Parsons & Prentice, 1981; Prentice & Webb, 1986):

$$y_{ik} = \alpha_i x_{ik} + w_i \quad (4)$$

It is built on the assumption that the amount of background pollen (w_i) divided by the amount of immediate source pollen (y_{ik}) is a constant that is only a function of taxon i . The assumption is valid if the total absolute pollen deposition is constant between sites and is approximately valid as long as the amount of background pollen is minor compared to the amount of immediate source pollen ($w_i \ll y_{ik}$). The results showed that a clear differentiation between the regional and extraregional contribution of main pollen taxa can be established using a ML method (Parsons & Prentice, 1981).

3. The closed sum problem

It can be expected that there are variations in representation values from site to site, which means that these variations likely also occur in time. The solution of Davis (1963) and Andersen (1970) was to relativise the R-values to a reference taxon, under the assumption that the R-values relative to each other are constant. High deviations in R_{rel} for taxon i may then be caused by real variations in representation (α). However, it cannot simply be excluded that differences occur due to a certain amount of sampling error and/or interdependencies between taxa, because the ratios in the percentage diagram must add up to 100% (the closed sum problem).

To quantify spatial differences between R-values, a multivariate analysis (e.g. cluster analysis

or principal component analysis) can be applied where the entirety of R-values at a certain site (an R-spectrum) is compared with R-spectra of other sites based on their similarity. This method was able to identify geographic variations in R-spectra and identify taxa that profoundly determined the grouping pattern (Parsons & Prentice, 1981).

The closed sum problem relates to the expectation of nonlinearity caused by interdependencies among taxa, also called the *Fagerlind effect* (Prentice & Webb, 1986). Nonlinearity here means that changes in the pollen diagram are not always linearly (or even monotonously) related to vegetational changes (Fagerlind, 1952). [Figure 1](#) shows that this effect is expected to be especially strong for either very strong or very weak pollen contributors. If the α -diversity² of contributing taxa is high (no single taxon exceeds 20-30% abundance) the Fagerlind effect is expected to be negligible (Webb et al., 1981). Nevertheless, the ERV-model implements a parameter to correct for the Fagerlind effect:

$$p_{ik} = \alpha_i v_{ik} f_k + z_i \quad (5)$$

where z_i is the background component and f_k is a site-specific term that is dependent on abundance-weighted slope coefficients ($\alpha_j v_{jk}$) of all taxa represented in the vegetation, which is used for modelling the Fagerlind effect (Prentice & Webb, 1986; Jackson et al., 1995). In a theoretical and empirical exploration, it was found that the Fagerlind effect indeed caused tendencies to nonlinearity for taxa with extreme relative representation coefficients (Prentice & Webb, 1986). However, the overall magnitude of the Fagerlind effect was found to be rather moderate (Prentice & Webb, 1986) and the expected nonlinearity was not seen in weak pollen producers (Calcote, 1995). Nevertheless, ERV models allow for the identification of the Fagerlind effect and its magnitude and separating it from other sources of variation in pollen spectra (Jackson et al., 1995)

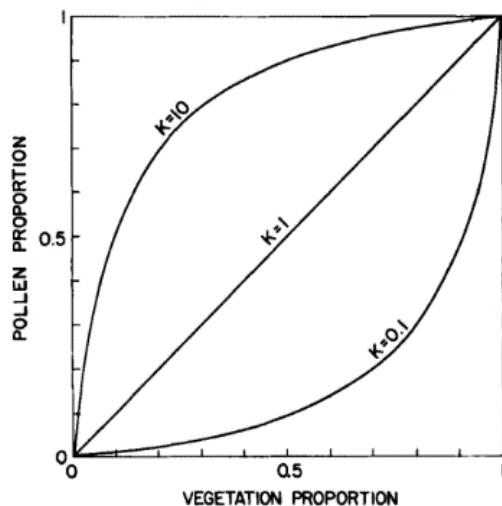


Figure 1: The Fagerlind effect with two taxa. The graph shows that the relationship between the pollen and vegetation proportion starts to deviate from linearity if a taxon 1 produces K times as much pollen as a taxon 2 (in this example graphed for $K = 10, 1$ and 0.1). If the number of taxa increases, the relationship becomes less straightforward, because the different abundances of all taxa cause scatter. Nevertheless, high pollen producers should yield curves similar to $K=10$, whereas low pollen producers should show curves more similar to $K=0.1$, unless the proportion of the taxon in the vegetation remains below 30%; then the curves deviate only minimally from linearity. From Prentice & Webb (1986)

² Not to confuse with the pollen representation factor (α). α -diversity is a biological metric of diversity that includes the number of taxonomic groups as well as the distribution of their abundances (Willis, 2019).

1.3 Pollen source area and basin size

Jacobson & Bradshaw (1981) defined the pollen source area as the area from which a fixed percentage of the pollen sampled at a site originate from. Due to differential dispersion effectiveness of different pollen types, the range of dispersal varies per taxon. Prentice (1985) simulated the source area for individual taxa by varying the pollen deposition velocity v_g (which in general is higher for heavier and/or larger pollen types). It followed that increasing deposition velocities decreased the potential dispersal area. When pollen types with various deposition velocities occur in the same pollen spectra, one can assume that these taxa have different source areas (Theuerkauf et al., 2013). Prentice (1985) also reasoned that the pollen source area was dependent on the basin size of the sampling point. Both assumptions are exemplified in the simulation shown in [Figure 2](#). Here, the lake radius is plotted against the percentage of pollen originating from different distances under different deposition velocities (v_g/u). It is shown that light pollen types (upper panel) have larger source areas compared to heavy pollen types (lower panel). For instance, when the lake radius is 100m, about 75% of pollen from a pollen type with $v_g/u = 0.005$ originate from beyond a 2km radius from the lake, whereas this proportion is about 50% with $v_g/u = 0.01$, about 25% when $v_g/u = 0.02$ and less than 10% when $v_g/u = 0.04$. Furthermore, [Figure 2](#) shows that the individual source areas are expected to increase with basin size and that this effect is more pronounced for heavy pollen types than for light pollen types (Prentice, 1985). Simulations by Sugita (1993) predict that the differences in deposition velocity between pollen types could lead to differences in source radius between the light and heavy pollen types by a factor 100 ([Figure 5](#)).

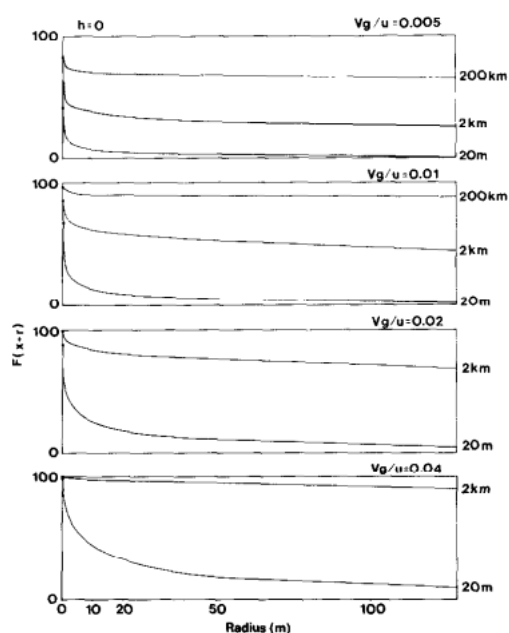


Figure 2: Simulated pollen source areas for individual taxa. The graphs form a series from light pollen types ($v_g/u = 0.005$) to heavy pollen types ($v_g/u = 0.04$). The graphs show the proportion of pollen (left y-axis) originating from given distances (right y-axis) under different basin radii (x-axis). It is shown that source area for light pollen types is much larger than for heavy pollen types. For instance in the upper graph, the proportion of pollen originating from within 2km is around 30% (basin radius = 100m), whereas for the heaviest pollen type (lower graph) and the same basin radius this is more than 90%. From Prentice (1985)

A typical pollen assemblage of course consists of multiple taxa with different source areas. If light pollen types become progressively better represented in larger basins compared to heavy pollen types, the 'average' source area also increases with basin size (Prentice, 1985). Jacobson & Bradshaw (1981) predicted that pollen influx in small lakes would be dominated by trunk space transport and

gravity transport (Figure 7; Tauber, 1965). However, even in small lakes (radius < 20m), most of the pollen appeared to have a regional origin, supporting the sole consideration of above canopy flow in the Prentice model (Jackson, 1990). Figure 3 illustrates that the effect of basin radius on the pollen assemblage quickly reaches an asymptote (Prentice, 1985). Jackson (1990) elaborates that only moss pollsters or humus samples under closed canopy accurately represent mainly (extra)local pollen origins and that even small openings in the canopy cause significant fluxes of pollen with a regional origin.

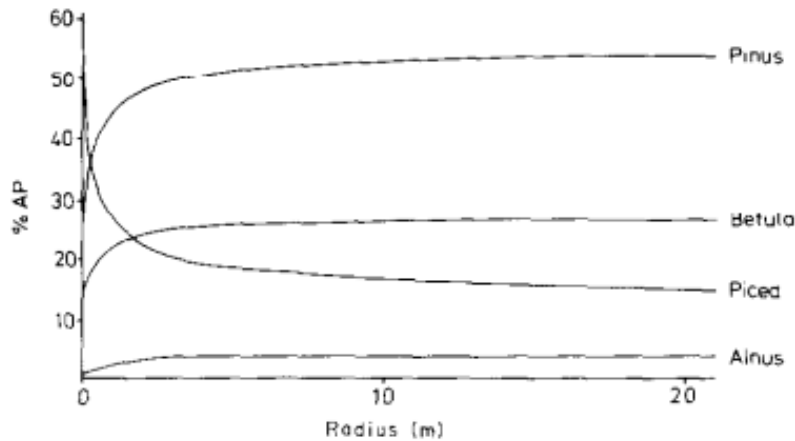


Figure 3: Simulated effect of basin size on the pollen assemblage in a Fennoscandian boreal forest. The simulation shows that even in small basins, pollen percentages from grains with different deposition velocities quickly reach asymptote values. Relative source strengths were modelled in a way to give 60% Pinus, 30% Betula, 6% Picea, and 4% Alnus (site radius = 400m), comparable to a typical pollen spectrum of a Finnish bog. V_d/u values were rounded to 0.03 for Picea and 0.01 for the other taxa. From Prentice (1985)

1.4 Pollen source area and basin type

The Prentice model (Prentice, 1985) calculates source areas on a point at the centre of a basin (Prentice, 1985). While this may be accurate for bogs, where pollen become immobile after being deposited, pollen in lake basins behave differently, being mixed in the water before sedimentation and being susceptible to resuspension afterwards (Sugita 1993). Consequently, for lakes the Prentice model needs to be slightly adapted so that the entire lake surface is regarded as the deposition area instead of only the centre of the lake (Sugita, 1993). To obtain the basin area, one first needs to draw two concentric circles from a certain pollen point source (A) and calculate the area of $(Z + R) - (Z - R)$ (Figure 4). If it is then assumed that the ratio of pollen deposition over the lake surface is equal to the ratio of pollen deposition over the area between $Z+R$ and $Z-R$, the pollen deposition on the lake from point A can be estimated. Integrating this for the source area between the edge of a lake and distance z gives:

$$Y_{i,\text{lake}}(z,R) \equiv \int_R^z y_{i,\text{lake}}(z,R) dz. \quad (6)$$

This formula is equivalent to the ERV model from Prentice (1985) for the centre of a basin.

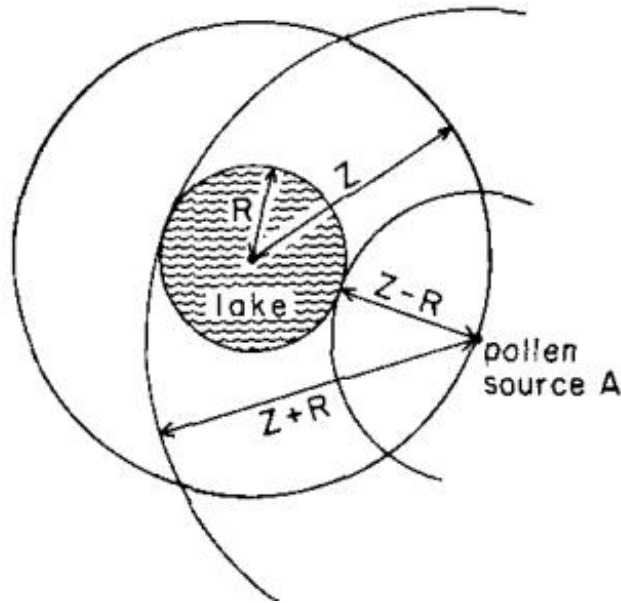


Figure 4: Schematic diagram of a lake and the pollen source. The lake radius is denoted by R ; Z represents the distance between the lake centre and a given pollen source A and is also the radius of a circular pollen source. The pollen deposition on the lake from source A can be approximated by calculating the ratio of surface area between the lake area and the remaining area between the concentric rings of $Z-R$ and $Z+R$, assuming that the ratio of pollen deposition is equal to the surface area ratio. From Sugita (1993).

In [Figure 5](#) simulations of pollen deposition are displayed for four different categories of pollen grains (A = heavy, B = relatively heavy, C = relatively light, D = light). Furthermore, four basin radii are considered ($R = 2\text{m}$ (forest hollow), $R = 50\text{m}$ (small lake), $R = 250\text{m}$ (medium lake), $R = 750\text{m}$ (large lake)). The simulations compare the pollen source radii (Z_{center} for measuring over a point at the centre of the basin and Z_{lake} for measuring over the entire basin) and the ratio of pollen originating from outside that radius. The simulations show that the source radius over a basin surface is 10-30% shorter compared to the source radius over a basin centre. This difference in source radius is larger for heavy pollen types than for lighter pollen types and for larger basins compared to smaller basins. The pollen deposition patterns also suggested that the pollen deposition on an entire lake surface is more strongly influenced by local pollen dispersers than the deposition at the centre of the basin. Again, heavier pollen types here show a stronger trend than lighter pollen types (Sugita, 1993).

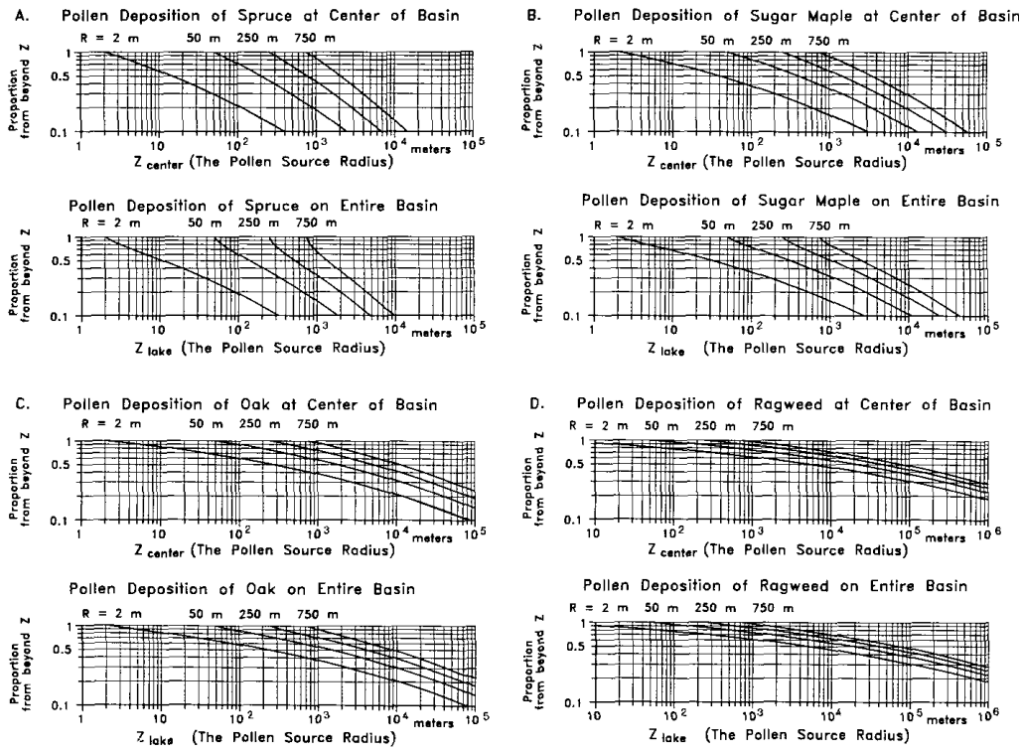


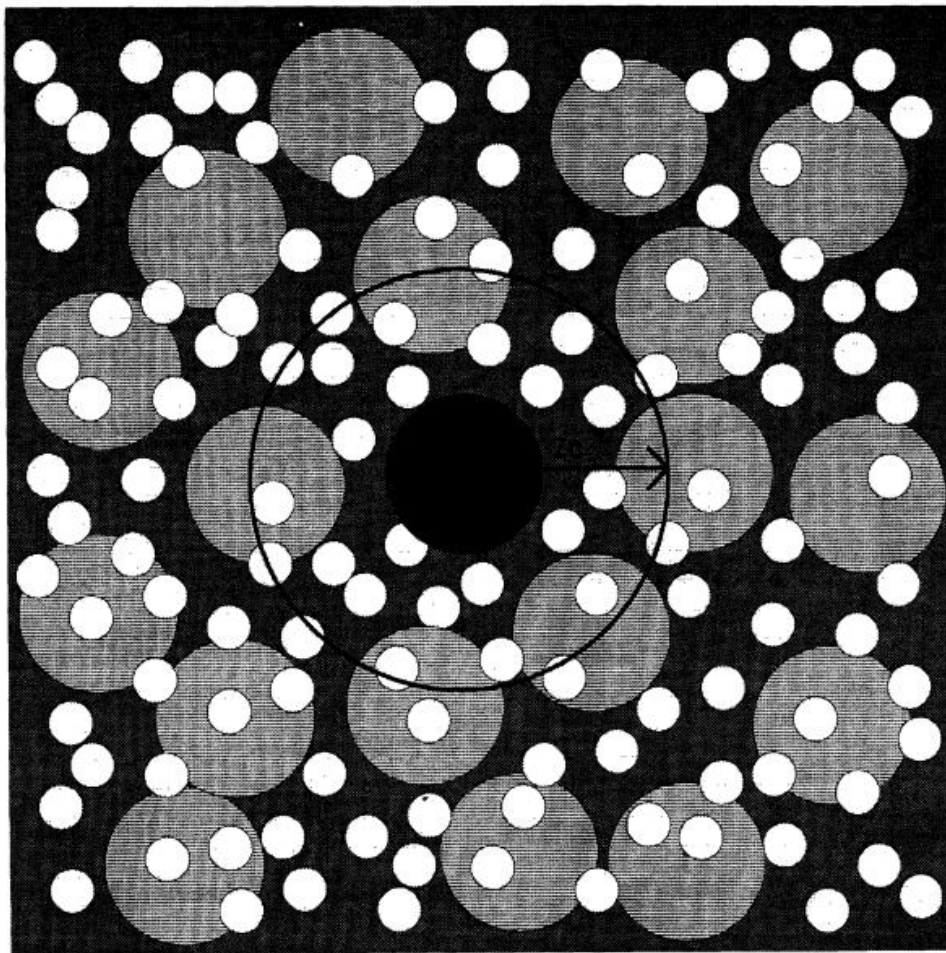
Figure 5: The relationship between the pollen source radius and the proportion of pollen of species i coming from beyond the source distance to the total pollen of species i . These plots show the differences in source area for Z_{center} (the source radius for a point at the centre) and Z_{lake} (the source radius for an entire lake surface) for four different basin radii (2m, 50m, 250m, 750m) are chosen and 4 different pollen types; A (spruce), B (sugar maple), C (oak) and D (ragweed). From Sugita (1993).

1.5 Relevant source area and vegetation heterogeneity

Now that there are estimations of source areas of pollen with different deposition velocities and average source areas for differently types and sizes of basins, a question remains over how large a distance vegetation needs to be surveyed to accurately correlate with pollen surface data (Sugita, 1994). The **Relevant Source Area of Pollen (RSAP)** is defined as the smallest area within which reliable estimates of parameter values and asymptotic r^2 or likelihood function scores can be obtained (*sensu* Sugita, 1994). Outside this area, the linear relationship between pollen and plant abundance would not further improve. In other words, the RSAP is the smallest value where the r^2 or minimum likelihood parameter plotted against the distance to the basin centre (Z_c) approaches an asymptote (given that ω_i (background pollen load) is nearly constant independent of site conditions).

Vegetation is often heterogeneously distributed (e.g. distributed in patches of different species composition, see Figure 6). For studying the relationship between pollen and vegetation around a basin, the basin size should ideally be equivalent or smaller than the patch sizes of the vegetation. If the size of the basin is larger than the recorded vegetation patches, the latter will appear homogenous and there are no reliable estimates to be made of pollen productivity and background pollen load (Sugita, 1994). Additionally, patch size is a key determinant for the width of the RSAP. Bunting et al. (2004) simulated the importance of individual factors relevant to pollen dispersal within the Prentice-Sugita model under equal basin size and atmospheric conditions. They found that under these conditions, the RSAP is primarily an expression of the patterning of the different vegetation elements within the landscape. This stressed the importance of vegetation structure in relation to the estimation of the RSAP and formed accordance with earlier studies (e.g. Sugita,

1999). The more unevenly the patches are distributed, the larger the RSAP needs to be (Hellman et al., 2009).



- Lake
- Hemlock
- Sugar Maple
- Yellow Birch

1 km

Figure 6: Schematic map of large patches of sugar maple (grey), hemlock matrix (black) and small patches of yellow birch (white) surrounding a lake, used for simulation experiments of pollen representation of patchy vegetation. Z_c (black circle) is the distance within which distance-weighted plant abundance is calculated and compared with pollen loading on a lake. The figure illustrates that vegetation patchiness is a big determinant of the width of the RSAP; high patchiness requires a larger source area to accurately capture the surrounding vegetation. From Sugita (1994).

Whereas Prentice et al. (1987) argued that the minimum likelihood score for the surveyed area coincided with a 70% pollen source area (a 30-50km radius for small to medium sized lakes), Sugita (1994) concludes that for a reliable estimation of parameters and high correlations between pollen loading and distance weighted vegetation abundance a 30-45% source area is sufficient, which translates into a significantly smaller RSAP (300-400m and 600-800m for small and medium sized lakes, respectively).

Chapter 2: Production and dispersal bias

2.1 Production bias

Pollen production bias is caused by differential pollen production between species. This means that a general estimate of pollen productivity for a given taxon is quite difficult to measure.

To deal with production bias, it is essential to quantify the pollen productivity of the most important species in the regional vegetation. Empirically, pollen productivity estimates (PPEs) can be estimated by comparing (modern) pollen surface samples with a plant cover parameter such as leaf cover, biomass or rooted frequency (Broström et al., 2008).³ With the development of more sophisticated statistical techniques since the 1980s regarding pollen deposition modelling (e.g. Prentice, 1985; Sugita, 1994)), it became possible to more accurately measure PPEs (e.g. Calcote, 1995; Sugita et al., 1999; Broström et al., 2004; Broström et al., 2008)).

Pioneering work by Andersen (1970) estimated pollen productivity by comparing (tree) basal area and canopy cover with surface samples from pollen traps and/or moss pollsters, providing a reasonably accurate measurement of pollen production from trees growing in the vicinity of approximately 100m from a sampling site at closed canopy. Absolute pollen productivity is here defined as the number of pollen grains produced per unit crown area per time unit

$$P = \frac{p - p_0}{a} \quad (7)$$

where p is the amount of pollen deposited on a unit area per year, a represents the crown area of the trees and p_0 the amount of pollen deposited from trees from a given species standing outside the source area. Ultimately, P is the degree with which the pollen deposition increases with the area of the species and depends heavily on the species pollen productivity.

When absolute pollen productivity estimates are applied on a model using relative pollen abundances – pollen percentage diagrams, pollen productivity estimates from taxa cannot be considered independently, because of the interdependence between taxa. As a result, productivity estimates first need to be relativized to a ‘reference’ taxon that is given a set productivity of 1, while the relative pollen productivity (P_{rel} or rPPE) of the other species can then be calculated as the species’ pollen productivity compared to the reference species (Fang et al., 2019).

Distance-weighted vegetation surveys further improved the accuracy of PPEs (Broström, 2004). It also became empirically evident that PPEs were highly regionally constrained, meaning that their informative value was constricted to the landscape type or area in which they were measured (Broström et al., 2004; Broström et al., 2008). Additionally, sampling methods play a role in different outcomes. For instance insect pollinated plant species are better represented in moss pollsters than in lake sediment due to their poor dispersal (Broström et al., 2008). Anthropogenic influences also possibly ‘contaminate’ the accuracy of PPE measurement of non-arboreal pollen (NAP) because the pollen production of grasses and other arable herbs is influenced by pesticide use and mowing practices, which may result in unreliable estimations of past pollen productivity (Theuerkauf et al., 2013).

³ Variation may also occur between specimens of the same species depending on environmental conditions, interannual variations not related to changes in plant cover (Fang et al., 2019) and whether plants are free standing or in closed stand (Iversen, 1941), but this is left out of consideration, because an assumption of PPE’s is that they are constant in space and time.

2.2 Dispersal bias

Dispersal bias can be caused by a variety of factors, mainly revolving around pollen grain morphology and weight. Heavy and non-aerodynamic pollen grains have higher fall speeds and as a result are rather ill-dispersed, while light pollen grains with smooth surfaces disperse better (Rempe, 1937; Prentice, 1985). Dispersal effectiveness correlates roughly with pollination syndromes. Insect-pollinated plants often produce relatively large pollen with sticky surfaces, which are optimal to be transported by visiting insects, while wind-pollinated species often produce light pollen with smooth surfaces or air sacs (*Gymnosperms*) that allow to travel large distances by wind transport.

When measuring pollen dispersal, there are considered to be five main modes of transport (Tauber, 1965; Jacobson & Bradshaw, 1981; Prentice, 1985):

1. Upper canopy pollen transport (C_c)
2. Trunk space transport (C_t)
3. Rainout/washout pollen transport (C_r)
4. Gravity pollen transport (C_g)
5. Water borne transport (C_w)

These modes are illustrated in [Figure 7](#). The predominant mode of transport depends partly on the landscape structure and the type of sampling site. At sampling sites directly under the canopy (e.g. forest litter surface samples or moss pollsters) the influence of the gravity component is expected to exceed the trunk space component strongly, leading to a very local representation of pollen spectra from such sites ([Figure 7](#) upper diagram; Prentice, 1985). Modes of pollen deposition in lakes or bogs, the sites where most fossil pollen spectra are extracted from, are more difficult to disentangle. There is especially a discussion about the relative contributions of upper canopy transport and trunk space transport, specifically related to basin size. The contributions of the different modes of transport are important regarding quantitative vegetation reconstruction because they represent pollen influx from different distances. In general, gravity transport can be considered to contain exclusively pollen from local origin (Prentice, 1985). Also trunk space transport is expected to contain predominantly pollen from local contributors, depending on the density of the vegetation. Upper canopy transport on the other hand contributes pollen from a regional distance of a few kilometres, though with decreasing efficiency at increasing distances. Pollen deposited by rainout/washout may have travelled very far distances and should be considered part of the background component. The same goes for water borne transport if the sampling site is in the vicinity of a river that could contain pollen from an entirely different region. However, in the case of an isolated lake, water borne transport probably only contains local (aquatic) species (Tauber, 1965; Prentice, 1985).

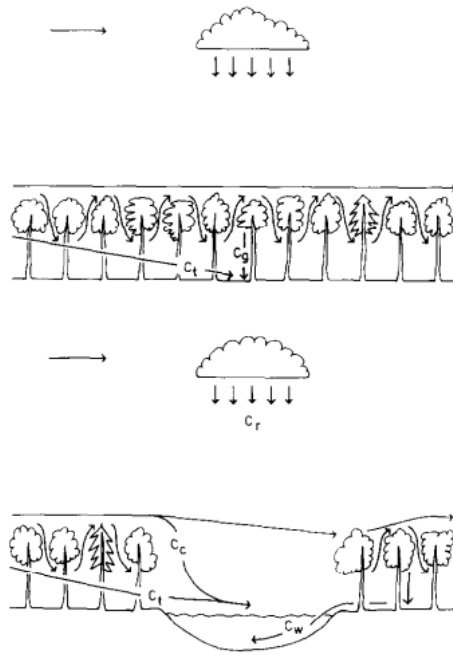


Figure 7: Components of pollen transfer from the forest canopy to a sampling site beneath the canopy (upper diagram) and a lake (lower diagram), after Tauber (1965, 1977) and others. C_t is the component carried in the trunk space by 0.5- 1.5 m sect' winds, C_c is the component carried above the canopy by 2-6 m set t winds, C_r is the component carried at higher altitudes and "rained out." C_g is "gravity" deposited, i.e., with falling leaves, raindrops, and C_w is waterborne, i.e. stream flow or flooding events. C_g contains large numbers of pollen grains intercepted from above the canopy (Andersen, 1974a) Within-forest sampling sites have highly local source areas; it is inferred that C_g "swamps" C_t (Andersen, 1974b; Jacobson and Bradshaw, 1981). Pollen deposition outside the forest is more complex. According to Tauber (1965) the regional airstream would be expected to "overshoot" small basins, where C_t would therefore be the main pollen source, but other work indicates that C_c dominates C_t even in small basins. The effects of C_r and C_w are mainly dependent on basin type and size but are overall considered to be of minor importance. From Prentice (1985).

2.2.1 Dispersal models

Pollen dispersal is commonly modelled using Sutton's equations for eddy diffusion, a special form of the Gaussian Plume Model (GPM) (Sutton, 1947, in: Tauber, 1965). This model was originally designed to model atmospheric diffusion of pollution particles but proved to be applicable to dispersion modelling in a broader sense. This model is appropriate for modelling the mass behaviour of 10-100 μ m particles (the size range of most pollen grains) released from a point source above or at ground level (Tauber, 1965; Prentice, 1985; Sugita, 1994, 2007a, b). There has been discussion whether a tree should be considered an elevated source. If a tree is considered an elevated source, it emits pollen from a height h , which would cause a 'skip distance' of $\sim 10h$ before the particles start being deposited (Prentice, 1985). However, according to Prentice (1985), trees cannot be considered true elevated sources, because in free stand, pollen deposition follows a monotonous path down and in closed stand, the canopy causes pollen to be either pulled down by gravitational transport or moved through upper canopy transport. Most authors followed the reasoning of Prentice (1985) and consider a ground level source the best approximation for the largest fraction of pollen transport (the combined upper canopy and gravitational transport). As there is evidence of a skip distance in at least a few taxa (Jackson & Lyford, 1999), some studies did incorporate a moderate elevation to create a small skip distance (Sjögren et al., 2008). [Figure 8](#) panel B shows the effect of elevation on the pollen dispersal. The optimal elevation assignment, however, is ultimately subjective. Jackson & Lyford (1999) conclude that both Gaussian models with and without skip distance probably inadequately reflect the actual dispersion pattern of pollen.

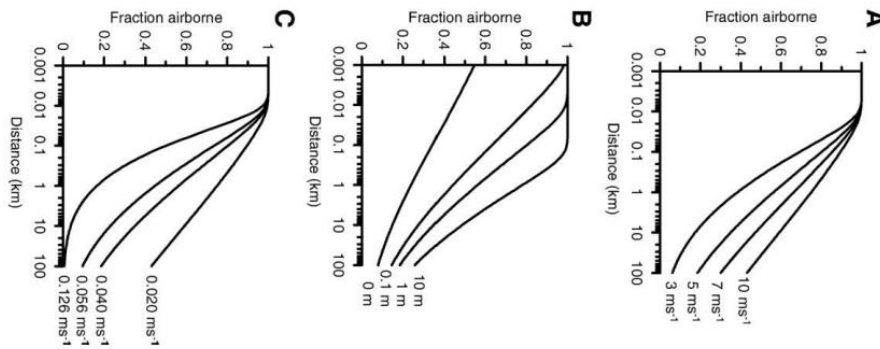


Figure 8: The effect of different parameters on Sutton's (1953) dispersal function for air borne particles. Default settings are wind speed of 5 ms^{-1} , height of pollen release (injection height) at 1 m , and a fall-speed of 0.04 ms^{-1} . These parameters are varied individually in the different graphs: A) Wind speed, B) Injection height, and C) Fall speed. From Sjögren et al. (2008).

An alternative to the GPM is a Lagrangian Stochastic dispersal model (LSM). Diffusion models such as the GPM approximate the entire dispersal pattern which makes them less suitable to incorporate irregular airflows that likely strongly influence pollen dispersal (Kuparinen et al., 2007). The LSM on the other hand primarily describes wind and turbulence patterns in the atmosphere, based on which individual particle trajectories can be simulated (Figure 9; Theuerkauf et al., 2013). There is no unambiguous evidence which model works best. Some studies show that the LSM works better in determining PPE's than the GPM (Theuerkauf et al., 2013; Mariani et al., 2016). Theuerkauf & Joosten indicate that the LSM is a more accurate dispersal model for long-distance pollen transport because it is more regulated by convective airstreams, instead of the neutral atmospheric conditions the GPM model works best for. Theuerkauf et al. (2013) indicated that the GPM model tended to underestimate the pollen contribution from $> 10 \text{ km}$ away and to overestimate the dispersal differences between light and heavy pollen grains. Fang (2018) points out that both models describe the general patterns of arboreal pollen quite well, but that the LSM might work slightly better on small and medium-sized lakes.

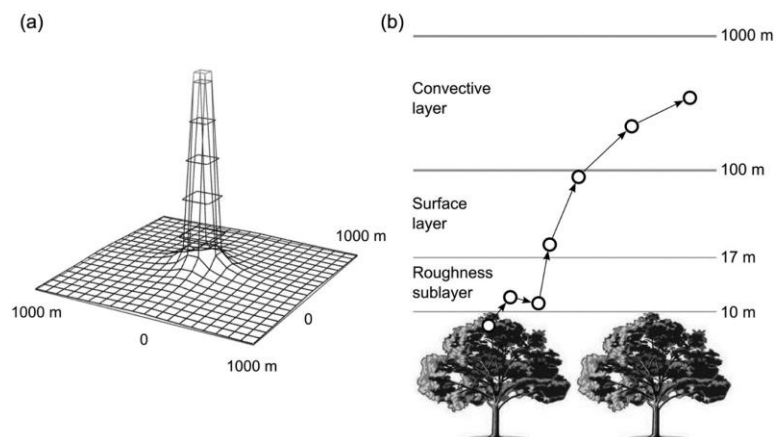


Figure 9: Basic approaches in pollen dispersal modelling. Gaussian plume models (a), such as the Prentice model, approximate the entire dispersal pattern at once. Lagrangian stochastic models (b); e.g. Kuparinen et al., 2007) simulate the trajectory of a single pollen grain within the atmospheric boundary layer (ABL) by calculating the three-dimensional position of that pollen grain for consecutive time steps from the moment the grain is released until it lands on the surface. Final dispersal pattern is produced by simulating a large number of individual pollen grains. From Theuerkauf et al. (2013).

Chapter 3: Comparing the reconstruction models

3.1 Landscape Reconstruction Algorithm

The Landscape Reconstruction Algorithm (LRA) is a multistep method to assess both regional and local vegetation dynamics (Figure 10; Sugita, 2007a, b). The first step of the LRA is REVEALS (Regional Estimates of VEgetation Abundance from Large Sites) and can be used for regional vegetation reconstruction (10^4 - 10^5 km²). The second step is LOVE (Local VEgetation Estimates) and uses the estimated background pollen signal generated by REVEALS for the reconstruction of local vegetation change ($<10^4$ km²) (Sugita, 2007a, b).

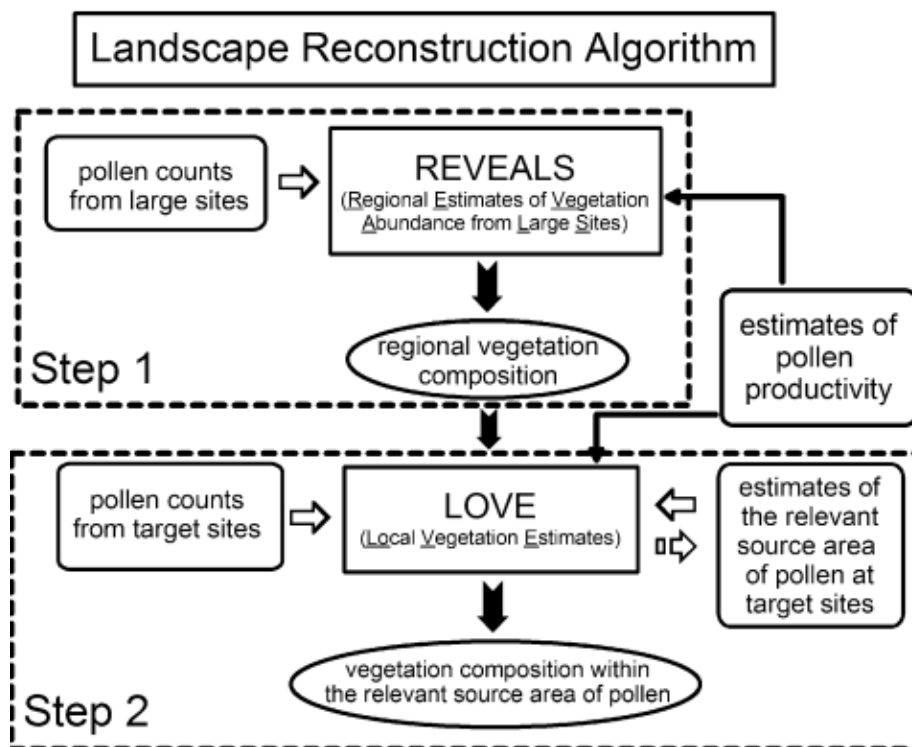


Figure 10: Framework of the Landscape Reconstruction Algorithm (LRA). The first step (REVEALS) uses large basins ($> 10^2$ ha), and PPE estimates to assess the regional vegetation composition which is then used as input for the second step (LOVE) that uses small sites (< 10 - 10^2 ha) together with the estimated RSAP of these sites in a region and PPE's. From Sugita (2007b)

3.1.1 REVEALS

Prentice (1985) hypothesized that pollen samples from a network of sites in a region could represent regional vegetation estimates if local variation could be filtered out. The idea behind it is that the pollen assemblages from large lakes or mires tend to be similar among sites in a region (Sugita, 1994). Therefore, REVEALS requires pollen counts from multiple 'large basins'. Figure 11 shows that lakes with a radius of 390m or larger do not show significant site-to-site differences ($\alpha = 0.05$) and are therefore suitable for REVEALS modelling (Sugita, 2007a). Ideally, even larger lakes (radius > 650 m) are used to minimize the effect of local differences (Sugita, 2007a).

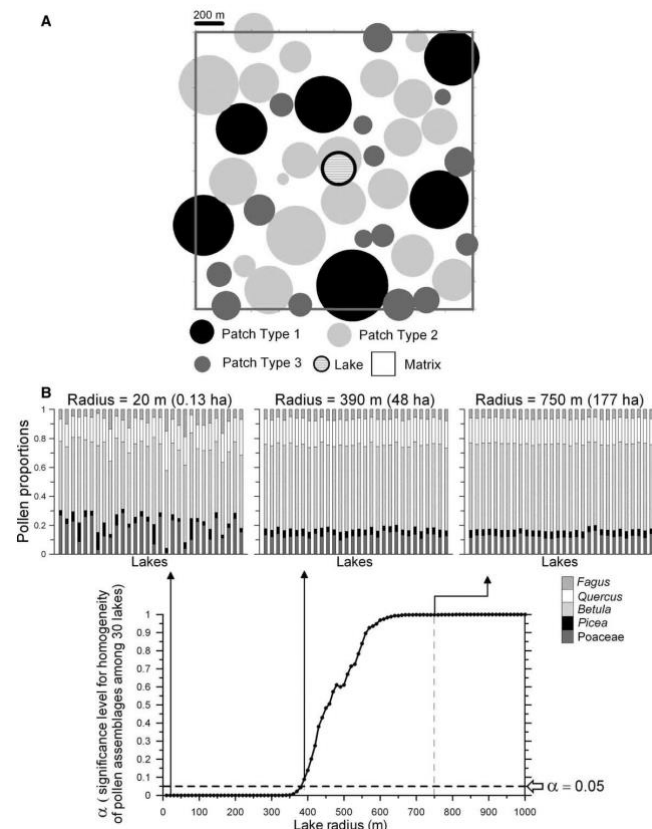


Figure 11: (A) Example of the landscape design to simulate pollen deposition on lakes. Three types of vegetation patches are randomly distributed in the matrix of a Poaceae-dominated vegetation type. Each stand type has a distinctive species composition. (B) Simulated variations of pollen assemblages among 30 lakes of different sizes. In this landscape, when lakes are > 48 ha (radius > 390m), pollen assemblages can be regarded as homogeneous among sites at the significance level $\alpha = 0.05$ using the likelihood ratio statistic G^2 . From Sugita (2007a).

Assumptions and potential sources of error

The REVEALS model is quite robust for violations of its assumptions. The first assumption is that basins are a circular opening in the canopy and that no source plants grow on the basin surface (Sugita, 2007a). Lake shape irregularity may violate the assumption in the Prentice-Sugita model that pollen loading in the lake is eventually evenly mixed in the water before being deposited (Bunting & Middleton, 2005). However, empirical evidence suggests that irregularly shaped basins still provide reliable estimates (Soepboer et al., 2010). Source plant growth on the basin surface is mostly a problem related to bogs and mires, even though Sugita (2007a) states that REVEALS should work for both lakes and mires. However, when small bogs are used, source plants (mainly heath, cypergrasses and birch) may cause systemic differences in output compared to lake basins (Mazier et al., 2012; Trondman et al., 2016).

Furthermore, the REVEALS model requires large lakes. Based on the scenario depicted in [Figure 11](#), it requires lakes with a radius of at least 390m (48ha), but optimally the lake size should be between 100-500ha to reduce the effect of large patches of vegetation and because at that point the homogeneity of pollen assemblages across sites does not further increase (Sugita, 2007a). However, not in every situation this requirement can be met due to absence of such large basins. If that is the case, a large number of small basins could provide reasonable estimates, but the standard errors will be larger (Sugita, 2007a; Trondman et al., 2016). Increasing the number of sites could improve the estimates of regional vegetation cover and reduce standard error (Sugita, 2007a; Mazier et al., 2012;

Trondman et al., 2016). When bogs or mires are studied instead of lakes, the required minimal basin size may be lower than 48ha, because the pollen source area is larger for bogs and mires than for lakes (Sugita, 1993), but this has not been formally simulated.

In general, the estimates for vegetation reconstruction improve with higher pollen counts. Under the simulation of Sugita (2007a), pollen counts of 1000 per sample were assumed, but if pollen counts are lower (e.g. 300-500), the standard errors for vegetation composition could be 10-20% larger (Sugita, 2007a). Increasing the pollen count decreases the standard error of REVEALS estimates for multiple small sites (Trondman et al., 2016).

The dispersal model that has been most frequently used in REVEALS is the Sutton equation derived from the Gaussian Plume Model (GPM) (Sugita, 2007a). Like the Prentice-Sugita model, it assumes neutral atmospheric conditions, no specific directionality in wind flow and only upper canopy transport is considered; all other modes of transport as described by Tauber (1965) are disregarded. However, several studies have indicated that these assumptions can often not be met (e.g. Theuerkauf & Joosten, 2009; Theuerkauf et al., 2013).

Theuerkauf et al. (2016) argue that REVEALS performs better if the GPM is replaced by a Lagrangian Stochastic Model (LSM), mainly because the estimation of PPE's is more accurate using the LSM compared to the GPM. Additionally, the LSM does not need to assume atmospheric neutral conditions, which might be inaccurate for pollen dispersal (Theuerkauf & Joosten, 2009). Case study evidence by Mariani et al. (2016) supports a more accurate estimation of PPE's using the LSM.

Implementing a wind rose for modelling wind directionality is not possible in the simulation program used by Sugita (2007a, b), but other programs such as HUMPOL (Bunting & Middleton, 2005) offer a more flexible handling of these parameters.

Applicability

REVEALS is only applicable on large vegetational scales (10^4 - 10^5 km²). Because it represents the pollen assemblage of a 100-200km radius around the lake (Sugita, 2007), it cannot account for abrupt vegetational changes caused by geomorphology, hydrology, elevation etc. It assumes a considerable homogenous ring of vegetation around the basin, which is inappropriate for cultural landscapes or landscapes with strong regional variation (Bunting & Middleton, 2005). Additionally, the suitability of REVEALS in mountainous areas is limited because the elevation differences require modelling pollen dispersal from different deposition heights (Bunting et al., 2008). Some of these problems might be alleviated using the LSM instead of the GPM as the dispersion model, because the LSM models individual particle trajectories instead of the entire dispersal pattern (Theuerkauf et al., 2013).

3.1.2 LOVE

The second part of the LRA as proposed by Sugita (2007a, b) is the LOVE model: LOcal Vegetation Estimates. Whereas REVEALS subtracts site-specific vegetation elements to measure regional-scale vegetation composition dynamics, the goal of LOVE is to calculate the vegetation composition within a reconstructed *relevant source area of pollen of past vegetations* (Figure 10). This **RSAP of past vegetation** works slightly different than the RSAP mentioned above. In the LOVE model, the RSAP is back calculated by prediction modelling under the assumption that all other parameters are known (see Figure 12).

The LOVE model makes use of a regional vegetation factor S_i ; a parameter estimated by the ratio between the total sum of distance-weighted plant abundance of all taxa within the relevant source area of pollen (RSAP) at site k and the distance-weighted regional plant abundance of species i . S_i is important in the LOVE model, because it represents the regional vegetation that is estimated by REVEALS and needs to be subtracted in order to identify the local

vegetation. To obtain S_i , the regional vegetation composition estimate V_i must be yielded from large lakes using the REVEALS model (Sugita, 2007a). This estimate, together with the relative pollen productivity, is then used in the LOVE model equation. For LOVE, it is not necessary to use large size lakes, it is preferable to use pollen counts from smaller lakes (radius = 50m *sensu* Sugita, 2007b).

Assumptions and potential sources of error

The most difficult part of the LOVE model is the reconstruction of past RSAP, a crucial parameter for the estimation of S_i . Sugita (2007b) proposed that backward modelling of LOVE could provide an accurate estimation of the relevant source area. Figure 12 illustrates how this backward modelling works. First, it is assumed that all parameters for the LOVE equation are known, except for the past RSAP (Z_{RSAP}). If Z_{RSAP} is then iterated, a threshold value is eventually yielded where the estimated ratio of every taxon in the landscape can be considered a realistic value (> 0 and < 1) (Sugita, 2007b). The threshold is in this case defined by 150 predictions for the vegetation estimates of 5 taxa in 30 lakes used in the simulation by Sugita (2007b). In panel (a) of Figure 12, first a past RSAP of 150m is predicted (C). This results in 32 'wrong' vegetation composition values (where abundances are < 0 or > 1 ; B). By iteratively increasing the past RSAP, the amount of wrongly predicted estimates reaches 0. The point where this happens can be considered the 'real' past RSAP according to Sugita (2007b).

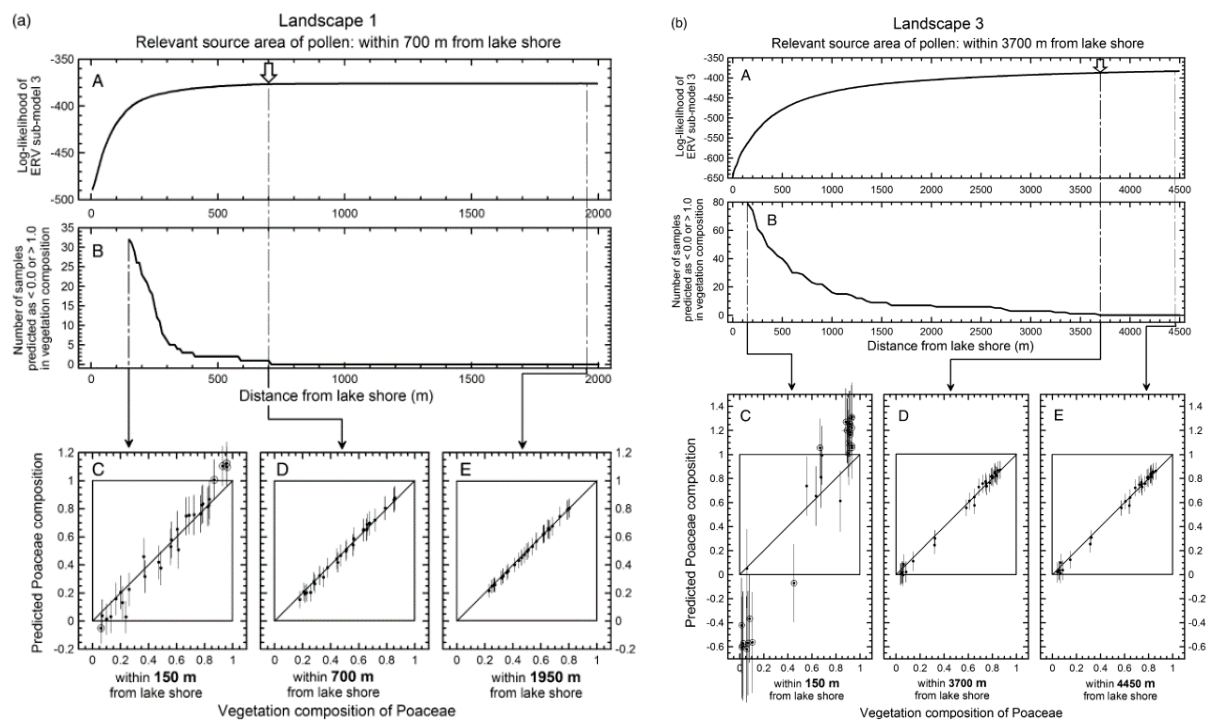


Figure 12: Effects of the selection of the relevant source area of pollen on the accuracy of vegetation reconstruction using the LOVE model. In both (a) and (b), graph A shows changes in log-likelihood of ERV submodel 3 with distance, and the relevant source area of pollen (open arrow); graph B shows changes in the total number of unrealistic predictions of vegetation proportions (i.e., < 0.0 or > 1.0) for all taxa at all sites; graphs C, D and E show the LOVE-based prediction of Poaceae proportions at 30 sites plotted against its abundance within the areas much smaller than, equal to and larger than the 'real' relevant source area of pollen, respectively. (a) Vegetation composition of Poaceae predicted by LOVE, assuming the area within 150 m, 700 m, and 1950 m to be the relevant source area of pollen (Landscape 1). The 'real' relevant source area of pollen in Landscape 1 is the area within 700 m from the lake shores. (b) Vegetation composition of Poaceae predicted by LOVE, assuming the area within 150 m, 3700 m, and 4450 m to be the relevant source area of pollen (Landscape 3). Landscape 3 includes very large (ie, 10 000 ha) patches of Patch type 1 (Table 2), and the 'real' relevant source area of pollen is the area within 3700 m from the lake shores. Examples in (a) and (b) show that differences in spatial structure of vegetation affect the size of the relevant source area of pollen. From Sugita (2007b).

However, this method is cumbersome because a lot of assumptions must be met to be able to perform the backward calculation (Sugita, 2007b). Recently, an R package for the LOVE model (LOVEoptim) has been developed in which past plant abundances can be estimated by numerical optimization (Theuerkauf & Couwenberg, 2024). This method makes fewer assumptions and is therefore mathematically less complex for vegetation reconstruction on a local scale.

Applicability

In terms of spatial scale, the applicability of the LOVE model *sensu* Sugita (2007b) depends on the RSAP, which mainly depends on vegetation patchiness. Figure 12 illustrates that a highly patchy landscape (panel (b) Landscape 3) has a much higher RSAP than Landscape 1 in panel (a). If Landscape 3 can be considered as extremely patchy, the spatial scale of the LOVE model should be >5km around a small (50m radius) site. The LOVEoptim model on the other hand makes use of the *Necessary Source Area of Pollen* (NSAP) (*sensu* Theuerkauf & Couwenberg, 2024) and is defined as the smallest local area for which LOVE does not produce vegetation cover estimates < 0 and > 1 (comparable to [Figure 12](#)). The advantage of the NSAP is that it is not mainly dependent on vegetation patchiness but rather on the pollen dispersal pattern.

The applicability of the LOVE model *sensu* Sugita is heavily hindered by the complexity of back calculating the past RSAPs. If one of the assumptions of the LOVE model is violated, back calculating is not informative (Sugita, 2007b). Additionally, reliable back calculation requires many similarly sized sites that may not always be available (Sugita, 2007b). Perhaps this is why the LOVE model, in comparison to REVEALS, has not been applied often. In studies that have applied LOVE (e.g. Hjelle et al., 2015; Fredh et al., 2019) the estimation of past RSAP proved indeed to be cumbersome. The applicability of the LOVEoptim model by Theuerkauf & Couwenberg (2024) has not yet been thoroughly tested, but it appears to be easier, because less assumptions have to be made.

3.2 Multiple Scenario Approach

The Multiple Scenario Approach (MSA) aims to reconstruct vegetation based on the landscape instead of the pollen assemblage (as does the LRA), whilst accounting for **equifinality**, the fact that multiple scenarios may result in a similar pollen assemblage (Bunting & Middleton, 2009). In a GIS environment multiple ecologically distinct landscape possibilities are simulated. Using a model of pollen deposition and dispersal (e.g. the Prentice-Sugita model), a pollen assemblage is simulated from each landscape scenario at a known sampling point. These assemblages are then compared to an actual pollen assemblage from the same sampling point to determine the likelihood of the possible vegetation compositions. The MSA works for single pollen records but becomes more robust when multiple sites with a strong chronological framework are included (Bunting & Middleton, 2009). It follows three steps (see also [Figure 13](#)):

1. Identify the to be reconstructed area and acquire data on (paleogeographic) invariant physical aspects of the landscape. Then identify relevant plant taxa based on e.g. PPE availability and identify their overall auto/synecological habitat. The combination of physical and ecological requirements generates different scenario maps. There are many combinations of landscape rules that can each generate a multitude of different scenario maps. Scenarios are ordered by landscape rule combination, which is called a *scenario family* (Bunting & Middleton, 2009).
2. For multiple members of each scenario family a pollen assemblage is simulated. These pollen assemblages are then compared to an actual pollen assemblage from the site.

Depending on the comparison mode, simulated assemblages should be within certain limits to be considered a 'good match'.

3. For each 'likely' family, a larger number of replicates is run to identify 'best fit' members. In terms of reconstruction, the family can be seen as a representative for broad-scale vegetation patterns, while the best-fit models shed light on the possible vegetations in the vicinity of the basin.

When one wants to reconstruct a timeline of vegetation development, one needs to form realistic successions between scenarios of different time slices.

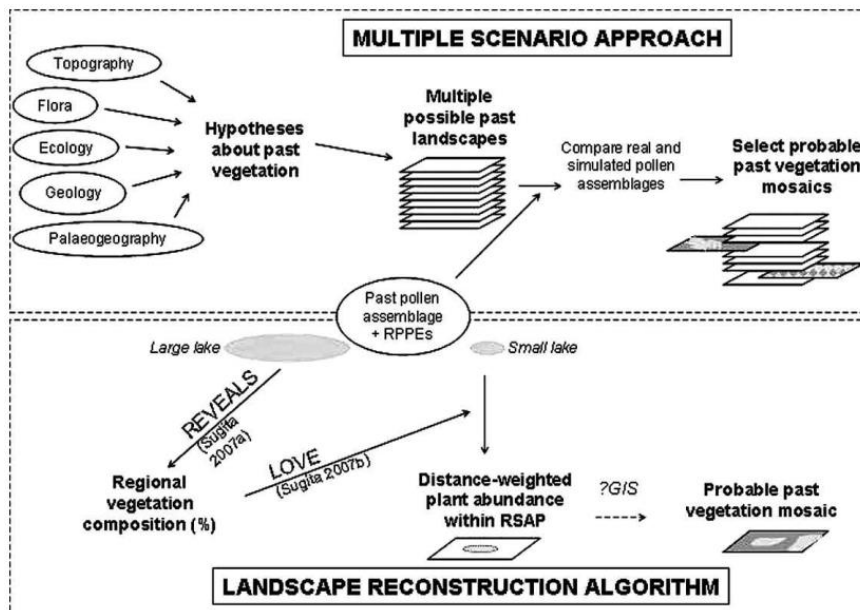


Figure 13: Schematic comparison of the Multiple Scenario Approach and the Landscape Reconstruction Algorithm. Oval boxes contain information to be input into the process, and labels along arrows (REVEALS, LOVE) indicate the software used during particular steps of the LRA. From Bunting & Middleton (2009).

Assumptions and potential sources of error

The MSA makes the following assumptions:

1. A palynomorph⁴ represents a set of plant taxa that is ecologically consistent over time.
2. It is possible to define the auto/synecology⁵ of the set of plant taxa and these are uniform over time.
3. The model of pollen deposition and dispersal is correct and invariant through time.

Assumptions about the ecological consistency of species and their interactions can be dangerous because it is known that these ecological requirements and interactions can temporarily change over time. However, the timespan over which these changes are likely to be conserved exceed the timespan Quaternary palynologists are interested in (Lososová et al., 2020). Therefore, it is relatively safe to consider these assumptions valid for 'recent' studies of past environments.

⁴ General term for pollen, spores and other microremains of plants, animals and Protista that give information about the landscape structure.

⁵ Autecology refers to the interaction of a single organism/species with its environment, whereas synecology describes the interaction between multiple organisms/species with each other and with the environment.

The robustness of the third assumption depends on the model for pollen deposition and dispersal that is used. Bunting & Middleton (2009) use the Prentice-Sugita model, but model is flexible to incorporate other deposition and dispersal models, even though up to present, there is no record of studies implementing the LSM in the MSA.

Applicability

Bunting et al. (2018) showed that the application of the MSA offered information on the landscape somewhere on a scale between REVEALS and LOVE. A large advantage of the MSA compared to algebraic reconstruction methods such as the LRA is that the MSA can incorporate (changes in) wetland vegetation instead of needing to omit wetland taxa from the calculations. Furthermore, the MSA, unlike the LRA, does not assume a uniform regional landscape, which opens many more possible landscape scenarios. Lastly, the MSA performs well in landscapes with different elevation levels (hills or mountains) and is through the modelling environment HUMPOL able to incorporate (multiple) wind roses, which enhances the realism of model outputs (Bunting et al., 2008).

However, reconstruction of paleoenvironments remains difficult from a map-based approach such as the MSA if palaeogeographical information such as sea level and lake/moor size is unknown (Bunting et al., 2018). Another weakness of the MSA is that some palynomorphs are only detectable on a high taxonomic level, which means a palynomorph could contain plant species living in very different ecological habitats. This makes it hard to assign ecological groups to such pollen taxa.

Currently, the application of the MSA is particularly related to the development of cultural landscapes and archaeology (Bunting et al., 2018, 2022). Theoretically, the MSA should be able to perform landscape reconstructions on a larger scale comparable to REVEALS, but the number of scenarios that are created if many sites are included increases exponentially, which means that eventually the computation process likely becomes a bottle neck (Bunting et al., 2018).

3.3 The Extended Downscaling Approach

The Extended Downscaling Approach (EDA) (Theuerkauf et al., 2014, 2017) is a combination of the downscaling approach (DA) (Theuerkauf & Joosten, 2009) and a simulation approach using a Lagrangian Stochastic dispersal model (LSM) (Theuerkauf et al., 2013). The DA aims to relate past pollen assemblages to present modern day landscape parameters. Like the MSA, the DA first creates the landscape in which abiotic soil controls are defined (Theuerkauf & Joosten, 2009). The area around each basin within a certain radius is divided in concentric circles of a certain size (Theuerkauf & Joosten (2009) use a radius of 50km and 1km wide rings). The substrate composition is then averaged per circle and distant rings are downweighed under the assumption that vegetation from distant rings contributed increasingly less pollen to the basin. Now the correlations between pollen taxa in past pollen assemblages and the present-day soil substrate are tested. This is visualized in [Figure 14](#). The left panel shows the expected (ideally linear) relationship between the pollen loading and the vegetation representation of pine. The right panel shows that if the abundance of pine is predominantly determined by sandy soils, their pollen abundance would increase with the area of sandy soils. The results show a strong correlation between present-day substrate cover and past pollen deposition ([Figure 14](#)), which indicates that present-day landscape parameters may prove a valuable proxy for determining past vegetation cover.

Combined with the simulation approach, a vegetation composition is randomly assigned to each landscape parameter category (e.g. landscape parameter = soil substrate, categories = sand, loam, clay) (Theuerkauf et al., 2013). From this vegetation cover, pollen deposition is then simulated in an iterative process for every distance-weighted concentric ring around the basin. Simulated

pollen assemblages are then compared to empirical assemblages and the vegetation composition per landscape parameter category is being adjusted until an optimal fit between the modelled and empirical assemblage is reached (Theuerkauf et al., 2014, 2017)

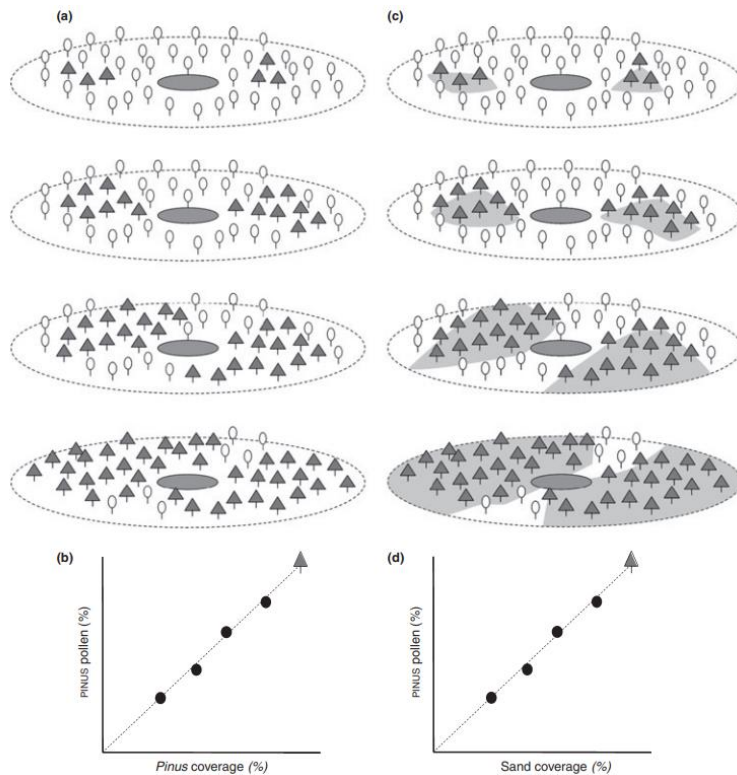


Figure 14: An illustration of the approach: a basic observation in pollen analysis is that with increasing abundance of a species, the abundance of its pollen also increases. Thus, in a landscape (a) with a central lake and two species, *Betula* (elliptical) and *Pinus* (triangular), the proportion of *Pinus* pollen in sediments of that lake increases (ideally linearly) with the proportion of *Pinus* in the surrounding vegetation (b). If the occurrence of *Pinus* is determined by that of an environmental factor, e.g. the presence of sand (c, grey shading), the proportion of *Pinus* in the vegetation and thus the proportion of *Pinus* pollen increases with the total proportion of areas with sand (d). A strong correlation between sand and *Pinus* pollen (d) thus indicates a strong substrate dependency of *Pinus*. This correlation can also be tested using fossil pollen data and modern substrate distribution. Close correlation then indicates that the distribution of species in the past might also have been determined predominantly by the substrate, provided the pattern of abiotic conditions has remained constant over time. From Theuerkauf & Joosten (2009)

Assumptions and potential sources of error

Like the MSA, the EDA also assumes that the landscape parameters that are studied are stable over time (Theuerkauf & Joosten, 2009) and that the relationship between landscape parameters and past vegetation cover is equal with the relationship between these parameters and present vegetation cover (Abraham et al., 2023). However, some landscape parameters (soil unit, soil wetness and soil nutrients) may have been more variable over time. This means that it is less certain whether the relationships between present landscape units and past vegetation cover is directly indicative for past landscape units. For example, Theuerkauf et al. (2013) associated the expansion of hazel after the Late glacial with present day high soil wetness, but it is doubtful if under drier climatic conditions, soil wetness was as high as it is today (even though the relative soil wetness may still have been higher than in the surrounding area).

Other assumptions depend on the dispersal model that is used. The EDA is flexible for the

implementation of multiple dispersal models, although the LSM provides the most accurate results (Theuerkauf & Couwenberg, 2017).

Like the MSA, the EDA defines landscape units based on invariant physical properties of the landscape and it compares empiric pollen deposition with simulations, but because the MSA produces multiple landscape scenarios, forward modelling for multiple sites exponentially increases the amount of possible scenarios and therewith the simulation time. The EDA is designed to use multiple sites and simulates only one 'optimal' match with the empirical pollen assemblages, which makes forward modelling aiming to determine the vegetation cover associated with landscape units more convenient (Theuerkauf & Couwenberg, 2017).

Applicability

The EDA is mostly applicable for scenarios with sharply contrasting substrates; with intermediate substrates the correlation between pollen type and substrate cover may become faded (Theuerkauf & Joosten, 2009). Additionally, when running complex EDA scenarios, the model may yield false-positive results of two landscape units are closely correlated but inhabited by very different vegetation types (Theuerkauf et al. (2017) give the example that oak and grasses were predicted to be present on stagnosols, while in fact, they are vegetation dominant on gleysoils, likely because stagnosols and gleysoils both occur frequently in flat morainic areas and thus are 'neighbouring' landscape types. The EDA can have difficulties assigning the correct vegetation cover to closely related landscape types.

The EDA is also less applicable for understorey plants, who are more dependent on light availability than to substrate cover. Lastly, Theuerkauf et al. (2014) emphasize that the EDA works best with landscape parameters that are as invariable as possible. The method is therefore less applicable for correlations with landscape parameters that varied strongly over time.

Discussion

The goal of this paper was threefold: First, I aimed to present an overview of the main developments in quantitative vegetation reconstruction through pollen analysis from roughly 1950-2000. It shows how the idea of R-value models gained shape and how the eventual Extended R-value Model (ERV) from Prentice (1985) cumulatively developed over the following decades with additions from e.g. Prentice & Webb (1986) and Sugita (1993, 1994).

Production & dispersal bias

In the second chapter, I elaborated on the importance of production and dispersal bias in pollen analysis and on methods to mitigate these biases. Pioneering work by Andersen (1970) and Tauber (1965) led ultimately to the development and refinement of pollen productivity estimates (PPEs) and dispersal models, respectively. The application of PPEs calls for caution, because their informativity is region dependent, the methodology of acquiring PPEs is not uniform (Broström et al., 2008) and modern estimates of pollen productivity might not always be accurate for past pollen productivity (Theuerkauf et al., 2013). On regional scales a solution could be to take the means of all PPEs from different regions while excluding clear outliers (Serge et al., 2023). However, the authors warn that on regional scales, the choice of the mean PPE dataset can be very determining of the eventual REVEALS estimates. This is because large taxonomic pollen groups (e.g. Ericaceae) can be comprised of region-specific species with quite different PPEs to which the REVEALS estimates are very sensitive (Serge et al., 2023). For smaller spatial scales, PPE estimates that have been measured in the concerning area should be more accurate (Mazier et al., 2012).

Recent studies highlight the dependency of PPEs on the chosen dispersal model. Liu et al. (2022) stress that the interaction between pollen productivity, pollen dispersal processes and the heterogeneity of the landscape is a main determinant on the representation of the vegetation surrounding basins. Theuerkauf & Couwenberg (2022) conclude that most errors of estimating PPEs are caused by the overestimation of the fall speed in the Gaussian Plume Model (GPM). Both authors agree that the GPM likely overestimates the contribution of local vegetation, affecting PPEs and that the Lagrangian Stochastic Model (LSM) could provide more accurate estimates. However, if pollen dispersal is driven mostly by disturbing effects (e.g. forest eddies), also the LSM may be unable to model the dispersal pattern accurately (Theuerkauf & Couwenberg, 2022). Liu et al. (2022) propose that pollen dispersal models should also make use of semi-mechanistic (statistical estimations of dispersal patterns based on empirical data).

All current dispersion models only consider the upper canopy flow C_c as mode for pollen transport. However, the findings by Tauber (1965) indicated that also trunk space transport C_t could be a considerable contributor to the pollen assemblage in case of small lakes and that also the density of the vegetation around the lake influences the eventual pollen deposition. Considering that in several areas in the world (for instance the Netherlands) the pollen record consists of mainly small and a few medium-sized lakes, it would be worthwhile to explore the impact of trunk-space transport further. If trunk space transport is indeed a significant contributor to the pollen assemblage, the (extra)local vegetation might be overrepresented in small lakes, which would affect the accuracy of regional vegetation reconstruction.

Intermediate scale landscape reconstruction

Lastly, I evaluated three currently available vegetation reconstruction models: (1) the Landscape Reconstruction Algorithm (LRA) (Sugita, 2007a, b), (2) the Multiple Scenario Approach (MSA) (Bunting & Middleton, 2009) and (3) the Extended Downscaling Approach (EDA) (Theuerkauf & Couwenberg, 2017). It is difficult to compare the approaches due to the limited number of case studies for the MSA, EDA and the LOVE model in the LRA. So far, the MSA and EDA have mainly been applied on relatively local scales in Great Britain and northeastern Germany (e.g. Bunting et al., 2018; Theuerkauf et al., (2013), whereas a few case studies of the LOVE model exist in Scandinavia and the Pyrenees (e.g. Hjelle et al., 2015; Marquer et al., 2020).

How well would the three models perform on an intermediate scale, in this case a Lateglacial and Holocene landscape reconstruction of the Netherlands?

LRA

The first step of the LRA, REVEALS, assumes the usage of a few large basins between 100ha and 500ha. There are few of such sites in the Netherlands with a continuous pollen record, although there are also only 2-5 basins required for the model to work. And according to Sugita (2007a) and case study evidence, the REVEALS model also works adequate when using multiple smaller sites instead. However, REVEALS may work less well in landscapes with frequent abrupt geomorphological or hydrological changes and major differences in elevation (Bunting et al., 2008). While the latter is not a big problem in the Netherlands, as a low altitude coastal country the hydrology is and has been very dynamic.

The second step of the LRA, LOVE, requires many small, similarly sized basins and a regional vegetation estimate calculated by REVEALS that then can be used in the LOVE model to back calculate the past RSAP (*sensu* Sugita, 2007b). The other newly available option is numerical optimization for the calculation of the NSAP (*sensu* Theuerkauf & Couwenberg, 2024). Both techniques require the regional vegetation input from REVEALS, so their applicability stands or falls with that. The original LOVE model furthermore assumes that the vegetation within the past RSAP is homogeneous and has a similar composition as the regional vegetation (Sugita, 2007b). However, as the Netherlands covers and covered an area with significant local environmental gradients (Cohen & Pierik, 2021), this assumption would not be valid. The LOVEoptim model from Theuerkauf & Couwenberg (2024) does not make this assumption and may therefore be more applicable in this setting.

MSA

The MSA uses environmental parameters as input to create multiple landscapes for one site in a GIS environment. In turn multiple potential pollen assemblages are simulated on this landscape, which are then compared to the 'actual' pollen assemblage of the site (Bunting & Middleton, 2009). The MSA requires detailed paleogeographic information from the to be reconstructed area as input. In the Netherlands, the National Geological Service (TNO, formerly RGD) has collected an extensive national dataset of various environmental parameters, so that would be beneficial for the usage of the MSA. Furthermore, the MSA, unlike the LRA, works well in patchy, cultural landscapes and does not have to omit wetland taxa, which means the landscape reconstruction would be completer and more realistic. This would be very helpful in a landscape reconstruction in the Netherlands; being in a coastal low altitude region and incorporating the large Rhine-Meuse delta made for the development of numerous and variable wetland areas during the Holocene.

A difficulty remains when assigning high taxonomic pollen groups to an ecological group

(Bunting et al., 2018). For instance Poaceae (grasses) are often not determinable on a lower taxonomic level, but they can inhabit both wetlands (e.g. reeds) and drylands.

EDA

The EDA also uses landscape parameters to simulate pollen assemblages, but the EDA sets up only one landscape based on the available abiotic parameters and uses a key to optimize the correlations between those parameters and vegetation in relation to the empirical pollen assemblage (Theuerkauf et al., 2014). The EDA works best with small to medium sized basins, which are numerous in the Netherlands. Additionally, it is mentioned that the EDA works optimal when landscape parameters are invariable over time (Theuerkauf & Joosten, 2009). For some (geomorphological) parameters, this would not be a problem, but other parameters (e.g. soil substrate, ground water level, distance to the sea) may have varied more strongly in a coastal region as the Netherlands as a result of sea level rise. While there is no case study available where this assumption is heavily violated, it may be more important that the pollen taxa are strongly dependent on an environmental parameter and to be wary for landscape parameters that are closely correlated but are inhabited by very different taxa, because under violation of those criteria, the EDA seems to generate more false-positive landscape-vegetation cover correlations (Theuerkauf & Couwenberg, 2017). In the Netherlands, a large part of the country has been shaped by peat formation and anthropogenic peat harvesting and land reclamations, which could make looking for correlations between the variable landscape parameters and the vegetation cover more informative compared to the invariable landscape parameters.

Conclusion

I conclude that in theory, all three models could be applicable to an intermediate scale landscape, in this case the Netherlands. With the LRA, the main bottleneck would be the availability of large basins for REVEALS, which in turn would make the execution of LOVE impossible, since the regional vegetation factor for REVEALS is essential input. Both the MSA and EDA benefit from the large Dutch database of abiotic soil parameters and are able to yield high resolution reconstructions. For the MSA, scaling up from a local to intermediate scale landscape reconstruction exponentially increases the computation times of all scenarios, so that may be an important consideration to keep in mind. For the EDA, it is mainly important to avoid false-positive landscape parameter-vegetation cover correlations and it could in that light be inconvenient that the vegetation cover in the north-western part of the Netherlands may have been more dependent on variable landscape parameters due to sea level rise from the mid Holocene onwards. Independently, all models would benefit from further research on dispersal models and renewed investigation towards the contributions of different modes of pollen transport.

Literature

- Abraham, V., Man, M., Theuerkauf, M., Pokorný, P., Bobek, P., & Novák, J. (2023). Spatially explicit, quantitative reconstruction of past vegetation based on pollen or charcoal data as a tool for autecology of trees. *Landscape Ecology*, 38(7), 1747–1763. <https://doi.org/10.1007/s10980-023-01652-8>
- Andersen, S. T. (1970). The Relative Pollen Productivity and Pollen Representation of North European Trees, and Correction Factors for Tree Pollen Spectra. Determined by Surface Pollen Analyses from Forests. *Danmarks Geologiske Undersøgelse. 2. Række*, 96, 1–99. <https://doi.org/10.34194/raekke2.v96.6887>
- Andersen, S. T. (1974a). Wind Conditions and Pollen Deposition in a Mixed Deciduous Forest. *Grana*, 14(2–3), 57–63. <https://doi.org/10.1080/00173137409429894>
- Andersen, S. T. (1974b). Wind Conditions and Pollen Deposition in a Mixed Deciduous Forest. *Grana*, 14(2–3), 64–77. <https://doi.org/10.1080/00173137409429895>
- Birks, H. J. B., & Berglund, B. E. (2017). One hundred years of Quaternary pollen analysis 1916–2016. *Vegetation History And Archaeobotany*, 27(2), 271–309. <https://doi.org/10.1007/s00334-017-0630-2>
- Broström, A., Nielsen, A. B., Gaillard, M., Hjelle, K., Mazier, F., Binney, H., Bunting, J., Fyfe, R., Meltsov, V., Poska, A., Räsänen, S., Soepboer, W., Von Stedingk, H., Suutari, H., & Sugita, S. (2008). Pollen productivity estimates of key European plant taxa for quantitative reconstruction of past vegetation: a review. *Vegetation History And Archaeobotany*, 17(5), 461–478. <https://doi.org/10.1007/s00334-008-0148-8>
- Broström, A., Sugita, S., & Gaillard, M. (2004). Pollen productivity estimates for the reconstruction of past vegetation cover in the cultural landscape of southern Sweden. *Holocene*, 14(3), 368–381. <https://doi.org/10.1191/0959683604hl713rp>
- Bunting, M. J., Farrell, M., Bayliss, A., Marshall, P., & Whittle, A. (2018). Maps From Mud—Using the Multiple Scenario Approach to Reconstruct Land Cover Dynamics From Pollen Records: A Case Study of Two Neolithic Landscapes. *Frontiers in Ecology And Evolution*, 6. <https://doi.org/10.3389/fevo.2018.00036>
- Bunting, M. J., Farrell, M., Dunbar, E., Reimer, P., Bayliss, A., Marshall, P., & Whittle, A. (2022). Landscapes for Neolithic People in Mainland, Orkney. *Journal Of World Prehistory*, 35(1), 87–107. <https://doi.org/10.1007/s10963-022-09166-y>
- Bunting, M. J., & Middleton, D. (2005). Modelling pollen dispersal and deposition using HUMPOL software, including simulating windroses and irregular lakes. *Review Of Palaeobotany And Palynology*, 134(3–4), 185–196. <https://doi.org/10.1016/j.revpalbo.2004.12.009>
- Bunting, M., & Middleton, R. (2009). Equifinality and uncertainty in the interpretation of pollen data: the Multiple Scenario Approach to reconstruction of past vegetation mosaics. *Holocene*, 19(5), 799–803. <https://doi.org/10.1177/0959683609105304>
- Bunting, M., Twiddle, C., & Middleton, R. (2008). Using models of pollen dispersal and deposition in hilly landscapes: Some possible approaches. *Palaeogeography Palaeoclimatology Palaeoecology*, 259(1), 77–91. <https://doi.org/10.1016/j.palaeo.2007.03.051>

- Butcher, C. L., Rubin, B. Y., Anderson, S. L., & Lewis, J. D. (2020). Pollen dispersal patterns differ among sites for a wind-pollinated species and an insect-pollinated species. *American Journal of Botany*, 107(11), 1504–1517. <https://doi.org/10.1002/ajb2.1554>
- Calcote, R. (1995). Pollen Source Area and Pollen Productivity: Evidence from Forest Hollows. *Journal Of Ecology*, 83(4), 591. <https://doi.org/10.2307/2261627>
- Cohen, K. M., & Pierik, H. J. (2021). Actualisatie en uitbreiding van begraven landschapskaarten voor Holoceen afgedekt Nederland: 2021 versie kaartlaag T0123 voor RCE's Kenniskaartportaal.
- Davis, M. B. (1963). On the theory of pollen analysis. *American Journal Of Science/~The American Journal Of Science*, 261(10), 897–912. <https://doi.org/10.2475/ajs.261.10.897>
- Fægri, K. (1966). Some problems of representivity in pollen analysis. *Journal Of Palaeosciences*, 15((1-3)), 135–140. <https://doi.org/10.54991/jop.1966.759>
- Fagerlind F. (1952). The real signification of pollen diagrams. *Bot Not* 105:185–224
- Fang, Y. (2018). *Relationships between surface pollen and vegetation in the Meiling Mountains, southeast China: an aid to reconstruct past vegetation dynamics*. <https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.769110>
- Fang, Y., Ma, C., & Bunting, M. J. (2019). Novel methods of estimating relative pollen productivity: A key parameter for reconstruction of past land cover from pollen records. *Progress in Physical Geography Earth And Environment*, 43(6), 731–753. <https://doi.org/10.1177/0309133319861808>
- Fredh, E. D., Lagerås, P., Mazier, F., Björkman, L., Lindbladh, M., & Broström, A. (2019). Farm establishment, abandonment and agricultural practices during the last 1,300 years: a case study from southern Sweden based on pollen records and the LOVE model. *Vegetation History And Archaeobotany*, 28(5), 529–544. <https://doi.org/10.1007/s00334-019-00712-x>
- Gaillard, M., Sugita, S., Bunting, M. J., Middleton, R., Broström, A., Caseldine, C., Giesecke, T., Hellman, S. E. V., Hicks, S., Hjelle, K., Langdon, C., Nielsen, A., Poska, A., Von Stedingk, H., & Veski, S. (2008). The use of modelling and simulation approach in reconstructing past landscapes from fossil pollen data: a review and results from the POLLANDCAL network. *Vegetation History And Archaeobotany*, 17(5), 419–443. <https://doi.org/10.1007/s00334-008-0169-3>
- Hellman, S., Bunting, M., & Gaillard, M. (2008). Relevant Source Area of Pollen in patchy cultural landscapes and signals of anthropogenic landscape disturbance in the pollen record: A simulation approach. *Review Of Palaeobotany And Palynology*, 153(3–4), 245–258. <https://doi.org/10.1016/j.revpalbo.2008.08.006>
- Hjelle, K. L., Mehl, I. K., Sugita, S., & Andersen, G. L. (2015). From pollen percentage to vegetation cover: evaluation of the Landscape Reconstruction Algorithm in western Norway. *Journal Of Quaternary Science*, 30(4), 312–324. <https://doi.org/10.1002/jqs.2769>
- Iversen, J. (1941). Landnam i Danmarks Stenalder. En pollenanalytisk Undersøgelse over det første Landbrugs Indvirkning paa Vegetationsudviklingen. *Danmarks Geologiske Undersøgelse. 2. Række*, 66, 1–68. <https://doi.org/10.34194/raekke2.v66.6855>

Jackson, S. T. (1990). Pollen source area and representation in small lakes of the northeastern United States. *Review Of Palaeobotany And Palynology*, 63(1–2), 53–76. [https://doi.org/10.1016/0034-6667\(90\)90006-5](https://doi.org/10.1016/0034-6667(90)90006-5)

Jackson, S. T., & Lyford, M. E. (1999). Pollen dispersal models in Quaternary plant ecology: Assumptions, parameters, and prescriptions. *The Botanical Review*, 65(1), 39–75. <https://doi.org/10.1007/bf02856557>

Jackson, S. T., Webb, T., Prentice, I., & Hansen, J. E. (1995). Exploration and calibration of pollen/vegetation relationships: a PC program for the extended R-value models. *Review Of Palaeobotany And Palynology*, 84(3–4), 365–374. [https://doi.org/10.1016/0034-6667\(94\)00113-x](https://doi.org/10.1016/0034-6667(94)00113-x)

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

Kuparinen, A., Markkanen, T., Riikonen, H., & Vesala, T. (2007). Modeling air-mediated dispersal of spores, pollen and seeds in forested areas. *Ecological Modelling*, 208(2–4), 177–188. <https://doi.org/10.1016/j.ecolmodel.2007.05.023>

Liu, Y., Ogle, K., Lichstein, J. W., & Jackson, S. T. (2022). Estimation of pollen productivity and dispersal: How pollen assemblages in small lakes represent vegetation. *Ecological Monographs*, 92(3). <https://doi.org/10.1002/ecm.1513>

Lososová, Z., Divíšek, J., Chytrý, M., Götzenberger, L., Těšitel, J., & Mucina, L. (2020). Macroevolutionary patterns in European vegetation. *Journal Of Vegetation Science*, 32(1). <https://doi.org/10.1111/jvs.12942>

Mariani, M., Connor, S., Theuerkauf, M., Kuneš, P., & Fletcher, M. (2016). Testing quantitative pollen dispersal models in animal-pollinated vegetation mosaics: An example from temperate Tasmania, Australia. *Quaternary Science Reviews*, 154, 214–225. <https://doi.org/10.1016/j.quascirev.2016.10.020>

Marquer, L., Mazier, F., Sugita, S., Galop, D., Houet, T., Faure, E., Gaillard, M., Haunold, S., De Munnik, N., Simonneau, A., De Vleeschouwer, F., & Roux, G. L. (2019). Pollen-based reconstruction of Holocene land-cover in mountain regions: Evaluation of the Landscape Reconstruction Algorithm in the Vicdessos valley, northern Pyrenees, France. *Quaternary Science Reviews*, 228, 106049. <https://doi.org/10.1016/j.quascirev.2019.106049>

Parsons, R., Gordon, A., & Prentice, I. (1983). Statistical uncertainty in forest composition estimates obtained from fossil pollen spectra via the R-value model. *Review Of Palaeobotany And Palynology*, 40(3), 177–189. [https://doi.org/10.1016/0034-6667\(83\)90035-0](https://doi.org/10.1016/0034-6667(83)90035-0)

Parsons, R., & Prentice, I. (1981). Statistical approaches to R-values and the pollen—vegetation relationship. *Review Of Palaeobotany And Palynology*, 32(2–3), 127–152. [https://doi.org/10.1016/0034-6667\(81\)90001-4](https://doi.org/10.1016/0034-6667(81)90001-4)

Prentice, I. (1985). Pollen Representation, Source Area, and Basin Size: Toward a Unified Theory of Pollen Analysis. *Quaternary Research*, 23(1), 76–86. [https://doi.org/10.1016/0033-5894\(85\)90073-0](https://doi.org/10.1016/0033-5894(85)90073-0)

Prentice, I. C., Berglund, B. E., & Olsson, T. (1987). Quantitative forest-composition sensing characteristics of pollen samples from Swedish lakes. *Boreas*, 16(1), 43–54. <https://doi.org/10.1111/j.1502-3885.1987.tb00753.x>

- Prentice, I. C., & Parsons, R. W. (1983). Maximum Likelihood Linear Calibration of Pollen Spectra in Terms of Forest Composition. *Biometrics*, 39(4), 1051. <https://doi.org/10.2307/2531338>
- Prentice, I. C., & Webb, T. (1986). Pollen percentages, tree abundances and the Fagerlind effect. *JQS. Journal Of Quaternary Science/Journal Of Quaternary Science*, 1(1), 35–43. <https://doi.org/10.1002/jqs.3390010105>
- Rempe, H. (1937). Untersuchungen über die Verbreitung des Blütenstaubes durch die Luftströmungen. *Planta*, 27(1), 93–147. <https://doi.org/10.1007/bf01939376>
- Serge, M., Mazier, F., Fyfe, R., Gaillard, M., Klein, T., Lagnoux, A., Galop, D., Githumbi, E., Mindrescu, M., Nielsen, A., Trondman, A., Poska, A., Sugita, S., Woodbridge, J., Abel-Schaad, D., Åkesson, C., Alenius, T., Ammann, B., Andersen, S., . . . Zernitskaya, V. (2023). Testing the Effect of Relative Pollen Productivity on the REVEALS Model: A Validated Reconstruction of Europe-Wide Holocene Vegetation. *Land*, 12(5), 986. <https://doi.org/10.3390/land12050986>
- Sjögren, P., Van Der Knaap, W., Huusko, A., & Van Leeuwen, J. F. (2008). Pollen productivity, dispersal, and correction factors for major tree taxa in the Swiss Alps based on pollen-trap results. *Review Of Palaeobotany And Palynology*, 152(3–4), 200–210. <https://doi.org/10.1016/j.revpalbo.2008.05.003>
- Soepboer, W., Sugita, S., & Lotter, A. F. (2009). Regional vegetation-cover changes on the Swiss Plateau during the past two millennia: A pollen-based reconstruction using the REVEALS model. *Quaternary Science Reviews*, 29(3–4), 472–483. <https://doi.org/10.1016/j.quascirev.2009.09.027>
- Sugita, S. (1993). A Model of Pollen Source Area for an Entire Lake Surface. *Quaternary Research*, 39(2), 239–244. <https://doi.org/10.1006/qres.1993.1027>
- Sugita, S. (1994). Pollen Representation of Vegetation in Quaternary Sediments: Theory and Method in Patchy Vegetation. *Journal Of Ecology*, 82(4), 881. <https://doi.org/10.2307/2261452>
- Sugita, S. (2007a). Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition. *Holocene*, 17(2), 229–241. <https://doi.org/10.1177/0959683607075837>
- Sugita, S. (2007b). Theory of quantitative reconstruction of vegetation II: all you need is LOVE. *Holocene*, 17(2), 243–257. <https://doi.org/10.1177/0959683607075838>
- Sugita, S., Gaillard, M., & Broström, A. (1999). Landscape openness and pollen records: a simulation approach. *Holocene*, 9(4), 409–421. <https://doi.org/10.1191/095968399666429937>
- Tauber, H. (1965). Differential pollen dispersion and the interpretation of pollen diagrams. With a contribution to the interpretation of the elm fall. *Danmarks Geologiske Undersøgelse. 2. Række*, 89, 1–69. <https://doi.org/10.34194/raekke2.v89.6880>
- Tauber, H. (1967). Investigations of the mode of pollen transfer in forested areas. *Review Of Palaeobotany And Palynology*, 3(1–4), 277–286. [https://doi.org/10.1016/0034-6667\(67\)90060-7](https://doi.org/10.1016/0034-6667(67)90060-7)
- Theuerkauf, M., Bos, J. A., Jahns, S., Janke, W., Kuparinen, A., Stebich, M., & Joosten, H. (2014). *Corylus* expansion and persistent openness in the early Holocene vegetation of northern central Europe. *Quaternary Science Reviews*, 90, 183–198. <https://doi.org/10.1016/j.quascirev.2014.03.002>

Theuerkauf, M., & Couwenberg, J. (2017). The extended downscaling approach: A new R-tool for pollen-based reconstruction of vegetation patterns. *Holocene*, 27(8), 1252–1258. <https://doi.org/10.1177/0959683616683256>

Theuerkauf, M., & Couwenberg, J. (2022). Pollen productivity estimates strongly depend on assumed pollen dispersal II: Extending the ERV model. *Holocene*, 32(11), 1233–1250. <https://doi.org/10.1177/09596836211041729>

Theuerkauf, M., & Couwenberg, J. (2024). LOVE Is in the R—Two R Tools for Local Vegetation Reconstruction. *Quaternary*, 7(2), 18. <https://doi.org/10.3390/quat7020018>

Theuerkauf, M., Couwenberg, J., Kuparinen, A., & Liebscher, V. (2016). A matter of dispersal: REVEALSinR introduces state-of-the-art dispersal models to quantitative vegetation reconstruction. *Vegetation History And Archaeobotany*, 25(6), 541–553. <https://doi.org/10.1007/s00334-016-0572-0>

Theuerkauf, M., & Joosten, H. (2009). Substrate dependency of Lateglacial forests in north-east Germany: untangling vegetation patterns, ecological amplitudes and pollen dispersal in the past by downscaling regional pollen. *Journal Of Biogeography*, 36(5), 942–953. <https://doi.org/10.1111/j.1365-2699.2008.02047.x>

Theuerkauf, M., Kuparinen, A., & Joosten, H. (2013). Pollen productivity estimates strongly depend on assumed pollen dispersal. *Holocene*, 23(1), 14–24. <https://doi.org/10.1177/0959683612450194>

Willis, A. D. (2019). Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.02407>