



Interpretation of the herbaceous pollen spectra in paleoecological reconstructions: A spatial extension of Indices of Association and determination of individual pollen source areas from binary data

Helen Shaw ^{*}, Ian Whyte

Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster LA1 4YW, UK

ARTICLE INFO

Article history:

Received 13 December 2019
Received in revised form 1 April 2020
Accepted 7 April 2020
Available online 11 April 2020

Keywords:

Paleoecology
Herbaceous pollen
Pollen dispersal,
Indices of Association
Spatially extended Indices of Association
Individual Pollen Source areas

ABSTRACT

To examine the relationship between plant presence and pollen presence, modern pollen samples were collected and counted from natural moss polsters in the UK uplands. The herbaceous pollen signal is difficult to interpret in paleoecological analysis as pollen presence is often rare for individual herb taxa. Indices of association have previously been used to determine association from over- and under-representation; however, previous surveys have used spatially limited single-source areas. Indices of Association (IOA) between pollen and plant presence at increasing distance from the pollen source, to 50 m, are presented in a new spatially extended IOA ($_{SE}$ IOA) method. For several pollen taxa, including *Plantago lanceolata*, *Ranunculus acris*-type, *Rumex* and the Cichorioideae family, association increased progressively, when plants at greater distance were included in the calculation, and over-representation was reduced. Other taxa become increasingly under-represented at the landscape scale, e.g., *Filipendula*, Apiaceae, *Cerastium* indicating true ultra-local source areas. A weak association and increasing under-representation indices, cautions that pollen absence of these types in paleoecological analysis is not a secure indicator of plant absence. Extended IOA values can aid the interpretation of the rare-type pollen spectra in paleoecological analyses. The method shows potential for developing individual pollen source areas (iPSAs) for herbaceous pollen types. These iPSAs together with the probability of correct application, determined by accuracy and association scores, are useful when applying pollen correction factors in pollen-derived vegetation community and landscape models. Further indices and significance testing are suggested to enhance understanding of the association and the rate of correct classification. The results can assist both interpretation of community composition from the herbaceous pollen spectra in local-scale paleoecological studies, and understanding the limits to this interpretation in paleoecological analysis.

© 2020 Published by Elsevier B.V.

1. Introduction

Pollen analysis is a widely used method employed to gain an understanding of past ecological environments (e.g., Edwards et al., 2017; Birks and Berglund, 2018; de Klerk, 2018). Pollen, extracted, identified and counted from stratigraphic subsamples from peat, soil or lake sediment cores, is used as a proxy for plant presence. This simple method belies considerable complexity in interpretation for paleoecological reconstructions. Pollen is produced in different quantities by different plant taxa. Pollen also undertakes differing modes of dispersal, facilitated variously by wind, insects and other vectors, as well as by pollen morphological differences between taxa (Faegri and Iversen, 2000). These factors drive taxon-specific variation in the amount of pollen

transported from plant to sampled substrate, as well as variation in the distance of dispersal. One of the enduring problems of paleoecological analysis is, therefore, the extraction of spatially and quantitatively accurate reconstructions of the vegetation in past environments from the analysis of this complex proxy data (e.g., Davis, 1963; Andersen, 1970; Prentice, 1985; Sugita, 1993, 1994; Davis, 2000; Sugita, 2007a, 2007b; Gaillard et al., 2008a, 2008b; Hellman et al., 2009a, 2009b; Nielsen et al., 2012; Felde et al., 2014; Abraham et al., 2014).

The pollen spectra in a paleoecological sample is usually dominated by a small selection of key pollen taxa. These are accompanied by a much wider variety of pollen types for which there may be single, or very few, pollen grains counted per taxon – referred to as rare-type pollen taxa. The differing quantities of pollen counted generally reflect the different pollen dispersal mechanisms: key taxa being produced abundantly and mostly wind-dispersed, and rare-type taxa principally deriving from insect pollinated species. There is, however, some overlap between dispersal mechanisms leading to considerable variation in pollen–vegetation relationships between taxa.

^{*} Corresponding author at: Maynooth University, ICARUS/Department of Geography, Maynooth, Ireland.

E-mail address: helen.shaw@mu.ie (H. Shaw).

Three issues require attention in order to produce quantified environmental reconstructions using taxon-specific calibrations.

1. How to adjust pollen abundance data to reflect, more accurately, the plant cover?
2. Over what radius from the site of pollen deposition is the environmental reconstruction relevant?
3. How consistent are the above factors across sites?

These issues have been subject to over 50 years of testing, leading to the development of a range of quantitative models.

Pollen–plant correlations as R-values (Davis, 1963) were expanded to include a background term for pollen derived from beyond the surveyed area and developed later into extended R-values (ERV) (Parsons and Prentice, 1981; Prentice and Parsons, 1983), using maximum likelihood methods. This paved the way for the Prentice–Sugita model (Prentice, 1985; Sugita, 1993) for the reconstruction of land cover (e.g., Nielsen and Odgaard, 2005). Based on this model, the Landscape Reconstruction Algorithm (LRA) (e.g., Sugita, 2007a, 2007b) is now a well-established method which drives more robust interpretations of paleoecological histories across regional landscapes (e.g., Hellman et al., 2008; Gaillard et al., 2010; Sugita et al., 2010; Fyfe et al., 2013; Nielsen and Odgaard, 2010; Hjelle et al., 2018). A key function of these models, reviewed in Gaillard et al. (2008), is recalibrating pollen data by adjusting the abundance by a Pollen Productivity Estimate (PPE) (e.g., Broström et al., 2005) allowing for application of the adjusted abundance data to a spatial scale defined by the relevant source area for pollen (RSAP). Testing of the models and development of PPEs and RSAPs is based on iterative datasets, repeatedly comparing modern pollen spectra with plant presence at increasing survey distance, for key pollen types which appear abundantly in the paleoecological record (e.g., Prentice et al., 1987; Broström et al., 2004; Nielsen and Sugita, 2005; Soepboer et al., 2007; Broström et al., 2008; von Stedingk et al., 2008; Hellman et al., 2009a, 2009b; Poska et al., 2011; Bunting et al., 2013a).

However, an increasingly important role for applied paleoecology is the evaluation of local-scale ecological change over recent centuries (Rull, 2014). This application of paleoecology provides essential information for nature conservation (e.g., Birks, 1993; Odgaard, 1999; Shaw and Tipping, 2006; Feeser and O'Connell, 2009; Fredh et al., 2012; Shaw and Whyte, 2013). Studies of this type need to contribute information on past ecological community structure and temporal shifts in species presence (as reviewed in Seddon et al., 2014), often at the limit of the taxonomic and interpretive resolution of paleoecology.

Rare-type pollen taxa appear in such low quantities in the pollen record that they cannot be analyzed by ERV methods, therefore important herbaceous types may be underestimated (e.g., Hellman et al., 2009b) and rare-types may be excluded from the LRA. Yet these taxa define vegetation communities and depict diversity – changes in which are key to a detailed understanding of long-term ecological dynamics (e.g., Marquer et al., 2014). It is in local-scale biodiversity and ecological change studies that the lack of ability to interpret quantitatively, or confidently, the array of rare-type pollen taxa is most frustrating. We need to understand the extent with which we can rely on rare herb pollen data as indicators of local or regional environmental presence.

Two main methods have been used to examine the rare-type pollen–vegetation signal. Comparative approaches and the modern analog technique (e.g., Delcourt and Delcourt, 1985; Overpeck et al., 1985; and reviewed in Birks and Gordon (1985) and Felde et al. (2014) allow ordination between pollen spectra diversity and plant community diversity, in order to look for modern analogs where the pollen spectra produces recognizable extant community types (e.g., Gaillard et al., 1992; Gaillard et al., 1994; Mazier et al., 2009). Whilst Indices of Association, introduced by Davis (1984), measure taxon-specific association, over and under-representation, based upon the development of indices derived from simple presence/absence data.

Comparative methods are useful to understand if paleoecologists can interpret plant community types from pollen spectra, however, given that rare pollen types are used in these analyses, we also need to be able to understand to what extent each rare pollen taxa can be securely interpreted; and, over what distance our community interpretation extends.

Since their introduction, Indices of Association (IOA) have been used in several studies of rare-type pollen. These indices have proved useful in interpreting studies of open pastoral and moorland landscapes, as they can help to determine which rare pollen-types provide a consistent representation of vegetation presence in areas where forb diversity can indicate arable, grazing and/or cultural activity. (e.g., Hjelle, 1997, 1998; Bunting, 2003; Shaw, 2006; Mazier et al., 2006; Brun et al., 2007; Schofield et al., 2007; Cañellas-Boltà et al., 2009; Nienartowicz et al., 2011; Fjordheim et al., 2018). IOA have also been compared with pollination mechanisms to determine the extent that we can predict over and under-represented types (e.g., Hjelle, 1997; Brun et al., 2007).

However, one limit to Davis Indices has been the lack of spatial scale in the analyses. Plant surveys associated with these studies have been limited; usually to 1 or 2 m around each site (e.g., Schofield et al., 2007; Bunting, 2003), under the assumption that pollen may only disperse a few meters from the source plant (Bunting, 2003; Bunting and Hjelle, 2010; Bunting et al., 2013b). Where IOA studies have related pollen presence to a wider plant source area, a single plot (e.g., 10 × 10 m for Hjelle, 1997; 2.5 m × 5 m for Fjordheim et al., 2018), or a vegetation sampling approach, was used to collect data on plant presence (e.g., Mazier et al., 2006; Brun et al., 2007). All of these IOA studies display high over-representation scores among many taxa; which, as observed by Davis (1984), may indicate a wider pollen source area. Whilst most researchers put this down to anemophilous pollen dispersal mechanisms (e.g., Schofield et al., 2007, p. 483, Brun et al., 2007), there has been no recourse to testing further the extent of the local source area for pollen in these simple binary studies. In addition, some intraspecific variation of indices values between studies is found (cf. Hjelle, 1997; Mazier et al., 2006; Schofield et al., 2007; Brun et al., 2007; Cañellas-Boltà et al., 2009; Nienartowicz et al., 2011; Fjordheim et al., 2018); however, it is unclear if this is due to site-specific variation, or influenced by the differing limits of the plant survey area.

The starting point for key taxa calibrations is based in the ERV approach. Plant survey data, as percentage, cover or abundance, are collected in concentric circles (e.g., Bunting et al., 2013a, and references therein) at increasing distance from a pollen source. Pollen abundance is then iteratively correlated against cumulative plant abundance at increasing survey distance to determine a relevant source area for pollen (Sugita, 1994). This approach, although a well-established method for key taxa, has had little application for rare taxa, where abundance data are often limited. However, the IOA indices would, logically, lend themselves well to a similar approach. Repeated IOA analysis can, theoretically, be carried out on cumulative plant presence, measured in binary form over increasing survey areas; with the results plotted to assess patterns of attenuation in the association score over increasing pollen dispersal distance. It is important to know over what scale in the landscape we can make a paleoecological interpretation of plant presence (Bennington et al., 2009), or apply PPEs and pollen representation factors on rare-pollen taxa (sensu Hjelle, 1998, 1999). Extending the IOA approach would seem, therefore, to be an area for fruitful research (Shaw, 2012), as it could be used to elucidate the area around the sample of best association between rare pollen types and their source plants and could be used to highlight any trade-offs between over- and under-representation.

This paper presents the results from two study areas in the UK where IOA analyses were applied against vegetation data at increasing source areas to test this new spatial extension of the IOA method. Additional measures, to include negative co-occurrences, as well as testing for significance, are implemented. The study provides an opportunity to examine some of the individual dispersal distances among the

herbaceous taxa and to test the potential for greater understanding of the pollen–vegetation relationship of rare taxa, and to derive independent pollen source areas (iPSA) for these rare herb types, through spatially extended IOA ($s_{E}IOA$). The $s_{E}IOA$ data also provides an opportunity to compare IOA values for several taxa across a range of European studies and to discuss the interpretation of pollen data from these taxa.

2. Case study areas

Two study sites were used, the main study is from samples in the pastoral uplands of Ribblesdale in Northern England, and additional samples are included for comparison, from an earlier study, in semi-open pine woodland and heathland environment (see Fig. 1) in Glen Affric, Scotland.

Nineteen samples were taken from upper Ribblesdale, North Yorkshire, UK, in May and June of 2009 and 2010. The area comprises pastoral upland grassland predominantly overlying limestone bedrock, with alternating beds of the Yoredale group comprising sandstone and shale bands making up the Ingleborough massif above (Johnson, 2008). The plant communities vary from limestone grassland of the CG9 *Sesleria albicans*–*Galium sternerii* grassland (Rodwell, 1992, 2006) to various mesotrophic and acid grassland and heathland communities where layers of loess, or glacial deposits, overlie the limestone and separate the plant communities from alkaline influence. Wet acid bog deposits have accumulated in a number of small blocked sink-holes.

The area is currently managed with grazing regimes of varying intensity with some areas left ungrazed for conservation management (Natural England, n.d.); grazing, where it occurs, is by sheep and beef

cattle. Arboreal cover is very limited with much of the area almost completely devoid of trees. A long history of pastoral land-use has influenced the vegetation. The $s_{E}IOA$ at this site forms part of a wider study to examine the impact of post-medieval land management in order to understand, in more detail, the relevance of restoring traditional management in maintaining biodiversity and environmental sustainability (Shaw and Whyte, 2013).

Ribblesdale may differ subtly, in ecological structure, from other areas previously studied with Indices of Association. Although valued for biodiversity, in reality the diversity in upland meadows and pastoral landscapes in the UK may be lower than in many areas of Europe. Wood pasture is scarce. There has been no recent period of abandonment and arboreal recovery in the UK uplands. Hay meadow management saw a period of earlier decline, with the subsequent expansion of improved grazing and rough pasture. Ecological community differences are subtle between sheep-grazed and cattle-grazed pastures within the system. There is a need, therefore, to understand the impact, on forb communities, of very subtle changes in management through time in these very open pastoral landscapes (Shaw and Whyte, 2008); and, a need to test whether paleoecology can resolve this type of detail, in order to link with ecological management. In particular, past grazing regimes, especially heavy sheep-grazing since the monastic period, tends to lead to very small phenotypes, flowering low to the ground. These may produce or disperse different quantities of pollen than the same forb taxa in the extensively managed pastures, wood pastures and hay meadows, where IOA studies have been undertaken in other areas of Europe (e.g., Hjelle, 1998, 1999).

A further 29 samples are also used in the analysis, from Glen Affric surveys undertaken in 2002 and 2003. The sampling was focused in and around the patchy native pine woodland and the more open moorland areas. The Glen Affric pine woods are subject to a rewilding strategy and the main focus of this study was to elucidate changes in pine forest cover through time (Shaw and Tipping, 2006). However, an understanding of the ability to detect openness in pollen diagrams was key to this study. To this end, modern pollen–plant surveys were undertaken in Glen Affric focussed on both key and rare-type pollen taxa (Shaw, 2006). Although in less detail than in Ribblesdale, the addition of the data collected from the Glen Affric study allows for some limited extension and testing here, of the Spatially Extended IOA model in different environments.

3. Methods

3.1. Sample collection

A moss polster was collected from the center of every survey area. Moss polsters act as a natural pollen trap in the landscape (e.g., Cundill, 1979). A variety of moss types were utilized. Whilst it would have been better to have used one moss type, for example, the sphagnum mosses found in small hollows, in reality this would have meant limiting the vegetation communities sampled. Where possible *Sphagnum* or other damp mosses were used, but, in some cases, other dry land moss/turf samples were collected. Previous studies have found differences in pollen accumulation between polster types (e.g., Boyd, 1986), but reasonable local representation (Hall, 1989) indicating that the range of samples should not create bias in this local scale study. Samples were sealed in plastic bags and stored in a refrigerator until processed for pollen extraction.

Pollen was extracted from a single moss polster. This method has varied between researchers; with some amalgamating several samples, within a 50 cm diameter, or within the quadrat to ensure mixing of any taxa over-represented due to the inclusion of anthers (e.g., Pardoe, 2001; Mazier et al., 2008); and, some preferring to extract a single sample (e.g., Räsänen et al., 2007; and see Bunting et al., 2013a, table 2, for a review of methods). It is reasoned here that the single sample approach is more analogous to the single sample used in paleoecological analysis.



Fig. 1. Locations of Glen Affric and Ribblesdale in the UK. Base map adapted from image: Uk_outline_map.png / CC BY-SA (<http://creativecommons.org/licenses/by-sa/3.0/>).

Subsamples from a peat core represent a very small surface area and thus reduce the probability that rare pollen taxa are found. Amalgamating subsamples may increase that probability to an unrealistic level for comparison with paleoecological samples. In addition, for other quantified analyses, a sample spiked with an anther (and thus producing a biased pollen spectra with an unusually high pollen count), can be easily identified and discarded, whereas, in an amalgamated sample, this increased count may be diluted to an abundance where bias is still present, but not evident to the analyst.

3.2. Field survey methods in Ribblesdale

Sample locations were selected using a stratified sampling approach to include study sites with a variety of grazing densities and vegetation communities. Surveys were carried out between April and July in 2010 and 2011. Plant surveys were carried out in concentric rings at 0–1, 1–2, 2–3, 3–4, 4–6, 6–8, 8–10, 10–15, 15–20, 20–30, 30–40, 40–50 m radius from the moss collection point. Each ring was marked into eight sections (octants) to aid survey using canes and 50 m tapes (Fig. 2).

An estimate of percentage plant cover was made, for key taxa (Poaceae, Cyperaceae, Ericaceae agg., *Calluna vulgaris*, herbaceous species agg.), within each octant. Herbaceous types and Pteridophytes were recorded using the DAFOR scale. Arboreal taxa presence was also recorded by trunk presence in each octant.

In contrast to many studies, which use a sampling approach to vegetation survey, surveying individual quadrats within each ring source or quadrant area (reviewed in Bunting et al., 2013a), an attempt was made here to record all taxa present within each octant. This is because rare-type pollen grains come from a variety of herbaceous species, which were observed to have a patchy presence within sub-communities at the micro-scale. The use of extrapolations of plant cover from quadrats within quadrants may be sensible for determining PPEs of key taxa; but it was considered here that it may not be adequate for detecting all of the rare plant taxa present in the mosaic of acid/mesotrophic/basic grassland and heathland communities present. The use of percentage cover derived from single quadrats within each ring source may also bias the measure of percentage cover of plants when scaled-up in a

patchy landscape, leading to over-estimation of the cover of some taxa and under-estimation of others depending on patchy presence captured in the quadrat.

3.3. Field survey methods in Glen Affric

Samples in Glen Affric were selected from a variety of open and semi-open pine woodland, as well as glades and open moorland communities. Surveys were carried out in May and June 2002 and 2003. Survey areas of 0–1, 1–2, 2–3, 3–4, 4–5, 5–10 and 10–20 m radius from the pollen source were used for herbaceous types, with trees and heaths recorded to 50 m. In Glen Affric, in contrast to the Ribblesdale survey method, percentage cover of key taxa was only carried out in random quadrats within the two widest source area rings. A whole octant walk-over approach was used for the recording of presence of rare-types. Note, the Glen Affric data for herbaceous types have a more limited survey extent with surveys to 20 m in contrast to 50 m in Ribblesdale.

3.4. Laboratory methods

Pollen was extracted using standard techniques (Moore et al., 1991) including HCl (10%), the addition of *Lycopodium* spores (Stockmarr, 1971), deflocculation in KOH (10%), retaining the sample partitioned between 6 and 180 µm sieves, followed by acetolysis, but without the use of HF. Pollen was stained with Safranin O stain and mounted in silicon fluid.

Pollen was counted using an Olympus BX40 microscope (Glen Affric samples) and a Leica DM2500 (Ribblesdale samples). In both datasets routine counting at ×400 magnification was supplemented with ×1000 oil immersion and phase contrast to distinguish between similar pollen taxa and clarify identification of difficult rare-types. In Glen Affric, samples were counted to at least 500 total land pollen grains (tlp); whilst, in Ribblesdale, each full slide was counted to ensure that any uneven dispersal on slides did not bias the count. This means that there is variation in tlp count, especially in the Ribblesdale samples, where samples contained between c. 400 and c. 1800 tlp. Glen Affric samples, counted to c. 500 tlp, often contained high counts of arboreal pollen taxa including *Pinus*, as well as heathland pollen taxa, the herbaceous pollen taxa counts are, therefore, low in most samples from this dataset. The open nature of the Ribblesdale landscape, with the correspondingly limited arboreal pollen counts, means that this dataset contains more extensive counts of herbaceous taxa.

Pollen identification for both sites utilized the key in Moore et al. (1991), with supplementary information on pollen types provided by various pollen taxonomic keys and descriptions (e.g., Punt et al., 1974–2014; and references therein) in addition to type slide collections at the University of Stirling (Glen Affric Samples) and a personal type slide collection (kindly donated by Dr. Margaret Atherden). Identification of pollen types from Ribblesdale was also supported by visits to use collections at Oxford Archaeology North and at the University of Hull as well as information and plates in Beug (2004).

3.5. Pollen and plant nomenclature and taxonomic grouping

Plant nomenclature follows Stace (2010). Stace (2010) conforms to APG III (2009) Stace is the reference most used by UK ecologists in conservation management and therefore provides a working model for communication between paleoecological research and the end user. Pollen nomenclature follows Bennett (1995–2007); but with some updates and exceptions:

1. If plant species names have been updated (according to Stace, 2010) the new name is used such that *Lychnis flos-cuculi* becomes *Silene flos-cuculi*.

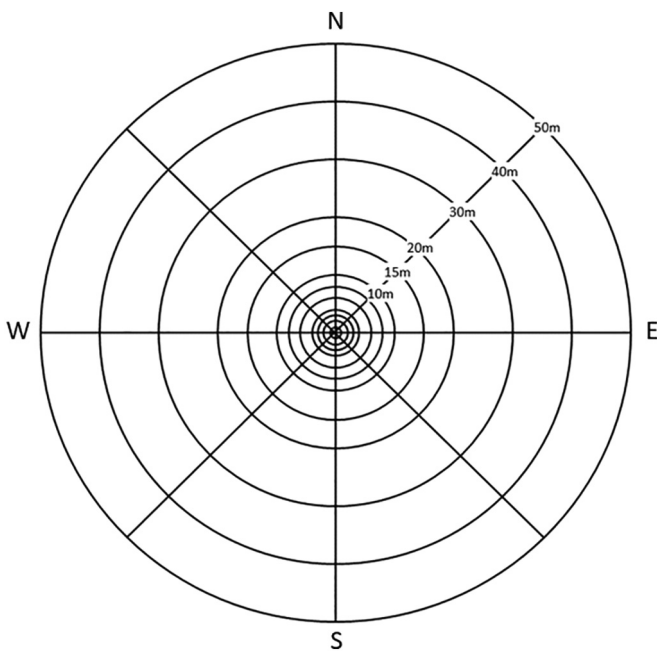


Fig. 2. Schematic of the vegetation survey sampling strategy showing the layout of quadrants used for the survey. Percentage cover of key plant types and presence, derived via random walkover of each quadrant, was recorded.

2. In several cases Moore et al. (1991) ascribe a genera-type and Bennett reduces this to a species name if there is only one species native to the UK. However, since these are surface samples, introduced non-native species are possible in some cases; therefore, genera-type is used here.
3. In many species splitting pollen taxa to species level requires good pollen preservation and a high level of precision in identification; at the edge of the capabilities of the light microscope (Moore et al., 1991). Where identification is uncertain I have split as far as possible but used Family/Sub-Family/Genera undiff. for those grains that cannot, confidently, be identified further. Undiff. is only used where further splitting of taxa should be possible with perfect preservation.
4. The aggregated term agg. is used to sum the pollen from the split taxonomic groups and the undiff. group in relevant cases. An example is for the fenestrate grains in the Cichorioideae (Lactuceae, of Moore et al., 1991) pollen, which can be identified to four pollen groups; however, the identification of fenestrate pattern in these grains can be obscured by folding, grain damage etc., in these cases grains are assigned to the Cichorioideae undiff. An analysis is carried out on each of the four pollen types where split identification was possible, then the pollen and associated plants are also summed to Cichorioideae agg., comprising those grains assigned to the four individual pollen taxon and those assigned to Cichorioideae undiff. Analysis is then repeated on this larger, less-refined, group.

Appendix Table A.1 contains a list of all of the plant and pollen taxa found, grouped into pollen taxa.

3.6. Data analysis

3.6.1. Indices of Association

IOA scores were derived for each species in the dataset using the formulae from Davis (1984). This gives three separate scores for association, over and under-representation.

Association score: $A = B_0 / (P_0 + P_1 + B_0)$

Over-representation score: $O = P_0 / (P_0 + B_0)$

Under-representation score: $U = P_1 / (P_1 + B_0)$

Where:

P_0 = number of sites with pollen taxon present and plant species absent;

P_1 = number of sites with plant species present where pollen taxon is absent;

B_0 = number of sites with plant present and pollen present.

Pollen counts and plant survey datasets were converted to presence/absence data. Plant survey for each ring survey area is cumulative so that presence in ring source 0–1 m returns a value of 1; and, even if absent in the subsequent ring sources, the value remains 1.

For s_{EIOA} , some general rules can be developed based upon the mathematical properties of the components of IOA calculations.

- 1) If the A score is low at proximal plant source areas, either there is not a strong pollen–vegetation relationship, or the relationship will only manifest when pollen from greater source distances is included (Davis, 1984). Increasing A scores with distance are possible where $P_0 > 1$; and likely when O values are high.
- 2) As plants are included from greater source distance;
 - a. P_0 reduces as B_0 increases, thus reducing O and increasing A (the sum of $P_0 + B_0$ is conserved).
 - or
 - b. P_1 increases thus increasing U .
 - c. If P_1 also increases as P_0 reduces A will remain weak.
- 3) B_0 may become saturated (pollen always occurs with plant presence) at which distance O becomes zero and after which distance any additional plants found will increase P_1 and the U value. A will then be reduced with increases in P_1 . At this distance a source area for pollen dispersal may be postulated.

- 4) IOA values may fluctuate between high and low A , O and U values at increasing distances. If this occurs dispersal may be so driven by stochastic influences that pollen dispersal/deposition is erratic.
- 5) There are insufficient data to determine any pattern, because pollen or plant presence is very rare across sites, in which case further data are required to clarify any relationship.

Theoretically, it would be possible to continue the IOA calculation at increasing source areas. High O values for some taxa demonstrate that this might be required. A final calculation is made at the end of each diagram for IOA scores using plant data from the wider Ribblesdale landscape. This distance is referred to as the Landscape scale. Provided that the plant associated with any pollen type is recorded in the Ribblesdale area P_0 , and therefore O , must eventually become zero. This additional data point allows a tentative visualization of the possible trajectory of the A score at the landscape scale of plant source area.

Pollen taxa subjected to IOA analysis have been assigned to groups based upon the degree of association, over- or under-representation (Bunting, 2003) Bunting defined these groups as: SAT – Strongly Associated Type; AT – Associated Type; WAT – Weakly Associated Type. Mazier et al. (2006, p. 99), determined these according to specific cut-off values. These are modified to take into account distance adjusted changes in IOA scores and used to compare data in this study.

3.6.2. Other indices

3.6.2.1. Negative values. The calculation of A is, essentially, the Jaccard similarity coefficient (Jaccard, 1901, 1912, 1928 in Kent and Coker, 1992). Jaccard similarity usually measures the similarity between two sites (samples) based on the number of species in each site and the number of species present at both sites. This coefficient does not, however, take into account values for absence in both datasets. The inclusion, or not, of negative co-occurrences in the coefficients of a contingency table analysis is an important consideration (Duarte et al., 1999; Kallio et al., 2011).

In ecological applications Lepš and Šmilauer (2003, p.79) note that “unlike similarity of samples, the d value is absolutely necessary for the similarity of species” where “ d value” refers to number of negative co-occurrences. IOA in palynology measures the similarity (matching) of pollen and plant from a single species (or pollen taxa) across a number of sites. The inclusion of negative co-occurrences may, therefore, be important in palynology; especially where plant and pollen prevalence is high.

The Simple Matching Coefficient, or accuracy score, used in medical trials and elsewhere (Sokal and Michener, 1958; Buckland and Elston, 1993; Dalirsefat et al., 2009), also known as the Correct Classification Rate (Fielding and Bell, 1997), considers both true positives and true negatives; i.e., the number of samples where both pollen and plant are present, and the number of samples where pollen and plant are absent; such that Correct Classification Rate (CCR):

$$CCR = Z_0 + B_0/N$$

Where Z_0 = Number of times plant and pollen are absent. N = Total number of samples.

The CCR, therefore provides a measure of overall accuracy across sites, as a proportion of incidences where pollen and plant correspond in absence or presence over the whole dataset. This, combined with analyses of U and O in the IOA, is tested to examine to what extent it adds useful information. The CCR is presented in this analysis together with the IOA analysis (Table A.2 and Fig. A.1).

Lepš and Šmilauer (2003, p. 79) demonstrate the difference between the Jaccard coefficient, which is relevant for evaluating sample similarity, and the V and Q coefficients, which compare similarity of species across sites by evaluating their presence in a given number of samples. They recommend V and Q (p. 78) coefficients, as absolute and complete

measures of association, both of these coefficients also include negative co-occurrences:

$$V(\text{MCC}) = (B_0 \times Z_0) - (P_0 \times P_1) / \sqrt{(B_0 + P_0)(P_1 + Z_0)(B_0 + P_0)(P_1 + Z_0)}$$

$$Q = (B_0 \times Z_0) - (P_0 \times P_1) / (B_0 \times Z_0) + (P_0 \times P_1).$$

These coefficients provide a stricter test of species independence. The V coefficient, or Matthews Correlation Coefficient (MCC), is also recommended by Chicco (2017) as a method to “avoid the dangerous illusion of inflated achievement” (Chicco, 2017, p. 2) in data mining and machine learning. Previous IOA analyses have not tested for significance and it is possible that pollen presences with weak association, or findings of a strong association based on limited data, may be no better at predicting vegetation presence than chance.

3.6.2.2. ROC and Precision-Recall. Two other common methods for examining the success of predictors in a contingency analysis are Receiver Operating Characteristic (ROC) and Precision-Recall (P-R). Used frequently in machine learning ROC has also been applied to pollen data to determine the best cut-off in squared chord distance between vegetation and pollen assemblages (e.g., Gavin et al., 2003; Wahl, 2004) and has also been used to develop a cut-off threshold for *Sporormiella* percentage as a grazing indicator (Gill et al., 2013).

ROC analysis uses a plot of sensitivity (True Positive Rate, or Sensitivity) against 1-specificity (False Positive Rate). The Area under the Curve (AUC) provides a valuable alternative to IOA, which can optimize the matches between pollen and plant, considering probability of detection (pollen indicating a plant presence as a proportion of times the plant is actually present) against the probability of false alarm (pollen indicating a plant presence as a proportion of times the plant is not present).

$$\text{Sensitivity} = B_0 / (B_0 + P_1) = 1 - U$$

$$1 - \text{Specificity} = P_0 / (P_0 + Z_0).$$

Examining the locations of the Sensitivity/1-Specificity scores plotted on a ROC graph provides an alternative and informative way to examine the data that minimizes type I and type II errors. Normally several points in ROC space are presented for different cut-off values in abundance data. In this study each ROC plot of the pollen-plant relationship is based on presence/absence data and there are no abundance data to create cut-of values. However, points on the ROC curve can be plotted for each plant source distance and their locations in ROC space assessed. In contrast to cut-off values in abundance data, which develop an incremental pattern in reduced error leading to a ROC curve plot, each data point at increased source distance is a unique curve. The single point with the highest AUC should correspond to most appropriate pollen source area for the pollen taxa.

One problem with ROC plots is that they can be susceptible to bias in imbalanced datasets (Davis and Goadrich, 2006). In application to spatially extended indices here, the ideal situation, of finding strong association as more plants are included at each site with distance, may shift the balance in the datasets. Precision-Recall (P-R) plots are recommended as an informative tool to highlight the susceptibility of ROC plots to imbalance (Davis and Goadrich, 2006; Saito and Rehmsmeier, 2015).

$$\text{Precision} = B_0 / (B_0 + P_0) = 1 - O$$

$$\text{Recall} = B_0 / (B_0 + P_1) = \text{Sensitivity} = 1 - U.$$

It can be seen from the Precision and Recall equations that P-R curves are a plot of 1-O and 1-U. P-R plots. These curves, unlike

ROC curves, do not utilize the important true negative data and so may be less informative of overall success; therefore both plots are explored with the Ribblesdale dataset. Furthermore, contrasts between each plot are made more difficult to read because ROC curves show positive relationships in the top left of the plot, with Sensitivity on the y axis and 1-Specificity on the x axis, whilst P-R plots show Recall (equivalent to sensitivity) on the x axis and Precision on the y axis. To render them visually comparable in this study an inverse of ROC is plotted, with Specificity (inverse of 1-Specificity) on the y axis and Sensitivity on the x axis. The curve and the AUC remains constant but the positive relationship plots in the top right-hand corner as for P-R plots.

Lastly, the IOA and CCR coefficients do not provide a test of significance. Significance testing for each taxon at each source area is provided by Fisher's exact test. Chi-square measures of association are not suitable due to the small dataset and sporadic nature of pollen and plant abundance; Fisher's test is more suitable in these circumstances.

3.6.3. Data handling

Pollen counts and plant surveys were converted to cumulative presence /absence data for each ring survey area so that presence in ring source 0–1 m returns a value of 1; and, even if absent in the subsequent ring sources, the value remains 1. IOA and CCR, as well as other values, were calculated using Excel for each at each vegetation survey distance according to the equations in Sections 3.6 and 3.7. Fisher's exact tests were calculated using an online calculator by Øyvind Langsrud (<https://www.langsrud.com/fisher.htm>). The right-tailed test statistic was applied, as, given that all pollen must come from a plant somewhere and the alternative to independence is therefore a positive association, a two-tailed test is unsuitable.

3.7. Correlation with plant floral traits and biodiversity

Several studies have correlated IOA score with pollination mechanisms. Hjelle (1997) matches IOA scores with flower morphology, an approach also taken by Brun et al. (2007). Hjelle (1997) further splits the pollination strategies into those flowers with few, or numerous exposed, anthers; and, those with concealed anthers. The separation in dispersal strategies, of wind and insect pollinated types is well-known and evident from previous studies.

There are, however, some other insect pollination strategies that may benefit from further exploration. Willmer (2011), for example, discusses that flowers offering no nectar reward have reduced insect visitations. If insect visitation rates influence pollen distribution we might expect “no nectar” insect pollinated flowers to disperse less pollen into the environment; unless this is counteracted by generalist strategies and open flower morphologies. In contrast, a generalist pollination strategy with lots of insect visitation, such as might be expected with nectar flowers with an open morphology, might spread more pollen over a wider area.

In addition, whilst specialist pollination strategies may be more conservative and waste less pollen grains to the environment, buzz-pollination seems to eject a cloud of pollen from plants, only some of which adheres to the bee and is transported away; this is illustrated well in several online video clips; including Smithsonian Channel (n.d.). Proctor et al. (1996) also describe a study of *Pedicularis* (Kindle book location 3337) in which they state that pollen is swept from the hairs of the bee on its flight between flowers. It seems possible that buzz pollinated flowers may deposit pollen more consistently in the ultra-local environment and, consequently, be more likely to have a high local association within a few meters.

The pollination strategies for each taxa were determined by a review of the literature. Pollination strategies together with a separate list of references consulted, are provided in Table A.1; these are

discussed in relation to the various results across plant families in Section 4.4.

4. Results

4.1. Data summary

The vegetation surveys in Ribblesdale yielded 147 non-arboreal higher plant taxa, these include, forbs (137), Poaceae (undiff.), Cyperaceae (undiff.) Juncaee (undiff.) and Ericaceous dwarf-shrub heath types (7). These plant taxa can be grouped into 82 pollen/spore types at the finest taxonomic resolution possible in palynological identification (see appendix Table A.1). The Ericaceae plant types found can be distinguished into five possible pollen types, plus two aggregated groups, representing all plant types found; except for *Arctostaphylos*

uva-ursi, which appeared in the pollen record but had no corresponding plant record. The 137 forb species can be distinguished as 75 possible pollen types at finest taxonomic resolution. There were 32 herbaceous plant taxa (with 21 equivalent pollen types) found in the Ribblesdale vegetation survey data, but with no corresponding pollen type in the pollen samples (excluding four orchid species, for which pollen is unlikely). Seventy-one herbaceous pollen types were found; 23 of which had no corresponding vegetation presence.

Of the 23 herbaceous pollen taxa with no corresponding plant present in any of the sites surveyed, plant taxa have been recorded in other surveys (Corkhill, 2003) in Ribblesdale for all but 8 pollen types. Plant types listed in other surveys in Ribblesdale are also listed in Table A.1. Pollen types with no plant recorded were *Chelidonium majus* (1,1), *Fumaria* (3,1), *Saxifraga stellaris*-type (2,1), *Illecebrum verticillatum* (2,1), Chenopodiaceae (2,2), *Anagallis tenella*-type (2,2), *Hottonia*

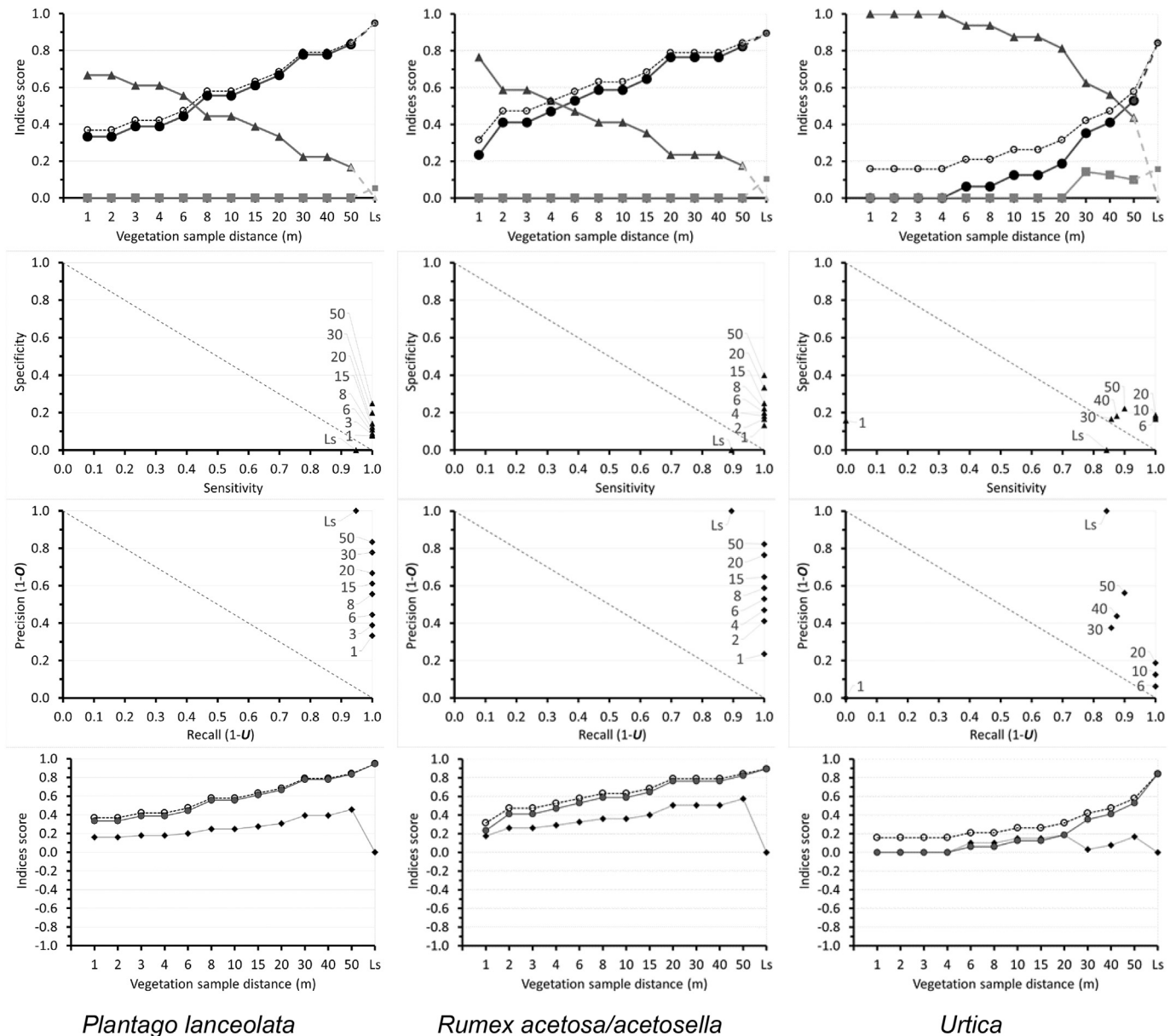


Fig. 3. Extended IOA plots for selected taxa showing the range of IOA over distance curves. Data are displayed from Ribblesdale only, additional data displays for all taxa with plant and pollen present are found in the Appendix. Significant associations (Fisher's Exact Test: $p \leq .05$) are denoted, for each distance where this occurs, by * at the top of the diagram. Row 1: IOA; Row 2: Inverse ROC; Row 3: Precision-Recall; Row 4: MCC (closed circles) compared to Association score and CRR. (a) Wind pollinated taxa where association continues to increase at 50 m or beyond. (b) Selected wind and insect pollinated taxa where association increases to a maximum within 50m (c) Selected insect pollinated taxa with a weak association that decreases at wider source distances (d) Selected taxa with a stochastic association, sometimes driven by limited data.

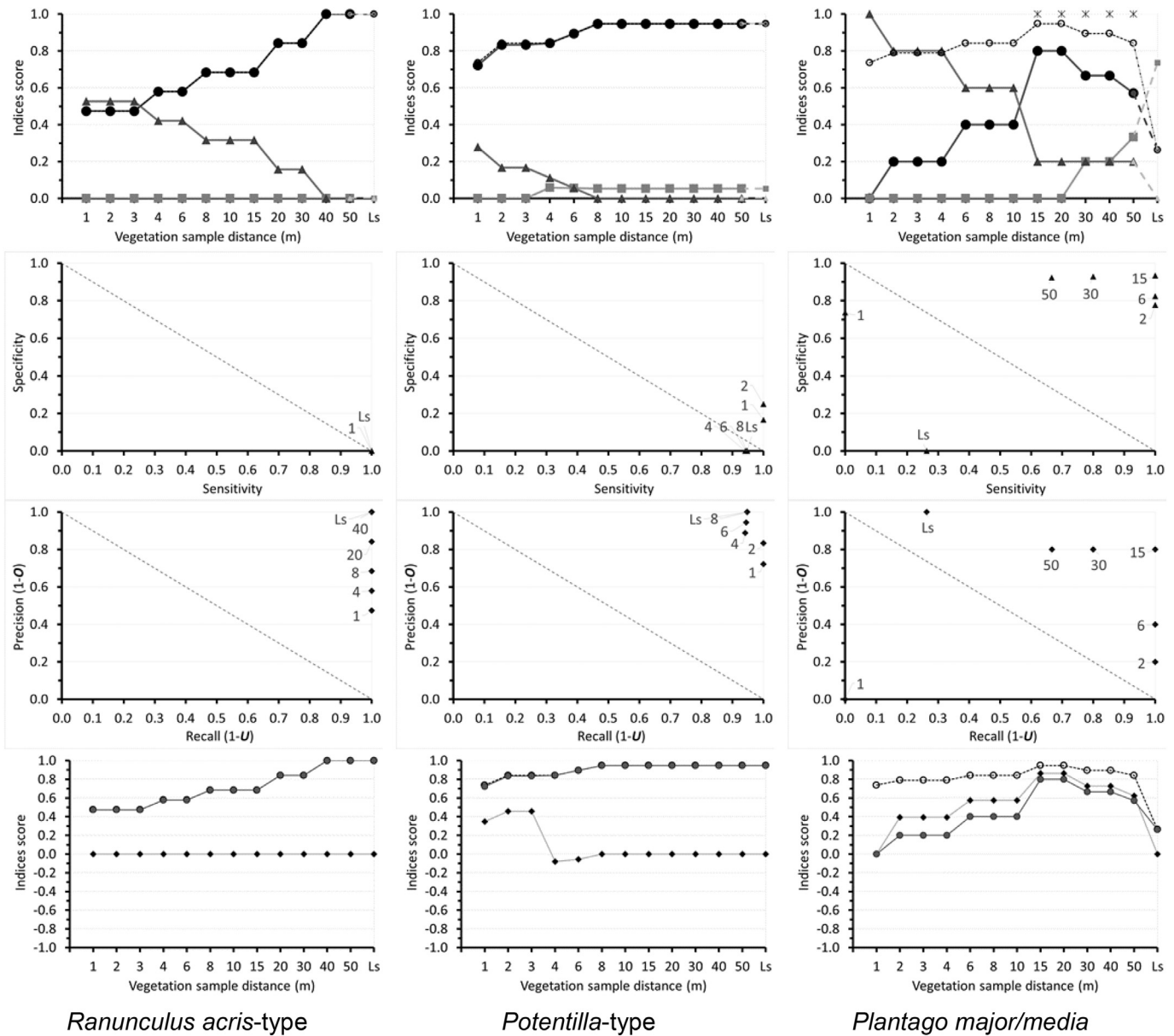


Fig. 3 (continued).

palustris (1,1), and *Valerianella* (1,1) (numbers in brackets refer to the number of grains and number of sites in which the pollen was found, respectively). Further consultation of the BSBI distribution maps (BSBI.org/maps) shows local records for all of these: apart from *Valerianella*, for which some older records occur within 10 km and also current survey records within 50 km to the west; and *Hottonia palustris*, which only has records within 50 km to the east and west. The *Hottonia palustris* pollen grain was not securely identified and so this single grain could be from a similar pollen type. The other major outlier in the data is *Arctostaphylos uva-ursi* (2,1), nearest records found (BSBI.org/Maps) are c. 50 km to the north, west and south. A more local source cannot be ruled out for this *A. uva-ursi*; but, given that the area has been well-surveyed, if present, this species is likely to be very rare.

Forty-eight herbaceous pollen taxa occurred as both pollen and plant data. An additional 12 aggregated pollen taxon groups were

created, where identification between particular pollen types is problematic.

A further 12 arboreal (tree and shrub) taxa are also recorded within the surveyed areas, corresponding to 10 possible pollen types. All 10 possible pollen types were found in the samples, although not all corresponded to those samples where plant was present. A further nine arboreal pollen types were found with no plant in the 50 m survey areas; although all occur within the wider landscape.

Fifteen Pteridophyte plant species were also present corresponding to eight possible spore types; only four of these were found in the pollen samples. However, several spores were found in each sample recorded as Pteropsida (monolete) indet. As an aggregated group for all Pteropsida types with spores minus the outer coat. An additional four Pteridophyte spore types were found in the pollen samples with no plant in the survey areas, although they are all found in the wider landscape.

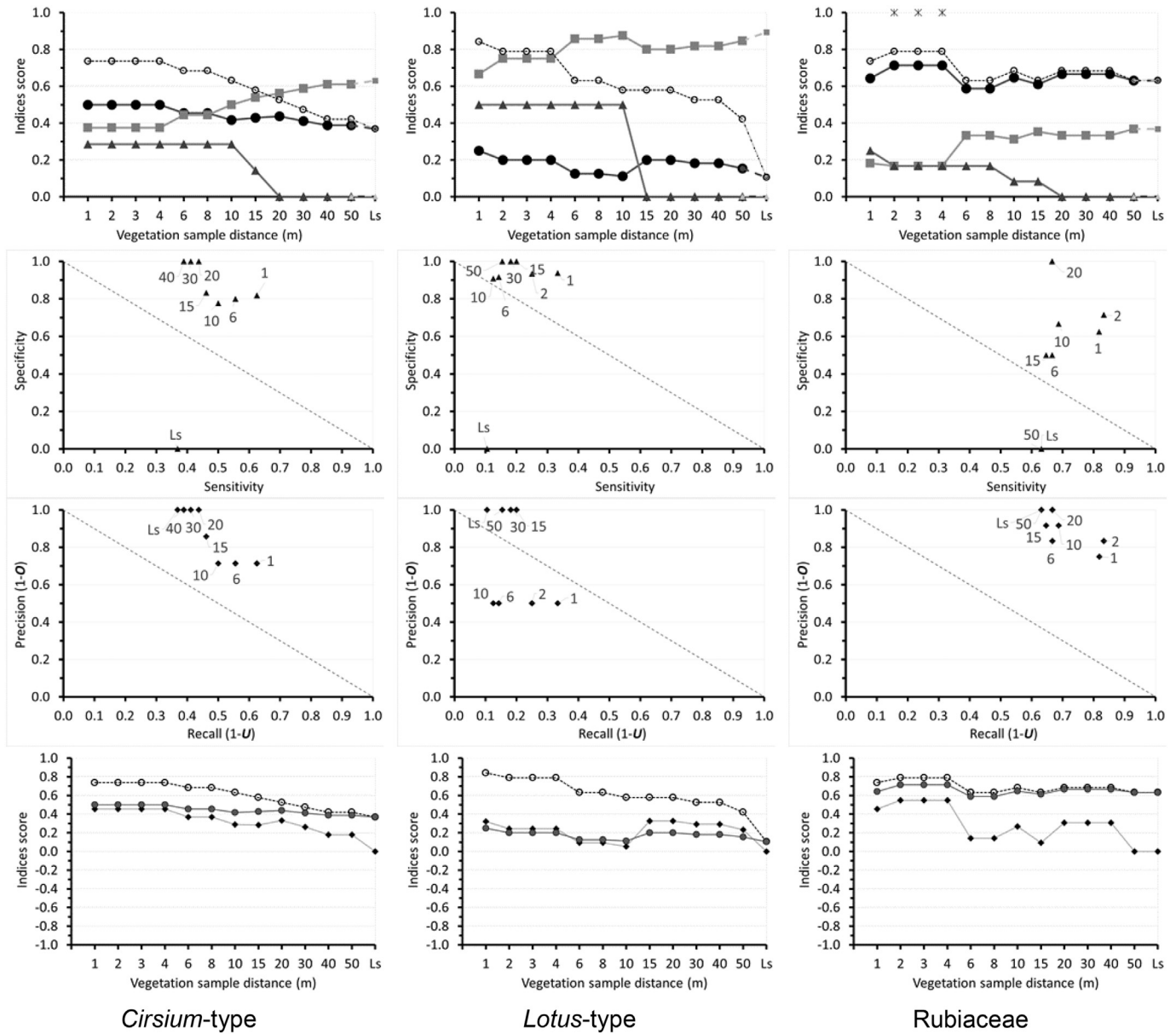


Fig. 3 (continued).

The Glen Affric dataset (Shaw, 2006) comprises 49 non-arboreal pollen/spore taxa present with 34 non-arboreal plant taxa recorded in the surveys and 12 taxa found in both pollen and plant surveys. A further 8 arboreal (trees and shrubs) taxa are also recorded as plant presence, with 13 arboreal taxa recorded in the pollen record.

4.2. Summary of IOA patterns with distance and individual pollen source areas

The relationship of $SEIOA$ scores with plant survey distance, for selected taxa, is displayed graphically in Fig. 3. A full set of $SEIOA$ graphs, together with Inverse ROC plots, P-R plots and MCC plots, for all taxa with pollen and plant presence (at each separate site; and, as the combined Ribblesdale and Glen Affric data), is provided in the Appendix Figs. A.1–3. Several taxa were present as single unassociated grains ($B_0 = 0$), or the number of sites with plant presences did not increase across the range of ring survey areas rendering spatially extended IOA

analysis impossible. Data for these types are included in appendix Table A.2, which provides the raw data for B_0 , P_0 , P_1 and Z_0 , together with the calculated IOA indices, CCR, Fisher’s exact test, and MCC, at each ring survey distance, for each pollen taxa where both plant and pollen were present.

The iPSAs can be assessed from examination of the IOA curve. Poaceae and Cyperaceae are SAT at all source distances (Fig. A.2). For types where the plant is ubiquitous in the environment the ability to sense distance of association with IOA is limited by prevalence. This becomes more evident in calculation of the MCC and False Positive Rate for use in ROC curves. For Poaceae since, from the 1 m radius, pollen and plant are always present the test lacks specificity and the False Positive Rate is unclassified, therefore a ROC curve cannot be plotted, the MCC is similarly unclassified (Table A.2). These taxa are more appropriate for ERV models using abundance data and are not really rare-type pollen. The need for quantified analysis of abundance is illustrated by the Cyperaceae; one site, in a hay meadow, did not have plant present

Table 1

Classification of pollen taxa IOA curves for Ribblesdale dataset (see Fig. 3 and Fig. A.1 for graphic plots of IOA values). Association cut-off values are used from Mazier et al. (2006). Abbreviations used are SAT – Strongly Associated type ($A > 65\%$); AT – Associated type ($50\% < A < 65\%$); WAT – Weakly Associated type ($A < 50\%$) UR – Under-represented; (low A values, high U values); OR – Over-represented type (low A values, high O values); UT – Unassociated Type: low A values and relatively high values for both U and O.

Strongly associated types (SAT): $A > 0.65$ A > 0.65 from 1 m source distance and maintains a strong association at all distances	Poaceae Cyperaceae	A = 0, high O and U values	<i>Caltha palustris</i> -type <i>Anemone nemorosa</i> <i>Thalictrum</i> <i>Vicia</i> -type <i>Mercurialis perennis</i> <i>Viola palustris</i> -type <i>Geranium</i> <i>Helianthemum nummularium</i> <i>Primula veris</i> -type <i>Veronica</i> <i>Plantago major</i> <i>Scabiosa</i> <i>Convallaria majalis</i> -type
An increase in A with distance to a SAT , U = 0 and O decreases as A increases	<i>Ranunculus acris</i> -type <i>Potentilla</i> -type <i>Rumex acetosa</i> <i>Rumex acetosa/acetosella</i> <i>Rumex</i> sp. agg. <i>Plantago lanceolata</i> <i>Plantago</i> sp. agg. <i>Plantago major/media</i>	Types occurring in the pollen set with no associated plant present in the vegetation surveys	Types occurring in the plant surveys with no associated pollen present (grouped by pollen type)
An increase in A with distance to a SAT followed by a decline in A , U = 0 and O decreases as A increases	<i>Cichorium intybus</i> -type Cichorioideae agg. Brassicaceae agg. <i>Erica tetralix</i> <i>Trollius europaeus</i> <i>Heracleum sphondylium</i>	<i>Arctostaphylos uva-ursi</i> * <i>Anagallis tenella</i> -type	<i>Allium</i> -type (<i>Allium ursinum</i>) <i>Arctium</i> -type (<i>Carlina vulgaris</i> and <i>Serratula tinctoria</i>) <i>Arum</i> (<i>Arum maculatum</i>) <i>Digitalis purpurea</i> <i>Drosera rotundifolia</i> -type <i>Epilobium</i> <i>Rhinanthus</i> -type (<i>R. minor</i> and <i>Euphrasia</i> sp.) <i>Geum</i> (<i>G. urbanum</i> & <i>G. rivale</i>) <i>Hypericum perforatum</i> -type (<i>H. perforatum</i>) <i>Linum</i> (<i>Linum catharticum</i>) <i>Mentha</i> -type (<i>M. aquatica</i> and <i>Thymus</i>) <i>Myosotis</i> (<i>Myosotis discolor</i>) <i>Pedicularis</i> (<i>P. sylvatica</i>) <i>Pinguicula</i> (<i>P. vulgaris</i>) <i>Polygala</i> (<i>P. serpyllifolia</i> and <i>P. vulgaris</i>) <i>Primula farinosa</i> <i>Ribes</i> <i>Sanguisorba officinalis</i> <i>Scutellaria</i> -type (<i>Ajuga reptans</i>) <i>Valeriana dioica</i> <i>Viola arvensis</i> -type (<i>Viola lutea</i>) (Total 21)
An increase in A with distance to a SAT Low but increasing U values O decreases as A increases Reaches SAT value at one source distance but data limitations prevent distance based analysis	<i>Ericaceae</i> undiff. (minus <i>Calluna</i>) <i>Urtica</i> Brassicaceae <i>Sinapis</i> -type Rubiaceae <i>Solidago virgaurea</i> -type Caryophyllaceae agg. Brassicaceae- <i>Hornungia</i> -type <i>Oxalis</i>	<i>Artemisia</i> <i>Callitriche</i> cf. <i>Cannabis</i> -type* Caryophyllaceae cf. <i>Arenaria</i> -type cf. <i>Chelidonium majus</i>	
Associated types (AT): $A 0.50 > 0.65$ A gradual increase in A with distance and decreasing O and U : both to <0.5	<i>Filipendula</i> <i>Poterium sanguisorba</i> <i>Rumex acetosella</i> <i>Rumex obtusifolius</i> -type Apiaceae <i>Potamogeton natans</i> -type <i>Blechnum spicant</i> <i>Dryopteris dilatata</i> <i>Dryopteris</i> sp. agg. <i>Calluna vulgaris</i> <i>Empetrum nigrum</i> <i>Silene flos-cuculi</i> <i>Mycelis muralis</i> -type <i>Stachys</i> -type <i>Cerastium</i> -type <i>Narthecium ossifragum</i> <i>Achillea</i> -type <i>Succisa pratensis</i> <i>Equisetum</i>	Amaranthaceae/Chenopodiaceae <i>Chrysosplenium</i>	
Weakly associated types (WAT): $0.0 < A < 0.5$ WAT A slight/gradual increase in A with distance	<i>Vaccinium</i> -type <i>Erica/Vaccinium</i> <i>Veronica</i> <i>Vicia/Trifolium</i> -type <i>Prunella</i> -type <i>Campanula</i> -type <i>Lotus</i> -type* <i>Trifolium</i> -type <i>Cirsium</i> -type	<i>Euphorbia</i> <i>Fumaria</i> * <i>Hedera helix</i> cf. <i>Hottonia palustris</i> <i>Illecebrum</i> <i>Nuphar</i> <i>Oxyria</i> -type cf. <i>Papaver radiculatum</i> cf. <i>Rosa</i> <i>Saxifraga granulata</i> -type <i>Saxifraga hirsuta</i> -type <i>Teucrium</i> <i>Valerianella</i> * <i>Vicia cracca</i> -type (Total 23)	
WAT(OR) A gradual increase in A with distance with high O			
WAT(OR) (Limited data) A slight/gradual increase in A with distance, high O moderate U			
WAT(UR) A slight/gradual increase in A with distance, high U moderate O			
WAT < 50 m) A weak peak between 1 and 50 m			
WAT (UT) A gradual decrease in A with distance with high O and U			
Under-represented types (UR): $O = 0.0$ and $U > A$ A slight/gradual increase in A with distance with high U			
A weak peak in A between 1 and 50 m with high U			
A gradual decrease in A with distance with high U			
Stochastic types A , O and/or U values fluctuate over distance	<i>Centaurea</i>		
Unassociated types UT: A = 0, high O	<i>Pteridium aquilinum</i> <i>Dryopteris filix-mas</i> -type		

within the 50 m survey, but did still have pollen in the corresponding moss polster sample leading to an **O** indices >0 and inferring a pollen source area beyond the 50 m survey area. However, this site did contain a lower number of Cyperaceae pollen grains; a factor which would be apparent in analysis of abundance data.

Other strongly associated types showing an increasing **A** value with survey distance are *Plantago lanceolata*, *Rumex acetosa/acetosella* and *Urtica* (Fig. 3); all are wind pollinated types or floral types with numerous exposed anthers (Table A.1). These types show steady increase in **A** to 50 m source area in Ribblesdale. Accuracy (CCR) scores also increase with distance and follow the **A** curve. Over-representation decreases as more plants are included; this indicates that source areas used for previous IOA studies are not sufficient to understand association of these types. Association values continue to increase at the landscape scale perhaps indicating a wider pollen source area. Although, at the landscape scale, and from 20 m for *Urtica*, under-representation is also introduced, demonstrating that, from these distances, there is a more stochastic relationship between pollen and plant. The ROC and P-R plots also demonstrate the improvement in both specificity and the positive predictive value of pollen as more plants are included with source distance and **O** reduces; however, the MCC is more conservative (Table A.2 and Fig. A.2) due to the high prevalence of positive co-occurrences and lack of negative co-occurrences.

Some insect pollinated types, especially those with open flowers and exposed anthers, such as *Ranunculus acris*-type and *Potentilla*-type, also

show a steady increase in **A** with increased plant source distances; however, association does not improve at the landscape scale (Fig. 3). These taxon are likely to have pollen sources distances of c. 50 m. In contrast, the association score for the *Plantago major/media* group peaks at 15–20 m and **U** increases from this distance, thus reducing **A** and indicating a defined iPSA of 15–20 m for this pollen type; even though it is generally thought to be pollinated by wind (Table A.1). The positive association at 15 m is also confirmed by Fisher's exact test ($p = .006$) and a high MCC of 0.86 as well as the location of the data point in the top right of Inverse ROC and P-R plots. However, the association is also strong and Fisher's test significant between 30 and 50 m source area, despite the inclusion of more plants at sites without pollen between these distances, which reduces both recall and sensitivity (Fig. 3b).

Other types show weaker, or gradually weaker association. For example, *Galium/Rubiaceae* pollen is strongly and significantly associated ($p = .029$; Table A.2) with plant presence occurring between 2 and 4 m from the pollen sample (Fig. 3c). Although association is still above 0.5 from 4 to 50 m the **U** score, the chance of missing the plant and inferring plant absence from pollen absence, increases slightly and the MCC also reduces. A similar pattern is displayed by *Cirsium*-type, another insect pollinated taxa with few exposed anthers, and the bee pollinated *Lotus*-type. For these types under-representation is higher than over-representation; and weak, but significant, associations between pollen and plant presence expand for no more than a few meters. Although association values are weaker, inclusion of negative co-occurrences through the addition of the accuracy score is useful. The accuracy score is high for these types at proximal source distances, gradually reducing at distance and supporting the assertion of a local source

area for pollen. However, for *Lotus*-type despite a higher association and MCC the Fisher's exact test shows no significance, a caution of interpretation also supported by the location of proximal source area points below the independence line in P-R plots.

In contrast, other types display fluctuation between pollen presence and plant presence, with high under-representation, or a peak in association at one source area, due often to limited data. *Centaurea nigra* (Fig. 3d) is associated between 4 and 15 m and the high accuracy score shows that it is rare in the landscape and pollen is more likely to be present with local plant presence. The MCC is high at 0.69 at the 4 m source distance plots in the zone of strong classification in ROC space (Fig. A.2). However, when examining the P-R plot it becomes obvious that, although the relationship between pollen and plant presence for *Centaurea* is highly specific (no false negatives), it is not very precise (false positives are likely). Indeed the area under the P-R curve is similar for 4 and 20 m source distance and, in fact, the 20 m source distance is significantly associated ($p = .035$), in comparison with $p = .105$ at the 4 m source distance.

4.3. Classification of extended Indices of Association

The summary results for each taxa, classified by cut-off values (Section 3.5.1 above and Mazier et al., 2006), are displayed in Table 1. The major difference from previous studies is that the classifications for each taxa shift across distance with inclusion of more plant presence in the more distant quadrants; for example from AT to SAT. Therefore Table 1 categorizes taxa by direction of change as well as initial association grouping. The data show a clear improvement in **A** values with distance for several taxa.

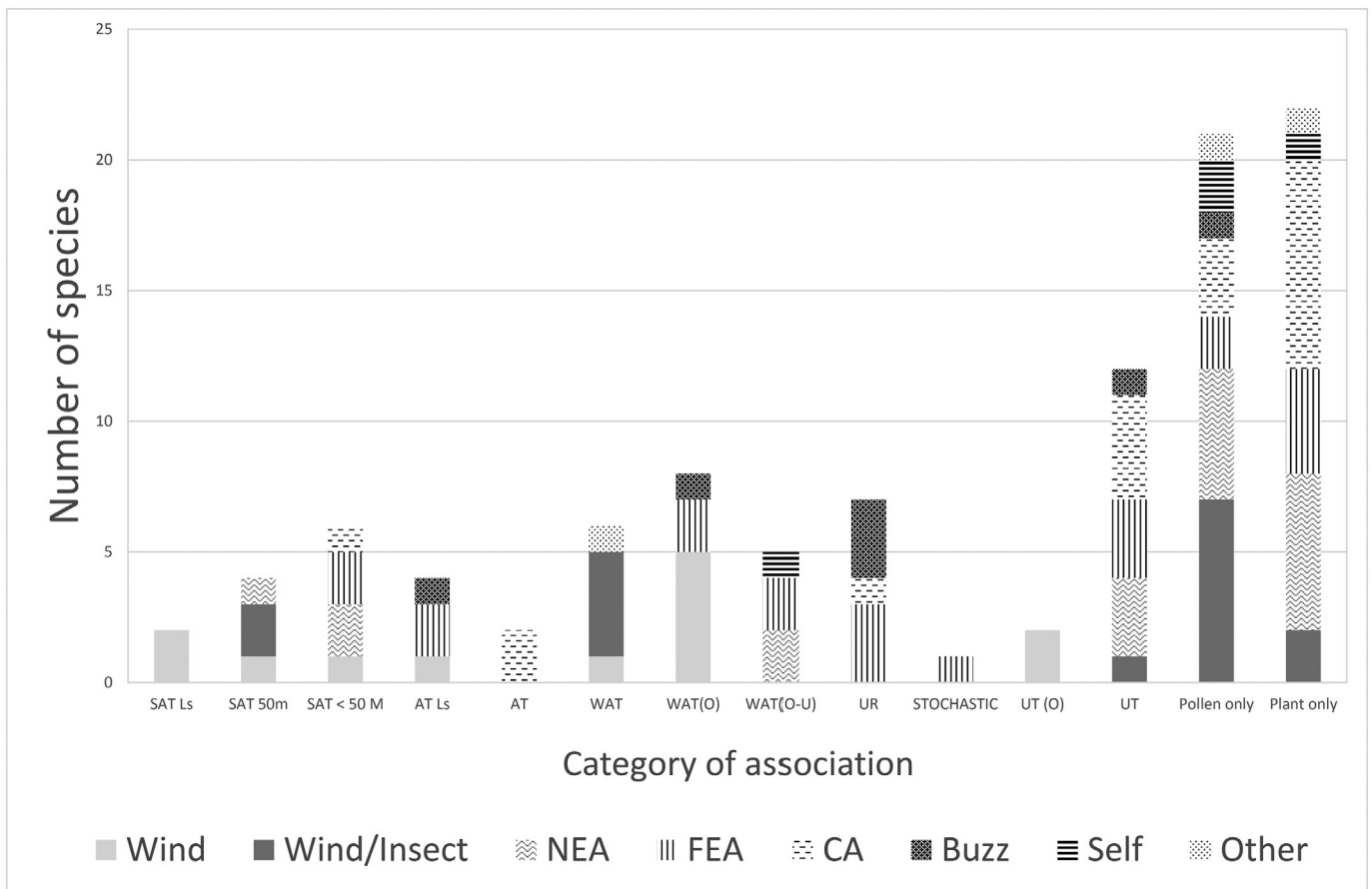


Fig. 4. Number of species (pollen taxa) in each association category by pollination mechanism. For categories of association see Table 1, and for pollination strategies refer to Table A.1.

4.3.1. Summary of IOA categories and pollination

The number of species in each pollination category is displayed in Fig. 4. There is a general trend for over-represented types, at all levels of association, to contain more wind-pollinated species, whilst buzz and self-pollinated species tend to be in the under-represented categories. However, a few species with concealed anthers, buzz pollination, and self-pollination, mechanisms are found as pollen-only, indicating some stochastic results. Surprisingly, among the herbaceous types no obligate wind-pollinated species are found in pollen samples only, however this could be due to ubiquity of plants of these types in the landscape.

4.4. Associations and pollen-plant relationships for all taxa

The pollen-plant relationships within each plant family are discussed below, highlighting some of the main associations and over- and under-represented types. Classification as SAT (Strongly associated type), AT (Associated type), WAT (Weakly associated type), UT (Under-represented type), ORT (Over-represented type) are used with the cut-off values of Mazier et al. (2006) and abbreviations are used throughout. In addition **A** – Association indices; **U** – Under-represented indices; **O** – Overrepresentation indices; CCR – Accuracy or correct classification rate. All data tables and figures are in the appendix. Refer to Fig. A.1 for groupings of taxa and pollination mechanisms; Table A.2 for detailed data and IOA scores, and Fig. A.1–3 for graphs of the IOA scores, Inverse ROC and P-R plots at the various sample distances. Plant families are ordered according to evolution, as ordered in Stace (2010).

4.5. Results and discussion for plant families

4.5.1. Pteridophyta

The test of association between plants of the Pteridophyta and spores is very limited in these datasets. Spores identifiable to family or species were few; and fern plants were rare in the surveys. Bradshaw (1981) found similar low counts of spores in his samples from woods in south-east England. In Ribblesdale *Dryopteris* and *Blechnum spicant* are WAT, with high **O** indicating widespread pollen dispersal. The Glen Affric dataset, in contrast, has no occurrence of plant of these taxa in the surveys for this study, and only sporadic wider plant occurrence in the Glen (BSBI, 2015: <https://bsbi.org/maps>). However, all pollen samples from both sites contained a number of undifferentiated monolet spores minus the outer coat, for which meaningful taxonomic distinction is impossible. Pteridophyte taxa which occur as plant species in sites with no identifiable spore present in the pollen sample are likely to be represented by these indeterminate monolet spores and the presence of these also indicates widespread spore dispersal.

The weak **A** for *Equisetum* in both Ribblesdale and Glen Affric is contrasted with a strong CCR driven by a high number of negative co-occurrences. In Ribblesdale the probability of a correctly identified association becomes reduced as further plants are included from beyond 10 m. However, the MCC of 0.05 cautions against reliance on accuracy in the small imbalanced dataset; whilst the data in Glen Affric are also too limited to assess over distance.

A very weak **A** for *Pteridium* increases slightly with distance in Glen Affric. However, the CCR shows greatest accuracy within a 3 m source area where over-representation is high; whereas, in Ribblesdale, the

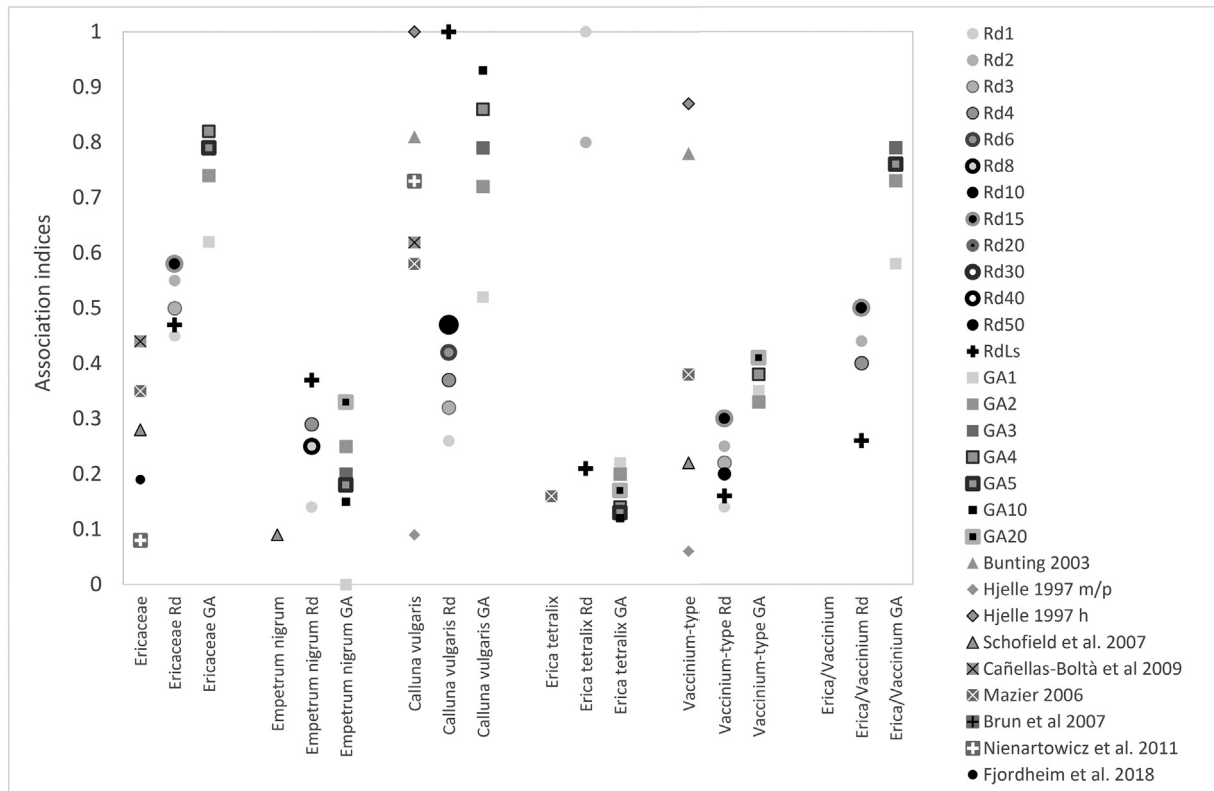


Fig. 5. A comparison of heath taxa association values with distance across Rd and GA sites and those from other studies in Europe, referenced in the text. Number after Rd (Ribblesdale) or GA (Glen Affric) is survey distance; for Hjellev, 1997 m/p = meadow/pasture, and h = heath (see original paper for details). Note for some papers only selected taxa values are available as other taxa are presented in ordination diagrams.

CCR at 0.58 (Fig. 3) is driven solely by negative co-occurrences, as there is actually no pollen found. The MCC indices, of zero, reflects the data accuracy more robustly here. In contrast, in Glen Affric, a local relationship is indicated by an MCC of 0.4 within a 2 m radius. A quantitative, abundance, approach may be required to test the representation of *Pteridium* spores, an important indicator of clearance of woodlands in past landscapes. However, this study supports the findings of Tinsley (2001) and Mazier et al. (2006) that *Pteridium* spore presence is difficult to interpret securely in pollen analyses.

From this evidence it seems unlikely that a local signal can be inferred securely from limited Pteridophyte spore counts in a paleoecological sample.

4.5.2. Ericaceae and Myricaceae

Several modes of pollination are observed in the Ericaceae, from anemophily through ambophily to obligate entomophily (see Table A.1). These different dispersal properties are illustrated in Ribblesdale by high **O** values in *Calluna vulgaris* and *Myrica gale* (wind-pollinated), and high **U** values in *Erica tetralix* (insect and self-pollinated) and *Vaccinium* type. Interestingly, Willmer (2011) notes that *Vaccinium* species may be buzz pollinated; which may allow for some local scale dispersal, since association (although, with high under-representation) does increase to 15 m. Across Ribblesdale, Glen Affric, and a range of other studies, association values in the Ericaceae tend to be varied and are higher in heathland sites (Fig. 5).

Pollen of wind-pollinated *Calluna vulgaris* was found in every sample in Ribblesdale despite only sporadic occurrence of the plant around each site. This high pollen prevalence indicates a landscape-scale source for the pollen. In Glen Affric, **A** increases to 0.93 due to widespread plant presence within the survey areas and **O** decreases correspondingly. Interpretation of Glen Affric data alone may lead to an assumption that *Calluna* is a SAT, in reasonable agreement with some other authors

who found *Calluna* to be a SAT (Bunting, 2003) or AT (e.g., Mazier et al., 2006; Nienartowicz et al., 2011) as shown in Fig. 5. However, the *p*-value from the Fisher's exact test, and the combined dataset, urge a more cautious interpretation. Whilst heathland sites have, as expected, a strong association, this is not matched in pastoral and agricultural sites where the association is weak. Weak **A** is driven by high over-representation making local-scale interpretation insecure.

Calluna is likely to be a SAT at the landscape scale as the Ribblesdale and Glen Affric combined datasets highlight considerable over-representation at the local scale. A larger dataset with sites at a greater range of distances from *Calluna* plant presence may help to clarify pollen dispersal. Correlations with plant-pollen abundance are generally used for this key pollen taxa, and may be more appropriate, although reported PPE values do vary (Broström et al., 2008). In addition, data from Glen Affric demonstrate an erratic relationship between pollen percentage and plant cover in *Calluna* over a 50 m source area (see Shaw, 2006). Since presence of *Calluna* pollen may indicate vegetation in the wider landscape the application of PPEs may require caution in local-scale interpretations, especially when trying to pinpoint the location of plant in the landscape across multiple local sites.

According to Knuth (1908) *Calluna* has a complex pollination strategy which comprises an early season insect-dominated mode and a late season wind dominated mode. Fluctuations in representation between sites and studies could potentially emerge, therefore, from variation in the seasonal timing of sample collection. Whilst the moss polster sample should contain multiple years of pollen deposition, it is possible that samples collected in the autumn might contain a higher seasonal pollen load from recent wind dispersal rendering comparisons across studies more complex.

Erica tetralix is a SAT in the Ribblesdale samples with significant association ($p < .001$ at 1 m, and MCC = 1.0). Limited data, with no additional plant presences occurring with increasing source area, preclude

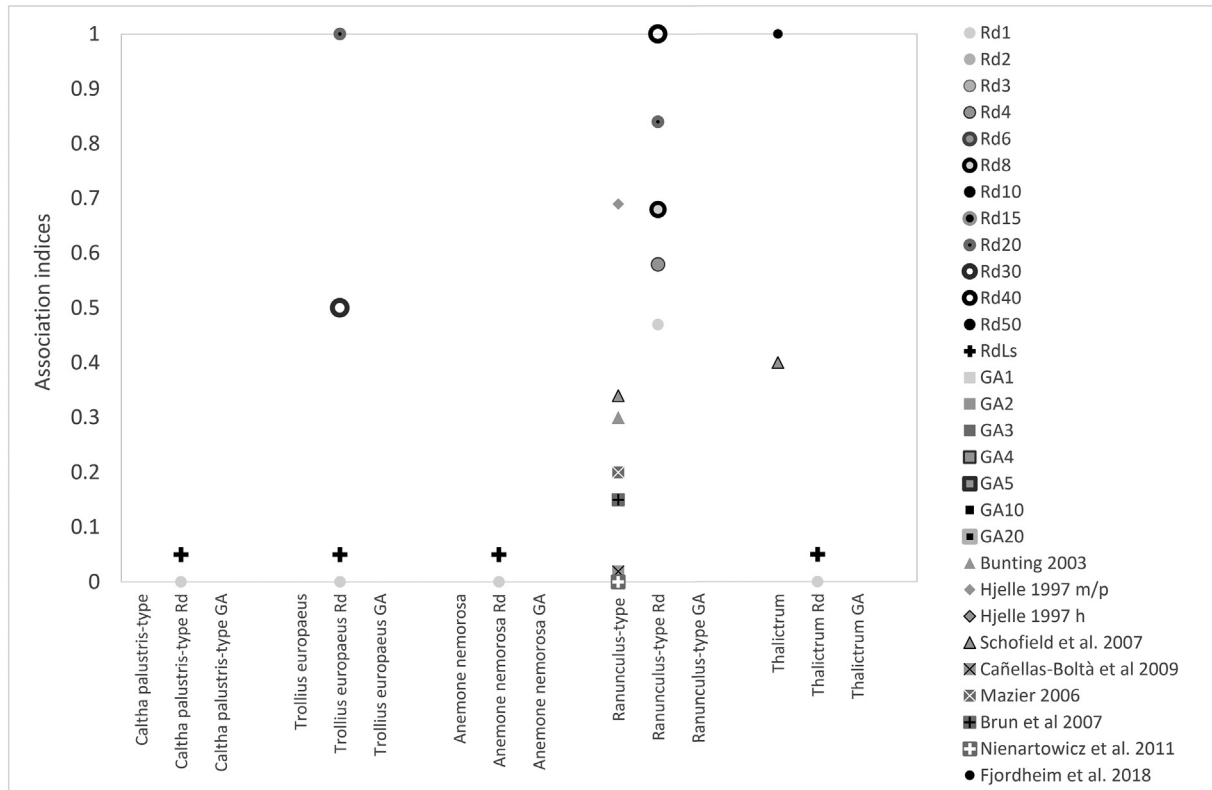


Fig. 6. A comparison of Ranunculaceae taxa association scores with distance across Rd and GA sites and those from other studies in Europe, referenced in the text. Number after Rd (Ribblesdale) or GA (Glen Affric) is survey distance; for Hjelle, 1997 m/p = meadow/pasture and h = heath (see original paper for details). Note for some papers only selected taxa values are available as other taxa are presented only in ordination diagrams.

the spatial extension of the analysis, but **A** reduces considerably at the landscape scale. The Glen Affric samples lack significance to the association. However, high values for **U** may be due to the limited pollen counts in this dataset; thus (especially in landscapes dominated by arboreal taxa) highlighting the importance of counting to a suitable pollen sum (Weng et al., 2006). It seems that a few grains of *Erica* pollen is a reasonably good predictor of very local plant presence (see also Shaw, 2006), but absence may be misleading. *Erica*, as with *Vaccinium* species, may be buzz pollinated (Willmer, 2011). This mechanism would match well with the somewhat sporadic, but local, dispersal patterns found.

The combined dataset shows a weak, but significant association between *Empetrum* pollen and plant presence at 2 m source area ($p = .026$). However, the Ribblesdale dataset shows WA but no significance and, similarly, Schofield et al. (2007) also finds *E. nigrum* to be a WAT in pastoral landscapes of Greenland (Fig. 5). In both datasets **A** becomes slightly stronger at the landscape scale; however, this is at the expense of increased **U** indices, indicating some stochasticity in dispersal. The inverse ROC and P-R plots also show weakening precision and recall at distance. The MCC, CCR indicate that the source area with most probability of successful pollen–plant matching is 3–4 m for *Empetrum*. *Empetrum* is an important species indicating disturbance and erosion (Tallis, 1997), under which circumstances pollen counts can become much increased. Given the low counts of pollen the local source area found in the Ribblesdale and Glen Affric studies the findings must be used with caution if *Empetrum* becomes abundant in the paleoecological record, as some extra-local dispersal is probable with increased plant cover.

Lastly, the taxonomic distinction between the Ericaceae is challenging for some grains. The combined *Erica/Vaccinium* agg. is **SAT** and has a high predictive value ((CCR > 0.6) and significance ($p < .01$) from 2 to 20 m); this may indicate that separation of some types of Ericaceae pollen may be subject to taxonomic error.

4.5.3. Ranunculaceae

The Ranunculaceae are represented in the Ribblesdale surveys by nine plant species and five pollen taxon (Table A.1). Data from Glen Affric are very limited; this is expected as the Ranunculaceae contains species that are regarded as anthropogenic indicators in pastoral environments and grazed forest (Behre, 1981), neither of which are communities currently found in Glen Affric.

Caltha palustris-type, *Anemone nemorosa* and *Thalictrum* are all characterized by simultaneously high **O** and high **U** values (Table A.2). Pollen and plant for these taxa are particularly scarce across the dataset leading to a lack of statistical power. However, the few occurrences of pollen in the data seem entirely stochastic in relation to plant presence; the zero, or in some cases negative, MCC confirms this. All three pollen types are represented by plant species that occur in the wider Ribblesdale environment. Some inference can be made therefore that a sporadic pollen presence of these taxa may indicate occurrence at the wider landscape scale, but not necessarily close to the site of deposition; whilst absence of pollen does not indicate absence of plant. Comparison with other datasets demonstrates how rare these pollen taxa are (see Fig. 6), as they seldom form part of the suite of plants analysed. *Thalictrum alpinum* is found associated with plants in Greenland and *Anemone nemorosa* is included in the IOA from Norway samples but data are only presented as an ordination plot where *Anemone* tracks the **O** value (Hjelle, 1997 p.6).

Trollius europaeus has a fluctuating **A** value: **A** = 1.0 at 20 m and **A** = 0.5 and **U** = 0.5 at 50 m; again, pollen and plant are scarce, occurring only in two of the survey sites in Ribblesdale, although the species is abundant in patches in the wider landscape. It seems possible to assume that where *T. europaeus* pollen is present there is likely to be plant presence in the local or extra-local area, since the Glen Affric data contains no plant and no pollen. However, in Ribblesdale the taxon is under-represented so absence of pollen does not indicate absence of plant. The Fisher's exact test is significant in Ribblesdale at 20 m ($p = .053$,

Table A.2) and strongly significant in the combined dataset with the addition of the Glen Affric negative co-occurrences. The MCC of 1.0 also confirms a high level of correlation at 20 m.

Ranunculus acris-type is **SAT** in Ribblesdale. **A** improves from 0.53 at the 1 m source distance to 1.0 at 40 m source distance. Pollen and plant are present in every survey; so data has high positive prevalence. The Glen Affric data show that *Ranunculus* pollen may be transported over wider distances: The pollen is present in 12 of the 29 Glen Affric sites, whilst the plant is present at only 1 site (which has no pollen). Although the plant survey distance is limited to 20 m in this dataset, other records for *Ranunculus* plants in Glen Affric, show only sporadic occurrence (BSBI 2015). Combining both datasets reduces the strength of association, but also increases the statistical power, leading to significant associations from 4 to 20 m. Several other studies have determined that *Ranunculus acris*-type or Ranunculaceae pollen is either a WAT, with high under-representation (Cañellas-Boltà et al., 2009), or an ORT (Nienartowicz et al., 2011; Brun et al., 2007; Schofield et al., 2007; Bunting, 2003) and (Fig. 6). Over-representation is consistent with a wider source area as demonstrated by the increase in **A** over distance in Ribblesdale. In the case of the study by Cañellas-Boltà et al. total pollen counts seem low, at a minimum of just 300 grains. Although their method was to count until the saturation of diversity, sparse grains may be found sporadically and this approach may still miss some grains found in much larger pollen counts per sample.

Mazier et al. (2009) recognize two types of pollen: taxa with a regional source (Regional Human Activities Pollen Indicators-RHAPI) group, and taxa with a more local source (Local Pastoral Pollen Indicators-LPPI); they allocate Ranunculaceae to the LPPI group with **A** values in their study of just 0.20 (Mazier et al., 2006, p. 99). In Ribblesdale empirical evidence of a strengthened association to 50 m, as more plants are included with distance, may delimit the maximum source area; and, therefore, support the LPPI classification. However, *Ranunculus* has high prevalence and further testing is required at wider source areas, across areas with no very local plant source, to test if dispersal to the pollen record occurs across greater distances. Ejarque et al. (2011), found that *Ranunculus*-type was present as plant in high-Alpine sites but over-represented in pollen samples from the valleys; thus, perhaps, indicating long-distance pollen transport. Further detailed vegetation survey for modern analogs, with a greater range of source areas, is required to clarify the pollen dispersal distance for the Ranunculaceae.

4.5.4. Fabaceae

In Ribblesdale the Fabaceae are represented by three pollen types and several plant species in the *Lotus*, *Lathyrus*, *Trifolium* and *Vicia* genera. The Fabaceae were not found in Glen Affric surface samples in plant or pollen taxa.

In Ribblesdale there was no *Vicia sylvatica*-type pollen present although the plant species associated with this pollen type appeared in six of the sampled areas. Since there is possibility of overlap between pollen grains of the Fabaceae some researchers have not differentiated pollen taxa within this family (e.g., Bunting, 2003). A variety of combinations are aggregated here: *Vicia sylvatica*-type and *Vicia cracca* (Moore et al., 1991) are combined to provide *Vicia*-type (cf. Beug, 2004, *Lathyrus-Vicia*-type, p. 359); *Vicia*-type and *Trifolium*-type are combined to form *Vicia/Trifolium* agg.; Fabaceae agg. is not used as *Lotus*-type should always be distinct due to its smaller pollen size.

All combinations of pollen taxa and grouped pollen taxa within the Fabaceae show consistently high under-representation. A comparison of data across published sites (Fig. 6) shows that association for all of the Fabaceae taxa is weak. In Ribblesdale high **O** indices in *Vicia* and *Vicia/Trifolium* agg. are due to two anomalous sites where pollen occurred as single grains, but no *Vicia* plants were present. Since both of these sites contained *Trifolium* plants, **A** increases in the *Vicia/Trifolium* agg. analysis and **O** indices reduces to zero; this, perhaps, indicates a misidentification in separating the Fabaceae.

The weak **A** values which decrease at the landscape scale indicate that a pollen source for Fabaceae is likely to be within 50 m. Several other studies have determined Fabaceae to be an URT (Bunting, 2003; Mazier et al., 2006; Nienartowicz et al., 2011), whilst others have found weak association with high **U** and **O** values (e.g., Brun et al., 2007, p. 34: *Lathyrus/Vicia*-type **A** = 0.5; *Lotus*-type **A** = 0.33 **U** = 0.33 **O** = 0.6 *Trifolium*-type **A** = 0.17; **U** = 0.8; **O** = 0.5). The classification as a WAT is corroborated by data from Ribblesdale. Whilst high **O** in Brun et al.'s samples may indicate possible association at a wider source area, it is not clear if **O** would be reduced and **A** increased by wider surveys. The high **O** could also, however, indicate contrasting dispersal between differing Fabaceae species. The Ribblesdale data, however, together with most other sites, infers local and somewhat stochastic dispersal as the CCR and **A** values reduce after the 10 m sample for *Vicia/Trifolium*. This very local, but somewhat sporadic, association may be consistent with buzz pollination mechanism in these taxa (Table A.1).

Hjelle (1999) finds some relationship between plant and pollen for *Lotus*-type, recommending a pollen correction factor of $\times 10$. Although this may assist in quantifying very local plant abundance where pollen is present, the high number of pollen absences with plant presence leading to a high **U** value (0.80 at 15 m), in the Ribblesdale dataset, indicates that application of pollen correction factors must be used with caution. Pollen absence does not mean plant absence; especially when counts of the herbaceous pollen sum are low, leading to sporadic detection of sparse grains across samples.

4.5.5. Rosaceae

Filipendula, *Potentilla*-type and *Poterium sanguisorba* (*Sanguisorba minor* sp. *minor*) are present in pollen and plant taxa in Ribblesdale. Each of these taxa show increasing **A** values with increased ring survey distances.

Potentilla-type is strongly associated in Ribblesdale; association increases with plant presence to 8 m where **A** = 0.95. One site contains the plant from the 3 m source area outwards, but with no pollen found in the corresponding sample. An examination of the DAFOR and percentage cover data for this particular sample site (not shown) indicates that *Potentilla* plants, although present, are extremely sparse at this particular site. *Potentilla*-type pollen is found often in pollen samples, despite being classed as insect-pollinated (Hanley et al., 2008). Wide dispersal may be driven by the open nature of the flower (Hjelle, 1997), which is common to the main plant types found in Ribblesdale, *Potentilla erecta*, *Potentilla anserina*, and *Fragaria vesca* are all classified as plants with exposed nectar and an open flower (Knuth, 1908). *Potentilla* is often visited by flies (Willmer, 2011) and, presumably, there may be some potential for local pollen spread by generalist insect activity.

Schofield et al. (2007) find, in Greenland, that *Potentilla*-type is associated and somewhat over-represented (**A** = 0.43 and **O** = 0.49) (2007, p. 483). This over-representation compares well with source areas of 1 or 2 m in Ribblesdale and Glen Affric (Fig. 7). However, the rapid reduction in **O** values at wider source areas in Ribblesdale and Glen Affric indicate that the plant survey area in Greenland may not have been wide enough to capture all plants dispersing pollen to the sample. The Ribblesdale study indicates that *Potentilla*-type pollen may disperse over a minimum of 8–10 m, perhaps further. In contrast, Bunting (2003) found a high **A** value within 2 m; however, this may indicate ubiquitous plant presence in those particular heathland samples, rather than a limit to pollen dispersal. Indeed the samples from similar heathland sites in Glen Affric, where *Potentilla* is common, show a weaker, but significant, relationship at 1 m.

Mazier et al. (2006) find that the presence of *Potentilla*-type pollen is a good indicator of grazing and the secure interpretation of local dispersal will be a useful tool in spatially delimiting the pollen signal. However, since this plant is also frequent in moorland (see above and Bunting, 2003) a cautious approach is required to interpretation across differing ecological communities. In the UK uplands *Potentilla erecta* can

occur in areas with very extensified grazing and may flower more freely in lower grazing pressure sites, whilst *Potentilla anserina* occurs in trampled areas, and *Fragaria vesca* in short grazed grassland as well as banks and scrub areas. In areas with more sporadic presence such as Ribblesdale the ability of the pollen to disperse across several communities is apparent.

Filipendula and *Poterium sanguisorba* are WAT. **A**, **O** and **U** scores tend to converge at 20–50 m following a reduction in **O** value with distance. However, *Filipendula* is weakly associated between these distances. At the landscape distance **A** values dip at some distance beyond 50 m because 10 sites have no pollen present as increasing numbers of plants are found at wider source areas. In contrast, only 3 sites have pollen present with no plant present by 50 m. For *P. sanguisorba*, pollen is present at 11 sites and absent at 8 sites; there are no data recorded for this species in other studies. In this case the landscape distance IOA values show **A** values continuing to increase to 0.58. *P. sanguisorba* may, therefore, be an associated type at some distance beyond 50 m and potentially have a wider pollen source area. This finding seems to support the pollination ecology of these plant taxa, with *Filipendula ulmaria* being an insect and wind pollinated species (Yıldırım and Turker, 2009), with regular pollination by flies (Willmer, 2011), but *P. sanguisorba* strictly anemophilous (Cox and Grubb, 1991), perhaps leading to dispersal over a wider area.

4.5.6. Urticaceae

Calculations at proximal ring source areas show high **O** values for *Urtica* pollen; however, as pollen source plants are included at greater distance the **O** value reduces and **A** increases. There is good agreement between **A** and CCR scores with distance because P_1 is limited. From 20 m one site includes the plant, but not the pollen. However, since three more sites containing pollen became associated with plants between 20 and 50 m, it would seem likely that *Urtica* pollen can be transported from distances greater than 20 m.

Hjelle (1998) omits *Urtica* from R-value estimates as it is found in the pollen but not the vegetation surveys in her study, as in Glen Affric for this study. Mazier et al., 2006) and Nienartowicz et al. (2011), find *Urtica* to have very weak **A** values and high **O** (Fig. 7). The application of a wider source area approach and the trajectory of increasing **A** values over distance in Ribblesdale, indicate that pollen-plant association for *Urtica* improves from an AT to SAT with the incorporation of wider plant source areas. In Glen Affric, *Urtica* pollen is found in nine samples and is present in the wider landscape in some limited plant recordings (BSBI 2015), but it is unclear how close beyond the 20 m surveyed area it occurs for any one of the sites containing pollen. The anemophilous nature of pollen dispersal (Greig-Smith, 1948; Stebbins, 1970) indicates a wider source area, but the Ribblesdale data with a small **U** value from (0.14) at 20 m indicates that dispersal may become more stochastic beyond this distance.

4.5.7. Oxalidaceae

Oxalis acetosella pollen and plants are sparse, occurring across only three sites. No occurrence of the plant was found within 20 m of the sample point; consequently, a wider range of samples with plant present at different source areas would help to clarify the association of this taxon. In contrast to **A**, the CCR score is high for this taxa; however, a high **U** indices value continues to increase at the landscape distance. It would seem, from this limited evidence, that *Oxalis* pollen is sporadically and locally dispersed; this would match with the cleistogamous pollination strategy of the taxa (Packham, 1978; Hjelle, 1997).

4.5.8. Euphorbiaceae, Violaceae, Geraniaceae, Cistaceae and Primulaceae

These plant families are represented by plant presence of *Mercurialis perennis*, *Geranium robertianum*, *Viola* sp. *Helianthemum nummularium*, and *Primula vulgaris*, *P. veris* and *P. farinosa*. Presence is sparse in the Ribblesdale sites and no relationship between pollen and plant is found. MCC at most distances is negative. In contrast Court-Picon et al.

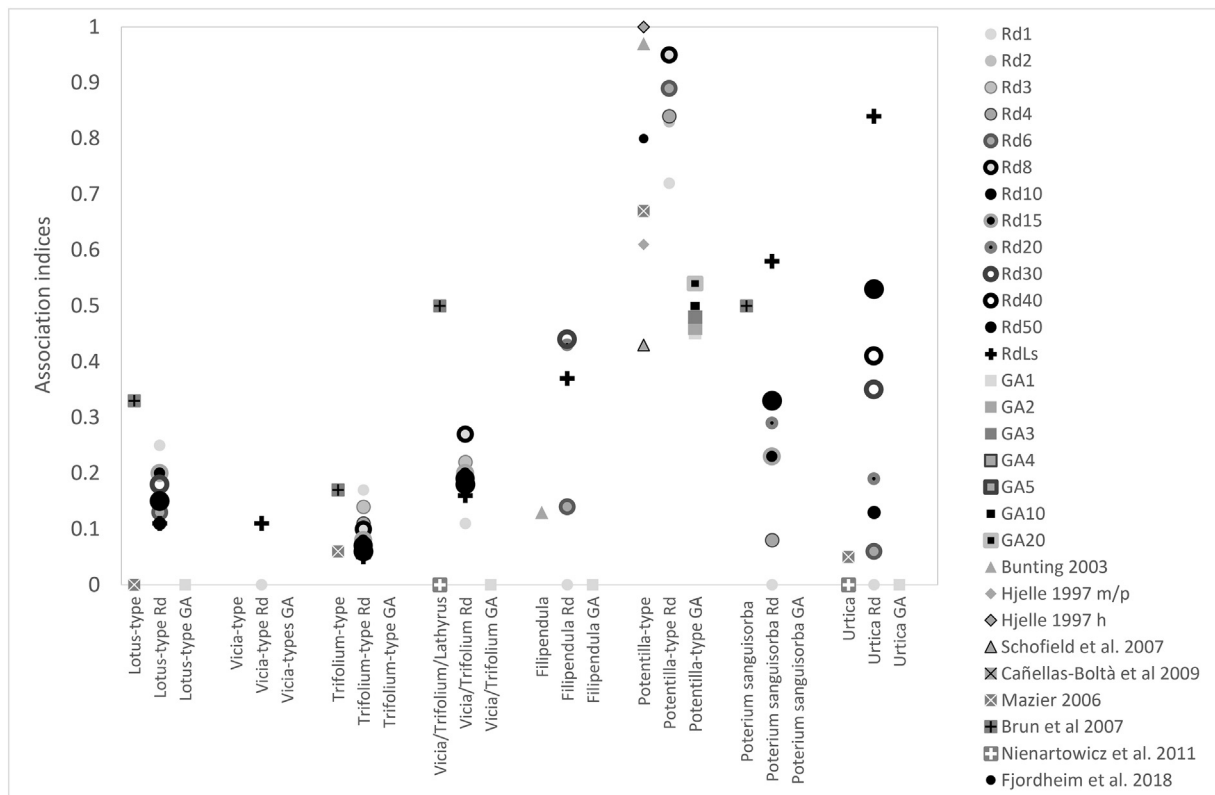


Fig. 7. A comparison of Fabaceae, Rosaceae and Urtica taxa association scores with distance across Rd and GA sites and those from other studies in Europe, referenced in the text. Number after Rd (Ribblesdale) or GA (Glen Affric) is survey distance; for Hjelle, 1997 m/p = meadow/pasture and h = heath (see original paper for details). Note for some papers only selected taxa values are available as other taxa are presented only in ordination diagrams.

(2005) found *Geranium* and *Mercurialis* pollen in their modern samples from the Champsaur valley in the French Alps. However these samples were probably associated with a greater number of flowers as they tended to be from areas with plants flourishing in hedgerows. Hedgerows are limited in Ribblesdale, where fields tend to be separated by dry-stone walls. Although, in the Champsaur valley *Helianthemum* pollen was also found associated with dry pasture. Site-specific factors seem important in the representation of these rare herb pollen types.

4.5.9. Brassicaceae

The tentative split between *Hornungia*-type and *Sinapis*-type pollen (Moore et al., 1991) may reveal some slightly more local, 8–15 m source area, for dispersal of *Hornungia*-type. However, data are very limited (Table A.2). As an aggregated group the overall pattern is for an increase in *A* indices value with distance thus a wider pollen dispersal area is suggested.

Brun et al. (2007) find a strong association in the Brassicaceae in their study of arable weed flora. Whereas Cañellas-Boltà et al. (2009) find a slightly weaker association with a high *O* value and Hjelle also finds a high *O* value. This difference might be due to the more widespread occurrence of Brassicaceae plants in the arable areas, and could also, perhaps, be due to different species within this taxa contributing to the pollen sample. In Ribblesdale the association score continues to increase to 50 m and beyond. However, the weak association within 20 m in the combined Glen Affric and Ribblesdale datasets does, have a high significance ($p < .01$ between 2 and 20 m) and *A* decreases at the landscape scale.

The findings from arable weed flora (Brun et al., 2007) may be supported by the combined data here, however the Ribblesdale dataset demonstrates caution in assuming a very local pollen signal from plant and pollen presence in the local area. If prevalence is high (areas with no local plants have not been tested), we must assume that

dispersal could occur over wider source areas. The MCC and Fisher's exact tests should be routinely applied to IOA studies to prevent over-analysis of the data.

4.5.10. Polygonaceae

Rumex pollen occurs in 17 of the 19 Ribblesdale samples. The taxon is mainly anemophilous (Knuth, 1906; Hjelle, 1997; Brun et al., 2007), and not strictly a rare-type pollen as it appears frequently in pollen diagrams; however, it is usually included in IOA studies. The increasing curve of *A* value with increased plant presence at further ring source areas confirms the wider dispersal distance. At 50 m in Ribblesdale *Rumex acetosella/acetosa* has an *A* value of 0.88. Brun et al. (2007, p. 34) find that *Rumex acetosa*-type has a Weak *A* (0.25, p34) and high *O* value, whilst and Bunting (2003) and Nienartowicz et al. (2011) also, find very low *A* values and high *O* (see Fig. 8). It seems clear that these values reflect the source area of the survey rather than a maximum association; the Ribblesdale data, where plants are infrequent within 20 m and pollen is over-represented, corroborate this. The s_{EIOA} indicate that a minimum of 50 m is realistic for pollen dispersal for this taxon; but, given high prevalence at this distance, the true source distance may be in excess of 50 m.

4.5.11. Droseraceae

Droseraceae plant and pollen are only found in Glen Affric; this limited dataset does, however, show a significant and strong association between plant and pollen to 3 m reducing from $A = 1$ to $A = 0.5$ at 3 m, revealing the possibility of a very local pollen source area. The local, but significant ($p = .03$), association with high *U* is probably supported by ambophily in this taxa. Wind-pollination can occur sporadically, when flowers open in this taxon (Crowder et al., 1990), but pollen size is likely to limit transport. However, the limited dataset here give no option to test dispersal over further source distances, as

no more plants are found. In comparison Fjordheim et al., 2018 record a weak association of 0.19 in their 2.5 m × 5 m plots in rich fen communities; however, the very high *U* of 0.94 also implies that wider source distances would not increase the association.

4.5.12. Caryophyllaceae

The Caryophyllaceae are generally found to be WAT in Ribblesdale. The strongest association is found with *Silene flos-cuculi*, where *A* increases gradually over distance to 0.33 at 20 m. However, at this distance there is still high over-representation which is outweighed by under-representation beyond the 50 m survey area. The data are supported by comparison of Inverse ROC and P-R plots which show that the pollen representation is specific, but not very precise (Fig. 3).

Silene flos-cuculi is entomophilous (Biere, 1991), with an open flower morphology; and is a generalist with some pollination by flies. Some limited dispersal might, therefore, be expected, as reflected in the weak association. Ribblesdale displays a steady increase in *A* to 50 m. Whilst the CCR indicates a high level of accuracy, this is driven by a high level of negative co-occurrences leading to a lack of significance in the Fisher's exact test; although the MCC is moderately positive. The predicted decrease in *A* at the landscape scale, together with high *U* indices values, indicates a limit to source area. There is no evidence of a stochastic nature to pollen dispersal from the limited IOA analysis in Ribblesdale, although data from other studies are very limited (Fig. 8).

4.5.13. Rubiaceae

Galium species are widespread in Ribblesdale, and *Cruciata laevipes* is also found. Rubiaceae grains are found frequently in the pollen spectra, despite the flower having only few exposed anthers (Brun et al., 2007, p. 34). Rubiaceae pollen is strongly associated at 2–4 m pollen dispersal distance *A* = 0.71 at 2 m source area (*p* = .03, Table A.2). From

4 m *A* reduces slightly and tends to fluctuate a little whilst *O* continues to decline to zero at 20 m ring source area. These results reflect a similar pattern to that of Brun et al. (2007, p. 34), in agricultural landscapes, where a survey area of 10 × 10 m yields an association of 0.58 with some over-representation (*O* = 0.36). There is a spread of *A* values in other studies (Fig. 8), perhaps reflecting inadequate source area. This, again, shows the potential of the *seIOA* method in defining source area for some taxa. Beyond the 4 m source area plants are still dispersing pollen into some sites; however, by the 20 m source area plants are always found where pollen is present, but the chance of missing pollen in the record where the plant is present has increased to 33%. The relationship, therefore, becomes more erratic at sources over 20 m, whereas at the 5 m source area surveyed by Brun et al., and the 2 m source area of Bunting (2003), there is still over-representation.

Differences in the species from differing communities must also be considered. *Galium* sp. of pasture environments are generally insect-pollinated (including by *Erebia epiphron*, the small mountain ringlet butterfly), with open flowers; whilst *Galium aparine*, a plant of disturbed habitats and tall-herb communities is self-pollinated (Grime et al., 1990) and, perhaps, less likely to disperse pollen. Grime et al. (1990, p. 178), note that *Galium* “tends to be suppressed in derelict grassland”; suppressed growth beneath the grass layer may presumably lead to some under-representation in this species, as the pollen may be trapped in the grassland canopy space. A reduction in *Galium*, but not in other grassland species might potentially, therefore, be interpreted as a change to a more extensive grazing system with a greater sward height; although this assertion requires field testing.

4.5.14. Plantaginaceae

Plantago species produce abundant pollen and form key indicators of pastoral and ruderal environments. *Plantago lanceolata* is mainly anemophilous and is frequent in many pollen diagrams. Bunting (2003,

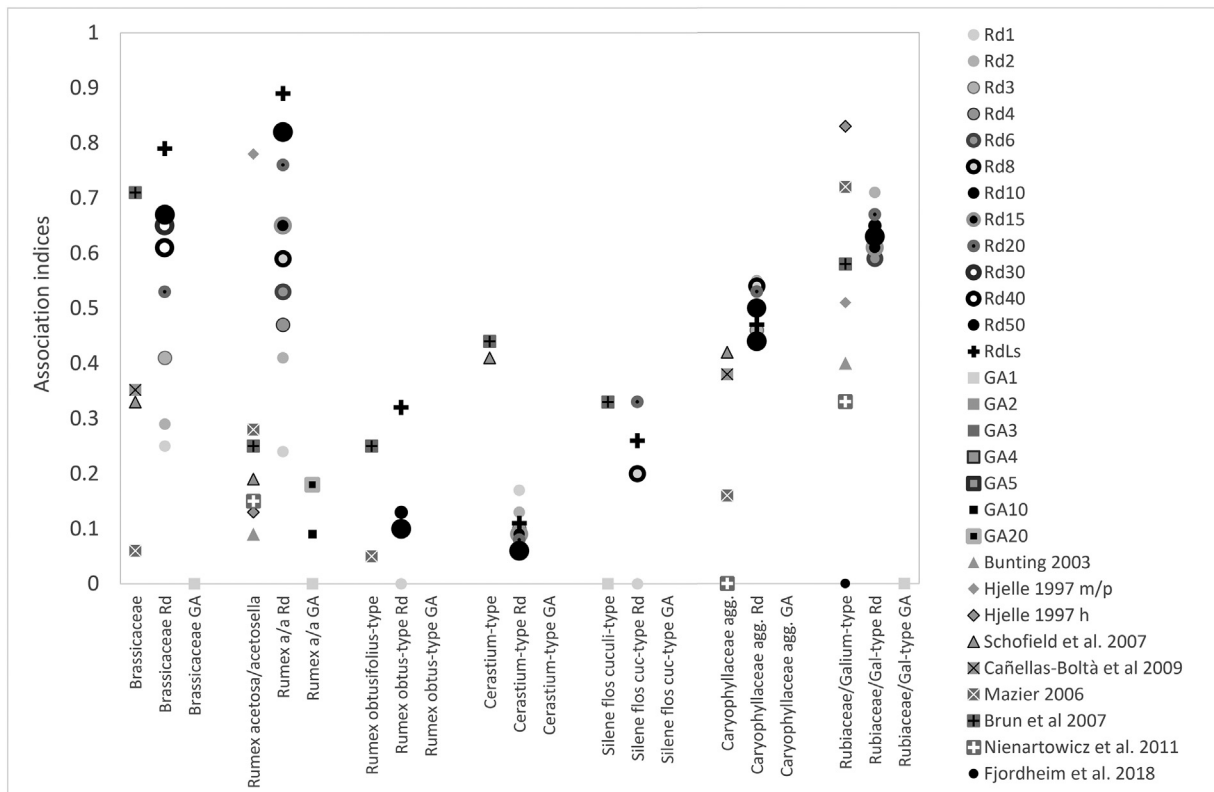


Fig. 8. A comparison of Brassicaceae, Caryophyllaceae and Rubiaceae taxa association scores with distance across Rd and GA sites and those from other studies in Europe, referenced in the text. Number after Rd (Ribblesdale) or GA (Glen Affric) is survey distance; for Hjelle, 1997 m/p = meadow/pasture and h = heath (see original paper for details). Note for some papers only selected taxa values are available as other taxa are presented only in ordination diagrams.

p. 292) uses a source area of 2 m and finds that *P. lanceolata* is over-represented: $O = 0.81$. O is similarly high in the Glen Affric samples. In Ribblesdale the A and O values at 2 m source area are 0.33 and 0.67 respectively (similar to values of Cañellas-Boltà et al., 2009, and Brun et al., 2007). However, the value of an extended plant survey area can be seen again, with the marked increase in A value for this taxa to 0.83 at 50 m, and 0.95 for the landscape distance (see Fig. 9 for comparison across sites and source distances). This confirms pollen dispersal of *Plantago lanceolata* over wider source areas.

Another possible consideration in comparing association between sites might be morphological plasticity: Knuth (1909, Poll. III, p. 322–4) discusses differing growth forms of *Plantago lanceolata* distinguishing insect-pollinated and wind-pollinated morphs. The two growth forms may correspond to areas of differing grazing pressure and may potentially lead to different levels of pollen dispersal at different sites.

For *Plantago media/major* agg. the A indices curve increases to SAT at the 10–20 m survey area and then starts to decrease slightly at all distances from 30 to 50 m. The association is highly significant at 10 and 20 m (Fisher's exact test $p = .001$) and significant to 50 m ($p = .02$; see Table A.2). This perhaps indicates an interpretive source area for this taxa within 10–30 m, at which distance the pollen–plant association demonstrates a high positive accuracy and a 100% negative predictive rate. Brun et al. (2007) group *P. major/media* and find an association of 0.5 (Fig. 9) but also over-representation of 0.5 in their 10×10 m plots. Whilst Nienartowicz et al. (2011) found complete over-representation within 20×20 m plots. *P. major* and *media* have numerous exposed anthers (Brun et al., 2007), but may also be wind-pollinated (Sagar and Harper, 1964); indeed, Knuth (Poll. III, p. 322–4), again recognizes two forms of *Plantago media* similarly divided into anemophilous and entomophilous forms. At greater distances it is more likely that plant will be found unassociated with pollen in the

corresponding sample, it seems likely that wind pollination is not effective at consistently transporting pollen over distance in excess of 40 m.

The A indices value increases greatly in the combined *Plantago media* and *P. major* dataset over the values for the individual species. This probably indicates misidentification between these two similar pollen morphologies. Although the two pollen types are separated by the key of Moore et al. (1991), the taxa are combined by Beug (2004). It would seem, on this evidence, that there is little merit to attempting a separation.

4.5.15. Lamiaceae and Campanulaceae

Stachys-type and *Prunella*-type within the Lamiaceae and *Campanula* show only weak association. However, comparison across IOA plots (Fig. A.1) shows possible differences in iPSA for each of these taxon. *Stachys*-type has a high O indices value and A increases to a peak of 0.4 at 20 m (Table A.2) as further plants are recorded at distance, whereas the IOA relationship in *Prunella*-type is more erratic in response to increased plant presence at wider source areas, leading to a high U . Similar findings for high U in *Prunella* and high O in *Stachys* are presented by Hjelle (1997, p. 6, fig. 2 and p. 10, table 2). Examination of the CCR and MCC scores for Ribblesdale confirm that *Prunella* has a very local iPSA of 2 m, whereas *Stachys* seems to have an iPSA of 20 m. Clarification of the pollen–plant relationship may be improved by larger datasets as the Fisher's exact test does shows significance for these taxa in the combined Ribblesdale and Glen Affric dataset (Table A.2). The slightly erratic relationship between pollen and plant in these taxa, together with the significant local association may be consistent with wastage during insect pollination as the zygomorphic flower type is evolved for bee pollination.

Campanula also shows a peaked IOA pattern with distance; however, this is based upon a single pollen grain presence and limited plant presence. Under-representation is high. The correct classification rate or

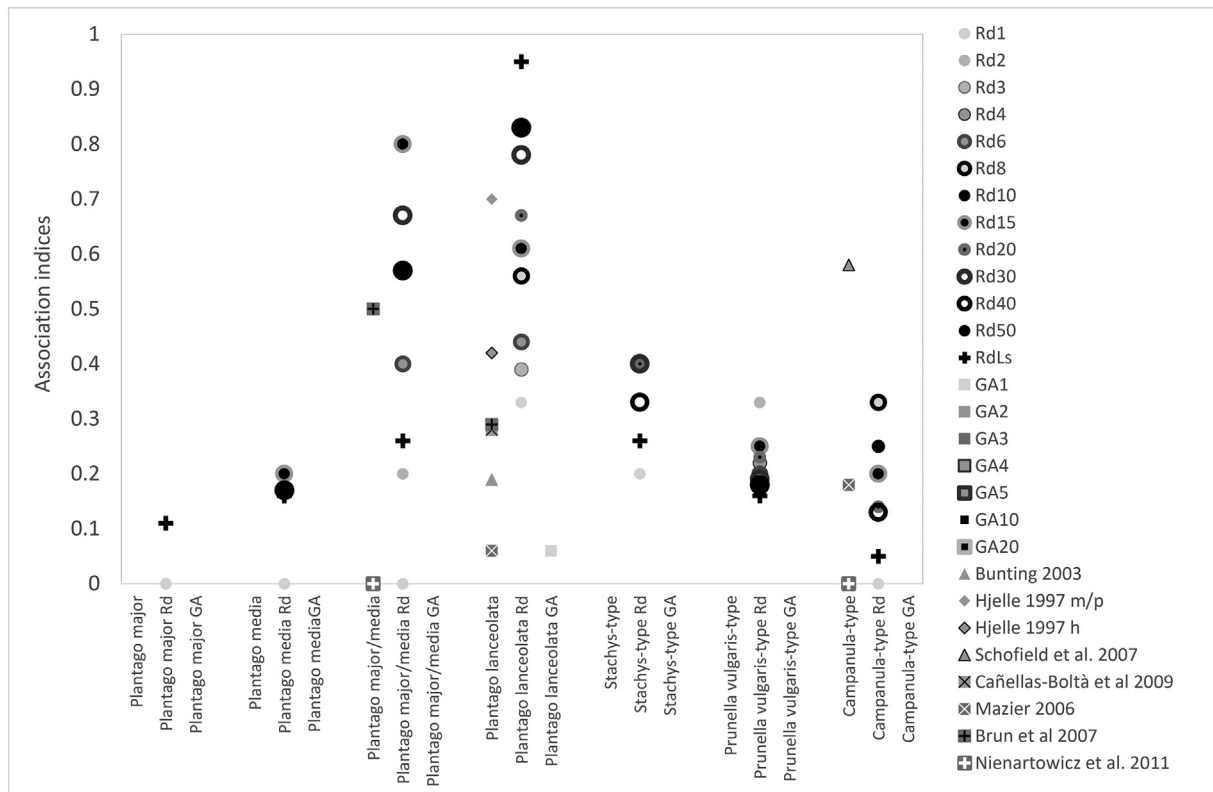


Fig. 9. A comparison of Plantaginaceae, Lamiaceae and Campanulaceae taxa association scores with distance across Rd and GA sites and those from other studies in Europe, referenced in the text. Number after Rd (Ribblesdale) or GA (Glen Affric) is survey distance; for Hjelle, 1997 m/p = meadow/pasture and h = heath (see original paper for details). Note for some papers only selected taxa values are available as other taxa are presented in ordination diagrams.

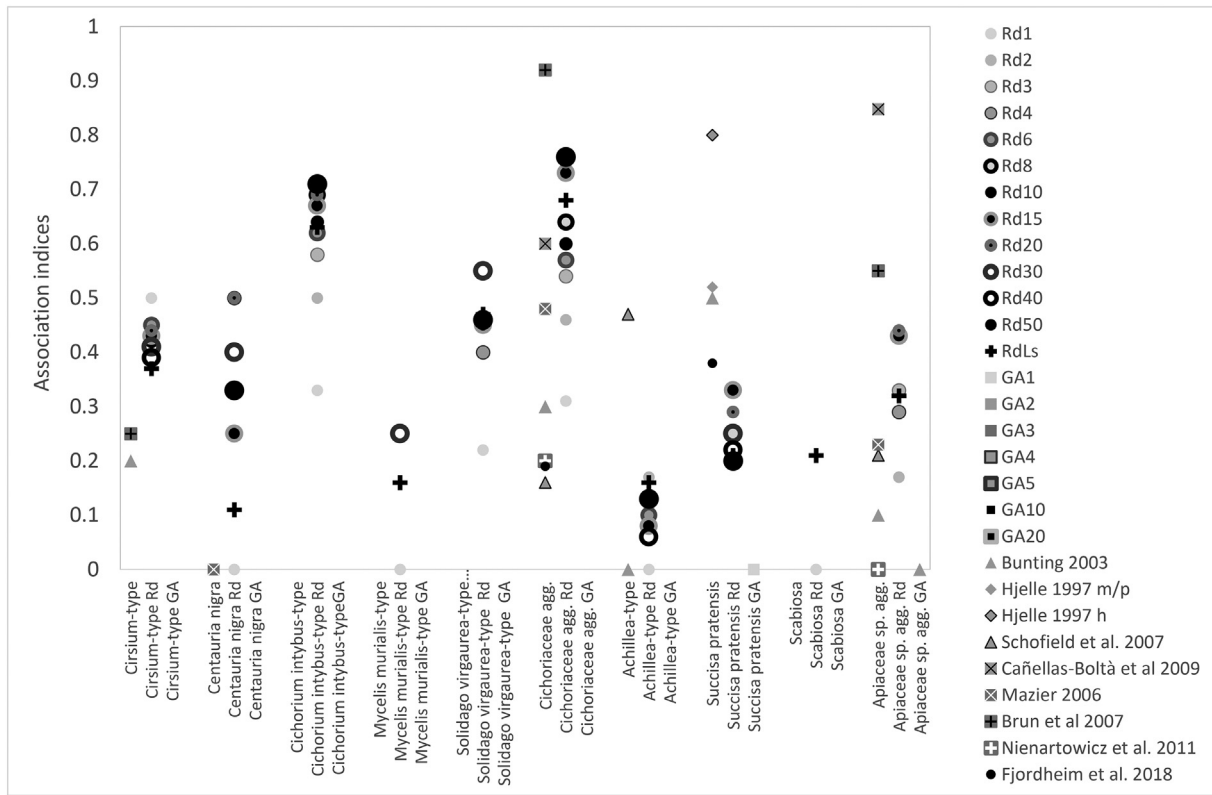


Fig. 10. A comparison of Asteraceae to Apiaceae taxa association scores with distance across Rd and GA sites and those from other studies in Europe, referenced in the text. Number after Rd (Ribblesdale) or GA (Glen Affric) is survey distance; for *Hjelle, 1997* m/p = meadow/pasture and h = heath (see original paper for details). Note for some papers only selected taxa values are available as other taxa are presented only in ordination diagrams.

accuracy is useful here. In the first 8 m source distance this remains high due to several negative co-occurrences but reduces over distance due to finding plants present without pollen representation; a cut-off iPSA with a weak association therefore appears to be 8 m. A spread of association values in three other studies (Fig. 9) corroborates this uncertainty, although *Hjelle (1997)* finds a strong association with some under-representation (*Hjelle, 1997, p. 6, Fig. 2*) where *Campanula* plants are frequent in the vegetation. *Campanula* is also bee pollinated, although may, according to *Stevens et al., 2012*, exhibit some wind pollination later in the season, leading to further potential for stochasticity.

4.5.16. Asteraceae

The Asteraceae show a variety of different association and dispersal patterns. *Solidago virgaurea*-type, and *Cichorium intybus*-type and *Cichorium sp. agg.* (driven primarily by data from *Cichorium intybus*-type), display an increasing association with distance (see Fig. A.1 and Fig. 10). For *Cichorium intybus*-type, although the association value is weaker at proximal plant source areas, the association is significant at 2, 3, 4 and 8 m, with *A* increasing from AT to SAT. Although *A* continues to increase slightly to 50 m high *A* values at greater distances are subject to a corresponding increase in *U*; whilst *A* indices dips at the landscape scale. *Hjelle (1997, p. 6)* finds *Solidago virgaurea*-type to be over-represented, and in the Ribblesdale data the reduction in *O* with increasing plant survey is evident and the highest *A* and CCR scores between 30 and 40 m indicate a likely source area for this pollen type; this was not evident from previous studies with smaller vegetation survey areas.

For the aggregated Chichoriaceae taxon group *A* score increases to 50 m and reduces at the landscape scale, thus also indicating a 50 m iPSA for interpreting community composition from pollen data. *Brun et al. (2007, table 2)* find Chichorioidaeae as a group to be SAT with some *O* in their 10 m plots; extending the source area to 50 m in Ribblesdale shows that *U* overtakes *O* at 15 m source distance. Extended

IOA therefore show that dispersal over wider source areas is more erratic.

Centaurea nigra demonstrates a stochastic pattern of fluctuating association; however, the second *A* peak of 0.5 at 20 m is significant ($p < .04$). In contrast, *Mazier et al. (2006)* find *C. nigra* to be completely dissociated (*A* = 0.0, *p. 99*), indicating a need for larger data sets to fully determine stochasticity from pattern.

Other taxa in the Asteraceae are *WATs*. None of the taxa in this family seem, from the available data, to exceed dispersal within the 50 m source area. The Asteraceae tend to be pollinated by generalist insects, with frequent visits by flies (*Knuth, 1908; Willmer, 2011*). The pollination process may be messy and wasteful via this generalist mechanism, and may, therefore, facilitate local dispersal from insect activity around the site, but preclude wider dispersal.

4.5.17. Dipsacaceae

Succisa pratensis is an important taxa in plant community analysis; but, shows only a weak association with plant presence. The *A* value peaks at 0.33 at the 15 m survey distance in Ribblesdale and does not increase after this point. *U* increases from 15 m as more plants are recorded in sites without pollen presence. The source area for interpretation would seem to be 15 m based upon this limited data; however, a larger dataset is required to provide the statistical power to test the significance of this relationship. An exaggeration factor relative to Poaceae is recommended to be $\times 30$ for this taxa (*Hjelle, 1999*) and *Hjelle (1997)* finds an association of 0.52 in meadow and 0.8 in heath sites (see Fig. 10); however, the plant is more prevalent appearing at 18/26 sites (*Hjelle, 1997, p. 14*) rather than the 8/19 sites in Ribblesdale. The high *U* values in Ribblesdale indicate that there is a substantial miss rate, as pollen may be absent altogether even though the plant is present; this is also supported by *Fjordheim et al. (2018)* in a rich fen community in Norway. In contrast, high *O* indices (0.5 at

15 m) in Ribblesdale, a pattern also present in Norway (Hjelle, 1997, p. 6), indicate some stochasticity to the pollen–plant relationship in this taxa. It would be interesting to test if the National Vegetation communities where *Succisa pratensis* appears in greater abundance and frequency across quadrats are the communities where pollen presence is more consistent. Some decision has to be made for application of the PPE, to choose between: (1) Assuming an extended source area beyond 50 m to reduce the false discovery rate to 0 (whereby the chance of missing the taxa is, according to the false negative rate, very high at 79%); or, (2) Interpreting at a local source area with chance of false discovery.

4.5.18. Apiaceae

Apiaceae are presented as an aggregated group (sensu Bunting, 2003) due to the difficulties in determining between taxa in the pollen of this family. The aggregated pollen taxa does not include *Heracleum sphondylium* as this taxon is more confidently identified.

Apiaceae is a **WAT** with similar **A**, **O** and **U** indices values demonstrating a limited relationship between pollen and plant. The **A** value does increase with distance and then fluctuates between 3 and 20 m source areas. At 3 m there is no under-representation but at 20 m association and **U** are both higher (0.44 and 0.43, respectively). Additional data would help to clarify this, and there may be differences between species in different communities given the broad spread of association values across studies (Fig. 10). However, a source area for maximal association between 3 and 20 m seems likely. Again, applying the PPE of 0.7 (Hjelle, 1999) to this taxa would seem to be subject to a high level of misclassification and this needs to be considered especially if applying the PPE to a different community structure to that of the Norway study.

H. sphondylium is SAT at 1 m for Ribblesdale; however whilst the **A** indices is high, the data remains untested for a range of local plant presence values and the Fisher's exact test reveals weak significance in this dataset ($p = .053$).

5. General discussion and conclusions

5.1. Appraisal of extended IOA analysis

Presence/absence plant survey data, collected in concentric rings around a pollen source, have been used in this study to provide a first attempt at a Spatially Extended Indices of Association, in order to understand better the relationship of rare-type pollen taxa to surrounding vegetation. Distance related curves of **A**, **O** and **U** values, provided here for several taxa from the UK uplands, demonstrate the ability of this method to refine our understanding of the source area for rare-type pollen using repeated indices calculations iterated across increased plant source areas. The range of patterns in the IOA curves show that extended IOA analysis can provide an effective tool to examine association and dispersal distance for rare-type pollen and to determine stochasticity from pattern, highlighting where there is a less secure basis for paleoecological interpretation.

The additional information gained by spatially extended IOA values helps to clarify some of the earlier issues with IOA noted by Davis (1984) for over-represented types. Where association is weak at proximal source areas the strength of association can be examined at increased source areas, across a range of sites with plant presence at different source distances, to examine whether there is a defined iPSA, with decreased **O** and increased **A**, or a more stochastic relationship with a trade-off at distance to increases in **U**. This method and data presented in this study provide additional support and clarification for the determination of local and regional pastoral pollen indicators (sensu Mazier et al., 2009); defining ultra-local (>10 m), local (to 50 m) and landscape scale source areas for different taxa.

There seems little doubt, from the data presented, that the source area for pollen differs between groups of rare-type pollen taxa; and, that source area may be wider, in some cases, than previously thought

(cf. Bunting (2003) and Bunting and Hjelle (2010) who infer a pollen source area of under 2 m in their studies). ERV modelling in REVEALS, which uses the most parsimonious single source distance for key taxa may overestimate source area for some of these herb species, especially the insect pollinated and ambophilous types. In cases where distribution is patchy at the landscape scale considerable over-estimation of cover could potentially occur. Presence of pollen grains from single paleoecological sites may not necessarily infer wider distribution in the landscape, or the ability to apply PPE calibrations to all taxa.

There seems to be a clear relationship with pollen dispersal strategy, in support of the findings of Hjelle (1997) and others, despite the differing landscapes. For several anemophilous species, where over-representation was found in previous IOA studies using smaller plant survey areas, association seems to be stronger in some cases within a 50 m radius of the pollen sample. The method therefore allows identification of truly weakly associated pollen types from those where a greater association can be found by including a wider area for plant presences. Once **U** values continue to increase, from increased plant presence with no pollen, a maximal association may be confirmed. However, several taxa demonstrate a clear iPSA in excess of 50 m and for these types interpolations between multiple locally distributed sites cannot provide a secure interpretation of wide-scale distribution in the landscape.

The value of the newly applied CCR score is clear in application to PPEs and further quantification of the herbaceous flora. The CCR indicates, for rare taxa, the proportion of times a PPE can be correctly applied to adjust the plant cover from pollen abundance for multiple sites in a landscape. By including negative co-occurrences it provides a more accurate and relevant measure of the errors associated with applying a PPE than the association score alone. For example, *Trifolium*-type is only weakly associated at 1–2 m source, when the association is considered against a small dataset of plant presences; but the chances of a correct classification of presence or absence, across the variety of sites in Ribblesdale, is 74%. If PPEs are to be applied to local scale data we must have an understanding of the probability of correctly applying the PPE to gain a quantified cover estimate when the plant is actually present as well as an understanding of how many times we might miss applying an exaggeration factor because the pollen is absent even though the plant is in the community of interest.

However, the CCR can similarly be biased by low prevalence, and other measures used in contingency analysis may be useful (e.g., Fielding and Bell, 1997; Allouche et al., 2006; Chicco, 2017). ROC and P-R plots can be helpful in determining trade-offs between under-representation, and over-representation and specificity. The example of *Centaurea nigra* discussed above shows that it can be easy to read a high association at a proximal source area based on single positive association as significant, but examination of the ROC, P-R and significance testing show a more significant but weaker association at a wider source area due to trade-offs between over and under-representation.

Many strongly or weakly associated types in previous analyses have been classified as such on very limited data. High **A** values with low prevalence have been interpreted as indicating a good association; however, this is not always the case. In contrast high **A** values with high prevalence can also be misleading. Comparing the high association of highly prevalent types, e.g., *Ranunculus* and *Potentilla* with high associations of very rare types such as *Drosera*, demonstrates some limitations of IOA calculations. The addition of the CCR or accuracy can be important for fuller understanding of the dataset. However, significance testing of the data via Fisher's exact tests, and the addition of the Matthews Correlation Coefficient, demonstrate the need for a more rigorous approach to presenting IOA analysis. Imbalanced datasets may lead to false assumptions of the predictor value of the association score (e.g., Chicco, 2017); and therefore to over-application of PPEs in future models; using and comparing results from a range of indices can help guard against this tendency.

A more rigorous approach to reporting sample size, including the number of negative co-occurrences, as well as significance testing of the association, is required. Ideally sample sizes would be numerous enough to determine significance across a range of vegetation types; however, in smaller datasets, ROC and P-R plots can also indicate where **A** score may be high, but data are less secure.

5.2. Implications for the application of Pollen Productivity Estimates

REVEALS models (Sugita, 2007a) determine a single RSAP for all of the pollen taxa from a site, within which vegetation reconstructions can be quantitatively readjusted using PPEs. Indeed, using a quantitative abundance approach and the ERV model in an open landscape Broström et al. (2005) indicate a c. 400 m RSAP for taxa including *Rumex*. However this was calculated using 20 taxa including 9 tree species as well a range of anemophilous herbaceous pollen types with frequent plant presence. Smaller iPSAs are inferred from this study for many taxa. For some rare herbaceous taxa the pollen source areas is much reduced and even for *Rumex*, Ranunculaceae and *Plantago* much of the pollen is likely to be coming from within a 50 m source area. Shaw (2006, p. 228) observes that “several different RSAPs may exist at differing sample radii with different values for z (the background pollen proportion)” and Abraham et al. (2014, p. 3) observe that, for optimal application, the REVEALS model requires “a matrix of different radii for all taxa.”

This study demonstrates that we can obtain dispersal radii as iPSAs for herbaceous and other rare-type taxa, via binary pollen counts and plant surveys. The use of an assumed 1 m radii for rare type taxa (Abraham et al., 2014) can be updated via this approach to include such a matrix of taxon-appropriate radii. Although this approach may have limited application in regional scale models, the application to local-scale ecological paleoecology would be invaluable; especially for determining the landscape location, across multiple local sites, of the rare pollen types, which define plant community composition in paleoecology.

5.3. Further analysis of contingency table probabilities

The addition of sensitivity–specificity analysis in ROC space allows some additional measure of pollen representation, however several other measures are useful in analysis of rates and probabilities in the confusion matrix (Fielding and Bell, 1997; Chicco, 2017), including likelihood ratios and odds ratios. The misclassification rate may be especially useful. In addition paleoecologists need to pay particular attention to Type I errors; as, within a large set of taxa there are likely to be false positives due to stochasticity. This can also be promulgated down core in a pollen diagram where presences and absences of rare types may have little actual meaning. Of course paleoecologists tend to interpret intuitively around these datasets, ignoring single occurrences and looking for continuous occurrences of rare pollen types over several stratigraphic levels, however, when there may be 20 or even 100 years between subsamples we must apply caution. If we find 10 single grains over a 100 stratigraphic samples we need to know the likelihood of false positives at any level in more detail. There may be opportunities to develop Bayesian probabilities of true presence based on the contingency matrix and the association with other taxa. This added to PPEs may complement the MSA approach of Bunting and Middleton (2009), which uses likely community composition derived from landscape factors to place plants in the environment, but observes that “A Bayesian mathematical approach may be an appropriate alternative means of choosing between reconstructions” (Bunting and Middleton, 2009, p. 801). Such an approach that draws on binary presence absence data from a range of communities could use appropriate priors, an approach being taken for key taxa in the LRA where different source areas and PPEs are defined in different layers of the stratigraphy dependent on the shifting pollen spectra across a range of background sites (Sugita et al., 2010).

5.4. Data collection methods

5.4.1. Field survey

The collection of binary plant presence data provides a relatively quick and efficient method of recording data that can be specifically implemented for IOA analysis, or collected alongside data for PPE analysis of key pollen types using the previously agreed methods of Bunting et al. (2013a).

A suggested update to the methods of Bunting et al. (2013a) is the separation of key types, for which abundance data must be collected; and rare types, where presence can be recorded via a random walk of each quadrat until no more new species are found. Whilst it may be rational to scale up percentage cover of grass, or *Calluna* or any key type from a quadrat to a quadrant, in reality the area of scaling increases with the area of each ring source. A rare plant recorded within the quadrat in a quadrant could be subject to unrealistic scaling up to arrive at an overestimation of cover in the quadrant in a patchy landscape. Given the lack of ability to quantify a relationship with stochastic and rare type pollen grains it is perhaps more advantageous to estimate cover for key pollen taxa within the quadrat/quadrant system and to simply record all rare taxa as presence/absence data. A random walkover is analogous to the species–area curve method in the Braun–Blanquet system (Kent and Coker, 1992) (add quadrats until no new species are found); but, is more rapid as it precludes the need for additional quadrat surveys, and reduces the complex recording effort for data across many quadrats. The random walk method may be more appropriate as a survey tool for pollen–diversity measures. Additionally this method avoids potential bias from selecting and amplifying cover data from potentially unrepresentative single quadrats within quadrants in the complex mosaics of local-scale ecological communities.

In addition survey method is limited by the temporal scale. Deposition of pollen into moss polsters may be over multiple years, vegetation survey was carried out on 1 day and this may impact on the plant record. It is likely that some plants are not discovered due to seasonal and annual variability in cover abundance. Of the 23 pollen taxa found with no local plant presence it is unclear how many may be found within the 50 m by repeat survey throughout the summer season. Repeat surveys would be ideal but are unlikely due to time constraints.

5.4.2. Pollen identification and counting

Lastly, some pollen analysts tend to be “splitters” and attempt to achieve very high taxonomic precision in their identification of pollen grains, whilst some tend to use aggregated groups for difficult types. Significant association between pollen and plant in the aggregated groups for some families (e.g., Ericaceae) might indicate limits to pollen identification; IOA groupings can form a tentative test of likely pollen misidentification in surface samples and thus highlight source of potential error in the paleoecological samples, although larger datasets, across multiple analysts, are required to test this further. Significant effort is required to split pollen types securely in order to link with ecological studies and accuracy of our pollen identifications is worth assessing continuously for bias and error.

5.4.3. Statistical power

The data presented here form a proof of method, with some results to guide interpretation of iPSA for rare type pollen. However, the use of the MCC and Fisher's exact test also highlight an important issue for modern analogue research. Often our datasets are small and this can lead to low statistical power. The number of samples in Ribblesdale was 19 and this needs to be expanded. Hjelle (1997, 1998, 1999) collected 187 samples over 40 sites for studies in Norway, this seems exceptional with most studies using 30–50 samples (e.g., Bunting, 2003; Court-Picon et al., 2005; Schofield et al., 2007; Cañellas-Boltà et al., 2009). A greater range of sites may improve the reliability of the iPSA.

Although the interpretation of rare type pollen grains remains difficult, it is interesting to note that all but 7 of the 82 pollen types found

were present in the wider Ribblesdale landscape; and, of these, some could have been present just outside of the 50 m surveyed areas. This means that the pollen coming from long distance transport is likely to be limited and, at the landscape scale, we have a genuine underestimation problem for diversity rather than an over-estimation one. The under-estimation is far worse at this landscape scale than the 23 plants with no pollen suggests, as there are many plants listed in the Ribblesdale area (e.g., Corkhill, 2003) which were not sampled within each of the 19 survey areas and do not contribute to the pollen spectra; If we start to broaden out our interpretation to the landscape scale these add considerably to under-representation of diversity.

Expansion of the quantity of vegetation surveys, the spatial extent, and the number of pollen grains counted per site may all add to a more rigorous understanding, although all, of course, all are time consuming for paleoecological research.

6. Conclusion

The analysis of this surface pollen sample dataset and the newly applied analyses of extended Indices of Association together with ROC, P-R, Correct Classification Rate, and tested with the Mathews Correlation Coefficient and Fisher's exact test both demonstrates success, and implies caution. Contingency indices can provide useful measures of association and under-representation for rare-type taxa in pollen diagrams. Using simple binary survey data over distance allows a more nuanced and rigorous understanding of pollen source area for local pollen types than IOA at measured in single surveys. However, more rigorous testing of our data is required to allow for bias due to rarity or prevalence. The opportunity to determine the individual Pollen Source Area (iPSA) for local-scale studies should prove valuable to future paleoecological interpretations. Also important is consulting a range of indices to establish the probabilities of misclassification in the use of PPEs. A larger dataset, with variation in vegetation presence at all distances from pollen source is required to test the relationships more fully. Comparisons across different landscapes is also essential to fully understand the utility of the rarer fraction of the pollen spectra.

Data Availability

Some pollen data from this project was lodged with the EPD surface data project (Davis et al., 2013). All presence/absence data are provided in the Appendix Table A2. Vegetation data and pollen data are available via the corresponding author. Pollen data will be lodged with the NEOTOMA and EPD databases once fully published.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The Ribblesdale data for this project was funded by a Research Project Grant from the Leverhulme Trust (2009–2013): Post-medieval vegetation change and grazing management in an upland landscape (PI: Professor Ian Whyte). All pollen counting and data analyses were carried out by Helen Shaw. HS is grateful to Dr. Terry Whitaker for tireless assistance with plant survey fieldwork in sometimes trying weather conditions. The project was carried out in and around the Ingleborough National Nature Reserve and thanks are extended to Natural England staff and the farmers and land owners who granted permission to access land. Specific locations are not given to protect privacy.

The data from Glen Affric was collected whilst the corresponding author was undertaking a PhD funded by the University of Stirling, UK and Forest Research, Northern Research Station, UK. The fieldwork to collect

the vegetation data was outside the original scope of the PhD and was funded by a bursary gratefully received from the Inverness Field Club. Thanks are due to several colleagues at Stirling who helped with fieldwork.

Michelle Farrell and Jane Bunting shared helpful discussions on pollen identification on occasional visits to Hull during the research, as well as facilitating use of the type slide collection at the University of Hull. The international POLLANDCAL network supported by NordForsk 2001–2005 also provided Helen Shaw with a supportive forum for discussions of pollen representation during her PhD. Michelle Farrell and David Wilkinson kindly made helpful comments on a much earlier draft of this manuscript.

Final data analysis and writing was carried out whilst at Maynooth University, where a balance between research and teaching has been possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.revpalbo.2020.104238>.

References

- Abraham, V., Oušková, V., Kuneš, P., 2014. Present-day vegetation helps quantifying past land cover in selected regions of the Czech Republic. *PLoS One* 9 (6), e100117. <https://doi.org/10.1371/journal.pone.0100117>.
- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43 (6), 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>.
- Andersen, S.Th., 1970. The relative pollen productivity and pollen representation of north European trees, and correction factors for tree pollen spectra. *Danmarks Geologiske Undersøgelse Series II* 96, 1–99.
- APG III, 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161 (2), 105–121. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>.
- Behre, K.-E., 1981. The interpretation of anthropogenic indicators in pollen diagrams. *Pollen Spores* 23, 225–245.
- Bennett, K.D., 1995–2007. Catalogue of pollen types. [unpublished]. <http://www.chrono.qub.ac.uk/pollen/pc-intro.html>.
- Bennington, J.B., Dimichele, W.A., Badgley, C., Bambach, R.K., Barrett, P.M., Behrensmeier, A.K., Bobe, R., Burnham, R.J., Daeschler, E.B., Van Dam, J.V., Eronen, J.T., Erwin, D.H., Finnegan, S., Holland, S.M., Hunt, G., Jablonski, D., Jackson, S.T., Jacobs, B.F., Kidwell, S.M., Koch, P.L., Kowalewski, M.J., Labandeira, C.C., Looy, C.V., Lyons, S.K., Novack-Gottshall, P.M., Potts, R., Roopnarine, P.D., Strömberg, C.A.E., Sues, H.-D., Wagner, P.J., Wilf, P., Wing, S.L., 2009. Critical issues of scale in paleoecology. *Palaeos* 24 (1), 1–4. <https://doi.org/10.2110/palo.2009.S01>.
- Beug, H.J., 2004. *Leitfaden der Pollenbestimmung für Mitteleuropa Und Angrenzende Gebiete*. Verlag, München.
- Biere, A., 1991. Parental effects in *Lychnis flos-cuculi*. II: Selection on time of emergence and seedling performance in the field. *J. Evol. Biol.* 4 (3), 467–486.
- Birks, H.J.B., 1993. Quaternary palaeoecology and vegetation science—current contributions and possible future developments. *Rev. Palaeobot. Palynol.* 79 (1), 153–177.
- Birks, H.J.B., Berglund, B.E., 2018. One hundred years of Quaternary pollen analysis 1916–2016. *Vegetat. Hist. Archaeobot.* 27 (2), 271–309. <https://doi.org/10.1007/s00334-017-0630-2>.
- Birks, H.J.B., Gordon, A.D., 1985. *Numerical Methods in Quaternary Pollen Analysis*. Academic Press, London.
- Boyd, W.E., 1986. The role of mosses in modern pollen analysis: The influence of moss morphology on pollen entrapment. *Pollen Spores* 28 (2), 243–256.
- Bradshaw, R.H., 1981. Modern pollen-representation factors for woods in south-east England. *J. Ecol.* 69 (1), 45–70.
- Broström, A., Sugita, S., Gaillard, M.-J., 2004. Pollen productivity estimates for the reconstruction of past vegetation cover in the cultural landscape of southern Sweden. *The Holocene* 14 (3), 368–381. <https://doi.org/10.1191/0959683604hl713rp>.
- Broström, A., Sugita, S., Gaillard, M., Pilesjö, P., 2005. Estimating the spatial scale of pollen dispersal in the cultural landscape of southern Sweden. *The Holocene* 15 (2), 252–262. <https://doi.org/10.1191/0959683605hl790rp>.
- Broström, A., Nielsen, A.B., Gaillard, M.J., Hjelle, K., Mazier, F., Binney, H., Bunting, M.-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. *Vegetat. Hist. Archaeobot.* 17 (5), 461–478. <https://doi.org/10.1007/s00334-008-0148-8>.
- Brun, C., Dessaint, F., Richard, H., Bretagnolle, F., 2007. Arable-weed flora and its pollen representation: A case study from the eastern part of France. *Rev. Palaeobot. Palynol.* 146 (1–4), 29–50. <https://doi.org/10.1016/j.revpalbo.2007.02.001>.
- Buckland, S.T., Elston, D.A., 1993. Empirical models for the spatial distribution of wildlife. *J. Appl. Ecol.* 30 (3), 478–495.

- Bunting, M.-J., 2003. Pollen-vegetation relationships in non-arboreal moorland taxa. *Rev. Palaeobot. Palynol.* 125 (3), 285–298. [https://doi.org/10.1016/S0034-6667\(03\)00005-8](https://doi.org/10.1016/S0034-6667(03)00005-8).
- Bunting, M.-J., Hjellev, K.L., 2010. Effect of vegetation data collection strategies on estimates of relevant source area of pollen (RSAP) and relative pollen productivity estimates (relative PPE) for non-arboreal taxa. *Vegetat. Hist. Archaeobot.* 19 (4), 365–374. <https://doi.org/10.1007/s00334-010-0246-2>.
- Bunting, M.J., Middleton, R., 2009. Equifinality and uncertainty in the interpretation of pollen data: The Multiple Scenario Approach to reconstruction of past vegetation mosaics. *The Holocene* 19 (5), 799–803. <https://doi.org/10.1177/0959683609105304>.
- Bunting, M.-J., Schofield, J.E., Edwards, K.J., 2013a. Estimates of relative pollen productivity (RPP) for selected taxa from southern Greenland: A pragmatic solution. *Rev. Palaeobot. Palynol.* 190, 66–74. <https://doi.org/10.1016/j.revpalbo.2012.11.003>.
- Bunting, M.-J., Farrell, M., Broström, A., Hjellev, K.L., Mazier, F., Middleton, R., Nielsen, A.B., Rushton, E., Shaw, H., Twiddle, C.L., 2013b. Palynological perspectives on vegetation survey: A critical step for model-based reconstruction of Quaternary land cover. *Quat. Sci. Rev.* 82, 41–55. <https://doi.org/10.1016/j.quascirev.2013.10.006>.
- Cañellas-Boltà, N., Rull, V., Vigo, J., Mercadé, A., 2009. Modern pollen-vegetation relationships along an altitudinal transect in the central Pyrenees (southwestern Europe). *The Holocene* 19 (8), 1185–1200. <https://doi.org/10.1177/0959683609345082>.
- Chicco, D., 2017. Ten quick tips for machine learning in computational biology. *BioData Mining* 10 (1), 35. <https://doi.org/10.1186/s13040-017-0155-3>.
- Corkhill, P., 2003. A provisional check list of the flora of Ingleborough. *Engl. Nat. (Unpublished)*.
- Court-Picon, M., Buttler, A., de Beaulieu, J.L., 2005. Modern pollen-vegetation relationships in the Champsaur valley (French Alps) and their potential in the interpretation of fossil pollen records of past cultural landscapes. *Rev. Palaeobot. Palynol.* 135 (1), 13–39. <https://doi.org/10.1016/j.revpalbo.2005.02.003>.
- Cox, P.A., Grubb, P.J., 1991. Abiotic pollination: An evolutionary escape for animal-pollinated angiosperms. *Phil. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 333 (1267), 217–224. <https://doi.org/10.1098/rstb.1991.0070>.
- Crowder, A.A., Pearson, M.C., Grubb, P.J., Langlois, P.H., 1990. *Drosera* L. *J. Ecol.* 78 (1), 233–267.
- Cundill, P.R., 1979. Contemporary pollen spectra on the North York Moors. *J. Biogeogr.* 6, 127–131.
- Dalirzefat, S.B., da Silva Meyer, A., Mirhoseini, S.Z., 2009. Comparison of similarity coefficients used for cluster analysis with amplified fragment length polymorphism markers in the silkworm, *Bombyx mori*. *J. Insect Sci.* 9 (1). <https://doi.org/10.1673/031.009.7101> (article 71).
- Davis, M.B., 1963. On the theory of pollen analysis. *Am. J. Sci.* 261 (10), 897–912.
- Davis, O.K., 1984. Pollen frequencies reflect vegetation patterns in a Great Basin (USA) mountain range. *Rev. Palaeobot. Palynol.* 40, 295–315. [https://doi.org/10.1016/0034-6667\(84\)90013-7](https://doi.org/10.1016/0034-6667(84)90013-7).
- Davis, M.B., 2000. Palynology after Y2K - understanding the source area of pollen in sediments. *Annu. Rev. Earth Planet. Sci.* 28 (1), 1–18. <https://doi.org/10.1146/annurev.earth.28.1.1>.
- Davis, J., Goadrich, M., 2006. The relationship between precision-recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, June 2006, pp. 233–240. <https://doi.org/10.1145/1143844.1143874>.
- Davis, B.A.S., Zanon, M., Collins, P., Mauri, A., Bakker, J., Barboni, D., Barthelme, A., Beaudouin, C., Bjune, A.E., Bozilova, E., Bradshaw, R.H.W., Brayshaw, B.A., Brewer, S., Brugiapaglia, E., Bunting, M.-J., Connor, S.E., de Beaulieu, J.-L., Edwards, K., Ejarque, A., Fall, P., Florenzano, A., Fyfe, R., Galop, D., Gardini, M., Giesecke, T., Grant, M.J., Guiot, J., Jahns, S., Jankovská, V., Juggins, S., Kahrman, M., Monika, Karpínišková-Kořáček, Kořáček, P., Kühn, N., Kuneš, P., Lapteva, E.G., Leroy, S.A.G., Leydet, M., López Sáez, J.A., Masi, A., Matthias, I., Mazier, F., Meltsov, V., Mercuri, A.M., Miras, Y., Mitchell, F.J.G., Morris, J.L., Naughton, F., Nielsen, A.-B., Novenko, E., Odegaard, B., Ortu, E., Overballe-Petersen, M.V., Pardoe, H.S., Peglar, S.M., Pidek, I.A., Sadori, L., Seppä, H., Severova, E., Shaw, H., Święta-Musznicka, J., Theuerkauf, M., Tonkov, S., Veski, S., van der Knaap, W.O., van Leeuwen, J.F.N., Woodbridge, J., Zimny, M., Kaplan, J.O., 2013. The European Modern Pollen Database (EMPD) project. *Vegetat. Hist. Archaeobot.* 22 (6), 521–530. <https://doi.org/10.1007/s00334-012-0388-5>.
- de Klerk, P., 2018. The roots of pollen analysis: The road to Lennart von Post. *Vegetat. Hist. Archaeobot.* 27 (2), 393–409. <https://doi.org/10.1007/s00334-017-0626-y>.
- Delcourt, H.R., Delcourt, P.A., 1985. Comparison of taxon calibrations, modern analogue techniques, and forest-stand simulation models for the quantitative reconstruction of past vegetation. *Earth Surf. Process. Landf.* 10 (3), 293–304. <https://doi.org/10.1002/esp.3290100311>.
- Duarte, M.C., Santos, J.B., Melo, L.C., 1999. Comparison of similarity coefficients based on RAPD markers in the common bean. *Genet. Mol. Biol.* 22 (3), 427–432. <https://doi.org/10.1590/S1415-4757199900300024>.
- Edwards, K.J., Fyfe, R.M., Jackson, S.T., 2017. The first 100 years of pollen analysis. *Nat. Plants* 3, 17001. <https://doi.org/10.1038/nplants.2017.1>.
- Ejarque, A., Miras, Y., Riera, S., 2011. Pollen and non-pollen palynomorph indicators of vegetation and highland grazing activities obtained from modern surface and dung datasets in the eastern Pyrenees. *Rev. Palaeobot. Palynol.* 167 (1–2), 123–139. <https://doi.org/10.1016/j.revpalbo.2011.08.001>.
- Faegri, K., Iversen, J., 2000. In: Faegri, K., Kaland, P.E., Krzywinski, K. (Eds.), *Textbook of Pollen Analysis, IV edition* The Blackburn Press.
- Feeser, I., O'Connell, M., 2009. Fresh insights into long-term changes in flora, vegetation, land use and soil erosion in the karstic environment of the Burren, western Ireland. *Journal of Ecology* 97 (5), 1083–1100. <https://doi.org/10.1111/j.1365-2745.2009.01533.x>.
- Felde, V.A., Bjune, A.E., Grytnes, J.-A., Birks, H.J.B., 2014. A comparison of novel and traditional numerical methods for the analysis of modern pollen assemblages from major vegetation-landform types. *Rev. Palaeobot. Palynol.* 210, 22–36. <https://doi.org/10.1016/j.revpalbo.2014.06.003>.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24 (1), 38–49. <https://doi.org/10.1017/S0376892997000088>.
- Fjordheim, K., Moen, A., Hjellev, K.L., Bjune, A.E., Birks, H.H., 2018. Modern pollen-vegetation relationships in traditionally aown and unmanaged boreal rich-fen communities in central Norway. *Rev. Palaeobot. Palynol.* 251, 14–27. <https://doi.org/10.1016/j.revpalbo.2018.01.001>.
- Fredh, D., Broström, A., Zillén, L., Mazier, F., Rundgren, M., Lagerås, P., 2012. Floristic diversity in the transition from traditional to modern land-use in southern Sweden AD 1800–2008. *Vegetat. Hist. Archaeobot.* 21 (6), 439–452. <https://doi.org/10.1007/s00334-012-0357-z>.
- Fyfe, R.M., Twiddle, C., Sugita, S., Gaillard, M.-J., Barratt, P., Caseldine, C., Dodson, J., Edwards, K., Farrell, M., Froyd, C., Grant, M., Huckerby, E., Innes, J., Shaw, H., Waller, M., 2013. The Holocene vegetation cover of Britain and Ireland: overcoming problems of scale and discerning patterns of openness. *Quat. Sci. Rev.* 73, 132–148. <https://doi.org/10.1016/j.quascirev.2013.05.014>.
- Gaillard, M.-J., Birks, H.J.B., Emanuelsson, U., Berglund, B.E., 1992. Modern pollen/land-use relationships as an aid in the reconstruction of past land-uses and cultural landscapes: an example from south Sweden. *Vegetat. Hist. Archaeobot.* 1 (1), 3–17. <https://doi.org/10.1007/BF00190697>.
- Gaillard, M.-J., Birks, H.J.B., Emanuelsson, U., Karlsson, S., Lageras, P., Olausson, D., 1994. Application of modern pollen/land-use relationships to the interpretation of pollen diagrams - reconstructions of land-use history in south Sweden, 3000–0 BP. *Rev. Palaeobot. Palynol.* 82 (1–2), 47–73. [https://doi.org/10.1016/0034-6667\(94\)90019-1](https://doi.org/10.1016/0034-6667(94)90019-1).
- Gaillard, M.-J., Sugita, S., Bunting, M.-J., Middleton, R., Broström, A., Caseldine, C., Giesecke, T., Hellman, S., Hicks, S., Hjellev, K., Langdon, C., Nielsen, A.-B., Poska, A., von Stedingk, H., Veski, S., POLLANDCAL members, 2008a. The use of modelling and simulation approach in reconstructing past landscapes from fossil pollen data: A review and results from the POLLANDCAL network. *Vegetat. Hist. Archaeobot.* 17 (5), 419–443. <https://doi.org/10.1007/s00334-008-0169-3>.
- Gaillard, M.-J., Sugita, S., Bunting, J., Dearing, J., Bittmann, F., 2008b. Human impact on terrestrial ecosystems, pollen calibration and quantitative reconstruction of past land-cover. *Vegetat. Hist. Archaeobot.* 17 (5), 415–418. <https://doi.org/10.1007/s00334-008-0170-x>.
- Gaillard, M.-J., Sugita, S., Mazier, F., Trondman, A.-K., Broström, A., Hickler, T., Kaplan, J.O., Kjellström, E., Kokfelt, U., Kuneš, P., Lemmen, C., Miller, P., Olofsson, J., Poska, A., Rundgren, M., Smith, B., Strandberg, G., Fyfe, R., Nielsen, A.B., Alenius, T., Balakauskas, L., Barnekow, L., Birks, H.J.B., Bjune, A., Björkman, L., Giesecke, T., Hjellev, K., Kalina, L., Kangur, M., van der Knaap, W.O., Koff, T., Lagerås, P., Latalowa, M., Leydet, M., Lechterbeck, J., Lindbladh, M., Odegaard, B., Peglar, S., Segerström, U., von Stedingk, H., Seppä, H., 2010. Holocene land-cover reconstructions for studies on land cover-climate feedbacks. *Clim. Past* 6, 483–499. <https://doi.org/10.5194/cp-6-483-2010>.
- Gavin, D.G., Oswald, W.W., Wahl, E.R., Williams, J.W., 2003. A statistical approach to evaluating distance metrics and analog assignments for pollen records. *Quat. Res.* 60 (3), 356–367. [https://doi.org/10.1016/S0033-5894\(03\)00088-7](https://doi.org/10.1016/S0033-5894(03)00088-7).
- Gill, J.L., McLauchlan, K.K., Skibbe, A.M., Goring, S., Zerbe, C.R., Williams, J.W., 2013. Linking abundances of the dung fungus *Sporormiella* to the density of bison: Implications for assessing grazing by megaherbivores in palaeorecords. *J. Ecol.* 101 (5), 1125–1136. <https://doi.org/10.1111/1365-2745.12310>.
- Greig-Smith, P., 1948. Biological Flora of the British Isles: *Urtica dioica* L. *J. Ecol.* 36 (2), 339–355. <https://doi.org/10.1111/j.1365-2745.2009.01575.x>.
- Grime, J.P., Hodgson, J.G., Hunt, R., 1990. *The Abridged Comparative Plant Ecology*. Chapman Hall.
- Hall, V., 1989. A comparison of grass foliage, moss polsters and soil surfaces as pollen traps in modern pollen studies. *Circaea* 6 (1), 63–69.
- Hanley, M.E., Franco, M., Pichon, S., Darvill, B., Goulson, D., 2008. Breeding system, pollinator choice and variation in pollen quality in British herbaceous plants. *Funct. Ecol.* 22 (4), 592–598. <https://doi.org/10.1111/j.1365-2435.2008.01415.x>.
- Hellman, S., Gaillard, M.-J., Broström, A., Sugita, S., 2008. The REVEALS model, a new tool to estimate past regional plant abundance from pollen data in large lakes: Validation in southern Sweden. *J. Quat. Sci.* 23 (1), 21–42. <https://doi.org/10.1002/jqs.1126>.
- Hellman, S., Bunting, M.-J., Gaillard, M.-J., 2009a. Relevant source area of pollen in patchy cultural landscapes and signals of anthropogenic landscape disturbance in the pollen record: A simulation approach. *Rev. Palaeobot. Palynol.* 153 (3), 245–258. <https://doi.org/10.1016/j.revpalbo.2008.08.006>.
- Hellman, S., Gaillard, M.-J., Bunting, M.-J., Mazier, F., 2009b. Estimating the relevant source area of pollen in the past cultural landscapes of southern Sweden - a forward modelling approach. *Rev. Palaeobot. Palynol.* 153 (3), 259–271. <https://doi.org/10.1016/j.revpalbo.2008.08.008>.
- Hjellev, K.L., 1997. Relationships between pollen and plants in human-influenced vegetation types using presence-absence data in western Norway. *Rev. Palaeobot. Palynol.* 99, 1–16. [https://doi.org/10.1016/S0034-6667\(97\)00041-9](https://doi.org/10.1016/S0034-6667(97)00041-9).
- Hjellev, K.L., 1998. Herb pollen representation in surface moss samples from mown meadows and pastures in western Norway. *Vegetat. Hist. Archaeobot.* 7 (2), 79–96. <https://doi.org/10.1007/BF01373926>.
- Hjellev, K.L., 1999. Use of modern pollen samples and estimated pollen representation factors as aids in the interpretation of cultural activity in local pollen diagrams. *Norweg. Archaeol. Rev.* 32 (1), 19–39. <https://doi.org/10.1080/002936599420885>.
- Hjellev, K.L., Halvorsen, L.S., Prösch-Danielsen, L., Sugita, S., Paus, A., Kaland, P.E., Mehl, I.K., Overland, A., Danielsen, R., Hoeg, H.L., Midtbø, I., 2018. Long-term changes in regional vegetation cover along the west coast of southern Norway: The importance of human impact. *J. Veget. Sci.* 29 (3), 404–415. <https://doi.org/10.1111/jvs.12626>.
- Johnson, D., 2008. *Ingleborough: Landscape and History*. Carnegie Publishing Ltd.

- Kallio, A., Puolamäki, K., Fortelius, M., Mannila, H., 2011. Correlations and co-occurrences of taxa: The role of temporal, geographic, and taxonomic restrictions. *Palaeontol. Electron.* 14 (1), 4–14.
- Kent, M., Coker, P., 1992. *Vegetation Description and Analysis. A practical approach*. John Wiley and Sons, Chichester.
- Knuth, P.E.O.W., 1906. *Handbook of Flower Pollination. Vol. I*. Clarendon Press, Oxford (English translation by JR Ainsworth Davis).
- Knuth, P.E.O.W., 1908. *Handbook of Flower Pollination. Vol. II*. Clarendon Press, Oxford (English translation by JR Ainsworth Davis).
- Knuth, P.E.O.W., 1909. *Handbook of Flower Pollination. Vol. III*. Clarendon Press, Oxford (English translation by JR Ainsworth Davis).
- Lepš, J., Šmilauer, P., 2003. *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge University Press.
- Marquer, L., Gaillard, M.-J., Sugita, S., Trondman, A.-K., Mazier, F., Nielsen, A.-B., Yffe, R.M., Van Oedgaard, B., Alenius, T., Birks, H.J.B., Bjune, A.E., Christiansen, J., Dodson, J., Edwards, K.J., Giesecke, T., Herzschuh, U., Kangur, M., Lorenz, S., Poska, A., Schult, M., Seppä, H., 2014. Holocene changes in vegetation composition in northern Europe: Why quantitative pollen-based vegetation reconstructions matter. *Quat. Sci. Rev.* 99, 199–216. <https://doi.org/10.1016/j.quascirev.2014.02.013>.
- Mazier, F., Galop, D., Brun, C., Buttler, A., 2006. Modern pollen assemblages from grazed vegetation in the western Pyrenees, France: A numerical tool for more precise reconstruction of past cultural landscapes. *The Holocene* 16 (1), 91–103. <https://doi.org/10.1191/0959683606hl908rp>.
- Mazier, F., Broström, A., Gaillard, M.-J., Sugita, S., Vittoz, P., Buttler, A., 2008. Pollen productivity estimates and relevant source area of pollen for selected plant taxa in a pasture woodland landscape of the Jura Mountains (Switzerland). *Vegetat. Hist. Archaeobot.* 17 (5), 479–495. <https://doi.org/10.1007/s00334-008-0143-0>.
- Mazier, F., Galop, D., Gaillard, M.-J., Rendu, C., Cuny, C., Legaz, A., Peyron, O., Buttler, A., 2009. Multidisciplinary approach to reconstructing local pastoral activities: An example from the Pyrenean Mountains (Pays Basque). *The Holocene* 19 (2), 171–188. <https://doi.org/10.1177/0959683608098956>.
- Moore, P.D., Webb, J.A., Collinson, M.E., 1991. *Pollen Analysis. second edition*. Blackwell Science Reprinted 1997.
- Yıldırım, A.B., Turker, A.U., 2009. In vitro adventitious shoot regeneration of the medicinal plant meadowsweet (*Filipendula ulmaria* (L.) Maxim). *In Vitro Cellular & Developmental Biology-Plant* 45 (2), 135–144. <https://doi.org/10.1007/s11627-009-9194-x>.
- Natural England. n.d. Ingleborough National Nature Reserve. <http://publications.naturalengland.org.uk/publication/31034?category=59026> [accessed 28 Aug 2018].
- Nielsen, A.-B., Odgaard, B.V., 2005. Reconstructing land cover from pollen assemblages from small lakes in Denmark. *Rev. Palaeobot. Palynol.* 133 (1), 1–21. <https://doi.org/10.1016/j.revpalbo.2004.08.002>.
- Nielsen, A.-B., Odgaard, B.V., 2010. Quantitative landscape dynamics in Denmark through the last three millennia based on the Landscape Reconstruction Algorithm approach. *Vegetat. Hist. Archaeobot.* 19 (4), 375–387. <https://doi.org/10.1007/s00334-010-0249-z>.
- Nielsen, A.-B., Sugita, S., 2005. Estimating relevant source area of pollen for small Danish lakes around AD 1800. *The Holocene* 15 (7), 1006–1020. <https://doi.org/10.1191/0959683605hl874ra>.
- Nielsen, A.B., Giesecke, T., Theuerkauf, M., Feeser, I., Behre, K.E., Beug, H.J., Chen, S.-H., Christiansen, J., Dörfler, W., Endtmann, E., Jahns, S., de Klerk, P., Kühl, N., Latałowa, M., Odgaard, B., Rasmussen, P., Stockholm, R.J., Voigt, R., Wiethold, J., Wolters, S., 2012. Quantitative reconstructions of changes in regional openness in north-central Europe reveal new insights into old questions. *Quat. Sci. Rev.* 47, 131–149. <https://doi.org/10.1016/j.quascirev.2012.05.011>.
- Nienartowicz, A., Filbrandt-Czaja, A., Piernik, A., Jabłoński, P., Kunz, M., Deptuła, M., 2011. Modern pollen rain in heathlands and adjacent forest phytocoenoses in Northern and Central Poland. *Ecol. Quest.* 15 (1), 127–151. <https://doi.org/10.12775/v10090-011-0044-1>.
- Odgaard, B.V., 1999. Fossil pollen as a record of past biodiversity. *J. Biogeogr.* 26 (1), 7–17. <https://doi.org/10.1046/j.1365-2699.1999.00280.x>.
- Overpeck, J.T., Webb, T.L.I.I., Prentice, I.C., 1985. Quantitative interpretation of fossil pollen spectra: Dissimilarity coefficients and the method of modern analogs. *Quat. Res.* 23 (1), 87–108. [https://doi.org/10.1016/0033-5894\(85\)90074-2](https://doi.org/10.1016/0033-5894(85)90074-2).
- Packham, J.R., 1978. *Oxalis acetosella* L. *J. Ecol.* 66 (2), 669–693.
- Pardoe, H.S., 2001. The representation of taxa in surface pollen spectra on alpine and sub-alpine glacier forelands in southern Norway. *Rev. Palaeobot. Palynol.* 117 (1–3), 63–78.
- Parsons, R.W., Prentice, I.C., 1981. Statistical approaches to R-values and the pollen-vegetation relationship. *Rev. Palaeobot. Palynol.* 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).
- Poska, A., Meltsov, V., Sugita, S., Vassiljev, J., 2011. Relative pollen productivity estimates of major anemophilous taxa and relevant source area of pollen in a cultural landscape of the hemi-boreal forest zone (Estonia). *Rev. Palaeobot. Palynol.* 167 (1–2), 30–39. <https://doi.org/10.1016/j.revpalbo.2011.07.001>.
- Prentice, I.C., 1985. Pollen representation, source area, and basin size: Toward a unified theory of pollen analysis. *Quat. Res.* 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., Parsons, R.W., 1983. Maximum likelihood linear calibration of pollen spectra in terms of forest composition. *Biometrics* 39 (4), 1051–1057.
- 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>.
- Proctor, M., Yeo, P., Lack, A., 1996. *The Natural History of Pollination* (Collins New Naturalist Library, Book 83) Kindle Edition.
- Punt, W., Blackmore, S., Hoen, P., Stafford, P., Clarke, G., 1974–2014. *The Northwest European Pollen Flora. Vols 1–8*. Elsevier.
- Räsänen, S., Suutari, H., Nielsen, A.-B., 2007. A step further towards quantitative reconstruction of past vegetation in Fennoscandian boreal forests: Pollen productivity estimates for six dominant taxa. *Rev. Palaeobot. Palynol.* 146 (1–4), 208–220. <https://doi.org/10.1016/j.revpalbo.2007.04.004>.
- Rodwell, J.S. (Ed.), 1992. *British Plant Communities. Grasslands and Montane Communities Vol. 3*. Cambridge University Press, Cambridge.
- Rodwell, J.S., 2006. *National Vegetation Classification: Users' Handbook*. JNCC http://jncc.defra.gov.uk/pdf/pub06_NVCUsersHandbook2006.pdf 2006.
- Rull, V., 2014. Time continuum and true long-term ecology: From theory to practice. *Front. Ecol. Evol.* 2, 75. <https://doi.org/10.3389/fevo.2014.00075>.
- Sagar, G.R., Harper, J.L., 1964. *Plantago major* L., *P. media* L. and *P. lanceolata* L. *J. Ecol.* 52 (1), 189–221.
- Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10 (3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>.
- Schofield, J.E., Edwards, K.J., McMullen, J.A., 2007. Modern pollen-vegetation relationships in subarctic southern Greenland and the interpretation of fossil pollen data from the Norse *landnám*. *J. Biogeogr.* 34 (3), 473–488. <https://doi.org/10.1111/j.1365-2699.2006.01607.x>.
- Seddon, A.W.R., Mackay, A.W., Baker, A.G., Birks, H.J.B., Breman, E., Buck, C.E., Ellis, E.C., Froyd, C.A., Gill, J.L., Gillson, L., Johnson, E.A., Jones, V.J., Juggins, S., Macias-Fauria, M., Mills, K., Morris, J.L., Nogués-Bravo, D., Punyasena, S.W., Roland, T.P., Tanentzap, A.J., Willis, K.J., Aberhan, M., van Asperen, E.N., Austin, W.E.N., Battarbee, R.W., Bhagwat, S., Belanger, C.L., Bennett, K.D., Birks, H.H., Bronk Ramsey, C., Brooks, S.J., de Bruyn, M., Butler, P.G., Chambers, F.M., Clarke, S.J., Davies, A.L., Dearing, J.A., Ezard, T.H.G., Feuerean, A., Flower, R.J., Gell, P., Hausmann, S., Hogan, E.J., Hopkins, M.J., Jeffers, E.S., Korhola, A.A., Marchant, R., Kiefer, T., Lamentowicz, M., Laroque-Tobler, I., Lopez-Merino, L., Liow, L.H., McGowan, S., Miller, J.H., Montoya, R.E., Morton, O., Nogue, S., Onoufriou, C., Boush, L.P., Rodriguez-Sanchez, F., Rose, N.L., Sayer, C.D., Shaw, H.E., Payne, R., Simpson, G., Sohar, K., Whitehouse, N.J., Williams, J.W., Witkowski, A., McGlone, M., 2014. Looking forward through the past: identification of 50 priority research questions in palaeoecology. *J. Ecol.* 102 (1), 256–267. <https://doi.org/10.1111/1365-2745.12195>.
- Shaw, H., 2006. *A palaeoecological investigation of long-term stand-scale ecological dynamics in semi-open native pine woods: Contributing to conservation management in east Glen Affric*. Unpublished PhD Thesis. School of Biological and Environmental Sciences. University of Stirling.
- Shaw, H., 2012. Interpreting the pollen signal to determine recent ecological change for nature conservation in the UK uplands. *Quat. Int.* 279–280, 444. <https://doi.org/10.1016/j.quaint.2012.08.1457>.
- Shaw, H., Tipping, R., 2006. Recent pine woodland dynamics in east Glen Affric, northern Scotland, from highly resolved palaeoecological analyses. *Forestry* 79 (3), 331–340. <https://doi.org/10.1093/forestry/cpl020>.
- Shaw, H., Whyte, I., 2008. Shifting ecosystem services through time in the North West uplands and implications for planning adaptation in the future. *Aspect. Appl. Biol. Shaping Vision Uplands*, 85, 99–106.
- Shaw, H., Whyte, I., 2013. *Land management and biodiversity through time in upper Ribblesdale, North Yorkshire, UK: understanding the impact of traditional management*. In: Rotherham, I.D. (Ed.), *Cultural Severance and the Environment: The Ending of Traditional and Customary Practice on Commons and Landscapes Managed in Common*. Environmental History vol 2. Springer, Dordrecht.
- Smithsonian Channel (n.d.). <https://www.smithsonianchannel.com/videos/slo-mo-foot-age-of-a-bumble-bee-dislodging-pollen/44877> [accessed July 2018].
- Soepboer, W., Sugita, S., Lotter, A.F., van Leeuwen, J.F., van der Knaap, W.O., 2007. Pollen productivity estimates for quantitative reconstruction of vegetation cover on the Swiss Plateau. *The Holocene* 17 (1), 65–77. <https://doi.org/10.1177/0959683607073279>.
- Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. *Univ. Kans. Scient. Bull.* 38, 1409–1438.
- Stace, C., 2010. *New Flora of the British Isles*. Ed 3. Cambridge University Press.
- Stebbins, G.L., 1970. Adaptive radiation of reproductive characteristics in angiosperms. I: Pollination mechanisms. *Annu. Rev. Ecol. Systemat.* 1, 307–326. <https://doi.org/10.1146/annurev.es.01.110170.001515>.
- Stevens, C.J., Wilson, J., McAllister, H.A., 2012. Biological flora of the British Isles: *Campanula rotundifolia* J. *Ecol.* 100 (3), 821–839. <https://doi.org/10.1111/j.1365-2745.2012.01963.x>.
- Stockmarr, J.A., 1971. Tablets with spores used in absolute pollen analysis. *Pollen Spores* 13, 615–621.
- Sugita, S., 1993. A model of pollen source area for an entire lake surface. *Quat. Res.* 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. *J. Ecol.* 82 (4), 881–897.
- Sugita, S., 2007a. Theory of quantitative reconstruction of vegetation I. Pollen from large sites REVEALS regional vegetation composition. *The 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. *The Holocene* 17 (2), 243–257. <https://doi.org/10.1177/0959683607075838>.
- Sugita, S., Parshall, T., Calcote, R., Walker, K., 2010. Testing the Landscape Reconstruction Algorithm for spatially explicit reconstruction of vegetation in northern Michigan and Wisconsin. *Quat. Res.* 74 (2), 289–300. <https://doi.org/10.1016/j.yqres.2010.07.008>.
- Tallis, J.H., 1997. The pollen record of *Empetrum nigrum* in southern Pennine peats: Implications for erosion and climate change. *J. Ecol.* 85 (4), 455–465.
- Tinsley, H., 2001. Modern pollen deposition in traps on a transect across an anthropogenic tree-line on Exmoor, southwest England: A note summarising the first three years of data. *Rev. Palaeobot. Palynol.* 117 (1–3), 153–158. [https://doi.org/10.1016/S0034-6667\(01\)00083-5](https://doi.org/10.1016/S0034-6667(01)00083-5).

- von Stedingk, H., Fyfe, R.M., Allard, A., 2008. Pollen productivity estimates from the forest-tundra ecotone in west-central Sweden: Implications for vegetation reconstruction at the limits of the boreal forest. *The Holocene* 18 (2), 323–332. <https://doi.org/10.1177/0959683607086769>.
- Wahl, E.R., 2004. A general framework for determining cutoff values to select pollen analogs with dissimilarity metrics in the modern analog technique. *Rev. Palaeobot. Palynol.* 128 (3–4), 263–280. [https://doi.org/10.1016/S0034-6667\(03\)00137-4](https://doi.org/10.1016/S0034-6667(03)00137-4).
- Weng, C., Hooghiemstra, H., Duivenvoorden, J.F., 2006. Challenges in estimating past plant diversity from fossil pollen data: statistical assessment, problems, and possible solutions. *Diver. Distrib.* 12 (3), 310–318. <https://doi.org/10.1111/j.1366-9516.2006.00230.x>.

Willmer, P.A., 2011. *Pollination and Floral Ecology*. Princeton University Press Ebook ISBN 9781400838943.

Glossary

- IOA *Indices of Association*
iPSA *Individual Pollen Source Area*
^{SE}IOA *Spatially Extended Indices of Association*
RSAP *Relevant Source Area for Pollen*