

# 1 Spatial scaling of pollen-based alpha and 2 beta diversity within forest and open 3 landscapes of Central Europe

---

4 Vojtěch Abraham<sup>1</sup>, Jan Roleček<sup>2,3</sup>, Ondřej Vild<sup>2</sup>, Eva Jamrichová<sup>3,4</sup>, Zuzana Plesková<sup>2,3</sup>,  
5 Barbora Werchan<sup>5</sup>, Helena Svitavská Svobodová<sup>1,4</sup>, Petr Kuneš<sup>1</sup>

6

7 1 Charles University, Faculty of Science, Department of Botany, Czech Republic, Benátská 2,  
8 CZ-12800 Praha, Czech Republic

9 2 Czech Academy of Sciences, Institute of Botany, Department of Vegetation Ecology,  
10 Lidická 25/27, CZ-602 00 Brno, Czech Republic

11 3 Masaryk University, Faculty of Science, Department of Botany and Zoology, Kotlářská 2,  
12 CZ-611 37 Brno, Czech Republic;

13 4 Czech Academy of Sciences, Institute of Botany, Laboratory of Palaeoecology, Zámek 1,  
14 CZ-252 43 Průhonice and Lidická 25/27, CZ-602 00 Brno, Czech Republic

15 5 German Pollen Information Service Foundation, Charitéplatz 1, 10117 Berlin, Germany

## 16 **KEYWORDS**

17 Pollen–vegetation relationship, source area of pollen, pollen richness, beta diversity,  
18 paleoecology, total variance

## 19 **Abstract**

20 Pollen is an abundant fossil and the most common proxy for plant diversity during the  
21 Holocene. Based on datasets in open, forest, and mixed habitats, we used the spatial  
22 distribution of floristic diversity to estimate the source area of pollen diversity and identify  
23 factors influencing the significance of this relationship.

24 Our study areas are Bohemian-Moravian Highlands and White Carpathians (the Czech  
25 Republic and the Slovak Republic).

26 Sampling 60 sites in forest and open habitats in two study regions with contrasting floristic  
27 diversity, we calculated taxonomic richness (alpha diversity) and total spatial variance (beta  
28 diversity) for pollen and floristic data along two transects, each 1 km long. Following this, we  
29 calculated the correlation between floristic and pollen diversity. We also assessed the  
30 consistency of the relationship in different habitats. Finally, we regressed local contributions  
31 of individual sites to the beta diversity of pollen and floristic data in each of the regions.

32 There was a positive correlation between pollen and floristic richness in both habitats in both  
33 regions; open and mixed datasets were significant. The highest correlation (adjusted R<sup>2</sup>)  
34 mostly occurred within the first tens of metres (1.5–70) and then within the first hundreds of  
35 metres (250–550). Variances of pollen data significantly correlated with variances of floristic

36 data between 100 and 250 m. Local contributions to beta diversity of pollen and plants  
37 significantly correlated in the forest and one of the mixed datasets.

38 Floristic richness at the pollen site and position of the site within the landscape structure  
39 determine the sequence of the appearing species in the increasing distance. The number of  
40 species sets the source area of pollen richness and dissimilarity of appearing species controls  
41 the source area of pollen variance. These findings, linking pollen and floristic diversity,  
42 provide an essential stepping-stone for the reconstruction of historic plant diversity.

## 43 **Declarations**

44 Funding – This study was financed by the Czech Science Foundation (Grant No. 16-10100S).  
45 Authors affiliated with the Institute of Botany were further supported by the long-term  
46 developmental project of the Czech Academy of Sciences (RVO 67985939).

47 Author contributions – PK, JR, VA conceived initial idea; JR, OV, VA developed  
48 methodology of vegetation data sampling; VA analysed of data and drafted the manuscript; all  
49 authors collected the data and commented the manuscript.

50 Conflicts of interest – The authors declare no competing interests.

51 Permit(s) – Permission to enter the nature reserves was given by the Nature Conservation  
52 Agency of the Czech Republic.

53 Data availability – pollen data will be available in the Neotoma Palaeoecological database  
54 Code availability – code to reproduce the numerical analysis will be available at  
55 <https://github.com/vojtechabraham/SpatialScalingPollenDiversity/>

## 56 **Acknowledgements**

57 The Nature Conservation Agency of the Czech Republic is acknowledged for granting  
58 permission to access the nature reserves. The authors are grateful to the following colleagues  
59 who kindly helped during any stage of the fieldwork: Přemysl Bobek, Zita Červenková, Pavel  
60 Daněk, Pavel Dřevojan, Michelle Farrell, Radim Hédl, Markéta Chudomelová, Kryštof  
61 Chytrý, Radka Kozáková, Pavel Novák, and Helena Prokešová.

62

## 63 **Introduction**

64 The ongoing, human-induced changes to biodiversity call for a progress in our understanding  
65 of past biodiversity and its long-term dynamics. Pollen is one of the most frequently used  
66 proxies of past plant diversity and composition (Birks 2019); therefore, a deeper  
67 understanding of pollen-vegetation relationship is essential. Numerous pollen-vegetation  
68 studies focusing on species composition (Davis 1963) have led to the development of spatially  
69 explicit models for plant abundances in the past (Sugita 2007, Theuerkauf and Couwenberg  
70 2017). The fundamental paleoecological proxy of plant diversity is pollen richness, i.e., the  
71 number of pollen types in the record, but available comparisons of current pollen and floristic  
72 richness, i.e., the number of plant species in the surrounding vegetation, resulted in a positive  
73 relationship in only a few studies (Birks 1973, Meltsov et al. 2011, Felde et al. 2015, Reitalu  
74 et al. 2019, Blaus et al. 2020).

75 Floristic data of these studies covered different spatial scales. The local scale, corresponding  
76 to alpha diversity in ecological studies, was captured by field surveys of the surrounding  
77 landscape, either in a defined radius from the pollen sample (250 m in Meltsov et al., 2011;  
78 500 m in Felde et al., 2015; 100 m in Blaus et al., 2020) or in clearly delimited plots (20 × 20  
79 m in Birks, 1973). Large-scale studies usually relied on floristic data available from databases  
80 and floras, which suffer from low spatial precision. Consequently, the resulting richness  
81 corresponds to gamma diversity rather than to alpha diversity as conceived in ecological  
82 studies (grid cells 50 × 50 km; (Reitalu et al. 2019). An empirical estimate of the relevant  
83 source area of pollen (RSAP; Sugita, 1994) is frequently used to detect the spatial scale  
84 captured by the pollen record. It usually spans from ten metres (Calcote 1995) to a few  
85 thousand metres (Abraham and Kozáková 2012) around the pollen sampling site. However,  
86 the RSAP concept is designed for frequent species in the pollen spectra (Bunting et al. 2013),  
87 while pollen richness is strongly affected by rare taxa, including i) herbs originating from the  
88 nearest surroundings (Bunting 2003), and ii) extra-regional pollen component, originating  
89 from distances greater than 100 km (Janssen 1973). The relevant spatial scale of floristic  
90 richness in reference to pollen richness is nevertheless mostly unknown.

91 Pollen richness is strongly influenced by an abundance of high pollen producers, such as  
92 *Pinus* or *Betula* (Odgaard 1999). These taxa tend to dominate the pollen rain and decrease the  
93 probability of detection of the rare taxa, which are often represented by a single pollen grain.  
94 Application of the correction factors (Andersen 1970) derived from pollen productivity  
95 estimates helps to equalise the representation of different taxa and may lead to a stronger  
96 positive relationship between pollen and floristic richness (Reitalu et al. 2019). Indeed, a  
97 study from an altitudinal transect in southern Norway showed that the strongest representation  
98 bias appears in the boreal forest biome, which is dominated by high pollen producers (Felde et  
99 al. 2015). In this respect, studies are needed in the regions where high pollen producers do not  
100 dominate the vegetation.

101 Landscape structure also significantly affects pollen diversity. Open landscapes have a larger  
102 source area of pollen than forest landscapes (Hellman et al. 2009a), though both types of  
103 landscape represent extremes in alpha diversity in temperate and boreal regions. Here, forests  
104 are generally species-poorer than open landscapes, and when sites from both types of  
105 landscape are included, pollen richness is significantly regressed to floristic richness (Meltsov  
106 et al. 2011). In general terms, landscape structure corresponds to beta diversity, i.e., the  
107 variation of species composition among sites within a study area (Whittaker 1960).

108 There were several attempts to estimate beta diversity from fossil pollen records using various  
109 beta diversity indices. Beta diversity can be understood as a compositional turnover across  
110 time or space. Both directional metrics can be estimated between a pair of pollen assemblages  
111 by dissimilarity coefficient as a rate-of-change (e.g. Figueroa-Rangel et al. 2010), or within  
112 the set of assemblages as gradient length in ordination space (e.g. Connor et al. 2019).  
113 Turnover represents an ecological process of species replacement; however, beta diversity can  
114 also be understood as nestedness (Baselga 2010), characterising the extent to which species  
115 assemblage is a subset of a species-rich sample (Blarquez et al., 2014).

116 Presumably, however, only a single study found a positive relationship between pollen and  
117 vegetation in the turnover calculated from data in a grid of 60 × 60 km (Nieto-Lugilde et al.  
118 2015). Other calibration experiments approximated beta diversity by different measures of  
119 landscape structure (Meltsov et al. 2013, Matthias et al. 2015). It therefore follows that studies  
120 treating open and forest habitats separately may provide a deeper understanding of the  
121 relationship between pollen diversity and different components of plant diversity.

122 Our two study areas in the temperate zone of Central Europe are dominated by semi-open  
123 landscape with forest occupied by oak or spruce, which do not belong among high pollen  
124 producers (Kuneš et al. 2019; see Appendix 1, Table A1 and Fig. A1 for further information.)  
125 We hypothesise significant regression in our datasets when open sites and forest sites are  
126 analysed together, much the same as the pattern reported by Meltsov et al. (2011). We aimed  
127 to explore the relationship between the alpha diversity (richness) and beta diversity (variance)  
128 of the pollen samples and surrounding vegetation within datasets of sites from forest, open, or  
129 both habitats together. Furthermore, we aimed to interpret the strength of the relationship  
130 whilst considering the radius of the vegetation survey.

## 131 **Material and methods**

### 132 **Study area**

133 The Žďárské vrchy Mts. are a part of the Bohemian-Moravian Highlands: the most extensive  
134 highland area in the Czech Republic. The landscape is covered mainly with *Picea abies*  
135 plantations, with patches of low-productivity grasslands and agricultural fields concentrated  
136 around villages. The area is relatively poor in plant species, and its Holocene environmental  
137 development is assumed to have been dominated by forests (Roleček et al. 2020). Sampling  
138 sites were placed over an area of 650 km<sup>2</sup> at altitudes between 569 and 760 m a.s.l.

139 The south-western White Carpathian Mts. are situated on the periphery of a forest-steppe  
140 region of the Pannonian Basin (Rasser et al. 2008, Chytrý 2012). The mildly undulating  
141 landscape is covered by a diverse mosaic of vegetation, including patches of broad-leaved  
142 forests dominated by *Quercus robur*, *Carpinus betulus* and *Fagus sylvatica*, as well as mown,  
143 semi-natural dry and mesic grasslands, fields, orchards, and vineyards. The area is considered  
144 a hotspot of fine-scale plant species richness (Wilson et al. 2012, Roleček et al. 2014) and  
145 harbours many rare species with disjunct distributional ranges (Hájková et al. 2011). It is a  
146 part of the White Carpathians Protected Landscape Area and Biosphere Reserve. Sampling  
147 sites were placed over an area of 250 km<sup>2</sup> at altitudes between 205 and 685 m a.s.l.

### 148 **Fieldwork**

149 We sampled 19 sites in the forest and 20 sites in open habitats (mesic and steppic meadows)  
150 of the White Carpathian Mountains region (hereafter WCM), and 10 sites in the forest and 11  
151 sites in open habitats (wet meadows) of the Bohemian-Moravian Highlands (hereafter BMH;  
152 see Fig. 1). Open sites were selected in continuous non-forest habitats with the minimum  
153 distance to a mature tree at least 10 m. Forest sites were located within a continuously  
154 forested area, in a forest gap of at least 1 m<sup>2</sup> to reduce the gravity component of pollen fallout  
155 without the contribution of wind dispersal (Sugita 1994). We tried to avoid overlapping of  
156 sampling sites and simultaneously kept the sampled area compact and homogeneous in terms  
157 of available vegetation types and environmental conditions.

158 We collected pollen samples from a moss cushion of at least 50 cm<sup>2</sup> in the central point of  
159 each site, while vegetation data were collected in a 1 km radius around the central point in the  
160 same year as the pollen data (Table A2). The effort of vegetation sampling was spread into  
161 three zones. Within the first 10 m, we recorded complete species lists in both regions;  
162 however, in WCM, 21 additional plots of 1 m<sup>2</sup> were sampled following a modified  
163 CRACKLES protocol (Bunting et al., 2013) within the 10 m circle to evaluate finer-scale  
164 relationships. Between 10 and 100 m, main vegetation types were mapped in the field with the  
165 help of aerial photographs. The occurrence of additional species not present in the first 10 m  
166 was recorded for each of the mapped polygons. Up to 1,000 m, we recorded additional plant

167 species, and vegetation types were mapped along two 20-metre-wide linear transects.  
168 Directions of the transects were chosen based on the aerial map to cover the highest possible  
169 habitat diversity. At the same time, the two transects had a minimum angular distance of 90°.  
170 Moreover, additional habitats not recorded in the transects were mapped and additional  
171 species recorded. Cultivated plants including ornamentals (e.g. *Thuja*, *Bergenia*) were also  
172 recorded.

173 Based on the collected data, we compiled six datasets: two “uniform datasets” with forest and  
174 open sites separated for each region, and one “mixed dataset” including both forest and open  
175 sites for each region.

176

### 177 **Laboratory and pollen counting**

178 Moss polsters were prepared for pollen analysis using standard procedures (Faegri et al.,  
179 1989). Moss samples were shaken in KOH during the night, and then acetolysed for 2 min.  
180 The pollen concentrate was stored in glycerine or silicone. Pollen slides were counted under  
181 the light microscope at 400× magnification; for selected taxa at 1,000× magnification. The  
182 original pollen sum includes all pollen and spores of vascular plants following the  
183 determination key of Beug (2004).

### 184 **Numerical methods**

185 Due to the varying pollen sum across samples (between 943 to ca. 4,000 grains), we unified  
186 the sum to 943 grains per sample by random selection without replacement and repeated this  
187 100 times. The median number of taxa across the selections was used for further calculation.  
188 The number of pollen taxa (pollen richness) was regressed against floristic richness. Although  
189 the distances spanned from 0.5 to 1,000 m, we considered pollen and floristic richness at this  
190 scale as an alpha diversity. The concept of beta diversity in ecology is less equivocal, and  
191 there are many definitions and corresponding ways as to how to calculate beta diversity  
192 (Anderson et al. 2011). Here, we adhered to the total variance of the site-by-species  
193 community table as a measure of beta diversity (Legendre and De Cáceres 2013). Total  
194 variance represented by BD<sub>total</sub> value is a sum of squares in the site-by-species community  
195 table. We used the Jaccard index on presence-absence data as its measure. The relative  
196 character of BD<sub>total</sub> ranging from 0 to 1 allows for different numbers of sites, thus also  
197 enabling comparison between mixed and uniform datasets. Pollen BD<sub>total</sub> values calculated  
198 for our six datasets were regressed against six floristic BD<sub>total</sub>s at different distances from  
199 sampling sites. Calculation of beta diversity in package *adespatial* (Legendre and De Cáceres  
200 2013, Dray et al. 2020) also provided measurements of the local contribution of each site to  
201 beta diversity (hereafter also “local contribution”) and its significance. We explored the  
202 relationship between pollen and floristic counterparts at different distances from sampling  
203 sites, again by linear regression. The strength of relationships (richness, BD<sub>total</sub>, and local  
204 contribution) was measured by adjusted R<sup>2</sup>. The correlation test proved its significance at the  
205 5% alpha level.

206 We used the software package R (version 3.4.3) for all statistical analyses (R Development  
207 Core Team 2017).

## 208 **Results**

### 209 **Trends and ranges of richness and variance values**

210 In both regions, we found 169 pollen types (95 in BMH and 151 in WCM) and 1,323 plant  
211 species (799 in BMH and 1,098 in WCM). Mean pollen richness per sample varied between  
212 50 pollen types in WCM meadows, 42 pollen types in WCM forest, 38 pollen types in BMH  
213 meadows, and 31 pollen types in BMH forest (Fig. 2). Floristic richness followed the same  
214 order as pollen richness at a distance between 600 and 1,000 m: in WCM, forest species  
215 gradually increased along the whole transect, with the lowest richness between 40 and 200 m.  
216 In other datasets, however, the increase was more irregular, with more than half of the species  
217 appearing already within the first 100 m (Fig. 3).

218 BDtotal in mixed datasets of pollen and plants was always higher than in uniform datasets.  
219 Pollen and floristic BDtotals in WCM were higher than those in BMH up to a distance of 700  
220 m. BDtotals in uniform pollen datasets varied from 0.21 in BMH meadows to 0.24 in WCM  
221 meadows (Fig. 2). Floristic BDtotal was highest in WCM forest and lowest in BMH meadows  
222 considering a distance of 40–200 m in uniform datasets. The general decreasing trend with  
223 increasing distance showed only minor exceptions, the most visible being the increase at 150  
224 m in the WCM forest dataset (Fig. 3).

225 Pollen and floristic richness values were lower in forests than in meadows; however, BDtotals  
226 were higher in the forest than in meadow datasets.

### 227 **Relationship between floristic and pollen richness, calibration of alpha diversity**

228 All datasets showed a positive correlation of pollen and floristic richness for at least some  
229 distances (Fig. 4a). Both mixed datasets and the dataset of WCM meadows showed highly  
230 significant correlations, while BMH meadows and BMH forest showed a less significant  
231 correlation, though a lower number of replications. WCM forest showed only a marginally  
232 significant correlation, despite a higher number of replications. The highest adjusted  $R^2$   
233 appeared on the distances between 1.5 and 550 m, depending on the dataset (Table A3). All  
234 datasets, except WCM forest, showed two maxima of adjusted  $R^2$ : the first one within tens of  
235 metres and the second one within hundreds of metres (450–550 m for BMH, 250–300 m for  
236 WCM). Mixed datasets had higher adjusted  $R^2$  than their uniform subsets, except for BMH  
237 regions between 40 and 200 m (Fig. 4a). The average distance of maximum adjusted  $R^2$  for  
238 six compared datasets was 286 m.

239 Adjusted  $R^2$  between pollen and floristic richness showed two general ranges of distances  
240 where the correlation was high. The first, within tens of metres from the central points,  
241 showed a correlation between pollen and plant richness values in all datasets except WCM  
242 forest (Fig. 4a). At this distance, most species naturally appear for the first time (Fig. 5). The  
243 highest correlation of pollen and plant richness values in WCM meadows is at 1.5 m within  
244 the habitat of species-rich meadows. BMH meadows had a local maximum of adjusted  $R^2$  at  
245 40 m where new habitats such as forests and forest roads commonly occur. BMH forest  
246 showed the best fit at 70 m, where species confined to forest roads frequently appear (Fig. 5).

247 The second range of maximum adjusted  $R^2$  values appeared between 400 and 550 in BMH  
248 meadows and between 250 and 300 m in WCM forest. Those distances match the appearance  
249 of a high number of new species in built-up areas and meadows or pastures, respectively. The  
250 local maximum of adjusted  $R^2$  in BMH forest at 450 m correlated with few species ( $< 10$ )  
251 from roads, clearings, and wet forests; however, a high number of species ( $> 30$ ) at 500 m was  
252 accompanied by a decrease of adjusted  $R^2$ .

253 Meadow datasets in both regions obtained most of the species within the first 100 m, whereas  
254 forest datasets received high numbers of species at greater distances. The floristic richness of  
255 BMH forest largely originated from human-made habitats (forest roads between 10 and 100  
256 m, and built-up areas usually at a distance above 500 m), whereas WCM forests are enriched  
257 by meadows and other semi-natural habitats, usually at a distance above 200 m (Fig. 5).

### 258 **Relationship between pollen and floristic variance, calibration of beta diversity**

259 The highest adjusted  $R^2$  between pollen and floristic BD<sub>total</sub> was identified at 150 m. The  
260 significant correlation appeared between 100 and 250 m, a remarkable high, but the  
261 insignificant correlation appeared between 300 and 600 m (Fig. 6a). The floristic BD<sub>total</sub> of  
262 WCM meadow at 150 m was lower than the floristic BD<sub>total</sub> of the rest of the datasets  
263 concerning the linear relationship to the pollen BD<sub>total</sub> (Fig. 6b). A distance of 150 m  
264 followed the steep decrease of floristic BD<sub>total</sub> at 10-100 m (Fig. 3), when most of the taxa  
265 appeared, and fell between both ranges of maximum adjusted  $R^2$  of the richness regression.

266 Local contributions of sites to the pollen and floristic BD<sub>total</sub>s correlated significantly in both  
267 forest datasets at 100–400 m and in the mixed dataset from the BMH region at 900–1,000 m.  
268 BMH meadows showed a positive but insignificant relationship, while meadow and mixed  
269 datasets from the WCM region did not show any relationship (Fig. 4b).

## 270 **Discussion**

271 Our results show that the spatial distribution of plants determines the relationship between  
272 pollen and floristic richness and between pollen and floristic variance. Furthermore, the  
273 number of newly appearing species controls the correlation of richness values. Different  
274 species with respect to the whole dataset determine the correlation of variances. The  
275 appearance of new habitats at a single site often leads to an abrupt increase in species  
276 numbers. Those species may also be new for the whole dataset; in this way, the source area of  
277 pollen richness and the source area of pollen variance are linked.

### 278 **Landscape structure determines the source area of pollen richness**

279 The source area of pollen richness, measured as adjusted  $R^2$  between pollen and floristic  
280 richness, is determined by numbers of new plant species appearing with increasing distance.  
281 The position of the pollen site within the landscape structure affects the order of the habitats  
282 and the sequence of appearing species. The grain size of the landscape structure is smaller in  
283 the WCM region than the BMH region; thus, the WCM region resembles a more even mosaic  
284 of habitats and the higher floristic-pollen relationship is reached at shorter distances. The  
285 same role of landscape structure was described in studies of RSAP (Bunting et al. 2004,  
286 Hellman et al. 2009a): areas with large patches of landscape have larger RSAP (Broström et  
287 al. 2005). This agreement is understandable, as both approaches seek the most appropriate  
288 distance with increasing areas of vegetation surveys. Floristic richness, assessed in our study,  
289 is a cumulative number of any new species, whereas vegetation data in RSAP approaches are  
290 cumulative abundances of main dominants, which are further distance- and dispersal-  
291 weighted. The RSAP identifies a distance at which all sites have sufficient vegetation cover of  
292 all taxa (Hellman et al. 2009b), or where all sites have all taxa present (Abraham and  
293 Kozáková 2012).

294 While a previous study estimated RSAP at 350–450 m (Kuneš et al. 2019), the WCM dataset  
295 allows comparison with the source area of pollen richness (250 m). Both values are similar,  
296 despite the conceptual differences mentioned above: while RSAP calculation was based on  
297 pollen/vegetation proportions of 17 taxa, here we deal with incidences of the whole spectra.

298 Moreover, vegetation structure for the range between 10 and 1,000 m was recorded  
299 independently in the two studies. We suggest that this robustness indicates the significant  
300 effect of landscape structure on the pollen–vegetation relationship.

301 A critical characteristic of the landscape mosaic, affecting the source area of pollen richness,  
302 appears to be the number of species per patch rather than its area (Fig. A4). The appearance of  
303 open habitats within forest led to the increase of species numbers and the approaching of local  
304 maxima of adjusted  $R^2$  in both regions. While in BMH forest the appearance of forest roads at  
305 about 70 m was crucial, meadows and orchards at about 250 m played a similar role in WCM  
306 forest. Other studies in semi-open landscapes which found a high correlation between pollen  
307 richness and landscape openness (Meltsov et al. 2013, Matthias et al. 2015) corroborate this  
308 idea. Source areas of pollen richness in forests of both regions fell close to the transition from  
309 species-poor to species-rich habitats. Meadow datasets, i.e., cases where pollen sites were in  
310 the species-rich habitats, provided additional insight regarding the compositional similarity  
311 between central habitats and those in the surrounding landscape. In BMH meadows, there was  
312 an increased correlation of floristic and pollen data at 400 and 550 m. This distance relates to  
313 the increase in the number of species due to the frequent transition of meadow complexes to  
314 shrubby habitats and built-up areas. We generalise the “landscape openness” finding of  
315 Matthias et al. (2015) and Meltsov et al. (2013) to the more general “species-rich patches”.

#### 316 **Effect of biodiversity hotspot on source area of pollen diversity**

317 Most of the species in WCM meadows appeared within the first 40 m in the dominant habitat  
318 of extremely species-rich steppic grasslands (Fig. 5). Even though habitats such as built-up  
319 areas and roads appearing beyond 40 m can be potentially species-rich and compositionally  
320 different to those grasslands, the number of new species appearing between 100 and 1,000 m  
321 is small (usually below 20). It is apparent that high fine-scale floristic diversity suppresses the  
322 influence of the surrounding landscape on pollen richness and decreases the source area of  
323 pollen richness. The relationship is stronger at 1.5 m (0.46) than at 250 m (0.3).

324 The strong effect of high pollen richness in WCM meadows is also visible in the comparison  
325 of pollen and floristic variance. At 150 m, WCM meadows had much lower floristic variance  
326 than the other datasets. Floristic variance in WCM meadows corresponding to the pollen  
327 variance and the pattern of the other datasets lay at 6 m (Fig. 6b). Again, this may be caused  
328 by the high fine-scale diversity of the meadows, which include most pollen types present in  
329 the surrounding landscape. Only a few new species appeared in broader surroundings. WCM  
330 meadow sites are too similar at 150 m than other analysed habitats. Indeed, a similar result,  
331 showing compensation of extremely high alpha diversity by low beta diversity, has already  
332 been reported from the White Carpathians (Michalcová et al. 2014).

#### 333 **Floristic reference of pollen-based beta diversity**

334 Estimating the source area of pollen variance as a regression of pollen and floristic variance  
335 implies that the resulting distance of 100-250 m represents all datasets. Though they differ in  
336 species richness, openness, and habitats of origin, the relationship between variances is fairly  
337 linear. The only exception is the biodiversity hotspot of WCM meadows mentioned above. It  
338 shows that the spatial scale at which the pollen variance corresponds to the floristic variance  
339 is dataset-specific. The linearity and the significance of the variance relationship within the  
340 rest of the datasets indicate certain robustness and possible applicability to a variety of fossil  
341 records from peat bogs.

342 The mechanism of establishing the source area of pollen variance was similar to that  
343 mentioned for the source area of pollen richness. The appearance of new habitats with new



344 species (Fig. 5), like open habitat for forest sites (WCM forest) or built-up areas for meadow  
345 sites (BMH meadow), caused small to negligible increases of floristic variance. Moreover, the  
346 high, yet insignificant relationship of the variances at the distance 250-600 m (Fig. 6a)  
347 corresponds to the distance of the second range of fit between richness (Fig. 4a).

348 The link between pollen and total floristic variance is underlined by the relationship between  
349 the amount of variance contributed by individual sites and the total variance. Indeed, these  
350 amounts in pollen and floristic data are significantly correlated. Distances of the high  
351 correlation of local contribution to total variance are related to the source area of pollen  
352 richness of individual datasets. In WCM forest and BMH forest, the increase of correlation of  
353 local contribution to beta diversity usually follows (or precedes) the richness correlation by a  
354 single ring (Fig. 4).

355 Beta diversity understood as directional turnover (temporal or spatial) belongs to more  
356 frequent and widespread measurements in pollen analysis (Figueroa-Rangel et al. 2010,  
357 Connor et al. 2019) than beta diversity as a non-directional variation. Indeed, the temporal  
358 dynamics is the main objective of pollen analysis. According to Nieto-Lugilde et al. (2015)  
359 pollen-based turnover correlates with forest-inventory-based turnover. We extend this finding  
360 from woody taxa to all species and from directional turnover to non-directional variance.  
361 Moreover, forest sites with high contributions to pollen beta diversity also show a high  
362 contribution to floristic beta diversity (Fig. 4b).

### 363 **Strengths and limitations**

364 We applied the presence-absence transformation of abundances of pollen data instead of  
365 widely used pollen proportions (Birks et al. 2016). This transformation may increase the  
366 effect of very rare taxa, sometimes appearing in a single grain due to long-distance transport.  
367 While we admit that our datasets contain such pollen components (*Ambrosia artemisiifolia*-  
368 type, *Ostrya*, *Castanea*), their presence was perhaps balanced by rare ornamental plants  
369 recorded in the plant survey, which were missing in the pollen record. In this light, the  
370 significance of the relationship between pollen-based beta diversity and floristic data is a  
371 remarkable result of this study and provides a rigorous reference for pollen-based  
372 macroecological studies (e.g. Šizling et al. 2016).

373 Disentangling the causal connection in any pollen–vegetation relationship requires the use of  
374 pollen–plant translation tables (Birks et al. 2016) and the comparison of pollen and vegetation  
375 patterns by individual taxa. Here, we report metrics inferred from all-species datasets.  
376 Without relying on any particular pollen dispersal model, we assumed that the correlation of  
377 pollen- and floristic-based metrics is caused by pollen transport. Studies reconstructing  
378 vegetation from pollen usually assume that the over-canopy component is the major  
379 component of pollen transport and do not consider the local herb layer in the forest (Prentice  
380 1985). However, we identified a positive correlation between pollen and floristic richness at  
381 70 m in BMH forest, matching the distance at which many open-habitat species, confined to  
382 forest roads and other openings, first appeared. If the plant individuals recorded in these  
383 habitats are direct sources of the recorded pollen, then pollen transported through the forest  
384 interior also has to be considered (Tauber 1967). Our data also indicated, however, that forest  
385 serves as a barrier against pollen transport. While many new plant species recorded in the  
386 built-up areas in the BMH region caused a decrease of adjusted  $R^2$  at 500 m in forest habitats,  
387 the opposite pattern was recorded in meadow sites in the same region. Open sites of WCM  
388 meadows also show a relationship between dispersal abilities of pollen and the source area of  
389 pollen richness. The distance of 1.5 m in the herb plants-dominated WCM meadows  
390 corresponds to the taxon-specific source area of pollen for herbs. In this regard, insect-

391 pollinated strategy with low pollen productivity makes source areas of individual plants very  
392 small (Shaw and Whyte 2020).

393 We had to consider the trade-off between data quality and the feasibility of the vegetation  
394 survey. For the range of radii 10–100 m, we mapped vegetation in the field with the help of  
395 aerial photographs; however, radial relevés for each ring might have been better (Shaw and  
396 Whyte 2020). In the case of homogeneous vegetation, our method produced a single or a few  
397 large polygons, and thus most species appeared “in waves” along the transects (from 10 to 40  
398 m in WCM meadows). Heterogeneous vegetation resulted in higher numbers of polygons and  
399 species appeared gradually along the transect there (BMH forest, Fig. 3).

400 Our survey methodology concentrates sampling effort around central points, which then  
401 decreases with increasing radius, though not smoothly. Change from polygon to transect  
402 method at 100 m is noticeable in the trend of floristic richness in the BMH region, which  
403 slightly slowed down beyond 100 m. Our survey methodology, emphasising a ‘pollen  
404 perspective’ of vegetation, may be contrasted with independent data sources, to illustrate the  
405 completeness of our data. The most complete database of Czech flora PLADIAS includes  
406 information on 1,477 species in 15 mapping squares covered by our survey in the BMH  
407 region and 2,045 species in 14 squares in the WCM region (Wild et al. 2019). This means that  
408 in both regions, we recorded close to 54% of the known regional species pool. This is quite  
409 good result, considering the incomplete spatial coverage of our survey. Moreover, the close  
410 agreement between the two regions speaks for consistency in data quality between the  
411 datasets.

## 412 **Conclusions**

413 There was a consistent positive relationship between pollen and floristic richness in both open  
414 and forest habitats in the two study regions. The strongest correlations mostly appeared at two  
415 spatial scales: within tens and hundreds of metres. These distances are controlled by numbers  
416 of newly appearing species along the transects, linked to landscape structure and the mean of  
417 the pollen transport. However, causality between the appearance of species and pollen taxa at  
418 individual sites remains tangled.

419 Our results support the hypothesis that a higher variance of datasets covering both open and  
420 forest habitats may contribute to a higher correlation within the forest and open-habitat  
421 subsets. Regarding the application value of these results for the interpretation of fossil  
422 records, we suggest that pollen richness reconstructions from sites with different openness or  
423 richness should be compared, since the non-significant relationships recovered in both forest  
424 datasets indicates some limitations of the method.

425 Total variation was already used as a measure of spatial beta diversity in paleoecology  
426 (Winegardner et al. 2017). We encourage its broader use in palynology and macroecology.  
427 Our field data prove that pollen-based beta diversity has a significant relationship to floristic  
428 beta diversity. The source area of pollen variance appears to be dataset-specific. Our results,  
429 pointing close to 150 m, represent the mean spatial reference for pollen-based beta diversity  
430 inferred from moss polsters at open, forest, species-poor, and species-rich sites. This result is  
431 also underlined by a significant correlation between pollen and floristic indices of local  
432 contributions of sites to the variation in both forest datasets. Decomposition of beta diversity  
433 on the level of individual sites or species is one of the advantages of the total variance  
434 (Legendre and De Cáceres 2013) and it has already been shown that it may provide additional  
435 insights to the fossil datasets (Connor et al. 2019).

436 **References**

- 437 Abraham, V. and Kozáková, R. 2012. Relative pollen productivity estimates in the modern  
438 agricultural landscape of Central Bohemia (Czech Republic). - *Rev. Palaeobot.*  
439 *Palynol.* 179: 1–12.
- 440 Andersen, S. T. 1970. The relative pollen productivity and representation of north European  
441 trees, and correction factors for tree pollen spectra. - *Dan. Geol. Unders. Række II* 96:  
442 1–99.
- 443 Anderson, M. J. et al. 2011. Navigating the multiple meanings of  $\beta$  diversity: a roadmap for  
444 the practicing ecologist: Roadmap for beta diversity. - *Ecol. Lett.* 14: 19–28.
- 445 Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. -  
446 *Glob. Ecol. Biogeogr.* 19: 134–143.
- 447 Beug, H.-J. 2004. Leitfaden der Pollenbestimmung in Mitteleuropa und angrenzenden  
448 Gebiete. - Verlag Dr. Friedrich Pfeil.
- 449 Birks, H. J. B. 1973. Past and present vegetation of the Isle of Skye: A palaeoecological  
450 study. - Cambridge University Press.
- 451 Birks, H. J. B. 2019. Contributions of Quaternary botany to modern ecology and  
452 biogeography. - *Plant Ecol. Divers.* 12: 189–385.
- 453 Birks, H. J. B. et al. 2016. Does pollen-assembly richness reflect floristic richness? A  
454 review of recent developments and future challenges. - *Rev. Palaeobot. Palynol.* 228:  
455 1–25.
- 456 Blaus, A. et al. 2020. Modern Pollen–Plant Diversity Relationships Inform Palaeoecological  
457 Reconstructions of Functional and Phylogenetic Diversity in Calcareous Fens. - *Front.*  
458 *Ecol. Evol.* in press.
- 459 Broström, A. et al. 2005. Estimating the spatial scale of pollen dispersal in the cultural  
460 landscape of southern Sweden. - *The Holocene* 15: 252–262.
- 461 Bunting, M. J. 2003. Pollen-vegetation relationships in non-arboreal moorland taxa. - *Rev.*  
462 *Palaeobot. Palynol.* 125: 285–298.
- 463 Bunting, M. J. et al. 2004. Vegetation structure and pollen source area. - *The Holocene* 14:  
464 651–660.
- 465 Bunting, M. J. et al. 2013. Palynological perspectives on vegetation survey: a critical step for  
466 model-based reconstruction of Quaternary land cover. - *Quat. Sci. Rev.* 82: 41–55.
- 467 Calcote, R. 1995. Pollen Source Area and Pollen Productivity: Evidence from Forest Hollows.  
468 - *J. Ecol.* 83: 591–602.
- 469 Chytrý, M. 2012. Vegetation of the Czech Republic: diversity, ecology, history and dynamics.  
470 - *Preslia* 84: 427–504.

- 471 Connor, S. E. et al. 2019. Humans take control of fire-driven diversity changes in  
472 Mediterranean Iberia's vegetation during the mid-late Holocene. - *The Holocene* 29:  
473 886–901.
- 474 Davis, M. B. 1963. On the theory of pollen analysis. - *Am J Sci* 261: 897–912.
- 475 Dray, S. et al. 2020. *adespatial: Multivariate Multiscale Spatial Analysis*.
- 476 Felde, V. A. et al. 2015. Modern pollen-plant richness and diversity relationships exist along  
477 a vegetational gradient in southern Norway. - *The Holocene*: 0959683615596843.
- 478 Figueroa-Rangel, B. L. et al. 2010. Cloud forest dynamics in the Mexican neotropics during  
479 the last 1300 years. - *Glob. Change Biol.* 16: 1689–1704.
- 480 Hájková, P. et al. 2011. Prehistoric origin of the extremely species-rich semi-dry grasslands in  
481 the Bílé Karpaty Mts (Czech Republic and Slovakia). - *Preslia* 83: 185–204.
- 482 Hellman, S. et al. 2009a. Relevant Source Area of Pollen in patchy cultural landscapes and  
483 signals of anthropogenic landscape disturbance in the pollen record: A simulation  
484 approach. - *Rev. Palaeobot. Palynol.* 153: 245–258.
- 485 Hellman, S. et al. 2009b. Estimating the Relevant Source Area of Pollen in the past cultural  
486 landscapes of southern Sweden -- A forward modelling approach. - *Rev. Palaeobot.*  
487 *Palynol.* 153: 259–271.
- 488 Janssen, C. R. 1973. Local and regional pollen deposititon. - In: Birks, H. J. B. and West, R.  
489 G. (eds), *Quaternary Plant Ecology*. 1st edition. Halsted Press, pp. 31–43.
- 490 Kuneš, P. et al. 2019. Relative pollen productivity estimates for vegetation reconstruction in  
491 central-eastern Europe inferred at local and regional scales. - *The Holocene* 29: 1708–  
492 1719.
- 493 Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data:  
494 dissimilarity coefficients and partitioning. - *Ecol. Lett.* 16: 951–963.
- 495 Matthias, I. et al. 2015. Pollen diversity captures landscape structure and diversity. - *J. Ecol.*  
496 103: 880–890.
- 497 Meltsov, V. et al. 2011. Palynological richness and pollen sample evenness in relation to local  
498 floristic diversity in southern Estonia. - *Rev. Palaeobot. Palynol.* 166: 344–351.
- 499 Meltsov, V. et al. 2013. The role of landscape structure in determining palynological and  
500 floristic richness. - *Veg. Hist. Archaeobotany* 22: 39–49.
- 501 Michalcová, D. et al. 2014. High Plant Diversity of Grasslands in a Landscape Context: A  
502 Comparison of Contrasting Regions in Central Europe. - *Folia Geobot.* 49: 117–135.
- 503 Nieto-Lugilde, D. et al. 2015. Close agreement between pollen-based and forest inventory-  
504 based models of vegetation turnover. - *Glob. Ecol. Biogeogr.* 24: 905–916.
- 505 Odgaard, B. V. 1999. Fossil pollen as a record of past biodiversity. - *J. Biogeogr.* 26: 7–17.

- 506 Prentice, I. C. 1985. Pollen representation, source area, and basin size: Toward a unified  
507 theory of pollen analysis. - *Quat. Res.* 23: 76–86.
- 508 R Development Core Team 2017. R: A language and environment for statistical computing. -  
509 R Foundation for Statistical Computing.
- 510 Rasser, M. W. et al. 2008. Paleogene and Neogene. - *Geol. Cent. Eur. Vol. 2 Mesoz.*  
511 *Cenozoic*: 1031–1139.
- 512 Reitalu, T. et al. 2019. Patterns of modern pollen and plant richness across northern Europe. -  
513 *J. Ecol.* 107: 1662–1677.
- 514 Roleček, J. et al. 2014. Understanding the extreme species richness of semi-dry grasslands in  
515 east-central Europe: a comparative approach. - *Preslia* 86: 13–34.
- 516 Roleček, J. et al. 2020. Conservation targets from the perspective of a palaeoecological  
517 reconstruction: the case study of Dářko peat bog in the Czech Republic. - *Preslia* 92:  
518 87–114.
- 519 Shaw, H. and Whyte, I. 2020. Interpretation of the herbaceous pollen spectra in  
520 paleoecological reconstructions: A spatial extension of Indices of Association and  
521 determination of individual pollen source areas from binary data. - *Rev. Palaeobot.*  
522 *Palynol.* 279: 104238.
- 523 Šizling, A. L. et al. 2016. Can people change the ecological rules that appear general across  
524 space? - *Glob. Ecol. Biogeogr.* 25: 1072–1084.
- 525 Sugita, S. 1994. Pollen Representation of Vegetation in Quaternary Sediments: Theory and  
526 Method in Patchy Vegetation. - *J. Ecol.* 82: 881–897.
- 527 Sugita, S. 2007. Theory of quantitative reconstruction of vegetation II: all you need is LOVE.  
528 - *Holocene* 17: 243–257.
- 529 Tauber, H. 1967. Investigations of the mode of pollen transfer in forested areas. - *Rev.*  
530 *Palaeobot. Palynol.* 3: 277–286.
- 531 Theuerkauf, M. and Couwenberg, J. 2017. The extended downscaling approach: A new R-tool  
532 for pollen-based reconstruction of vegetation patterns. - *The Holocene* 27: 1252–1258.
- 533 Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. - *Ecol.*  
534 *Monogr.* 30: 279–338.
- 535 Wild, J. et al. 2019. Plant distribution data for the Czech Republic integrated in the Pladias  
536 database. - *Preslia* 91: 1–24.
- 537 Wilson, J. B. et al. 2012. Plant species richness: the world records. - *J. Veg. Sci.* 23: 796–802.
- 538 Winegardner, A. K. et al. 2017. Diatom diversity patterns over the past *c.* 150 years across  
539 the conterminous United States of America: Identifying mechanisms behind beta  
540 diversity. - *Glob. Ecol. Biogeogr.* 26: 1303–1315.

541

## 542 **Figures**

543

544

545 Fig. 1: Map of the study areas showing a) position within Central Europe, b) BMH:  
546 Bohemian-Moravian Highlands, c) WCM: White Carpathians. Yellow and green circles  
547 indicate sites in meadow and forest, respectively. Blue squares show the area of the reference  
548 floristic data from the PLADIAS database (Wild et al. 2019). Grey indicates forested area.  
549 Short lines represent transects of the vegetation survey; d) circle 10–100 m and two transects  
550 of polygons recording the floristic diversity 100–1000 m, e) 21 plots within 0–10 m.

551

552 Fig. 2: Pollen alpha diversity (pollen richness, y-axis) and beta diversity (BD<sub>total</sub>, x-axis) in  
553 two study regions and their different habitats. Meadows (yellow), forest (green), and both  
554 habitats together (black).

555

556 Fig. 3: Spatial scaling of floristic alpha diversity (floristic richness) and beta diversity  
557 (BD<sub>total</sub>) in two study regions and their different habitats. The mean number of plant species  
558 appearing in the vegetation survey (top) and their total variance (bottom). Meadows are  
559 indicated by the yellow line, forest by the green line, and both habitats together by the black  
560 line.

561

562 Fig. 4: The strength of linear regression between a) pollen richness and floristic richness at  
563 different distances from sampling sites and b) local contributions of sites to pollen and  
564 floristic BD<sub>total</sub> at different distances from sampling sites. The black line shows the  
565 correlation for all sites, the orange line for meadow sites, and the green line for forest sites.  
566 Red dots indicate significant correlations. For more detail, see Fig. A2 and A3.

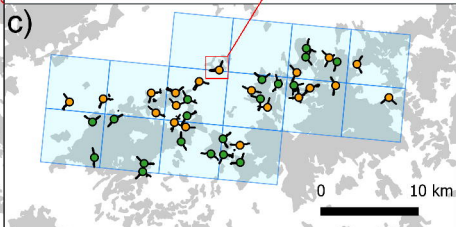
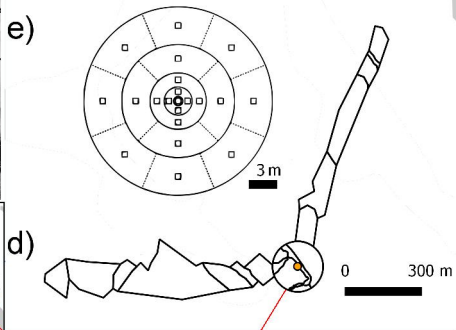
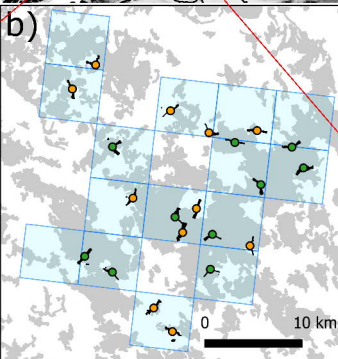
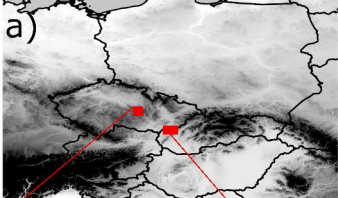
567

568

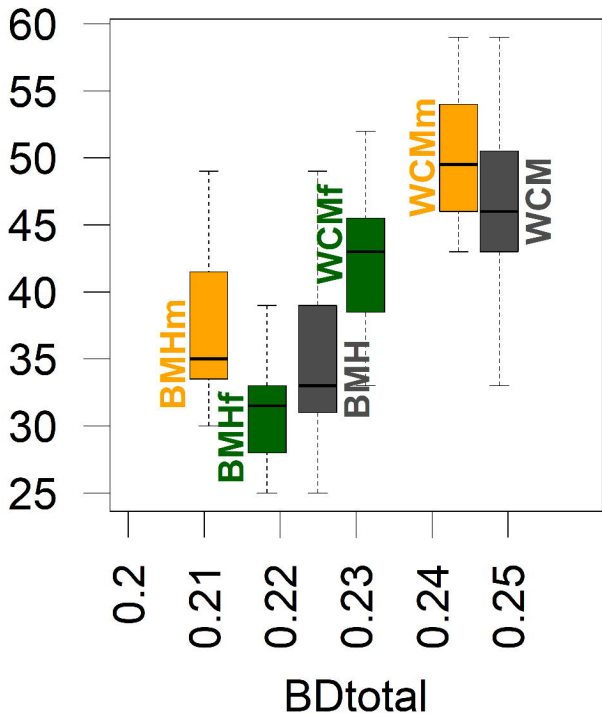
569 Fig. 5: Number of new species recorded with increasing distance in different study regions  
570 and habitats, coloured according to source habitats.

571

572 Fig. 6: Linear regression between pollen and floristic BD<sub>total</sub>: a) adjusted R<sup>2</sup> at different  
573 distances from sampling sites, red dots indicate significant correlations, and b) scatter plot of  
574 six datasets (text labels) for the distance of 150 m. Empty dots indicate floristic BD<sub>total</sub> at 6  
575 m of WCM meadows.

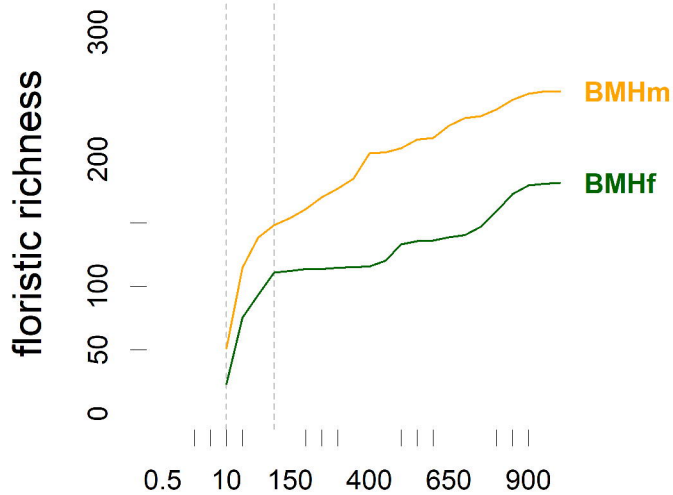


pollen richness

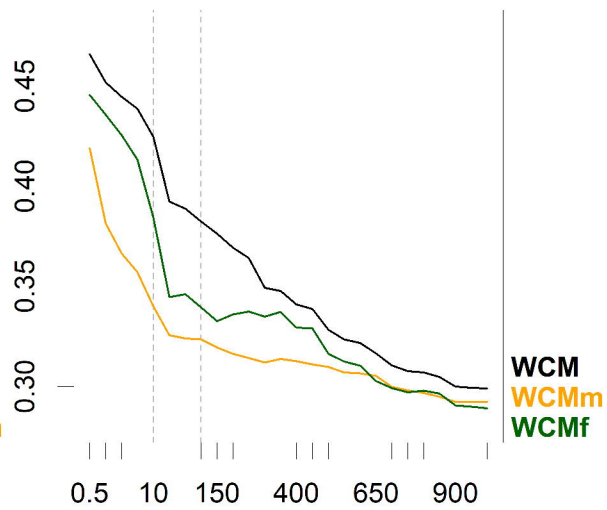
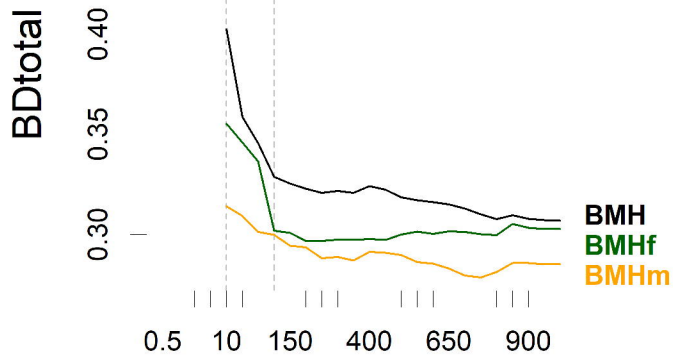
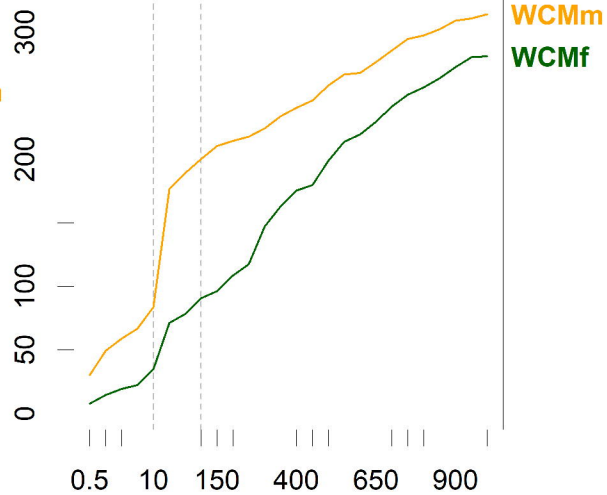




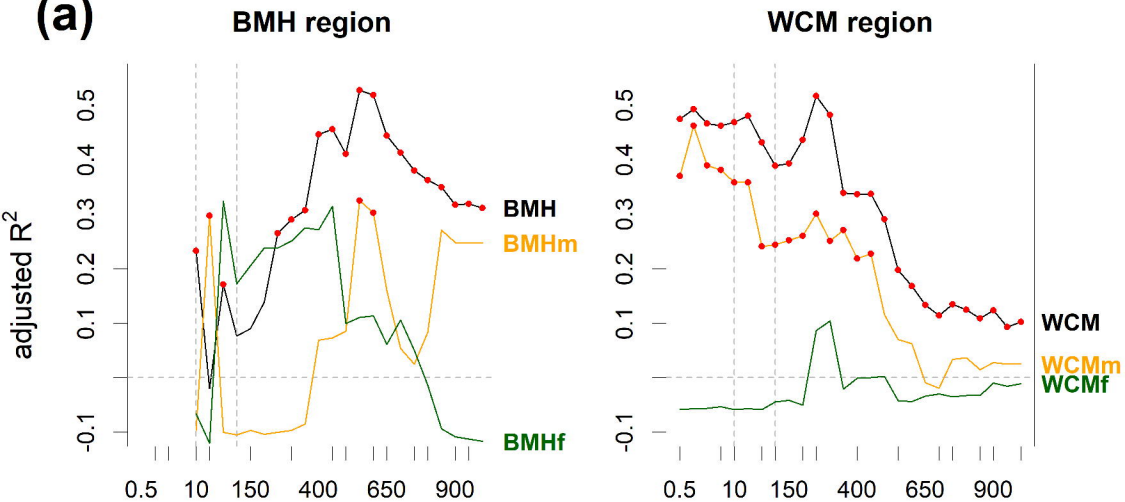
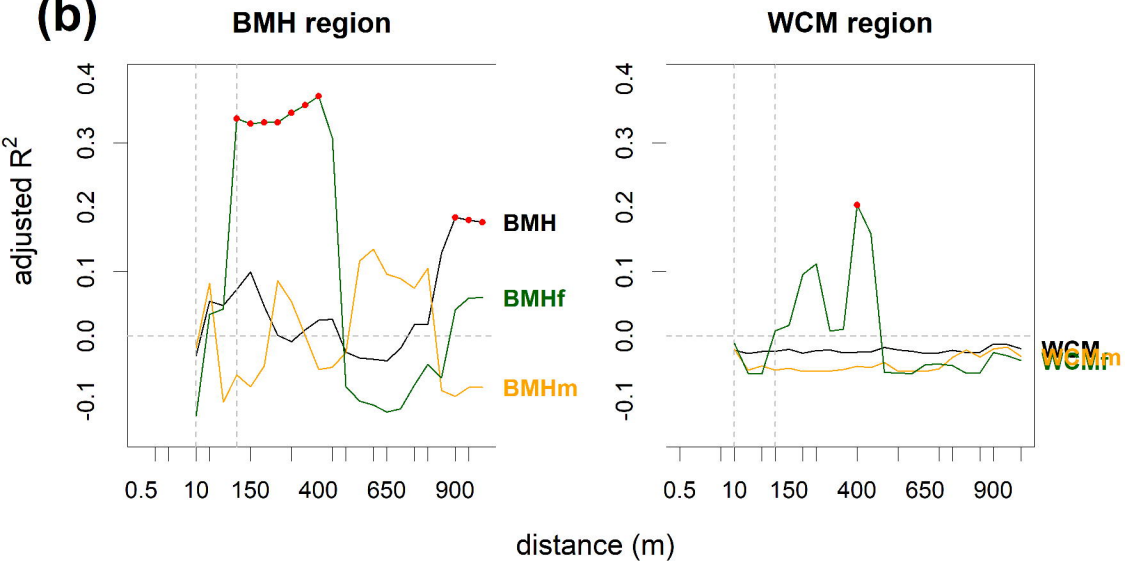
## BMH region



## WCM region

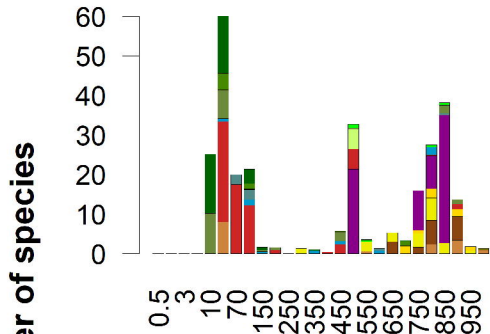


distance (m)

**(a)****(b)**

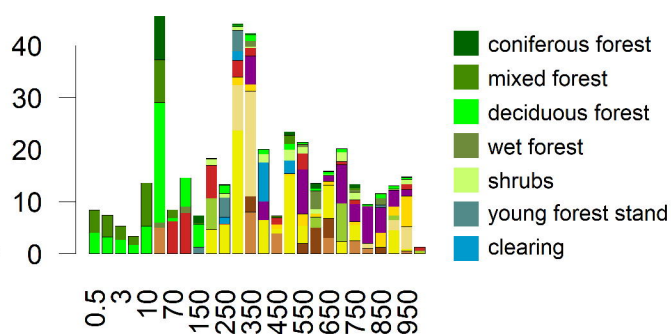
## BMH region

### forest



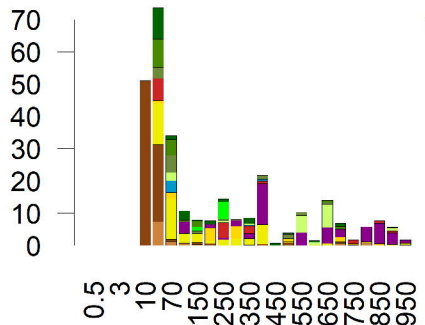
## WCM region

### forest

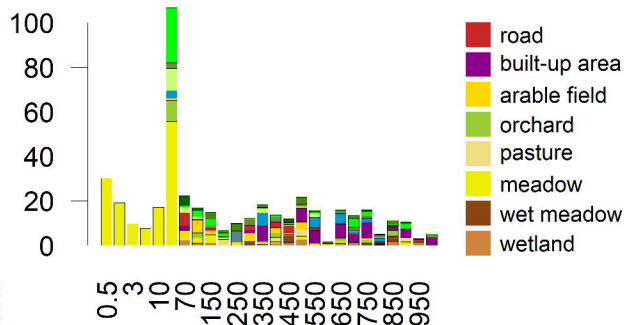


number of species

### meadow



### meadow



distance (m)

