CrossMark

ORIGINAL ARTICLE

Are pollen records from small sites appropriate for REVEALS model-based quantitative reconstructions of past regional vegetation? An empirical test in southern Sweden

Anna-Kari Trondman[1] · Marie-José Gaillard[1] · Shinya Sugita[2] · Leif Björkman[3] ·
Annica Greisman[1] · Tove Hultberg[4] · Per Lagerås[5] · Matts Lindbladh[4] ·
Florence Mazier[6]

Abstract In this paper we test the performance of the Regional Estimates of VEgetation Abundance from Large Sites (REVEALS) model using pollen records from multiple small sites. We use Holocene pollen records from large and small sites in southern Sweden to identify what is/are the most significant variable(s) affecting the REVEALS-based reconstructions, i.e. type of site (lakes and/or bogs), number of sites, site size, site location in relation to vegetation zones, and/or distance between small sites and large sites. To achieve this objective we grouped the small sites according to (i) the two major modern vegetation zones of the study region, and (ii) the distance between the small sites and large lakes, i.e. small sites within 50, 100, 150, or 200 km of the large lakes. The REVEALS-based reconstructions were performed using 24 pollen taxa. Redundancy analysis was performed on the results from all REVEALS-model runs using the groups within (i) and (ii) separately, and on the results from all runs using the groups within (ii) together. The explanatory power and significance of the variables were identified using forward selection and Monte Carlo permutation tests. The results show that (a) although the REVEALS model was designed for pollen data from large lakes, it also performs well with pollen data from multiple small sites in reconstructing the percentage cover of groups of plant taxa (e.g. open land taxa, summer-green trees, evergreen trees) or individual plant taxa; however, in the case of this study area, the reconstruction of the percentage cover of *Calluna vulgaris*, Cyperaceae, and *Betula* may be problematic when using small bogs; (b) standard errors of multiple small-site REVEALS estimates will generally be larger than those obtained using pollen records from large lakes, and they will decrease with increasing size of pollen counts and increasing number of small sites; (c) small lakes are better to use than small bogs if the total number of small sites is low; and (d) the size of small sites and the distance between them do not play a major role, but the distance between the small sites and landscape/vegetation boundaries is a determinant factor for the accuracy of the vegetation reconstructions.

Communicated by F. Bittmann.

✉ Marie-José Gaillard
marie-jose.gaillard-lemdahl@lnu.se

1 Department of Biology and Environmental Science, Linnaeus University, 391082 Kalmar, Sweden

2 Institute of Ecology, Tallinn University, Tallinn, Estonia

3 Viscum pollenanalys & miljöhistoria, c/o Leif Björkman, Bodavägen 16, 571 42 Nässjö, Sweden

4 Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences, Alnarp, Sweden

5 Swedish National Heritage Board, Lund, Sweden

6 CNRS GEODE UMR 5602, Toulouse University Le Mirail, Toulouse, France

Introduction

The landscape reconstruction algorithm (LRA) and its two models—Regional Estimates of VEgetation Abundance from Large Sites (REVEALS) (Sugita 2007a) and LOcal Vegetation Estimates (LOVE) (Sugita 2007b)—is an important step forward in the field of palaeoecology. Pollen-
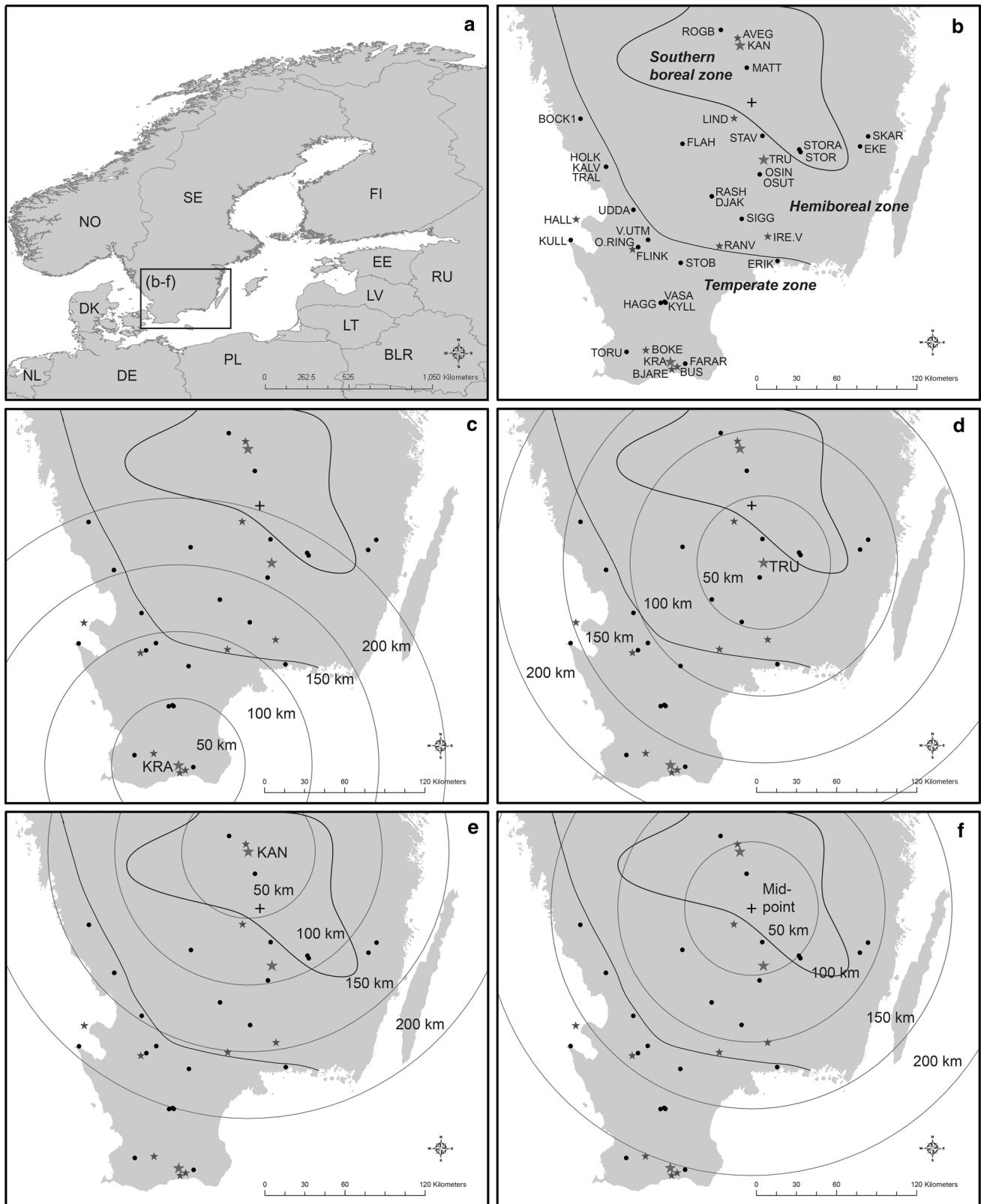
Fig. 1 Study area and location of all study sites for REVEALS RUNS 1–5 (see text for more explanations); **a** location of the study area in NW Europe; **b** location of all study sites and boundaries between vegetation zones (see Table 1 for metadata of the study sites) (RUNS 1); **c–f**: sites located within 50, 100, 150, and 200 km from the large lakes Krageholmsjön (**c**, RUNS 2), Trummen (**d**, RUNS 3), Kansjön (**e**, RUNS 4) and the midpoint on a segment joining the centres of lakes Trummen and Kansjön (**f**, RUNS 5)

Table 1 Metadata for all sites

Vegetation zone	Site name	Site code	Lat DD	Long DD	Elevation (m a.s.l.)	Site type	Radius (m)	TW 1	TW 2	TW 3	TW 4	TW 5
Hemiboreal	Avegöl	AVEG	57.6833	14.5	300	Lake	98	98	98	98	98	98
Hemiboreal	DjäknabyggdTUtmark	DJAK	56.6167	14.2	145	Bog	5	5	5	5	5	5
Hemiboreal	Ekenäs	EKE	56.95	16.0167	90	Bog	20	20	20	20	20	
Hemiboreal	Flahult	FLAH	56.9667	13.8333	188	Bog	16	16	16	16	16	
Hemiboreal	Ire: Västargylet	IRE.V	56.35	14.8833	100	Lake	69	69	69	69	69	69
Hemiboreal	**Kansjön**	**KAN**	57.6333	14.5333	308	**Lake**	505	505	505		505	505
Hemiboreal	Lindhultsgöl	LIND	57.1436	14.4661	200	Lake	150	150				
Hemiboreal	Mattarp	MATT	57.4833	14.6167	328	Bog	25	25	25	25	25	25
Hemiboreal	OsabyTInmark	OSIN	56.7667	14.7833	165	Bog	56	56	56	56	56	
Hemiboreal	OsabyTUtmark	OSUT	56.7667	14.7833	165	Bog	11	11	11	11	11	11
Hemiboreal	Ranviken	RANV	56.2833	14.3	81	Lake	138			138	138	138
Hemiboreal	RåshultTInmark	RASH	56.6167	14.2	145	Bog	16	16	16	16	16	
Hemiboreal	Rogberga	ROGB	57.7361	14.2886	228	Bog	40	40	40	40	40	40
Hemiboreal	Siggaboda	SIGG	56.4667	14.5667	148	Bog	10	10	10	10	10	
Hemiboreal	Skärsgölarna	SKAR	57.0167	16.1167	75	Bog	22	22	22	22	22	
Hemiboreal	Stavsåkra	STAV	57.0242	14.8131	187	Bog	80	80	80	80	80	80
Hemiboreal	Storasjö (a)	STOR	56.9167	15.2833	250	Bog	28	28	28	28		
Hemiboreal	Storasjö (b)	STORA	56.9333	15.2667	255	Bog	89	89	89	89	89	89
Hemiboreal	**Trummen**	**TRU**	56.8667	14.8333	161	**Lake**	492	492	492	492	492	492
Temperate	Bjäresjösjön	BJARE	55.45	13.75	50	Lake	80	80	80	80	80	
Temperate	Bocksten BP1	BOCK1	57.1167	12.5667	110	Bog	25	25	25	25		
Temperate	Bökesjön	BOKE	55.5756	13.4375	60	Lake	80	80	80	80	80	
Temperate	Bussjösjön	BUS	55.4667	13.8167	24	Lake	69	69	69	69	69	69
Temperate	Eriksberg	ERIK	56.1833	15	20	Bog	6	6	6	6	6	6
Temperate	Fårarpsmosse	FARAR	55.4883	13.9089	25	Bog	98		B: 98	L: 98	L: 169	L: 219
Temperate	Flinkasjön	FLINK	56.25	13.25	83	Lake	98	98	98	98	98	98
Temperate	Fulltofta: Häggenäs	HAGG	55.8936	13.6031	70	Bog	18	18	18	18		
Temperate	Fulltofta: Kyllingahus	KYLL	55.8953	13.6608	145	Bog	16	16	16	16	16	16
Temperate	Fulltofta: Vasahus	VASA	55.9022	13.6486	120	Bog	18	18	18	18		
Temperate	Hälledammen	HALL	56.44	12.5561	10	Lake	28		28	28	28	
Temperate	Holkåsen	HOLK	56.8	12.9	140	Bog	13	13	13	13	13	
Temperate	Kalvaberget	KALV	56.8	12.9	100	Bog	35	35	35	35		
Temperate	**Krageholmssjön**	**KRA**	55.5	13.7333	43	**Lake**	837	837	837	837	837	837
Temperate	Kullaberg	KULL	56.3	12.5	125	Bog	6	6	6	6	6	
Temperate	Östra Ringarp	O.RING	56.2667	13.3167	100	Bog	150	B: 150	L: 150	L: 150	L: 150	L: 150
Temperate	Stoby	STOB	56.1667	13.8333	35	Bog	80	80	80	80	80	80
Temperate	Torup	TORU	55.56	13.21	50	Bog	10	10	10	10	10	10
Temperate	Trälhultet	TRAL	56.8	12.9	118	Bog	25	25	25	25	25	
Temperate	Uddared	UDDA	56.5167	13.25	73	Bog	69	69	69	69	69	
Temperate	Värsjö Utmark	V.UTM	56.3167	13.4333	122	Bog	80	80	80	80		

Grey areas indicate lack of data for a specific site and time window. In the right hand five columns under "Time Window" the presence of a number indicates a time window with pollen data available and represents the site radius in metres. For two sites (Fårarpsmosse and Östra Ringarp) L corresponds to lake phase and B to bog phase and the values are the radius (in m) of the sites for the different time windows. Otherwise, the modern radius of the site (column labelled "site radius") was used for the five time windows. In bold: large lakes

based quantitative reconstructions of past vegetation at both regional and local spatial scales are necessary to answer palaeoecological questions for which the size of the area covered by a species/taxon or a vegetation unit plays a role. Such questions are for instance the effect of anthropogenic land use on climate or biodiversity in the past, or hypotheses related to the history of resource management by humans. REVEALS- and LOVE-based reconstructions

of past vegetation composition have already been shown to be useful for (i) the quantitative reconstruction of vegetation/plant cover at the local spatial scale (10–100 km^2) using Sugita's LRA (e.g. Nielsen and Odgaard 2010; Cui et al. 2013, 2014; Mazier et al. 2015; Hultberg et al. 2015), (ii) the study of regional spatio-temporal land-cover/landscape dynamics over the past millennia (e.g. Marquer et al. 2014; Fyfe et al. 2013; Nielsen et al. 2012; Trondman 2014), (iii) the evaluation of anthropogenic land-cover change scenarios (ALCCs) (Gaillard et al. 2010) and (iv) the study of land cover-climate interactions in the past (Strandberg et al. 2014). Recently, Pirzamanbein et al. (2014) developed a set of statistical models to create spatially continuous maps of past land cover by combining (i) pollen-based REVEALS "point estimates" of past land cover and (ii) spatially continuous estimates of past land cover, obtained by merging simulated potential vegetation (using the LPJ-GUESS model; Smith et al. 2001) with an anthropogenic land-cover change scenario (KK10; Kaplan et al. 2009). These maps (or datasets) of past land-cover can now potentially be used as alternative land-cover descriptions in climate modelling.

The REVEALS model was developed to reconstruct past vegetation composition in percentage cover at a scale of 100 by 100 km using pollen data from large lakes. It has been tested with numerous simulations (Sugita 2007a) and empirical studies in southern Sweden (Hellman et al. 2008a, b), on the Swiss Plateau (Soepboer et al. 2010), in Denmark (Nielsen et al. 2012) and in northern America (N Michigan and NW Wisconsin; Sugita et al. 2010). All the empirical studies except one—NW Wisconsin where pollen records from small forest hollows were the only data available—used pollen data from large lakes. Sugita (2007a) showed through simulations that, theoretically, pollen data from multiple small lakes within a region can be used instead of one or a few large lakes to estimate regional vegetation composition using REVEALS, but the error estimates on the results will generally be much larger than when pollen records from large lakes are used. Pollen data from small sites are better suited for vegetation reconstructions at the local spatial scale because they exhibit between-site differences within a mosaic landscape (Sugita 2007b). The variability of pollen assemblages from small sites is often high, which explains the large error estimates on REVEALS results based on pollen data from small sites. Empirical tests (Mazier et al. 2012; Fyfe et al. 2013) using pollen data from multiple small sites (lakes and/or bogs) support the theory, i.e. the results from Sugita simulations (2007a). However, these empirical tests are based on relatively few sites. Systematic tests of the REVEALS model using pollen data from small and large bogs have not been performed so far, but we know that such data will violate one of the assumptions of the model, namely that the surface of the deposition basin should not support

vegetation. Therefore, further testing of the performance of REVEALS with pollen data from small sites (lakes and bogs) is necessary not only to show that the REVEALS model works in practice as it does in simulations, but also to ensure that the application of REVEALS on pollen data from multiple small sites (bogs in particular) is a valid alternative to applications on pollen data from one or a few large lakes. The latter is particularly important for the regions of the world where large lakes are rare or absent. As argued in Sugita et al. (2010), we need to improve our understanding about the various factors that may affect the accuracy of the model, such as basin size and number of sites, and spatial complexity of vegetation patterns or gradients in dominant vegetation. The spatial scale of the REVEALS vegetation reconstructions is assumed to be ca. 100 by 100 km, based on the study of Hellman et al. (2008b) in southern Sweden. More tests of this kind are needed to evaluate whether this is true for other regions of Europe and the world.

In this paper we use Holocene pollen records from large lakes and small sites (lakes and bogs) in southern Sweden and numerical analyses to identify what is/are the most significant variable(s) affecting the REVEALS reconstructions based on

Table 2 Fall speed of pollen (FSP) and pollen productivity estimates (PPEs) for 25 plant taxa and the three corresponding land-cover types (LCTs)

Land-cover type (LCT)	Plant taxa/pollen-type	FSP (m/s)	PPE
Evergreen tree canopy (ET)	*Abies*	0.120	6.88 (1.44)
	Picea	0.056	2.62 (0.12)
	Pinus	0.031	6.38 (0.45)
	Juniperus	0.016	2.07 (0.04)
Summergreen tree canopy (ST)	*Alnus*	0.021	9.07 (0.10)
	Betula	0.024	3.09 (0.27)
	Corylus	0.025	1.99 (0.20)
	Fraxinus	0.022	1.03 (0.11)
	Quercus	0.035	5.83 (0.15)
	Carpinus	0.042	3.55 (0.43)
	Fagus	0.057	2.35 (0.11)
	Tilia	0.032	0.80 (0.03)
	Ulmus	0.032	1.27 (0.05)
	Salix	0.022	1.22 (0.11)
Open land (OL)	*Calluna vulgaris*	0.038	0.82 (0.02)
	Artemisia	0.025	3.48 (0.20)
	Cyperaceae	0.035	0.87 (0.06)
	Filipendula	0.006	2.81 (0.43)
	Poaceae	0.035	1.00 (0.00)
	Plantago lanceolata	0.029	1.04 (0.09)
	Plantago media	0.024	1.27 (0.18)
	Plantago montana	0.030	0.74 (0.13)
	Rumex acetosa-type	0.018	2.14 (0.28)
	Cerealia-type	0.060	1.85 (0.38)
	Secale cereale	0.060	3.02 (0.05)

pollen records from multiple small sites, i.e. site type, size, number, location in relation to vegetation zones, and/or distance between sites. The study region in southern Sweden (Fig. 1a) is well suited for the research objectives because of its clear division into two major vegetation zones (temperate and hemiboreal) due to differences in bedrock, soils, climate and land use, and the high number of existing pollen records from lakes and bogs. Moreover, pollen productivity estimates (PPEs)—a parameter required for the application of REVEALS—are available for the major plant taxa of the region (e.g. Broström et al. 2008; Mazier et al. 2012) and the REVEALS model was tested using modern pollen and vegetation data of that same region and was shown to perform satisfactorily (Hellman et al. 2008a, b).

Materials and methods

Study region and site selection

The study region covers southern Sweden up to 57.5°N (Fig. 1a). It is characterized by a major vegetation-zone boundary between the temperate and the hemiboreal zones (Ahti et al. 1968; Fig. 1b). Moreover, there is a small enclave of southern boreal vegetation (Ahti et al. 1968) within the hemiboreal zone (Fig. 1b). The temperate zone is characterized by deciduous trees with a dominance of *Fagus sylvatica* (beech) and/or *Quercus robur* (pendunculate oak). The hemiboreal zone is characterized by mixed forests of *Pinus sylvestris* (pine), *Picea abies* (spruce), and deciduous trees such as *Betula pubescens* and *B. pendula* (birch) and *Populus tremula* (aspen). Its southern boundary corresponds to the southern distribution limit of *P. abies* before the 20th century, and its northern boundary largely follows the northern distribution limit of *Q. robur*. Although the hemiboreal zone is more closely related to the boreal than the temperate zone, it has elements from both the coniferous forests of the boreal zone and the true deciduous forests of the temperate zone. The southern boreal zone is characterised by the prevalence of *Picea*, *Pinus* and *Betula*. There are some deciduous trees, but *Fagus* is very rare. Today, the hemiboreal and southern boreal zones of southern Sweden are characterized by modern forestry with plantations of *Picea* and *Pinus*. In the temperate zone, agricultural land represents ca. 30–40 % of the land, except in the southwestern part of the region where it is dominant (ca. 60–80 %).

The geology of the temperate zone is characterized by Cambro-Silurian and Cretaceous sedimentary bedrocks in

Table 3 "Extra fit" (i.e. explanatory power obtained by forward selection) and significance of environmental variables (tested by Monte Carlo permutations) for the RDA of all alternative runs (RUNS 1–5 separately and RUNS 2–5 together) for all five time windows

RUNS 1			RUNS 2			RUNS 3		
Variable	Extra fit, signif.		Variable	Extra fit, signif.		Variable	Extra fit, signif.	
Open land	0.4129	**	Open land	0.5444	**	Open land	0.4181	**
Number of sites	0.0730	**	Lake	0.0976	**	Number of sites	0.0867	**
Lake	0.0485	-	Bog	0.0899	**	Distance	0.0690	*
Mean basin size	0.0416	-	Number of sites	0.0669	*	Lake	0.0536	-
Bog	0.0386	-	Temperate zone	0.0614	*	Bog	0.0530	-
Lake+Bog	0.0079	-	Mixed veg. zone	0.0614	*	Mean basin size	0.0388	-
			Distance	0.0505	*	Hemiboreal zone	0.0365	-
			Mean basin size	0.0421	*	Mixed veg. zone	0.0365	-
			Lake+Bog	0.0033	-	Lake+Bog	0.0037	-
RUNS 4			**RUNS 5**			**RUNS 2·5**		
Variable	Extra fit, signif.		Variable	Extra fit, signif.		Variable	Extra fit, signif.	
Open land	0.4726	**	Open land	0.4841	**	Open land	0.3867	**
Number of sites	0.0791	**	Number of sites	0.0825	**	Number of sites	0.0607	**
Lake	0.0563	*	Lake	0.0528	*	Hemiboreal zone	0.0529	**
Distance	0.0516	*	Hemiboreal zone	0.0494	*	Temperate zone	0.0483	**
Bog	0.0492	*	Mixed veg. zone	0.0494	*	Lake	0.0401	**
Hemiboreal zone	0.0463	-	Bog	0.0472	*	Bog	0.0371	**
Mixed veg. zone	0.0463	*	Distance	0.0457	*	Mixed veg. zone	0.0269	**
Mean basin size	0.0290	-	Mean basin size	0.0272	-	Distance	0.0219	**
Lake+Bog	0.0019	-	Lake+Bog	0.0019	-	Mean basin size	0.0184	**
						Lake+Bog	0.0015	-

High and low values of "extra fit" correspond to high and low explanatory power of the specific variable. The significance of the variables in influencing the REVEALS estimates is indicated by * for P ≤ 0.05 and ** for P ≤ 0.01. Results that are not significant are indicated by –. Forward selection and the Monte Carlo permutation test were implemented by the computer program CANOCO 4.5 (Ter Braak and Šmilauer 2002)

the southwestern part, while Archaean gneiss and granite occur in the southeastern and northern parts of the zone. Silty-sandy till is dominant in the entire area, except in the southwestern part where clayey till is characteristic (Bergström 1988; Ekström 1946). The climate of the temperate zone is cold temperate and humid with mean January and July temperatures varying between −1 to −2 and 16–17 °C, respectively, and yearly precipitations between 500 and 600 mm over the years 2004–2013 (based on the climate data from the Swedish Meteorological and Hydrological Institute (SMHI)). Archaean granite and sandy-silty till are characteristic of the hemiboreal zone (altitudes below 200 m a.s.l.) (Daniel 1994) and the climate is slightly colder than in the temperate zone (mean January and July temperatures varying between −2 and −3 °C and between 15 and 16 °C, respectively over the years 2004–2013). The enclave of southern boreal vegetation is slightly different geomorphologically; it is characterized by a high richness in large boulders and gravelly till (Daniel 2002). Its climate is slightly colder (mean January and July temperatures of −3 and 15 °C, respectively) and more humid (mean annual precipitations of 680 mm) (mean values over the years 2004–2013) because of its altitudes >200 m a.s.l.

Sugita (2007a) defined "large lakes" as "lakes among which pollen assemblages are not statistically different within a given regional landscape". Hence, in theory, pollen records from any "large lake" can be used to reconstruct regional vegetation. Using simulations, Sugita (2007a) showed that pollen assemblages from lakes ≥48 ha were similar among sites in the specific landscape simulation used in that case. Sugita (2007a) also demonstrated using simulations that the minimum size of "large lakes" was positively correlated with the mean size of large vegetation patches. Therefore, "large lakes" should have a size similar to the largest vegetation patches in the region. "Small lakes" are defined as lakes among which pollen assemblages are statistically different within a given landscape. The maximum size of a small site depends on the type of regional vegetation, in particular the size and distribution of vegetation patches.
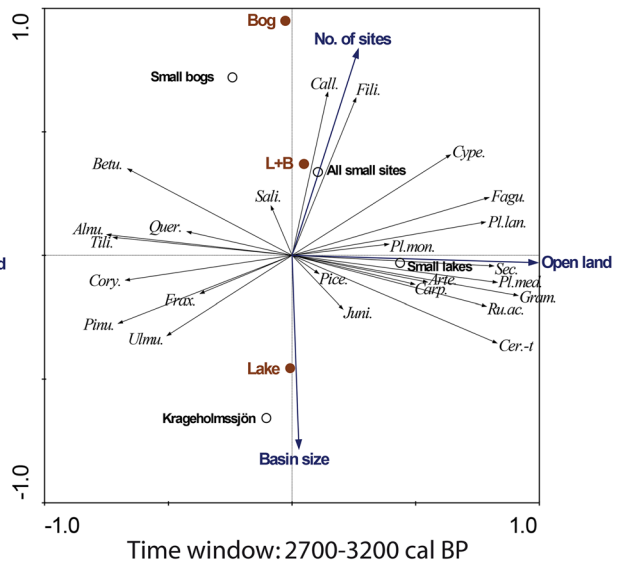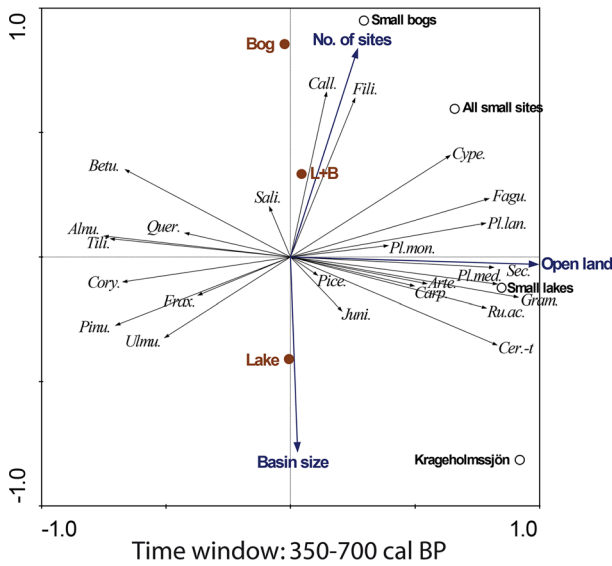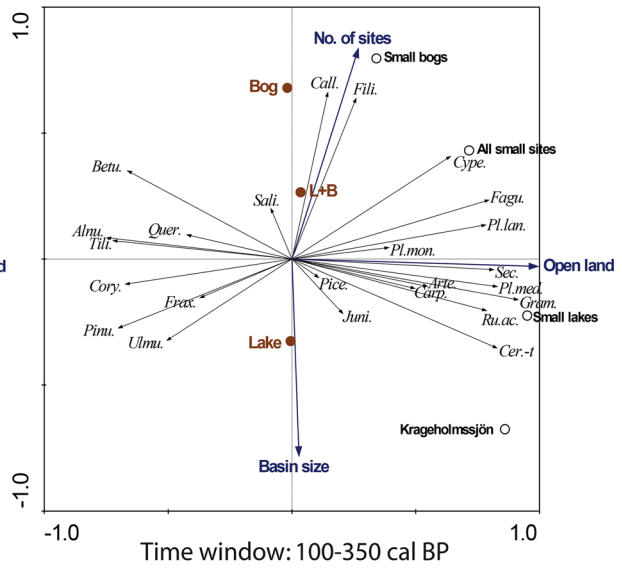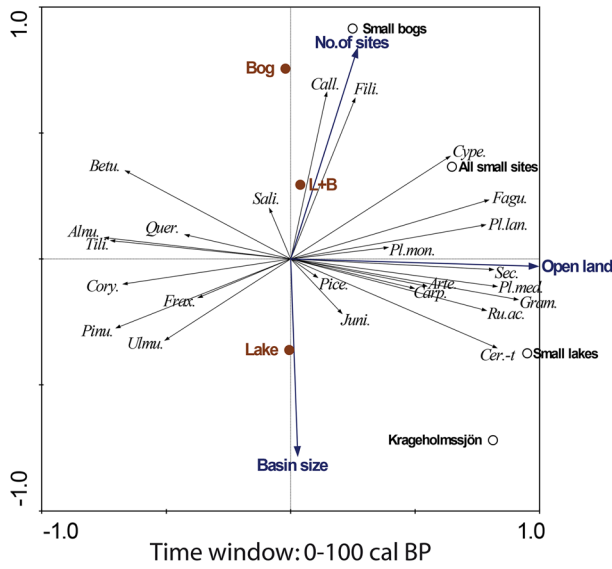
In this study we used 50 ha as the boundary between "small sites" and "large lakes". The minimum and maximum sizes of the selected small sites are 5 m radius (ca. 0.0025 ha) and 150 m radius (ca. 9 ha), respectively. We selected 40 Holocene pollen records with reliable chronologies over the last 6000 (6 k) calendar years before present (cal BP), 3 from large lakes (≥50 ha in size) and 37 from small sites (nine lakes and 28 bogs ≤10 ha in size; all lakes are larger than 25 m radius (from ca. 0.25 ha to max. 9 ha), 14 bogs are larger than 25 m radius, and 14 bogs are smaller than 25 m radius (Fig. 1; Table 1). All pollen
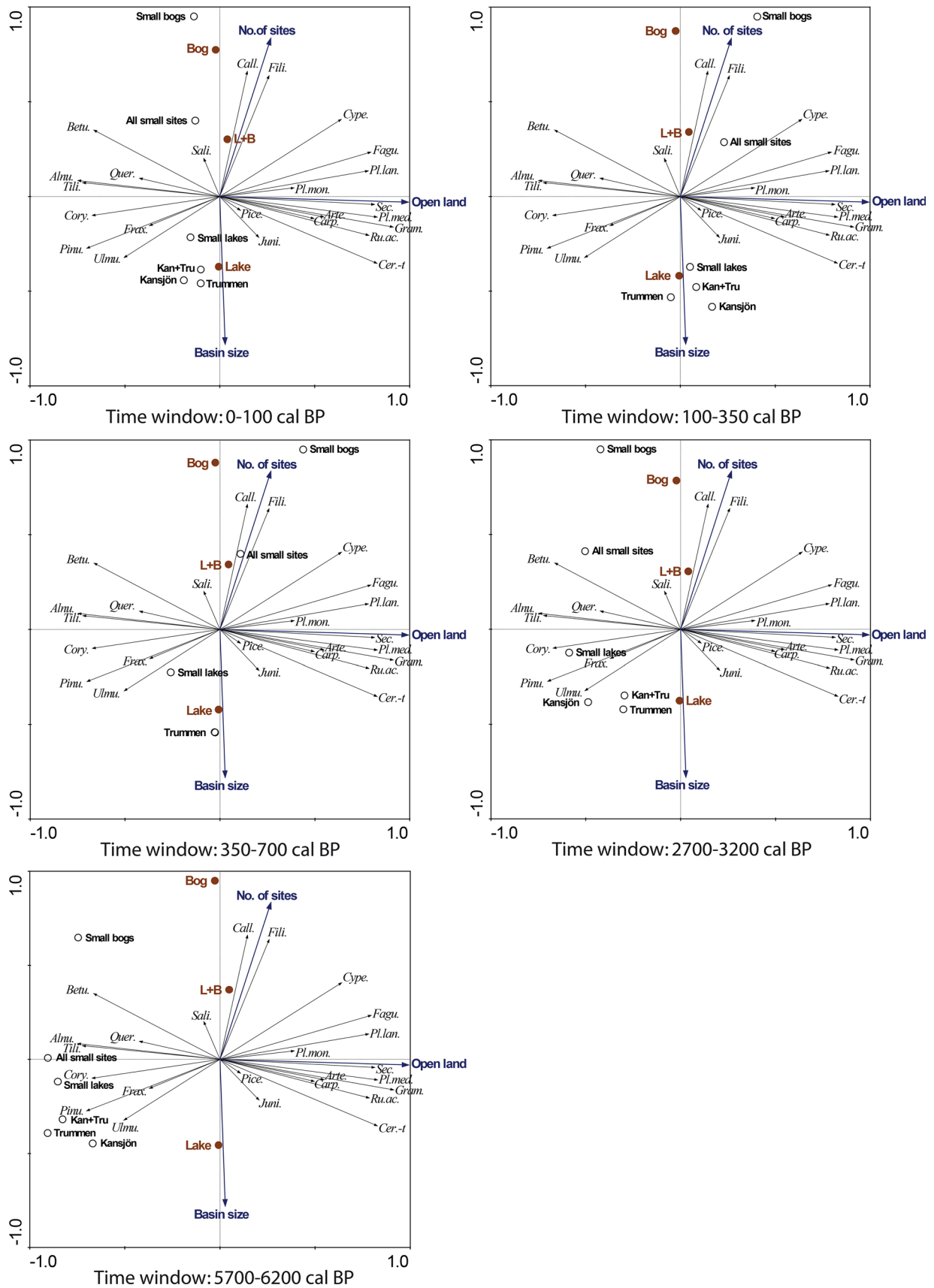
Fig. 2 RDA analyses of RUNS 1 REVEALS estimates for the ▶ temperate zone, for all five time windows (see text for more explanations). The eigenvalues of axes one and two for all RUNS 1 (this Figure and Fig. 3) are 0.42 and 0.061, respectively. The scores of the 24 plant taxa are shown with black arrows, the three quantitative variables "number of sites", "basin size" and "open land" (i.e. percentage cover of open land) with blue arrows, the three nominal variables (presence/absence) "site type" (lake and/or bog) with large brown dots, and the four nominal variables "small site", "small lake", "small bog", "Krageholmsjön" (large lake) with black circles. L + B = lakes + bogs

records are from the LANDCLIM pollen data archive (Trondman 2014); the sites are listed in Table 1 with their characteristics. The methods applied to build up the chronologies, and the criteria used to define a chronology as "reliable" are described in Trondman et al. (2015). There are 21 sites in the temperate zone, 13 sites in the hemiboreal zone, and six sites in the southern boreal zone (Fig. 1b). Because the hemiboreal zone has more in common with the southern boreal zone than the temperate zone in terms of geology, climate, vegetation, and land use, we chose to group the sites in the hemiboreal and southern boreal zones into a single category with a total of 19 sites.

Time windows and chronologies

We chose to work with the same five Holocene time windows as those used in the LANDCLIM project (Gaillard et al. 2010; Mazier et al. 2012; Trondman et al. 2015), i.e. x–100 cal BP (ca. 0.05 k, Recent past, with x = date of the core surface, i.e. year of coring, assuming that the upper peat or lake sediment were deposited during the year of coring), 100–350 cal BP (ca. 0.2 k, end of the Little Ice Age), 350–700 cal BP (ca. 0.5 k, end of Middle Ages–Modern time), 2700–3200 cal BP (ca. 3 k, Early/Late Bronze Age transition) and 5700–6200 cal BP (ca. 6 k, Mesolithic/Early Neolithic boundary). These time windows (referred to as 0.05, 0.2, 0.5, 3 and 6 k below) represent periods with contrasting human impact on the vegetation cover, which was an important criterion for the main objective of the LANDCLIM project, i.e. the study of the effect of anthropogenic land-cover change on climate. We also used the same chronologies for the pollen records as those applied in the LANDCLIM project (Mazier et al. 2012; Trondman et al. 2015) except for the sites Ekenäs and Skärsgölarna. For those two sites, new age-depth models were established using the computer program *clam* version 2.1 (Blaauw 2010) that implements the IntCal09 calibration curve of Reimer et al. (2009). A smoothing spline was applied for both Ekenäs (four radiocarbon dates + core top age) and Skärsgölarna (five radiocarbon dates + core top age) (not shown here).
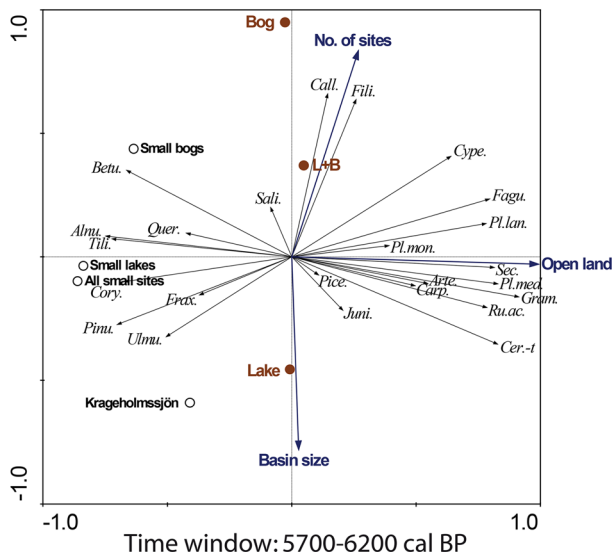
Time window: 0–100 cal BP

Time window: 100–350 cal BP

Time window: 350–700 cal BP

Time window: 2700–3200 cal BP

Time window: 5700–6200 cal BP

Time window: 0-100 cal BP

Time window: 100-350 cal BP

Time window: 350-700 cal BP

Time window: 2700-3200 cal BP

Time window: 5700-6200 cal BP

Regional vegetation reconstruction: the REVEALS model and its application

The REVEALS model estimates the regional vegetation abundance in percentage cover of individual plant taxa or groups of taxa for an area of 10^4–10^5 km^2. The REVEALS model requires (i) raw pollen counts, (ii) the radius of the site(s), (iii) pollen productivity estimates (PPEs) and estimates of fall speed of pollen (FSP) for each taxon to be reconstructed, (iv) an estimate of the maximum extent of the regional vegetation (Z_{max}), and (v) a pollen dispersal-deposition function (following Prentice (1985) for bogs and Sugita (1993) for lakes). The assumptions of the model are given in Sugita (2007a).

The total number of sites and the number of each site type may vary between time windows. Due to insufficient stratigraphical information for most sites, we assumed that the basin type and size were the same in the past as today except for two small bogs in the temperate zone (Fårarpmosse and Östra Ringarp) for which litho-stratigraphical information was available and indicated that these sites were lakes during the oldest time windows and that the size of Fårarpsmossen basin changed through time (Table 1). The REVEALS model was run separately for lakes and bogs using the pollen dispersal and deposition model of Sugita for lakes (Sugita 1993) and of Prentice for bogs (Prentice 1985). The mean REVEALS estimates were than calculated for small lakes and bogs together using the "Bog-Lake fusion program" (Sugita, unpublished computer program bog.lake.data.fusion.24Nov10.v5.exe). We used the same 25 plant taxa, PPEs, FSPs, and grouping of the 25 taxa into three land-cover types (LCTs) as the LANDCLIM project (Table 2; Mazier et al. 2012; Trondman et al. 2015); however, *Abies* (first taxon in Table 2) was not present in southern Sweden at any time in the Holocene and therefore only 24 taxa were reconstructed in this study. We kept the same taxon numbers for convenience; this implies that the 24 reconstructed taxa are numbers 2–25 in Table 2 and in Figs. 6 and 8.

Mazier et al. (2012) tested the effect of different estimates of Z_{max} (i.e. 50, 100 and 200 km) on REVEALS estimates using pollen records from the Czech Republic and found that the difference in Z_{max} did not affect the ranking of taxa in the REVEALS results. We performed the same test with a Z_{max} of 50 and 200 km using the pollen records selected for the present study and also found no significant differences in taxa ranking (results not shown here). Therefore, we set Z_{max} at 200 km, i.e. the value of the "characteristic radius" sensu Prentice (1988) calculated for large lakes and 24 plant taxa in southern Sweden by Hellman et al. (2008b) and defined as the distance from which ≥90 % of the pollen from all 24 taxa are coming.

REVEALS runs

The different groups of REVEALS runs (hereafter referred to as "RUNS 1–5", Online Resource 1) are as follows:

- *RUNS 1* the small sites were divided into two groups, 21 sites located in the temperate zone and 19 sites located in the hemiboreal/southern boreal zone. These REVEALS estimates were then compared with those of the large lake(s) in each vegetation zone (Fig. 1b).
- *RUNS 2–5* the small sites were divided into four groups including the sites within radii of 50 (RUNS 2), 100 (RUNS 3), 150 (RUNS 4) and 200 (RUNS 5) km from the center of each large lake (Fig. 1c–f). This grouping was made irrespectively of the vegetation zones.
- *RUNS 2* group of sites around the large lake (Krageholmssjön) in the temperate zone (Fig. 1c).
- *RUNS 3–4* group of sites around each of the two large lakes (Trummen and Kansjön) in the hemiboreal/southern boreal zone (Fig. 1d, e).
- *RUNS 5* group of sites around the midpoint of a segment joining the central points of the two large lakes Kansjön (K) and Trummen (T) (K + T) in the hemiboreal/southern boreal zone (Fig. 1f).

Hellman et al. (2008a) showed that the REVEALS model often performed better at estimating the cover of groups of taxa rather than individual taxa. In order to test this observation, we also calculated the mean REVEALS estimates for groups of taxa corresponding to the three land-cover types (LCTs) used in the LANDCLIM project, e.g. summer-green trees (ST), evergreen trees (ET), and open land (OL, i.e. non-forested land) (Gaillard et al. 2010, Table 2); this was performed for RUNS 1 only.

Multivariate analyses

Multivariate analysis implemented by the computer program CANOCO 4.5 (Ter Braak and Šmilauer 2002) was applied on the REVEALS RUNS 1–5 (see above) to evaluate the influence of a set of selected environmental/explanatory variables on the REVEALS estimates of regional vegetation composition. The explanatory variables include basin size, number and type (bog or lake) of sites,

the distance between small sites and large site(s), and vegetation characteristics (vegetation zone and vegetation openness; see also below). The length of the gradient in a detrended correspondence analysis (DCA) of the REVEALS dataset was short (<2.0 SD units) for all tested RUNS, which motivated the choice of a linear ordination method, e.g. redundancy analysis (RDA) for further analysis (Lepš and Šmilauer 2003).

RDA was applied on the REVEALS estimates from (i) each of the RUNS 1–5 separately, and (ii) all RUNS 2–5 together as one dataset, resulting in six RDAs. For RUNS 1 we selected four environmental variables for the analysis, i.e. number of sites, type of site (entered as a dummy variable (0, 1): lake, bog, or lake + bog; note that a group of small sites may consist of both bogs and lakes, therefore the variable "lake + bog"), proportion of open land (non-forested land expressed in percentage cover as estimated by REVEALS in each RUNS), and basin size (mean radius expressed in metres). For RUNS 2–5 six environmental variables were included in the analysis, the same four variables as for RUNS 1 with the addition of two variables, i.e. distance from the large lake (expressed in kilometres), and vegetation zone (entered as a dummy variable (0, 1): temperate zone, hemiboreal/southern boreal zone, or mixed zone; note that a group of small sites may consist of sites from the temperate zone and the hemiboreal/southern boreal zone, therefore the variable "mixed zone"). The variables "lake + bog" and "mixed zone" are not redundant because we are analysing REVEALS estimates from groups of sites and not from individual sites. The results of the RDAs are presented in tri-plots using CanoDraw 4 for Windows. The variables were evaluated using forward selection; the "extra fit" (indicating the importance, i.e. the explanatory power, of each variable) was determined for each environmental variable. The statistical significance of the variables was tested by Monte Carlo permutations (Table 3).
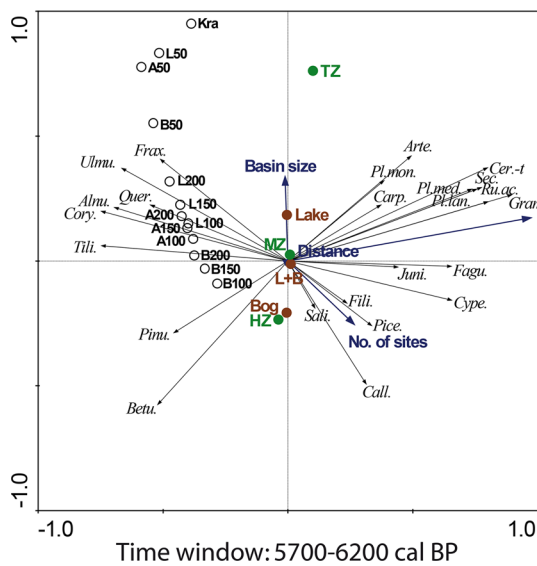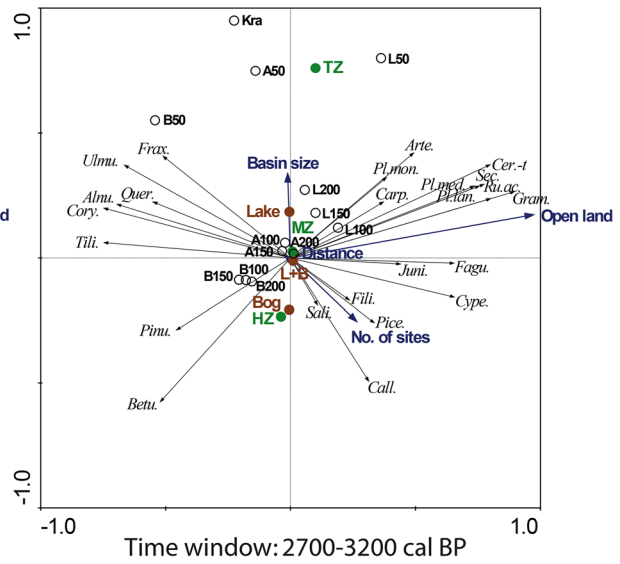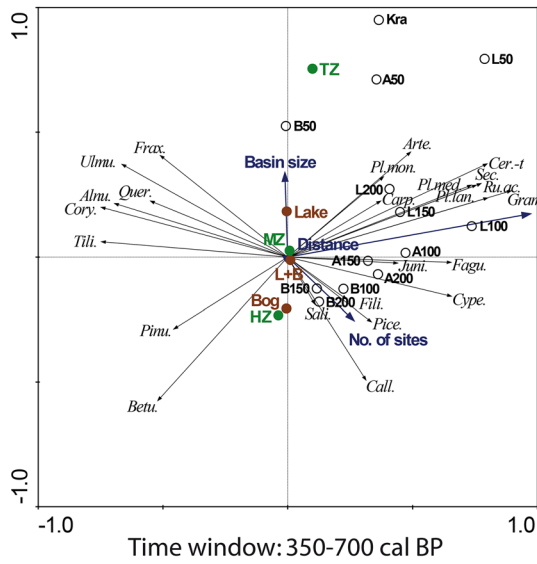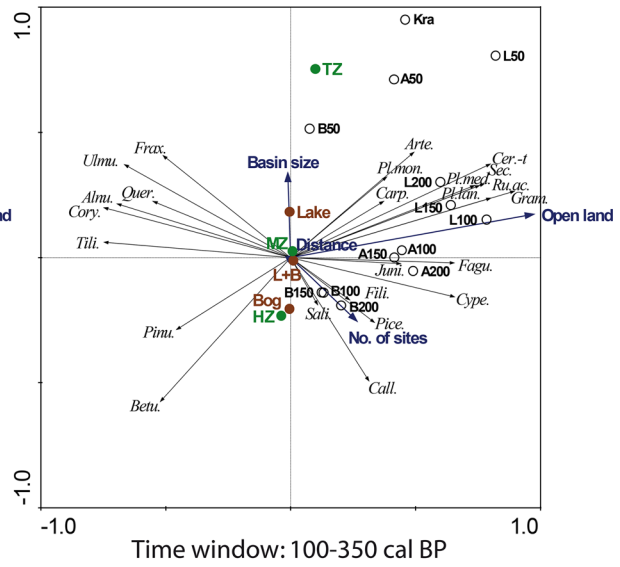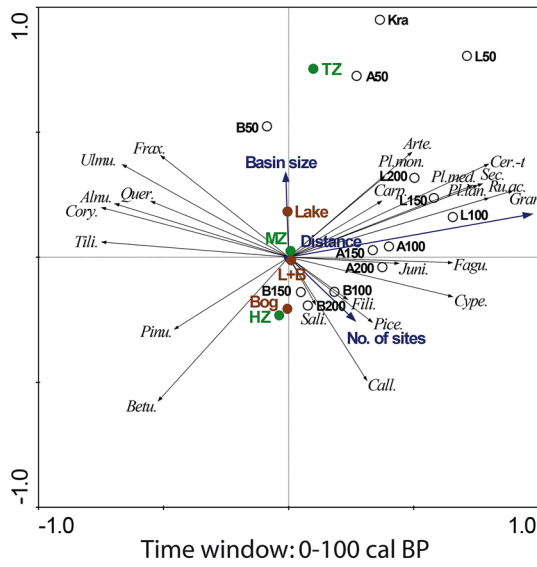
Results

Redundancy analysis (RDA): the effects of environmental variables on the mean REVEALS estimates based on multiple small sites

The results of the RDA are presented in Table 3 and in tri-plots for the different runs and time windows. In this paper we chose to present only the tri-plots for RUNS 1 (Figs. 2, 3) and RUNS 2-5 together as one dataset (Figs. 4, 5; Online Resource 2 Figs. 1, 2). In all tri-plots, the first axis contrasts taxa indicators of open land on the right with forested vegetation on the left, and the second axis contrasts Poaceae and indicator taxa of cultivated and ruderal

Fig. 4 RDA analyses of the REVEALS estimates from small sites ▶ within 50, 100, 150 and 200 km from large lakes (RUNS 2–5) for all five time windows (see text for more explanations): Lake Krage-holmssjön (temperate vegetation zone). The eigenvalues of axes one and two are 0.407 and 0.076, respectively. The scores of the three nominal variables "temperate" (TZ), "hemiboreal/southern boreal zone" (HZ) and "TZ + HZ" (= mixed zone, MZ) are shown with large green dots, and the scores of the nominal variables (presence/absence) all "small site i.e. lake + bog (A)", "small lake (L)", "small bog (B)" (within 50, 100, 150, 200 km from the large lake), and "Krageholmssjön" (large lake) with black circles. For other interpretation of the figure, see caption of Fig. 2

land (e.g. Cerealia-type, *Secale*, *Rumex acetosa*-type, *Artemisia*) with *Calluna vulgaris*, Cyperaceae and *Filipendula*; in RUNS 2–5, axis 2 also contrasts *Pinus* and *Betula* with broadleaved trees and *Alnus*. Moreover, *Calluna vulgaris* and *Filipendula* are related to the variables "number of sites" and "bog", while "basin size" and "lake" are related both to indicators of human impact and broadleaved trees. In addition, the second axis clearly contrasts the variable "lake" with the variable "bog". The RDA scores of "small lakes" (whatever the group of REVEALS runs) are almost always located closer to the score(s) of "large lake(s)" than the scores of "small bogs". In the RDA plots of RUNS 2–5 (Figs. 4, 5; Online Resource 2 Figs. 1, 2) the variable "distance" is not related to any of the other variables and groups of taxa mentioned above, which suggests that the distance may have an independent effect on the REVEALS estimates. In contrast, the variable "vegetation zone" is well correlated with the taxa characteristic for the two vegetation zones in the study region; "temperate" is correlated with broadleaved trees and taxa indicators of human-induced open land, while "hemiboreal/southern boreal" is associated in particular with *Pinus*, *Betula*, *Picea*, Cyperaceae, *Calluna vulgaris* and *Filipendula*.

The Monte Carlo permutation tests (Table 3) show that only "open land" and "number of sites" are significant variables when the RDA is implemented using the REVEALS estimate from RUNS 1 only. In contrast, all variables except "lake + bog" are significant when the RDA is implemented using the REVEALS estimates from RUNS 2–5 together. When the RDA is implemented separately for the REVEALS estimates from RUNS 2–5 (four RDA analyses), the variables "open land", "number of sites", and "distance" are the only ones that are always significant, and the variables "lake" and "bog" are significant in all except one test. The types of vegetation zone "temperate" and "mixed veg. zone" are significant for RUNS 2 (large site in the temperate zone), "hemiboreal/southern boreal" and "mixed veg. zone" for RUNS 5 (midpoint between two large lakes in the hemiboreal/southern boreal zone), and "mixed veg. zone" for RUNS 4 (large lake in the hemiboreal/southern boreal zone).

Time window: 0–100 cal BP

Time window: 100–350 cal BP

Time window: 350–700 cal BP

Time window: 2700–3200 cal BP

Time window: 5700–6200 cal BP

B	Small bogs
A	All small sites (i.e. Lake+Bog)
L	Small lakes
Kra	Lake Krageholmssjön
50	within 50 km from Krageholmssjön
100	within 100 km from Krageholmssjön
150	within 150 km from Krageholmssjön
200	within 200 km from Krageholmssjön
TZ	Temperate zone
HZ	Hemi-/South boreal zones
MZ	Temperate + Hemi-/South boreal zones

REVEALS estimates using pollen records from all small sites and 1–2 large lakes in each vegetation zone (RUNS 1)

Temperate zone

The mean REVEALS estimates from a total of 20 small sites (15 bogs and five lakes) were compared to the REVEALS estimates from one large lake, Krageholmssjön. The results are presented in bi-plots with standard errors for the five Holocene time windows (Fig. 6; Online Resource 3 Figs. 1, 2, 3, 4). When taking the standard errors of the REVEALS estimates into account there is a good agreement between the results from multiple small sites and the large lake except for a few taxa, in particular *Betula*, *C. vulgaris* and Cerealia-type. However, there is a better agreement between the small lakes and the large lake than between the small bogs and the large lake.

The REVEALS estimates obtained from pollen records of multiple small sites agree better with those obtained from the pollen record of the large lake Krageholmssjön for groups of taxa (i.e. LCTs) (Fig. 7; Online Resource 4 Figs. 1, 2, 3, 4) rather than for individual taxa (Fig. 6; Online Resource 3 Figs. 1, 2, 3, 4). The mean of all small sites LCTs shows a good correspondence with the large lake LCTs for all time windows. The small lakes LCTs are very close to the large lakes LCTs at 0.05, 0.2 and 0.5 k, while the small bogs LCTs slightly overestimate summergreen trees (ST), and underestimate open land (OL). In contrast, for the time windows 3 k and 6 k, the small-bogs LCTs are very close to the large lakes LCTs.

Hemiboreal/southern boreal zone

The mean REVEALS estimates from a total of 17 small sites (13 bogs and four lakes) were compared either with the REVEALS estimates from the two large lakes Trummen and Kansjön separately, or with the mean REVEALS estimates from the two lakes. Note that no data are available from Kansjön for the time window 0.5 k. The results are presented in bi-plots with standard errors (Fig. 6; Online Resource 3 Figs. 1, 2, 3, 4). There are some taxa that differ in abundance between the two large lakes in the hemiboreal/southern boreal zone, and this is also reflected by the large standard errors when the REVEALS estimates from the two lakes are used to calculate the mean REVEALS estimates. When taking the standard errors of the REVEALS estimates into account there is a good agreement between the results from multiple small sites and large lakes. As for the temperate zone, there is a better

Fig. 5 RDA analyses of REVEALS estimates from small sites within 50, 100, 150 and 200 km from large lakes (RUNS 2–5) for all five time windows (see text for more explanations): midpoint between the two large lakes Trummen (hemiboreal vegetation zone) and Kansjön (southern boreal vegetation zone). For interpretation of the figure, see captions of Figs. 2 and 4
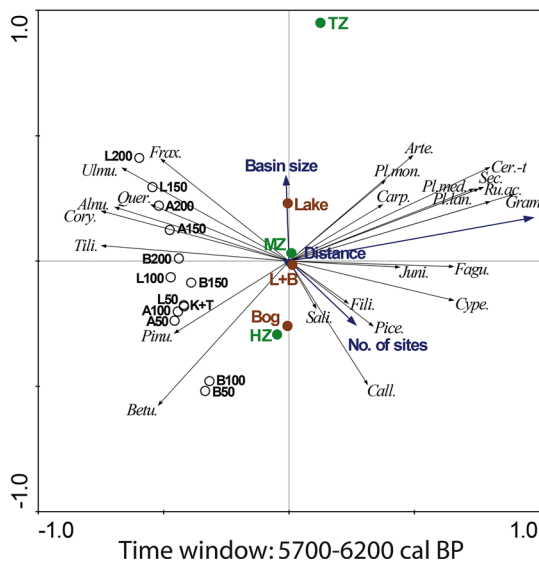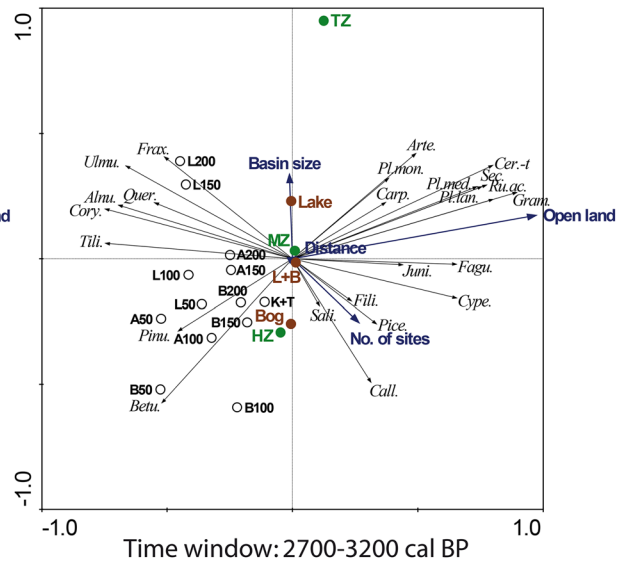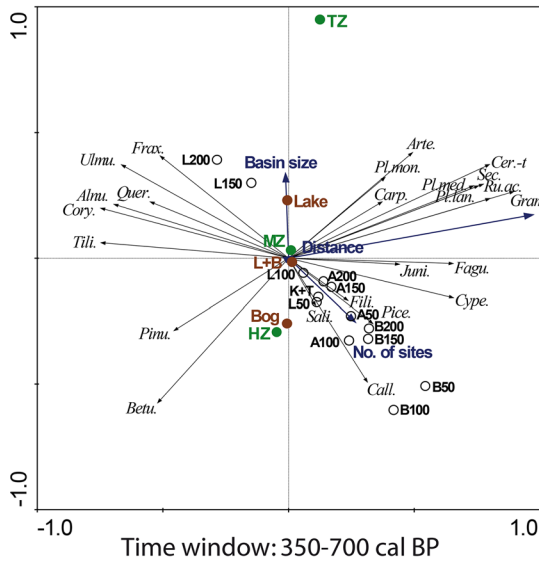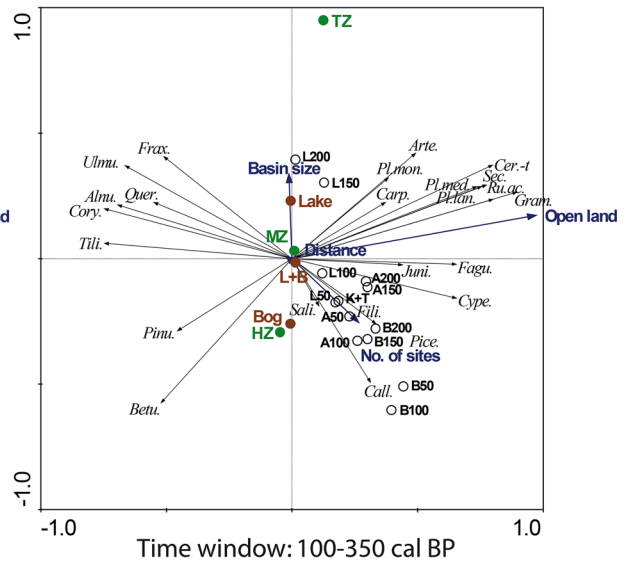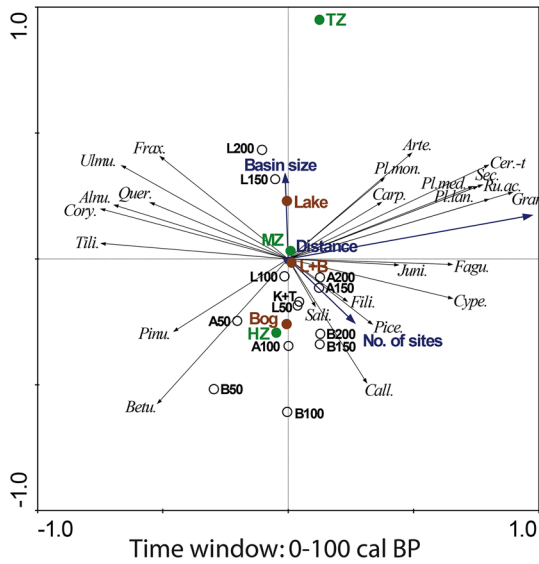
agreement with the large lakes estimates for the small lakes REVEALS estimates than for the small bogs estimates. There are also some taxa that differ substantially in their REVEALS estimates between the small sites and the large lakes, for example *Fagus*, *Betula* and *Picea*, when the results from the small sites are compared with those from the large lake Kansjön. The best agreement in REVEALS estimates is seen between the small lakes values and the Kansjön + Trummen mean values (Fig. 6; Online Resource 3 Figs. 1, 2, 3, 4). The mean values from small lakes + small bogs are also in good agreement with the Kansjön + Trummen mean values.

The comparison between the REVEALS estimates from small sites and the large lakes Trummen and Kansjön were also made for the three LCTs. These results mostly confirm those obtained in the temperate zone (Fig. 7; Online Resource 4 Figs. 1, 2, 3, 4).
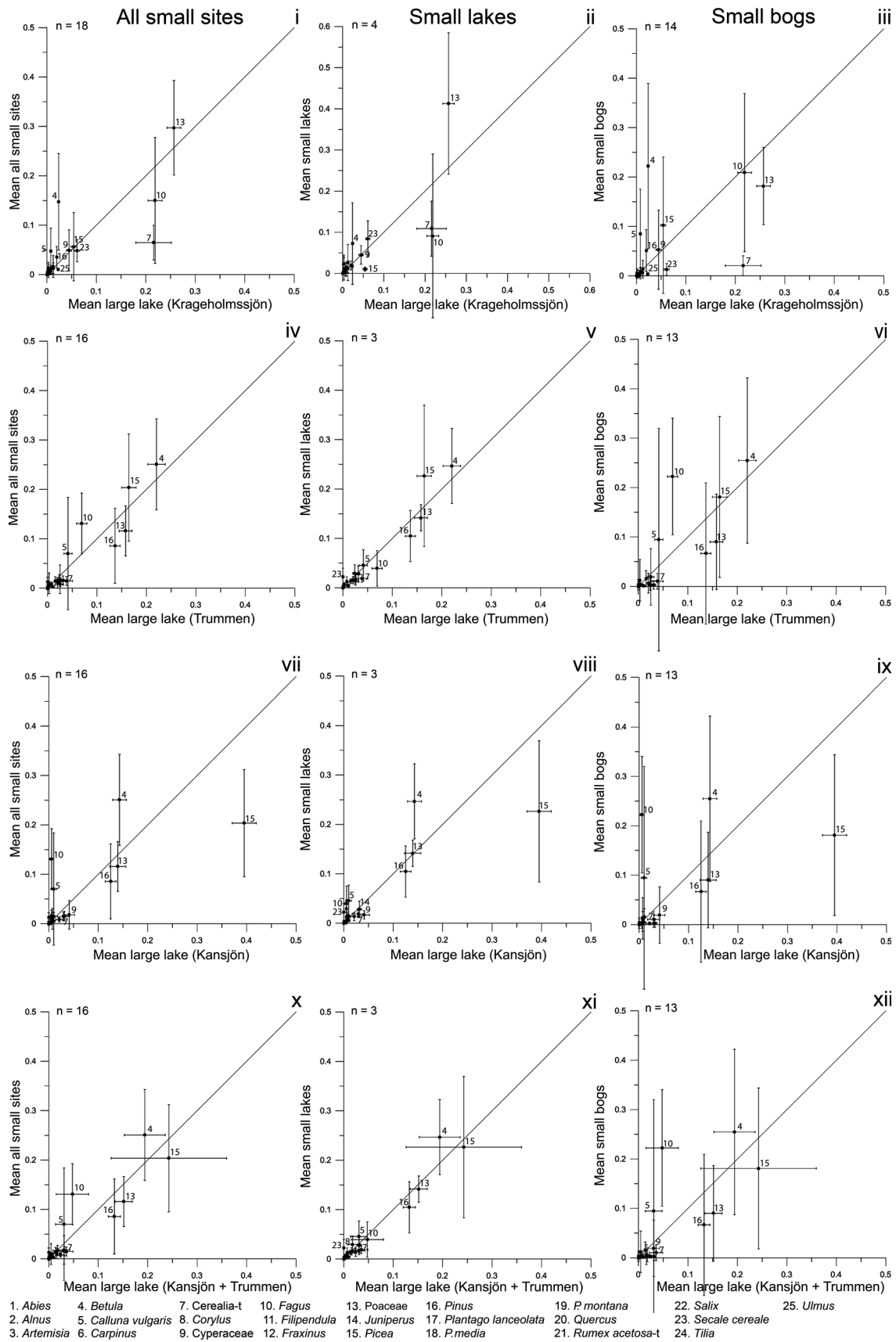
REVEALS estimates using pollen records from multiple small sites within different distances from the large lakes (RUNS 2–5)

Temperate zone: RUNS 2

All small sites available within the distances 50, 100, 150 and 200 km from the centre of the large lake Krageholmssjön (temperate zone) were grouped to produce four mean REVEALS estimates for each taxon to be compared with the REVEALS estimates from Krageholmssjön. The vegetation zone boundary is located approximately 100 km from the large lake, which implies that two groups of small sites include sites from the hemiboreal/southern boreal zone. The number of small sites (i.e. pollen records) varies from eight (within 50 km) to 32 (within 200 km) (Fig. 1c; Online Resource 1) and varies only slightly depending on the time window. The results are presented in bi-plots with standard errors (Fig. 8; Online Resource 5 Figs. 1, 2, 3, 4) for the five time windows. In the temperate zone, the multiple small-sites REVEALS estimates show similar trends whatever the distance between the large lake and the small sites. However, the distance for which the REVEALS estimates from the small sites and the large lake(s) are most similar differs somewhat depending on the taxon and time window. Nevertheless, the results for the major pollen taxa

1. *Abies*   4. *Betula*   7. Cerealia-t   10. *Fagus*   13. Poaceae   16. *Pinus*   19. *P. montana*   22. *Salix*   25. *Ulmus*
2. *Alnus*   5. *Calluna vulgaris* 8. *Corylus*   11. *Filipendula*   14. *Juniperus*   17. *Plantago lanceolata* 20. *Quercus*   23. *Secale cereale*
3. *Artemisia*   6. *Carpinus*   9. Cyperaceae   12. *Fraxinus*   15. *Picea*   18. *P.media*   21. *Rumex acetosa*-t 24. *Tilia*

Fig. 6 REVEALS estimates from RUNS 1 for 24 plant taxa at 0.05 k calibrated years BP (time window 0–0.1 k BP): mean REVEALS estimates from large lakes (Kragehomssjön—temperate vegetation zone (**i–iii**); Trummen—hemiboreal vegetation zone (**iv–vi**); Kansjön—southern boreal vegetation zone (**vii–ix**); K + T—mean REVEALS estimates from the two lakes Kansjön and Trummen (**x–xii**)) plotted against mean REVEALS estimates from all small sites (lakes + bogs) (*left*), small lakes (*middle*) and small bogs (*right*). *n* is the number of small sites used to produce mean REVEALS estimates for the specific site type and time window. For readability, only some plant taxa are labelled. Note that the numbers for plant taxa refer to the standard taxa numbers in the LANDCLIM project (see text for more explanations). *Abies* does not occur in the study region

suggest that groups of small sites within 50 km from the large lake generally provide the REVEALS estimates most similar to those from the large lake.

### Hemiboreal and southern boreal zones: RUNS 3–5

All small sites available within four different distances of 50, 100, 150 and 200 km from the centre of the large lakes Trummen (RUNS 3) and Kansjön (RUNS 4), and from the mid-point of a segment joining the central points of Trummen and Kansjön (K + T, RUNS 5), were grouped into 12 site groups to estimate REVEALS-based abundance of each taxon to be compared with those from the two large lakes separately and with the Trummen + Kansjön mean REVEALS values. For RUNS 3 (Fig. 1d) the number of small sites varies from nine (within 50 km) to 37 (within 200 km), for RUNS 4 (Fig. 1e) from three to 29, and for RUNS 5 (Fig. 1f) from four to 32. The number of sites may also vary slightly between time windows. The vegetation zone boundary is located ca. 70–150 km from Trummen, ca. 120–170 km from Kansjön, and ca. 110–150 km from the midpoint of the line joining the centres of the two lakes Trummen and Kansjön. The results for RUNS 3-5 are presented in bi-plots with standard errors (Fig. 8; Online Resource 5 Figs. 1, 2, 3, 4) for the five time windows; the REVEALS estimates show similar tendencies as those of RUNS 1 for the hemiboreal/southern boreal zone. However, as for RUNS 2, the distance from the large lake required for a good agreement of the small-sites REVEALS estimates with the large-lake ones differs somewhat between taxa, time windows, and runs. It indicates that not only the distance itself influences the REVEALS estimates but also other variables such as number and type of small sites and the distance to the vegetation zone boundary.

Discussion

Redundant analysis (RDA): the effect of site characteristics on the multiple small-sites REVEALS estimates

The results from the RDA analyses of RUNS 1 can be explained by the following characteristics of the dataset: (a) *Filipendula* and Cyperaceae are associated with bogs, (b) the number of bogs in the REVEALS runs is larger than the number of lakes, (c) a majority of the studied lakes are located in the open landscapes of the temperate zone and therefore have a tendency to be characterised by the occurrence of broadleaved trees in their surroundings and (d) lakes are generally larger than bogs in the temperate zone. The RDA analyses of RUNS 1 suggest that the type of basin (lake or bog) influences significantly the REVEALS estimates in this study for the reasons above. Further, the RDA analyses of RUNS 2–5 suggest that a few small sites located in another vegetation zone than the majority of the sites in a multi-site group can affect the mean REVEALS estimates significantly (see RUNS 2–5 for sites within 150 and 200 km of the large lakes).

The Monte Carlo permutation tests show that, after "open land", "number of sites" has the stronger explanatory power (except for RUNS 2) and is always a significant variable, either when it is implemented on RUNS 1 and on RUNS 2–5 separately, or on RUNS 2–5 together (Table 3). "Open land" has the stronger explanatory power in all RUNS 1–5, which is a common feature of the results from numerical analyses of pollen data with a strong gradient in values of non-arboreal pollen (NAP) (or arboreal pollen AP), axis one generally contrasting pollen assemblages with high NAP (or low AP) with pollen assemblages with low NAP (or high AP) (e.g. Gaillard et al. 1994; Mazier et al. 2006). The "site size" either is not significant or has a very low explanatory power. The latter suggests that, although the dataset in this study includes a higher number of small bogs than small lakes, and although the small lakes are generally larger than the small bogs, the "type of site" (bog or lake) explains the REVEALS estimates better than the site size. In other words, the size of the small bogs and lakes is not as important as their pollen assemblages or local vegetation (see discussion on Cyperaceae and *Calluna* above and below).
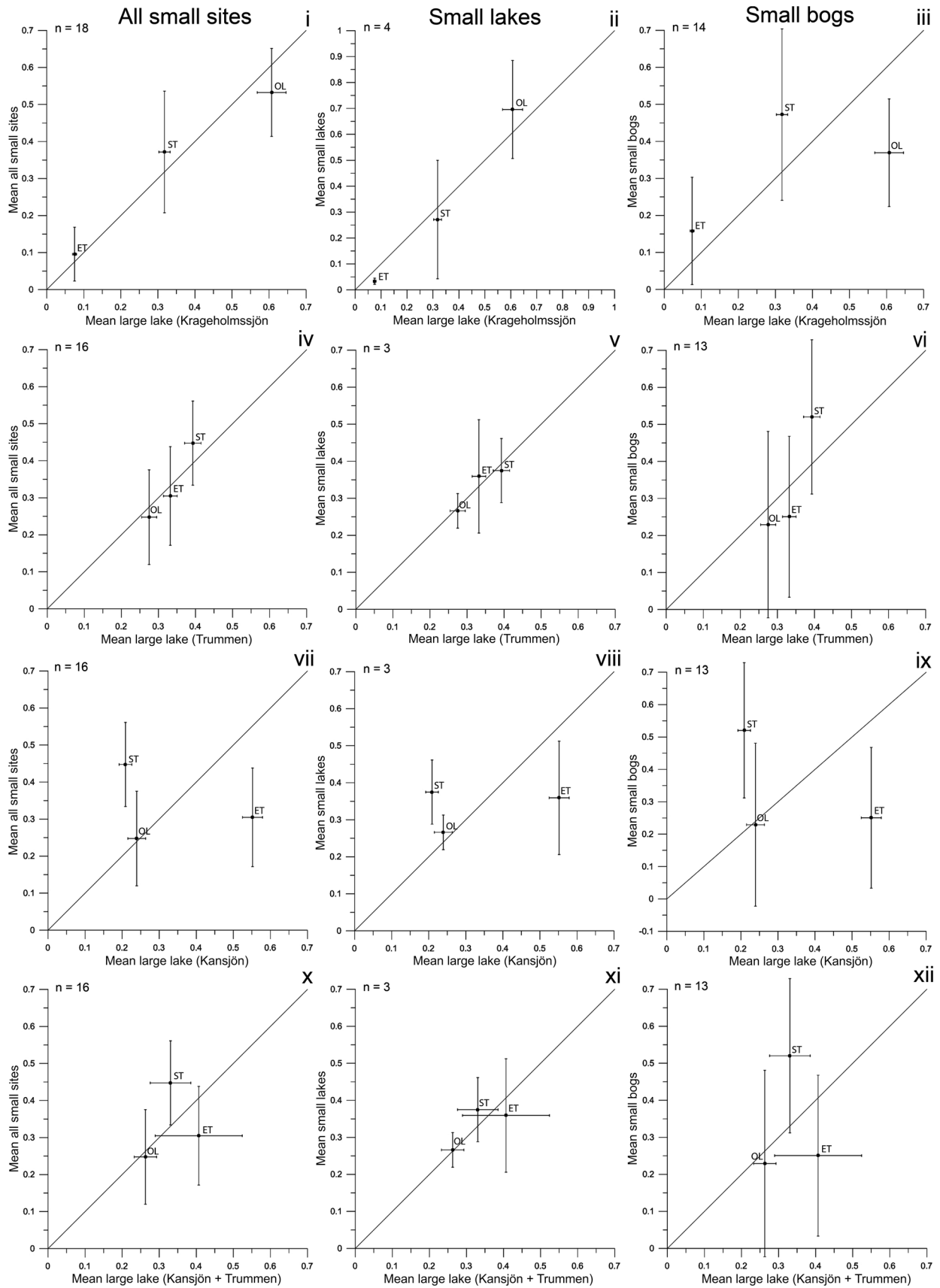
REVEALS estimates for individual plant taxa and for groups of taxa (land cover types)

Generally, REVEALS estimates for individual plant taxa are comparable between the small-sites estimates and the large-lakes estimates. However, some plant taxa that may be related to bog vegetation, such as *Calluna vulgaris* and Cyperaceae, tend to be slightly overestimated in some time windows when pollen data from small bogs are used, as also found by Nielsen et al. (2012). For example, *Calluna* values are higher for bogs than for lakes at 0.05 and 0.2 k in both vegetation zones, and Cyperaceae values are higher for bogs than for lakes at 0.2 and 0.5 k in the hemiboreal zone. However, the standard errors are generally very large and when taking these into account the results do not exhibit significant differences between runs. In the hemiboreal zone, *Betula* is also overestimated in the two most recent time windows when pollen data from bogs are used, and the latter is also valid when the standard errors are taken into account. This might be due to the fact that small bogs often were used in earlier times for grazing, hay making or wood (and were therefore more or less treeless), while they were not used in recent times and overgrown by birch, alder, pine and/or spruce. The Poaceae values are higher for the lakes than the bogs for all time windows except at 6 k in the temperate zone and at 0.2 and 0.05 k in the hemiboreal zone, although these differences are not significant when the error estimates are taken into account. Higher Poaceae values from lakes may be due to reed vegetation (*Phragmites australis*) that is a common feature of lake-shore vegetation. The possible overestimation of open land cover in REVEALS reconstructions using Poaceae as an indicator of open land is discussed further in Trondman et al. (2015).
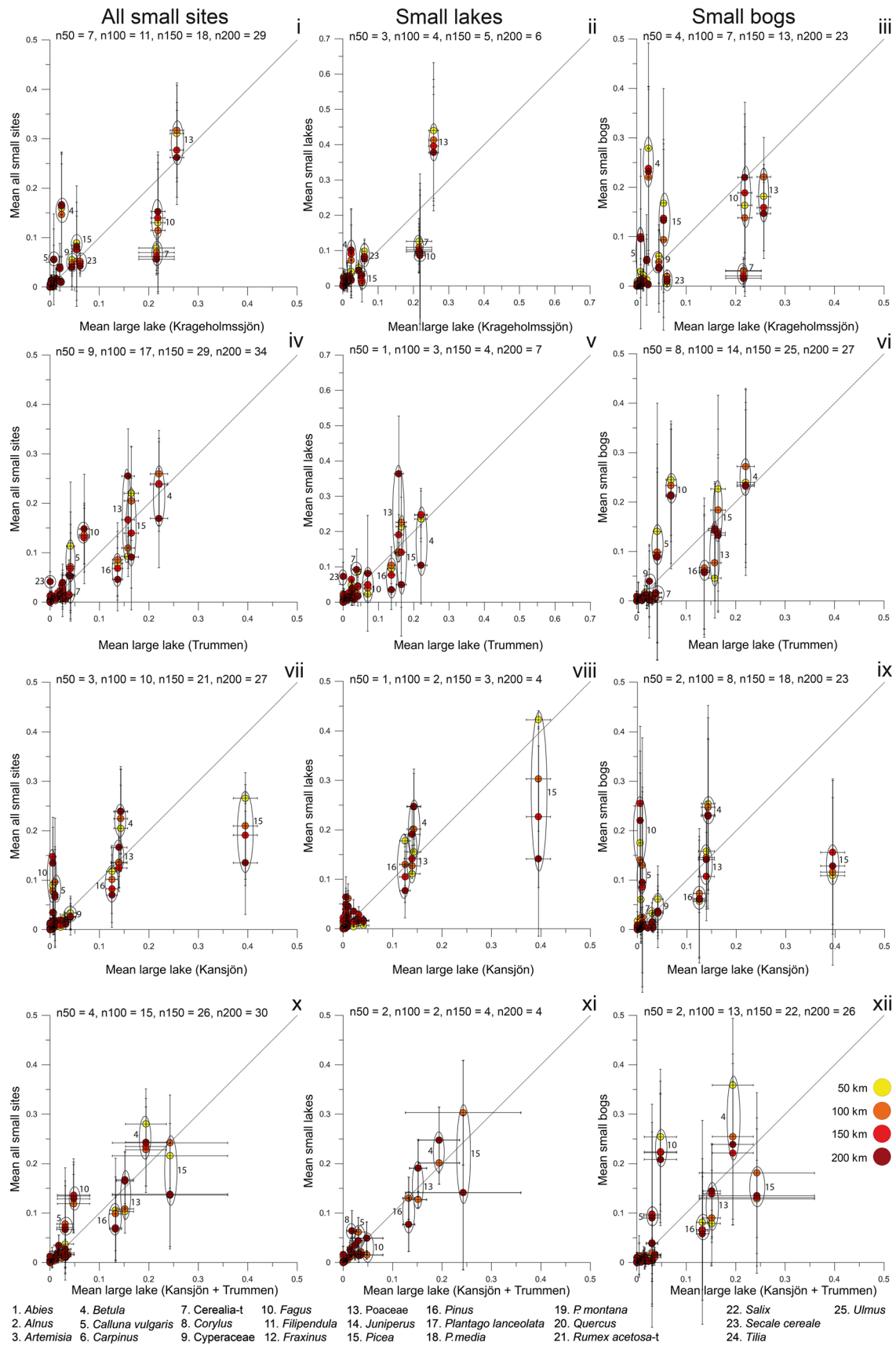
For the three land-cover types (LCTs) summer-green trees (ST), evergreen trees (ET), and open land (OL), REVEALS generally performs much better, which confirms the results of Hellman et al. (2008a). In that study, it was found that the agreement between the REVEALS reconstructions using pollen data from large lakes' surface sediments and the modern vegetation surveys was better for large taxa groups such as wooded (i.e. coniferous trees and broadleaved trees) and open land (i.e. grasses and cereals) than for individual taxa. Our study also shows that the small-lakes LCTs are generally closer to the large-lake LCTs than the small-bogs ones, except for the two oldest time windows (6 and 3 k) for which the estimates from

small lakes and small bogs are equally close to the large-lakes values. The latter might be due to the fact that, at 6 and 3 k, bogs were often mostly wooded and covered by alder and birch, two tree species also common by lake shores. Poaceae, Cyperaceae and *Calluna* were characteristic of small bogs later in the Holocene, when they were used by humans for e.g. grazing and hay making. The LCT results are of particular interest for REVEALS estimates of vegetation cover produced for climate modelling (e.g. Trondman et al. 2015; Pirzamanbein et al. 2014). Climate models generally use descriptions of land vegetation including few large units with comparable plant functional types, such as the LCTs used here (e.g. Strandberg et al. 2014). Pollen-based reconstructions over large areas for climate modelling will unavoidably require that small sites are used because there are too few pollen records from large lakes. Therefore, it is comforting that LCT REVEALS estimates from multiple small sites are generally comparable to those from large lakes.

Large lakes, small lakes and small bogs: what site type and how many sites should be used for the best REVEALS reconstructions?

The results in this study (RUNS 2–5) indicate that the REVEALS values from small lakes are most comparable to the REVEALS values from large lakes when small lakes are located within 50 km of large lakes, while for small bogs the REVEALS estimates within 150 and 200 km from the large lakes are closer to the estimates from the large lakes than the estimates within shorter distances (except in RUNS 2). Moreover, small-lakes and all-sites REVEALS values are often closer to the large-lakes values than the small-bogs values. All these results together suggest that the number of sites is more important than the type of site and the distance between the large and small sites. Nielsen et al. (2012) also found a low explanatory power for the variable "mean basin size" using the results from an RDA application on REVEALS reconstructions from small sites across northern Germany and Denmark for the time window 7500–8000 cal BP. Further, for all runs (RUNS 1 and 2–5), the tri-plots of the RDA analyses show that the variable "number of sites" is positively correlated to the variable "bog" and negatively correlated to the variable "lake", which implies that the number of site is an important variable for small bogs, but not necessarily for small lakes. Fyfe et al. (2013) also showed that a low number of small lakes often provide REVEALS estimates comparable to those from large lakes, while a higher number of bogs is needed to obtain good REVEALS estimates.

The best results from RUNS 1–5 in terms of the correspondence of the small-sites REVEALS estimates with the

1. *Abies*   4. *Betula*   7. Cerealia-t   10. *Fagus*   13. Poaceae   16. *Pinus*   19. *P. montana*   22. *Salix*   25. *Ulmus*
2. *Alnus*   5. *Calluna vulgaris* 8. *Corylus*   11. *Filipendula* 14. *Juniperus*   17. *Plantago lanceolata* 20. *Quercus*   23. *Secale cereale*
3. *Artemisia*   6. *Carpinus*   9. Cyperaceae   12. *Fraxinus*   15. *Picea*   18. *P. media*   21. *Rumex acetosa*-t   24. *Tilia*

large-lakes ones are those for the groups of sites located within the same vegetation zone as the large lake, except for RUNS 5. Therefore, the distance of small sites to vegetation boundaries is more important than the distance between small sites and large sites within a vegetation zone. In the case of RUNS 5, despite the fact that the small sites are located in two different vegetation zones (hemiboreal and south boreal) for most of the distances to the midpoint between the two large lakes, the small-sites REVEALS estimates agree reasonably well with the mean Trummen + Kansjön REVEALS estimates. Cui et al. (2013) assumed that the mean REVEALS estimates from Kansjön and Trummen was a good approximation of the regional vegetation influencing the pollen assemblages at three small study sites located in the hemiboreal zone (two sites) and the southern boreal zone (one site). This assumption is now supported by the results of this study. However, it should be stressed that the two vegetation zones do not differ much in their vegetation characteristics (see description of study area).

Although one large lake may be sufficient to obtain reliable REVEALS estimates of regional vegetation, this is true only if the regional vegetation is characterized by a regular, large-scale mosaic of vegetation patches (i.e. a mix of patches with comparable distances between them) (see Hellman et al. 2009a, b for further definition of the concepts of regular (homogenous) versus irregular (heterogenous) mosaic patterns). For instance, in the case of the temperate zone of southern Sweden, its southern part is—and has long been—characterized by a very open landscape (cultivated fields, pasture land and few trees) compared to its northern part (more wooded, less cultivated land and open pasture land) (Gaillard 1984; Regnéll 1989). Therefore, the large lake Krageholmssjön—located in the southern part of the vegetation zone—will tend to provide larger REVEALS estimates of open vegetation than the multiple small sites within ≥100 km from Krageholmsjön. The division of the temperate zone into two different landscape types may explain why the best correspondence in REVEALS estimates between Krageholmssjön and the multiple small sites (RUNS 2) is achieved with the small sites within 50 km from the large lake. The use of at least two large lakes, one in the southern part and one in the northern part of a particular vegetation zone, would most

probably provide better REVEALS estimates of the regional vegetation in the entire temperate zone. Hellman et al. (2008a) compared REVEALS estimates from surface-sediment pollen assemblages in ten large lakes of southern Sweden with the actual modern vegetation within 10^4 km². The results showed clear differences in pollen assemblages between the large lakes; some lakes were treated as outliers, for example Lake Vombsjön in the temperate zone. Although Vombsjön was the largest lake (1,224 ha) studied by Hellman et al. (2008a) and should theoretically be the best one to reconstruct regional vegetation, it is located in an area with planted *Pinus* woods which is atypical for the region. Therefore, Hellman et al. (2008a) suggested that a "multiple-site approach", also when large lakes are used, will always provide the most reliable REVEALS estimates, and greatly reduce the standard error of the estimates. The same is valid for small sites. If they were regularly distributed within the landscape, the mean REVEALS estimates would provide a reasonable reconstruction of the regional proportions of open and wooded land. However, the available small sites seldom represent a random sampling of the vegetation types in a region. In this study, most of the small sites in the temperate zone are situated in marginal areas that are (were) more wooded than other more central areas with open cultivated land and grazing land.

Conclusion

Pollen data from multiple small sites can be used to estimate regional vegetation abundance applying the REVEALS model; however, standard errors will generally be larger than those obtained using pollen records from large lakes. These results support the findings of Mazier et al. (2012) and Fyfe et al. (2013), and confirm that the theoretical framework of the REVEALS model (Sugita 2007a) can be applied in empirical situations. Monte Carlo permutation tests suggest that the number of small sites is the most significant variable explaining the REVEALS estimates, followed by the location within a regional vegetation unit, the type of site (bog or lake, i.e. the predominant local vegetation at the small site), and the distance to the boundary of the regional vegetation unit. The distance between the small sites within a regional vegetation unit and the size of the small sites are the least significant variables. Sugita (2007a) tested the effect of lake size (from 0.13 to 30,000 ha) on REVEALS reconstructions and showed that it is not possible to identify a minimum size of "small sites" to obtain reliable REVEALS-based reconstructions of regional vegetation, because it will vary depending on the characteristics of the landscape in terms of vegetation structure (size and distribution of vegetation patches). The

minimum sizes required for 30 lakes to achieve the best REVEALS reconstructions were 7, 80–100, and 800–1,000 ha when the largest vegetation patches were 15, 1,200, and 10,000 ha, respectively. In that study, the best REVEALS estimates were defined as those with the smallest error estimates possible and with REVEALS values that do not change with an increase in lake size. The major difference between the REVEALS estimates based on very small lakes (e.g. less than 7, 100 or 1,000 ha in this case) and those based on larger lakes was the size of the error estimate, i.e. largest for reconstructions using pollen data from the smallest sites. Moreover, the mean REVEALS estimates were slightly different (lower or higher value depending on the taxon) than the mean REVEALS estimates from larger lakes.

In our study, pollen records from multiple small lakes generally produce REVEALS estimates of vegetation abundance more similar to REVEALS estimates from large lakes than pollen records from small bogs, even when the number of small lakes is lower than the number of small bogs. One should always test both multiple small lakes and multiple small sites (lakes + bogs) reconstructions and consider both in the interpretation of results. Sugita (2007a) did not test the influence of the number of small sites used for REVEALS estimates. However, theoretically, for a given lake size and a given pollen count, the error estimates will decrease with increasing number of sites. The latter was demonstrated for large lakes >50 ha (Sugita 2007a). For the reasons mentioned above, we do not recommend a minimum number of small sites for REVEALS reconstructions. This number will depend on many variables, i.e. the size of the pollen counts for each pollen record, the type of site (lake or bog), the vegetation type (characteristics of the vegetation mosaic, i.e. number and size of patches), the distribution of the sites within the vegetation mosaic, and the size of the area for which vegetation cover is intended to be reconstructed. If the REVEALS model is used for reconstruction of the regional vegetation cover for a particular region, it is recommended to use as many small sites as possible, but obviously accounting for the limits of availability of sites, time and money. If the REVEALS model is used for reconstructions over very large areas, such as continents or subcontinents, for the purpose of e.g. large-scale vegetation descriptions for climate modelling (e.g. Trondman et al. 2015), a smaller number of sites (2–4 sites) can be used for 100 km × 100 km grid cells, because these reconstructions will be compared with adjacent ones based on a larger number of small sites or based on large sites.

References

Ahti T, Hämet-Ahti L, Jalas J (1968) Vegetation zones and their sections in north-western Europe. Ann Bot Fenn 5:169–211

Bergström J (1988) Bedrock map of Scania, Sveriges Geologiska Undersökningen. Swedish Geological Survey, Lund

Blaauw M (2010) Methods and code for "classical" age-modelling of radiocarbon sequences. Quat Geochronol 5:512–518

Broström A, Nielsen AB, Gaillard M-J et al (2008) Pollen productivity estimates of key European plant taxa for quantitative reconstruction of past vegetation: a review. Veget Hist Archaeobot 17:461–478

Cui QY, Gaillard M-J, Lemdahl G, Sugita S, Greisman A, Jacobson GL, Olsson F (2013) The role of tree composition in Holocene fire history of the hemiboreal and southern boreal zones of southern Sweden, as revealed by the application of the landscape reconstruction algorithm: implications for biodiversity and climate-change issues. Holocene 23:1,747–1,763

Cui QY, Gaillard M-J, Lemdahl G, Stenberg L, Sugita S, Zernova G (2014) Historical land-use and landscape change in southern Sweden and implications for present and future biodiversity. Ecol Evol 4:3,555–3,570. doi:10.1002/ece3.1198

Daniel E (1994) Beskrivning av jordartskartan Växjö SO [Description to the quaternary map Växjö SO]. Sveriges Geologiska Undersökning, Serie Ae 119. SGU, Uppsala

Daniel E (2002) Beskrivning av jordartskartan 5F Åseda SV [Description to the quarternary map 5F Åseda SV]. Sveriges Geologiska Undersökning, Serie Ae 149. SGU, Uppsala

Ekström G (1946) Jordartskarta över Skåne. Sveriges Geologiska Undersökning, Lund

Fyfe R, Twiddle C, Sugita S et al (2013) The Holocene vegetation cover of Britain and Ireland: overcoming problems of scale and discerning patterns of openness. Quat Sci Rev 73:132–148

Gaillard M-J (1984) A palaeohydrological study of Krageholmssjön (Scania, South Sweden)—regional vegetation history and water-level changes. Lundqua Report 13. Lund University, Department of Quaternary Geology, Lund

Gaillard M-J, Birks HJB, Emanuelsson U, Karlsson S, Lagerås 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 82:47–73

Gaillard M-J, Sugita S, Mazier F et al (2010) Holocene land-cover reconstructions for studies on land cover-climate feedbacks. Clim Past 6:483–499

Hellman S, Gaillard M-J, Broström A, Sugita S (2008a) 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 23:21–42

Hellman S, Gaillard M-J, Broström A, Sugita S (2008b) Effects of the sampling design and selection of parameter values on pollen-based quantitative reconstructions of regional vegetation: a case study in southern Sweden using the REVEALS model. Veget Hist Archaeobot 17:445–459

Hellman S, Bunting MJ, 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 153:245–258

Hellman S, Gaillard M-J, Bunting MJ, 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 153:259–271

Hultberg T, Gaillard M-J, Grundmann B, Lindbladh M (2015) Reconstruction of past landscape openness using the landscape reconstruction algorithm (LRA) applied on three local pollen sites in a southern Swedish biodiversity hotspot. Veget Hist Archaeobot. 24:253–266. doi:10.1007/s00334-014-0469-8

Kaplan JO, Krumhardt KM, Zimmermann N (2009) The prehistoric and preindustrial deforestation of Europe. Quat Sci Rev 28:3,016–3,034

Lepš J, Šmilauer P (2003) Multivariate analysis of ecological data using CANOCO. Cambridge University Press, New York

Marquer L, Gaillard M-J, Sugita S et al (2014) Holocene changes in vegetation composition in northern Europe: why quantitative pollen-based vegetation reconstructions matter. Quat Sci Rev 90:199–216

Mazier F, Galop D, Brun C, Buttler A (2006) Modern pollen assemblages from grazed vegetation in the western Pyrenean Mountains (France): a numerical tool for more precise reconstruction of past cultural landscapes. Holocene 16:91–103

Mazier F, Gaillard M-J, Kuneš P, Sugita S, Trondman A-K, Broström A (2012) Testing the effect of site selection and parameter setting on REVEALS-model estimates of plant abundance using the Czech Quaternary Palynological Database. Rev Palaeobot Palynol 187:38–49

Mazier F, Broström A, Bragée B, Fredh D, Stenberg L, Thiere G, Sugita S, Hammarlund D (2015) Two hundred years of changing land-use in the South Swedish Uplands: comparison of historical map-based estimates with a pollen-based reconstruction using the landscape reconstruction algorithm. Veget Hist Archaeobot. doi:10.1007/s00334-015-0516-0

Nielsen AB, Odgaard BV (2010) Quantitative landscape dynamics in Denmark through the last three millennia based on the landscape reconstruction algorithm approach. Veget Hist Archaeobot 19:375–387

Nielsen AB, Giesecke T, Theuerkauf M et al (2012) Quantitative reconstructions of changes in regional openness in north-central Europe reveal new insights into old questions. Quat Sci Rev 47:131–149

Pirzamanbein B, Lindström J, Poska A et al (2014) Creating spatially continuous maps of past land cover from point estimates: a new statistical approach applied to pollen data. Ecol Complex 20:127–141

Prentice IC (1985) Pollen representation, source area, and basin size: toward a unified theory of pollen analysis. Quat Res 23:76–86

Prentice IC (1988) Records of vegetation in time and space: the principles of pollen analysis. In: Huntley B, Webb T III (eds) Vegetation history. Handbook of vegetation science. Kluwer Academic Publishers, Dordrecht, pp 17–42

Regnéll J (1989) Vegetation and land use during 6000 years. Palaeoecology of the cultural landscape at two lake sites in Southern Skåne, Sweden. Lundqua Thesis 27. Lund University, Department of Quaternary Geology, Lund

Reimer PJ, Baillie MGL, Bard E et al (2009) IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. Radiocarbon 51:1,111–1,150

Smith B, Prentice IC, Sykes MT (2001) Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. Glob Ecol Biogeogr 10:621–637

Soepboer W, Sugita S, Lotter AF (2010) Regional vegetation-cover changes on the Swiss Plateau during the past two millennia: a pollen-based reconstruction using the REVEALS model. Quat Sci Rev 29:472–483

Strandberg G, Kjellström E, Poska A et al (2014) Regional climate model simulations for Europe at 6 and 0.2 k BP: sensitivity to changes in anthropogenic deforestation. Clim Past 10:661–680

Sugita S (1993) A model of pollen source area for an entire lake surface. Quat Res 39:239–244

Sugita (2007a) Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition. Holocene 17:229–241

Sugita (2007b) Theory of quantitative reconstruction of vegetation II: all you need is LOVE. Holocene 17:243–257

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 74:289–300

Ter Braak CJF, Šmilauer P (2002) CANOCO reference manual and CanoDraw for Windows user guide: software for canonical community ordination (version 4,5). Microcomputer Power, Ithaca

Trondman A-K (2014) Pollen-based quantitative reconstruction of land-cover change in Europe from 11,500 years ago until present—a dataset suitable for climate modelling. Linnaeus University Dissertations 193/2014. Linnaeus University Press, Växjö

Trondman A-K, Gaillard M-J, Mazier F et al (2015) Pollen-based quantitative reconstructions of Holocene regional vegetation cover (plant functional types and land-cover types) in Europe suitable for climate modelling. Glob Chang Biol 21:676–697. doi:10.1111/gcb.12737