Testing the effect of site selection and parameter setting on REVEALS-model estimates of plant abundance using th Czech Quaternary Palynological database

Florence Mazier, M.J. Gaillard, P. Kunes, S. Sugita, A.-K. Trondman, Anna Brostrom

Review of Palaeobotany and Palynology 187 (2012) 38–49

Contents lists available at SciVerse ScienceDirect

Review of Palaeobotany and Palynology

journal homepage: www.elsevier.com/locate/revpalbo

Research paper

Testing the effect of site selection and parameter setting on REVEALS-model estimates of plant abundance using the Czech Quaternary Palynological Database

F. Mazier [a,b,f,*], M.-J. Gaillard [b], P. Kuneš [c,d], S. Sugita [e], A.-K. Trondman [b], A. Broström [f]

[a] CNRS GEODE UMR 5602, Toulouse University Le Mirail, Toulouse, France
[b] School of Natural Sciences, Linnaeus University, Kalmar, Sweden
[c] Institute of Earth Sciences, Aarhus University, Aarhus, Denmark
[d] Department of Botany, Faculty of Science, Charles University in Prague, Prague, Czech Republic
[e] Institute of Ecology, Tallinn University, Tallinn, Estonia
[f] Department of Earth and Ecosystem Sciences, Lund University, Lund, Sweden

ARTICLE INFO

Article history:
Received 24 February 2012
Received in revised form 10 July 2012
Accepted 20 July 2012
Available online 30 August 2012

Keywords:
quantitative reconstruction of past land-cover
pollen data
Holocene
NW Europe
REVEALS model

ABSTRACT

REVEALS-based quantitative reconstruction of Holocene vegetation cover (expressed in plant functional types, PFTs) is used in the LANDCLIM project to assess the effect of human-induced land-cover change on past climate in NW Europe. Using the Czech Quaternary Pollen Database, this case study evaluates the extent to which selection of data and input parameters for the REVEALS model applications would affect reconstruction outcomes. The REVEALS estimates of PFTs (grid-cell based REVEALS PFT estimates, GB REVEALS PFT-s) are calculated for five time windows of the Holocene using fossil pollen records available in each $1°\times1°$ grid cell of the Czech Republic. The input data and parameters selected for testing are: basin type and size, number of ^{14}C dates used to establish the chronology of the pollen records, number of taxa, and pollen productivity estimates (PPE). We used the Spearman correlation coefficient to test the hypothesis that there is no association between GB REVEALS PFT-s using different data and parameter inputs. The results show that differences in the basin size and type, number of dates, number and type of taxa (entomophilous included or not), and PPE dataset do not affect the rank orders of the GB REVEALS PFT-s significantly, except for the cases when entomophilous taxa are included. It implies that, given careful selection of data and parameter and interpretation of results, REVEALS applications can use pollen records from lakes and bogs of different sizes together for reconstruction of past land cover at the regional to sub-continental spatial scales for purposes such as the study of past land cover–climate interactions. Our study also provides useful criteria to set up protocols for data compilation REVEALS applications of this kind.

1. Introduction

Quantitative estimates of land-cover changes during the Holocene have become increasingly important for a better understanding of vegetation/land-use–climate interactions on a long-time scale (Gaillard et al., 2010). Over the last few years several studies have attempted to estimate the past anthropogenic land-cover changes, providing information to assess the effects of land-cover changes on climate and carbon emissions in the past (Klein Goldewijk, 2001; Olofsson and Hickler, 2008; Pongratz et al., 2008; Kaplan et al., 2009; Lemmen, 2009; Klein Goldewijk et al., 2010; Kaplan et al., 2011). These estimates are primarily based on secondary information and assumptions such as the estimates of historical population growth, land suitability, and population size/land-use relationships. However, those scenarios of past anthropogenic land-cover change (ALCC) are inconsistent for several key time periods of the past (Gaillard et al., 2010; Boyle et al., 2011; Kaplan et al., 2011). Paleorecords, particularly pollen-based reconstructions, are thus invaluable to evaluate the existing ALCC scenarios (Gaillard et al., 2010; Boyle et al., 2011).

One of the major objectives of the LANDCLIM (LAND cover–CLIMate interactions in NW Europe during the Holocene) initiative (supported by the Swedish [VR] and Nordic [NordForsk] Research Councils) is to obtain pollen-based reconstructions of past land-cover at a spatial scale relevant to regional climate and vegetation modelling in order to assess the possible effects of land-cover/land-use changes on the past climate (see Gaillard et al., 2010 for description of the LANDCLIM project). The project uses the REVEALS model (Sugita, 2007a) to reconstruct the past land-cover changes in NW Europe (Fig. 1). The study area of the LANDCLIM project is NW and W Europe North of the Alps (Fig. 1), for which pollen productivity estimates (PPEs) of major plant taxa – one of the important parameters for REVEALS – are available

* Corresponding author at: CNRS GEODE UMR 5602, Toulouse University Le Mirail, 5 allée Antonio Machado, F-31058 Toulouse Cedex, France.
E-mail addresses: florence.mazier@univ-tlse2.fr (F. Mazier), marie-jose.gaillard-lemdahl@lnu.se (M.-J. Gaillard), petr@kunes.net (P. Kuneš), sugita@tlu.ee (S. Sugita), anna-kari.trondman@lnu.se (A.-K. Trondman), Anna.Brostrom@geol.lu.se (A. Broström).

0034-6667/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.revpalbo.2012.07.017

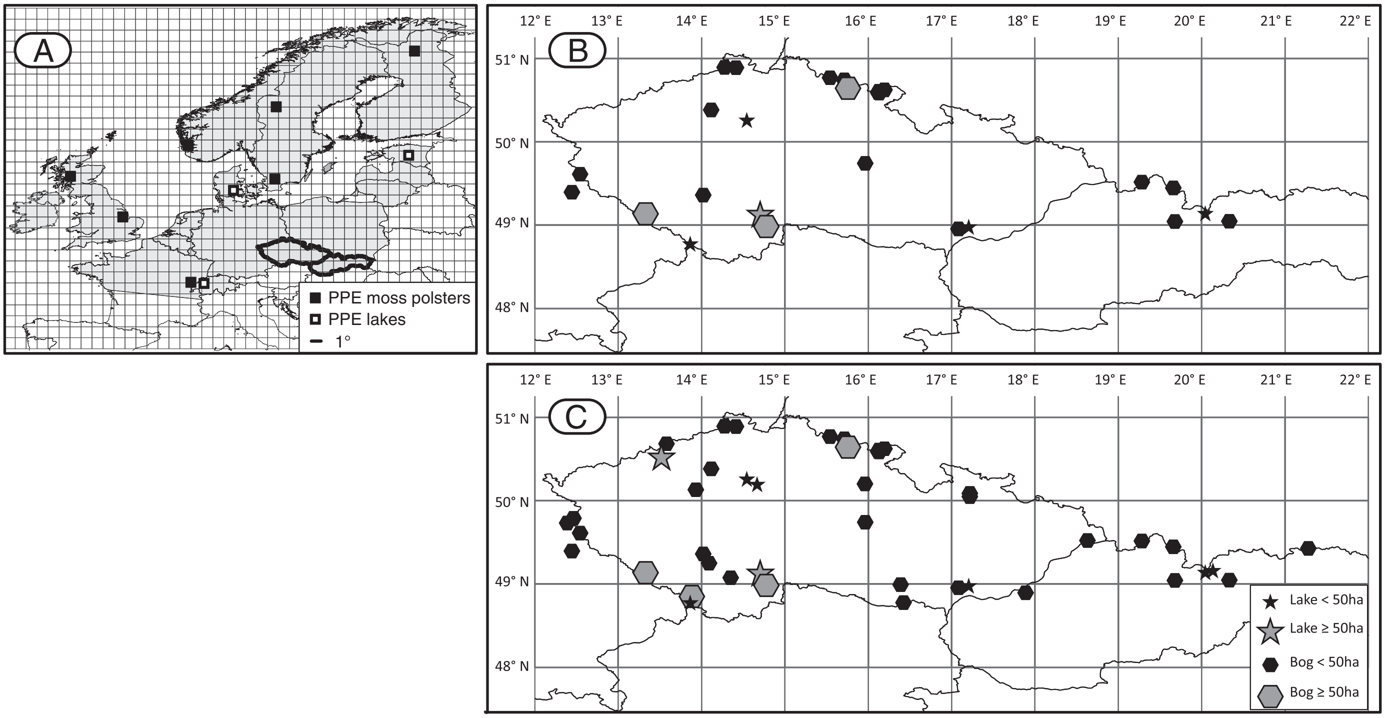

Fig. 1. Study area of A) LANDCLIM project (Gaillard et al., 2010) and B) the Czech Quaternary Palynological Database (PALYCZ) (Kuneš et al., 2009). Pollen productivity estimates (PPEs) for major plant types were obtained from datasets of modern pollen (from moss polsters, filled squares, or surface lake sediments, empty squares) and related vegetation data (Broström et al., 2008). Grid cell-based REVEALS estimates of 10–11 PFTs (GB REVEALS PFT-s) were estimated for 1°×1° grid cells covering the Czech Republic and neighbouring areas (B and C). The dots and stars show the location of bogs and lakes, respectively, when using a minimum of 5 ^{14}C dates (B) or 3 ^{14}C dates (C) to establish the depth-age model. The black and grey colours represent small (<50 ha) and large (≥50 ha) sites, respectively.

(Broström et al., 2008). In a first stage, the project focuses on two time-windows, 100–350 cal. BP (Little Ice Age) and 5700–6200 cal. BP (Early Neolithic), because they represent contrasting anthropogenic land-cover climate characteristics (Gaillard et al., 2010). In addition, land-cover reconstructions from the recent past (x–100 cal. BP, with x = date of the surface of the core), 350–700 cal. BP (Black Death) and 2700–3200 cal. BP (Late Bronze Age) are used for data-model comparison within another ongoing project (LUCCI, Lund University, Broström et al., in prep.).

The pollen–vegetation relationship is influenced by factors such as the size of sedimentary basins and inter-taxonomic differences in pollen productivity and dispersal characteristics. The REVEALS model corrects for biases caused by these factors and provides regional vegetation composition and land-cover within a 10^4–10^5 km^2 area (Sugita, 2007a; Hellman et al., 2008b; Sugita et al., 2010a). There are a number of parameters necessary for the application of the REVEALS model: basin size, basin type (lake or bog), pollen productivity estimates (PPEs) and fall speed of pollen (FSP) for individual modelled taxa, and the maximum extent of the regional vegetation (Sugita, 2007a). Hellman et al. (2008a) have shown that different selection of input data and parameters may affect the REVEALS results. Therefore, it is critical to establish a protocol for parameter setting to ensure consistency of results in research projects such as LANDCLIM.

In order to evaluate the extent to which selection of input data and parameters for REVEALS runs would influence the REVEALS results, this study uses the Czech Quaternary Palynological Database, PALYCZ (Kuneš et al., 2009). Ideally it would have been more logical to use all pollen data available in the study area for testing; however, the collection of pollen records for such a large region is time-consuming, and the data quality and quantity vary from region to region. We thus selected PALYCZ for testing and setting up the protocol of parameter selection for the LANDCLIM project, because it was the best database available in the region in terms of the quality control of chronology, metadata necessary for REVEALS, and ease of use at the start of the project (2009). Similar tests and complementary evaluations of data and parameter selection for REVEALS using data from different parts of the LANDCLIM study area are in progress and will be published elsewhere (e.g. Nielsen et al., 2012). In this paper, we present the rationale behind the tests and discuss the implications of the results for the application of the REVEALS model in general, as well as for the LANDCLIM project itself.

2. Rationale

2.1. The REVEALS model for reconstructing regional vegetation and land-cover

The LANDCLIM project uses the REVEALS model to estimate the percentage cover of plant functional types (PFTs) (Table 1) at a 1°×1° spatial resolution. REVEALS (Sugita, 2007a), a generalized form of the R-value model (Davis, 1963), expresses the regional vegetation composition as "the ratio of the pollen counts of each taxon weighted by its pollen productivity and dispersal term to the total sum of those for all taxa" (Sugita et al., 2010a). The "dispersal term" is a function that includes factors and parameters such as size and type of sedimentary basin (i.e. lake and bog), taxon-specific pollen dispersal and deposition, and spatial extent of regional vegetation (Z_{max} in eq. (1) of Sugita et al., 2010a). In essence, Z_{max} is the spatial resolution of the regional vegetation and land-cover to be reconstructed. Hellman et al. (2008b) showed that the correlation between the plant cover estimates from available vegetation survey data and those modelled by REVEALS in southern Sweden was satisfactory at the spatial scales of 0.25×10^4 km^2 and 10^4 km^2, the correlation being slightly better for the latter. On the Swiss Plateau, the REVEALS estimates of plant cover matched fairly well with those in the actual vegetation within a 200-km radius around the study sites (Soepboer et al., 2010). In addition, models of vegetation dynamics (LPJ-Guess: Smith et al., 2001, 2011) and regional climate (RCA3: Samuelsson et al., 2011), which will be used for data-model

Table 1
Plant functional types (PFTs) and corresponding pollen morphological types. Three data sets of pollen morphological types (GP_1, GP_2, and GP_3) and three data sets of PPEs (PPE.st_1, PPE.st_2, and PPE.st_3) are used to test the effect of different selections of data and parameter inputs on REVEALS estimates. GP_1 (all taxa), GP_2 (all entomophilous taxa excluded), and GP_3 (entomophilous taxa seldom found excluded) include 35, 25 and 28 taxa, respectively. Fall speed of pollen (FSP) and means of PPEs for standards 1, 2 and 3 (with their SE) are also listed (see text for more information and Table 3 for PPE values included in the calculation of the mean).

PFT	PFT definition	Plant taxa/pollen-morphological types	GP1	GP2	GP3	FSP (m/s)	PPE.st1	PPE.st2	PPE.st3
TBE1	Shade-tolerant evergreen trees	*Picea*				0.056	3.24 (0.10)	2.62 (0.12)	2.62 (0.12)
TBE2	Shade-tolerant evergreen trees	*Abies*				0.120	6.88 (1.44)	6.88 (1.44)	6.88 (1.44)
IBE	Shade-intolerant evergreen trees	*Pinus*				0.031	5.12 (0.35)	6.38 (0.45)	6.38 (0.45)
TSE	Tall shrub, evergreen	*Juniperus*				0.016	2.07 (0.04)	2.07 (0.04)	2.07 (0.04)
IBS	Shade-intolerant summergreen trees	*Alnus*				0.021	9.07 (0.10)	9.07 (0.10)	9.07 (0.10)
		Betula				0.024	3.99 (0.17)	3.09 (0.27)	3.99 (0.17)
		Corylus				0.025	1.99 (0.20)	1.99 (0.20)	1.99 (0.20)
		Fraxinus				0.022	1.03 (0.11)	1.03 (0.11)	1.03 (0.11)
		Quercus				0.035	5.83 (0.15)	5.83 (0.15)	5.83 (0.15)
TBS	Shade-tolerant summergreen trees	*Acer*				0.056	0.80 (0.23)	0.80 (0.23)	0.80 (0.23)
		Carpinus				0.042	3.55 (0.43)	3.55 (0.43)	3.55 (0.43)
		Fagus				0.057	3.43 (0.09)	2.35 (0.11)	3.43 (0.09)
		Tilia				0.032	0.80 (0.03)	0.80 (0.03)	0.80 (0.03)
		Ulmus				0.032	1.27 (0.05)	1.27 (0.05)	1.27 (0.05)
TSD	Tall shrub, summergreen	*Salix*				0.022	1.22 (0.11)	1.22 (0.11)	1.79 (0.16)
LSE	Low evergreen shrub	*Calluna vulgaris*				0.038	1.79 (0.17)	0.82 (0.02)	1.09 (0.03)
		Empetrum				0.038	0.11 (0.03)	0.11 (0.03)	0.11 (0.03)
LSD	Low summergreen shrub	Ericaceae				0.038	0.07 (0.04)	0.07 (0.04)	0.07 (0.04)
GL	Grassland – all herbs	Apiaceae				0.042	0.26 (0.009)	0.26 (0.009)	0.26 (0.009)
		Artemisia				0.025	3.48 (0.20)	3.48 (0.20)	3.48 (0.20)
		Comp. SF Cichorioideae				0.051	0.16 (0.02)	0.16 (0.02)	0.16 (0.02)
		Cyperaceae				0.035	0.83 (0.04)	0.87 (0.06)	0.96 (0.05)
		Filipendula				0.006	2.81 (0.43)	2.81 (0.43)	2.81 (0.43)
		Poaceae				0.035	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
		Leucanthemum (Anthemis-t)				0.029	0.10 (0.008)	0.10 (0.008)	0.10 (0.008)
		Plantago lanceolata				0.029	3.97 (0.46)	1.04 (0.09)	1.04 (0.09)
		Plantago media				0.024	1.27 (0.18)	1.27 (0.18)	1.27 (0.18)
		Plantago montana				0.030	0.74 (0.13)	0.74 (0.13)	0.74 (0.13)
		Potentilla-t				0.018	1.19 (0.14)	1.19 (0.13)	1.72 (0.20)
		Ranunculus acris-t				0.014	1.96 (0.36)	1.96 (0.36)	1.96 (0.36)
		Rubiaceae				0.019	2.61 (0.23)	2.61 (0.23)	3.71 (0.34)
		Rumex acetosa-t				0.018	2.14 (0.28)	2.14 (0.28)	0.85 (0.05)
		Trollius				0.013	2.29 (0.36)	2.29 (0.36)	2.29 (0.36)
AL	Agricultural land – cereals	*Cerealia-t*				0.060	1.85 (0.38)	1.85 (0.38)	1.18 (0.04)
		Secale-t				0.060	3.02 (0.05)	3.02 (0.05)	3.02 (0.05)

comparisons in the LANDCLIM project, commonly use 0.25°–1° × 0.25°–1° as a spatial unit. Accordingly, the LANDCLIM project selected a 1° × 1° grid cell as an appropriate spatial scale for both the REVEALS reconstructions and the vegetation- and climate-model runs.

2.2. Selection of basin size and type

In theory the REVEALS model was developed for pollen records from large sites (>100 ha), but extensive simulation studies showed that multiple small-sized sites (<1 ha) can be used when large sites are not available in a region (Sugita, 2007a). However, the performance of REVEALS using pollen from multiple small-sized sites has not been empirically evaluated yet. Moreover, the minimum number of small sites required to obtain reasonable outcomes depends on spatial patterns of vegetation. Empirical testing of REVEALS using small sites is currently underway in southern Sweden and Scotland (Trondman et al., in progress; Fyfe et al., in progress). REVEALS also assumes that no pollen-bearing plants grow on a sedimentary basin; thus pollen records from bogs could be problematic. This study is a first step in clarifying the robustness and reliability of the REVEALS results when fossil pollen data from differently-sized sites and/or different site types are used for data analysis.

2.3. Selection of dated pollen records and number of pollen records

The chronological control of the pollen records is essential for every paleoecological study. The pollen records with the best chronological control should be selected as much as possible. However, the stricter the criteria used for chronological quality, the lower the number of sites available. In addition, the number of pollen records from large sites is limited in Europe and elsewhere; thus, we are bound to use multiple small-sized sites to increase the spatial cover of the REVEALS reconstructions. It is important to choose a reasonable compromise between quality of chronology as well as pollen data, and number of sites used (i.e. spatial cover of reconstruction achieved). This study evaluates how differences in chronological control, and number and type of sites affect the REVEALS results.

2.4. Selection of pollen taxa and pollen productivity estimates (PPEs)

At the start of the LANDCLIM project (2009), pollen productivity estimates (PPEs) and fall speed of pollen (FSPs) were available for 36 taxa (15 tree and 21 herb taxa) from nine study areas in Europe (Hjelle, 1998; Sugita et al., 1999; Broström et al., 2004; Nielsen, 2004, 2005; Bunting et al., 2005; Räsänen et al., 2007; Broström et al., 2008; Mazier et al., 2008; von Stedingk et al., 2008) (Fig. 1). REVEALS assumes that the major agent of pollen transport is wind; thus inclusion of entomophilous taxa may affect the results because of the departure from the assumption. This study also tests and evaluates the extent to which selection of plant taxa, and the inclusion of entomophilous taxa in particular, would affect the REVEALS outcomes.

Hellman et al. (2008a) showed that the selection of PPE values influences the REVEALS estimates in southern Sweden; the PPEs obtained in Denmark for cereals and *Plantago lanceolata* provided REVEALS estimates of vegetation abundance more similar to the survey data than the PPEs obtained in southern Sweden. A review of the PPEs available in Europe (Broström et al., 2008) showed that, in some cases, PPEs for a given taxon could differ significantly between the study areas. Some of those discrepancies might be related to differences in climate, soils, land-use practices, or field methodologies of pollen sampling and vegetation survey (Broström et al., 2008; Bunting and Hjelle, 2010, Bunting et al., in progress). However, the number of PPEs available for individual taxa is still too limited to ascribe and discriminate between causes with confidence. Thus, this study uses means of the PPEs available and evaluates the extent to which selection of a set of PPE values influences the reconstruction results and comparability among the outcomes.

3. Material and methods

3.1. The Czech Quaternary Palynological Database (PALYCZ)

The Czech Quaternary Palynological Database (PALYCZ) contains Holocene pollen records from the Czech Republic with some additional records from Slovakia and bordering areas in Germany (Kuneš et al., 2009). In the database, a pollen record is termed “entity” and is defined as a series of pollen assemblages from a set of levels/depths in a sediment stratigraphy; thus the same site (lake, mire, bog or soil) may have several entities. Hereafter, we use “pollen record” as a synonym of “entity”. The chronology and pollen taxonomy of the 152 unrestricted pollen records are carefully checked, updated and corrected. The pollen nomenclature in PALYCZ is standardized to conform to Beug (2004). The database stores the raw pollen counts and the metadata related to each pollen record, i.e. the name of the author(s), a description of the locality, information on the lithology, the basin type (lake or bog) and size (ha), the radiocarbon dates, the type of age-model, and the number of ^{14}C dates retained in the age-depth mode. The software *clam* (Blaauw, 2010) is used to construct the age-depth models. *Clam* is applied for calibrated BC/AD radiocarbon dates and the most common age-depth model selected is the linear interpolation between the midpoints of the dates using 2 sd calibrated ranges.

3.2. The REVEALS-model runs

The REVEALS model requires raw pollen counts and parameter inputs including site radius (m), fall speed of pollen (FSP, $m.s^{-1}$), pollen productivity estimates (PPEs) and their standard errors (SEs). This study uses a computer programme, reveals.v2.2.2.exe (Sugita, unpublished; hereafter referred to as “REVEALS programme”), assuming neutral atmospheric conditions and wind speed of 3 $m.s^{-1}$ as in Prentice (1985) and Sugita (Sugita, 1993, 1994, 2007a, 2007b). Z_{max}, the maximum spatial extent of the regional vegetation from the centre of the site is set to 50 km, roughly corresponding to a 1°×1° grid cell. Although Z_{max} can be larger, a first test run showed that different values of Z_{max} (i.e. 50, 100 and 200 km) did not affect the REVEALS estimates using the PALYCZ data (results not shown).

For the REVEALS runs pollen counts from all samples in a given time-window are aggregated; the time windows are, (1) modern (age of core top) to 100 cal. BP, (2) 100–350 cal. BP, (3) 350–700 cal. BP, (4) 2700–3200 cal. BP, and (5) 5700–6200 cal. BP. Hereafter we use BP with the meaning “cal. BP” for all ages. The number of pollen records available for the reconstruction varies between grid cells and time windows (Fig. 1, Table 2). The REVEALS programme estimates the mean regional vegetation composition and its standard error for each taxon (Sugita, 2007a, 2007b) in each grid cell with available pollen data for a given time window (hereafter referred to as the grid cell-based REVEALS taxon estimate(s) or GB REVEALS taxon (-s)).

The grid-cell based REVEALS taxon-s are then grouped according to the definition of plant functional types (PFTs) used in the LANDCLIM project (Table 1) and converted to grid cell-based REVEALS estimates for PFTs (hereafter referred to as GB REVEALS PFT (-s)). The PFT classification system for this study and the LANDCLIM project follows Wolf et al. (2008) with a few modifications. The standard error of each GB REVEALS PFT is calculated using the delta method (Stuart and Ord, 1994).

3.3. Tests of the influence of different input data and parameters on the GB REVEALS PFT-s

3.3.1. Effect of basin size and type on GB REVEALS PFT-s

Pollen records from all sites regardless of their size are used to maximize the number of pollen records per grid cell. The site radius (m) is calculated from the total area (ha) provided in the PALYCZ database, assuming that the site shape is circular. The database does not include information on changes in basin size through time, and the lithologies available do not indicate clear transitions from lake sediments to peat over the last 6000 years. Thus, we assume for this study that the basin type and area-based estimate of the radius are consistent through time.

Simulations have shown that pollen assemblages from lakes > 48 ha (mean radius > ca. 390 m) can be used reliably for regional vegetation estimates even when the vegetation and land-cover are patchy

Table 2
Number of pollen records from small and large sites (bogs and lakes) per 1°×1° grid cell used to run REVEALS for each time window. Large sites (≥50 ha) are indicated in bold character. Each grid cell is named by its upper left corner position in *WGS84*. Degrees Minutes Seconds—DMS.

	Criteria number of dates ≥ 5										Criteria number of dates ≥ 3									
	x–100 BP		100–350 BP		350–700 BP		2700–3200 BP		5700–6200 BP		x–100 BP		100–350 BP		350–700 BP		2700–3200 BP		5700–6200 BP	
	Bogs	*Lakes*	*Bogs*	*Lakes*	*Bogs*	*Lakes*	*Bogs*	*Lakes*	*Bogs*	*Lakes*	*Bogs*	*Lakes*	*Bogs*	*Lakes*	*Bogs*	*Lakes*	*Bogs*	*Lakes*	*Bogs*	*Lakes*
N51:E13											1		1	**1**	2	**1**	2		2	**1**
N51:E14	3		3	1	3	1	3	1	1		3		3	1	3	2	3	1	1	
N51:E15	2 + **1**		2 + **1**		2 + **1**		1				6 + **1**		4 + **1**		6 + **1**		4			
N51:E16	2		2		1		2		2		4		3		3		4		4	
N51:E17											2		2		2		1			
N50:E12	2		2		2		2		2		4		3		3		3		4	
N50:E13	**1**		**1**		**1**		**1**		**1**		**1**		**1**		**1**		**1**		**1**	
N50:E14	1		1 + **1**		1 + **1**		1 + **1**		1 + **1**	**1**	2		1 + **1**		2 + **1**		2 + **1**		3 + **1**	**1**
N50:E15	1		1		1		1				1		1		1		1			
N50:E18											1		1		1				1	
N50:E19	2		2		2		3		2		2		2		2		3		2	
N50:E20						1		1				1		1		2	1	2		1
N50:E21											1		1		1		1		1	
N49:E13		1		1		1		1			**1**	1	**1**	1	**1**	1	**1**	1	**1**	
N49:E16											2		2		2		2		1	
N49:E17	1		1		1	1	1	1	1	1	2		2		2	1	2	1	1	1

(Sugita, 2007a). Accordingly "large" sites are defined in this study by sizes ≥ 50 ha.

Depending on the type of site, a different model of pollen dispersal and deposition is used, Sugita's model for lakes and ponds (Sugita, 1993) or Prentice's model for bogs and mires (Prentice, 1985, 1988). For the grid cells that include pollen data from both lakes and bogs, we apply REVEALS separately for the lake and bog data, and then calculate the GB REVEALS taxon-s and PFT-s for lakes, bogs, and lakes + bogs.

In this study, we use the grid cells including pollen records from both large and small sites to evaluate differences in GB REVEALS PFT-s when using i) several pollen records from "small" sites, or ii) one or two pollen records from "large" sites. We also use the grid cells with pollen records from bog and lake sites to test how GB REVEALS PFT-s would differ using data from one basin type or the other.

3.3.2. Effect of number of pollen records and the inclusion of pollen records with three and four ^{14}C dates on GB REVEALS PFT-s

We selected all pollen records with at least three or five ^{14}C dates, regardless of the time length covered by the pollen record, with the restriction that it should at least cover one of the five selected time windows studied. There are in total 47 (3 dates), respectively 23 (5 dates) pollen records in PALYCZ that meet these criteria (Fig. 1, Appendices 1A and 1B). The number of pollen records available in each grid cell and each time window differs depending on whether a threshold of three (d_3) or five (d_5) dates is chosen (Table 2).

We evaluate the extent to which the inclusion of pollen records that have only three or four ^{14}C dates to support their chronology (i.e. 24 pollen records, in total 47 pollen records) affects the GB REVEALS PFT-s when compared to the GB REVEALS PFT-s obtained using only the pollen records with ≥5 ^{14}C dates, i.e. 23 pollen records. Therefore, significant differences between results will be caused either by the difference in the site number and/or in the quality of the chronology. Separating the two causes is not possible.

3.3.3. Effect of type and number of pollen taxa and PPE datasets on GB REVEALS PFT-s

We ascribed the available PPEs and related fall speed of pollen (FSPs) (Tables 1 and 3) to the corresponding pollen morphological types included in the PALYCZ database by taking into account plant morphology, biology, and ecology of the species that are included in the pollen-morphological type. For instance, pollen types such as *Cichorium* type (-t), Compositae SF Cichorioideae, and *Crepis*-t, are classified as Compositae SF Cichorioideae, for which a PPE is available. The pollen sum for e.g. Compositae SF Cichorioideae is then calculated by summing up the counts of the grouped pollen types (Table 1). Appendix 2 lists all pollen morphological types found in the database and their attributed PPE (expressed as a taxon name).

We evaluate how the GB REVEALS PFT-s would differ when three sets of different combinations of pollen types are used (Table 1):

- GP_1 includes all taxa (35) for which reliable PPEs exist.
- GP_2 (25 taxa) excludes all entomophilous taxa (i.e., *Acer*, Apiaceae, Compositae SF Cichorioideae, *Empetrum*, Ericaceae, *Leucanthemum* (*Anthemis*-t), *Potentilla*-t, *Ranunculus acris*-t, Rubiaceae, and *Trollius)*. *Tilia*, *Artemisia* and *Filipendula* are not excluded because they are assumed to be partly anemophilous.
- GP_3 (28 taxa) excludes the entomophilous taxa except three common pollen taxa in NW Europe: *Empetrum* and two human-impact indicators (i.e., Compositae SF Cichorioideae and *Leucanthemum* (*Anthemis*-t)).

The LSD (Low summergreen shrub) PFT (see Table 1) is reconstructed only in GP_1 in which Ericaceae are included. Significant differences in results between the three alternatives above will imply that the inclusion of entomophilous taxa and/or the number of taxa have an effect on the GB REVEALS PFT-s. Discriminating between the two causes will not be possible.

This study also uses three sets of different combinations of PPEs (Table 1) to evaluate the effect of PPE selection on the GB REVEALS PFT-s (Table 1):

- Standard 1 ($PPE.st_1$). For each taxon, calculation of the mean PPE value and its SD uses all available PPE values for that taxon (Table 3) except PPE values that are not significantly different from zero considering the lower bound of its SE, and those values that were considered as uncertain in the original publications (e.g., *Vaccinium* for Finland (Räsänen et al., 2007), *Pinus* for Central Sweden (von Stedingk et al., 2008)). The PPE values that are not used in $PPE.st_1$ are also excluded from $PPE.st_2$ and $PPE.st_3$ below.
- Standard 2 ($PPE.st_2$). For each taxon, calculation of the mean PPE and its SE applies the following rules: (i) when 5 or more estimates of pollen productivity (N ≥ 5) are available for a pollen type, the largest and the smallest estimates (generally outlier values) are excluded, and the mean is calculated using the remaining estimates, (ii) when N = 4, the most deviating value is excluded, and the mean is calculated using the other three, and (iii) when N ≤ 3, the mean is based on all values available.
- Standard 3 ($PPE.st_3$). For each taxon, calculation of the mean PPE and its SE excludes PPE values that are assumed to be outliers or unreliable based on experts' knowledge on the plants involved and the field characteristics and conditions of the related pollen/vegetation studies. For example, the PPEs for Cyperaceae, *Potentilla*-t and Rubiaceae obtained in SW Norway (Hjelle, 1998) and those for *Salix* and *Calluna vulgaris* from Central Sweden (von Stedingk et al., 2008) were assumed to be too low compared to the values obtained in other study areas. Therefore, they were not used in the calculation of the mean PPEs.

The standard errors (SEs) of the mean PPEs are estimated using the delta method (Stuart and Ord, 1994). When individual PPEs did not have any SE (i.e. *Picea* and *Pinus* in southern Sweden, Table 3), SE was set to zero. In such cases, the SE of the mean PPE is expected to be underestimated. All the covariances of the PPEs between individual taxa are set to zero for REVEALS calculations.

3.4. Data analysis: the Spearman rank-order correlation test

The non-parametric Spearman rank-order correlation test measures the degree of association between two pairs of data series that are ranked in two ordered series. The calculated correlations and their significance are used to test whether there is an association or not (Siegel and Castellan, 1988). We chose the Spearman correlation instead of the Pearson product-moment correlation coefficient because significance tests of the latter is entirely appropriate only when two sets of data compared are bivariate normal variables (Sokal and Rohlf, 1981), which is not likely for the REVEALS results.

We compare, in pairs, the following sets of GB REVEALS PFT-s: (i) basin type (bogs—B or lakes—L, two sets), ii) minimum number of ^{14}C dates (3 dates—d_3 or 5 dates—d_5, two sets), iii) number of taxa used (35 taxa — GP_1, 25 taxa — GP_2, or 28 taxa — GP_3, three sets), and iv) PPE selection ($PPE.st_1$, $PPE.st_2$ or $PPE.st_3$ (see Table 1), three sets), i.e. ten pairs of sets in total. For instance, when comparing GP_1 and GP_2 at a given grid cell, time window and basin type, we get the following six combinations of GB REVEALS PFT-s sets: $GP_1.PPE.st_1.d_3$ vs $GP_2.PPE.st_1.d_3$, $GP_1.PPE.st_2.d_3$ vs $GP_2.PPE.st_2.d_3$, $GP_1.PPE.st_3.d_3$ vs $GP_2.PPE.st_3.d_3$, $GP_1.PPE.st_1.d_5$ vs $GP_2.PPE.st_1.d_5$, $GP_1.PPE.st_2.d_5$ vs $GP_2.PPE.st_2.d_5$, and $GP_1.PPE.st_3.d_5$ vs $GP_2.PPE.st_3.d_5$.

We then test the null hypothesis (H_0) that there is no association between two sets of GB REVEALS PFT-s, and use a two-tailed test with the significance probability $p = 0.05$. H_0 is rejected if there is an association between the two sets, i.e. the rank orders of the two sets of GB REVEALS PFT- are not independent.

Table 3

PPEs (with their standard errors SEs) of 21 herb and 15 tree pollen taxa obtained from 9 study areas. The type of surface sample and the ERV submodel used to calculate the PPEs are indicated. The PPE values used for the calculation of the mean PPEs and their corresponding SE for the three PPE datasets st1, st2, and st3 (Table 1) are indicated by the numbers 1 to 3. For comparison with all other PPEs, the British PPEs (originally calculated in relation to Quercus = 1) were recalculated assuming that Quercus PPE (in relation to Poaceae 1.00) in Britain is the same as the mean PPE of the other regions (Table 1). In order to estimate the SEs of the GB REVEALS PFT-s, the SEs of PPEs (unpublished data) are necessary. No SEs were published for the Norwegian PPEs, they were therefore calculated using the original dataset (Hjelle, 1998). Values for the FSP (Table 1) were obtained from the literature (Eisenhut, 1961; Sugita et al., 1999; Broström et al., 2004; Mazier et al., 2008) or, when not available (i.e. Apiaceae and Leucanthemum (Anthemis-type)), they were estimated from size measurements (Beug, 2004) of the pollen types belonging to the plant species included in the Norwegian pollen–vegetation dataset in Hjelle (1998), and using the Stoke's law (Gregory, 1973).

	Moss polster sites used to calculate PPEs						Lake sites used to calculate PPEs		
Country ERV submodel	Finland ERV 3	C Sweden ERV 3	S Sweden ERV 3	Norway ERV 1	England ERV 1	Swiss Jura ERV 1	Estonia ERV 3	Denmark ERV 1	Swiss Plateau ERV 3
Poaceae (Reference taxa)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)		1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Herb taxa									
Apiaceae				0.26 (0.009)					
Artemisia							3.48 (0.20)		
Calluna vulgaris		0.30 (0.03)[3]	4.70 (0.69)[2,3]	1.07 (0.03)				1.10 (0.05)	
Cerealia-t			3.20 (1.14)[3]				1.60 (0.07)	0.75 (0.04)	0.00076 (0.0019)[1,2,3]
Comp. SF. Cichorioideae			0.24 (0.06)	0.06 (0.004)					0.17 (0.03)
Cyperaceae	0.002 (0.0022)[1,2,3]	0.89 (0.03)	1.00 (0.16)	0.29 (0.01)[2,3]		0.73 (0.08)	1.23 (0.09)[2]		
Empetrum	0.07 (0.06)[1,2,3]	0.11 (0.03)							
Ericaceae		0.07 (0.04)							
Filipendula			2.48 (0.82)	3.39 (missing, 0.00)[1,2,3]			3.13 (0.24)		
Leucanthemum (Anthemis-t)				0.10 (0.008)					
Plantago lanceolata			12.76 (1.83)[2,3]	1.99 (0.04)				0.90 (0.23)	0.24 (0.15)
Plantago media						1.27 (0.18)			
Plantago montana						0.74 (0.13)			
Potentilla-t			2.47 (0.38)	0.14 (0.005)[3]		0.96 (0.13)			
Ranunculus acris-t			3.85 (0.72)	0.07 (0.004)					
Rubiaceae			3.95 (0.59)	0.42 (0.01)[3]		3.47 (0.35)			
Rumex acetosa-t			4.74 (0.83)[3]	0.13 (0.004)				1.56 (0.09)	
Secale-t			3.02 (0.05)						
Trollius						2.29 (0.36)			
Vaccinium	0.01 (0.01)[1,2,3]								
Tree taxa									
Abies						3.83 (0.37)			9.92 (2.86)
Acer			1.27 (0.45)			0.32 (0.10)			
Alnus			4.20 (0.14)		8.74 (0.35)		13.93 (0.15)		
Betula	4.6 (0.70)	2.24 (0.20)	8.87 (0.13)[2]		6.18 (0.35)		1.81 (0.02)[2]		2.42 (0.39)
Carpinus			2.53 (0.07)						4.56 (0.85)
Corylus			1.40 (0.04)		1.51 (0.06)				2.58 (0.39)
Fagus			6.67 (0.17)[2]			1.20 (0.16)		5.09 (0.22)	0.76 (0.17)
Fraxinus			0.67 (0.03)		0.70 (0.06)				1.39 (0.21)
Juniperus		0.11 (0.45)[1,2,3]	2.07 (0.04)						
Picea		2.78 (0.21)	1.76 (missing, 0.00)			8.43 (0.30)[2,3]	4.73 (0.13)	1.19 (0.42)	0.57 (0.16)[2,3]
Pinus	8.4 (1.34)	21.58 (2.87)[1,2,3]	5.66 (missing, 0.00)				5.07 (0.06)		1.35 (0.45)[2,3]
Quercus			7.53 (0.08)		5.83 (0.00)		7.39 (0.20)		2.56 (0.39)
Salix		0.09 (0.03)[3]	1.27 (0.31)		1.05 (0.17)		2.31 (0.08)		
Tilia			0.80 (0.03)						
Ulmus			1.27 (0.05)						
Number of taxa	6.00	10.00	26.00	12.00	6.00	11.00	11.00	7.00	13.00

Table 4
Synthesis of the Spearman's rank correlation test results: number of significant (H_0 accepted) and not significant (H_0 rejected) correlations in each grid cell (see text for more explanations on the tests performed). Each set of GB REVEALS PFT-s was tested individually: basin type (two sets), number of dates (two sets), number of taxa (GPs, three sets), and PPE data set (PPE.sts, three sets). The grey cells indicate the few cases where Ho is accepted.

H_1/H_0	N51:E13	N51:E14	N51:E15	N51:E16	N51:E17	N50:E12	N50:E13	N50:E14	N50:E15	N50:E18	N50:E19	N50:E20	N50:E21	N49:E13	N49:E16	N49:E17
d_3 vs d_5		18/–	36/–	45/–		45/–		27/–				18/–				36
GP_1 vs GP_2	33/–	37/2	3/21	26/4	3/9	30/–	12/3	36/–	12/–	12/–	15/–	27/–	15/–	33/6	15/–	51/–
GP_1vs GP_3	33/–	39/–	3/21	30/–	3/9	30/–	12/3	36/–	12/–	12/–	15/–	27/–	15/–	33/6	15/–	51/–
GP_2 vs GP_3	33/–	39/–	24/–	30/–	12/–	30/–	15/–	36/–	12/–	12/–	15/–	27/–	15/–	39/–	15/–	51/–
$PPE.st_1$ vs $PPE.st_2$	33/–	39/–	24/–	30/–	12/–	30/–	15/–	36/–	12/–	12/–	15/–	27/–	15/–	39/–	15/–	51/–
$PPE.st_1$ vs $PPE.st_3$	33/–	39/–	24/–	30/–	12/–	30/–	15/–	36/–	12/–	12/–	15/–	27/–	15/–	39/–	15/–	51/–
$PPE.st_2$ vs $PPE.st_3$	33/–	39/–	24/–	30/–	12/–	30/–	15/–	36/–	12/–	12/–	15/–	27/–	15/–	39/–	15/–	51/–
B vs L	23/4	27/9						18/–				9/–		36/–		45/–

4. Results

4.1. Tests of the influence of different combinations of input data and parameters on the GB REVEALS PFT-s

Table 4 summarizes the results of all coefficients calculated for all pairs of GB REVEALS PFT-s sets for each of the 10 to 11 PFTs, grid cells, and 5 time windows, e.g. a total of 2826 coefficients. The table does not present all individual coefficients, but the number of combinations for which H_0 is rejected vs accepted. In most cases, differences in data and parameter inputs do not affect the correlations GB REVEALS PFT-s, i.e. there is an association between the GB REVEALS PFT-s and H_0 is rejected. Out of 2826 test cases, there are only 97 cases for which there are no associations between the GB REVEALS PFT-s, i.e. H_0 is accepted. Of these 97 cases, 46%, 40% and 13% are related to the correlations GP_1 vs GP_2, GP_1 vs GP_3, and bogs vs lakes, respectively. Moreover, ca. 50% of those cases occur in grid cell N51:E15.

In conclusion, for most pairs of GB REVEALS PFT-s sets, the rank orders of two sets are not independent, i.e. there is an association and H_0 is rejected. It implies that the rank orders are generally insensitive to the differences in number and type of taxa, number of dates, and PPE datasets except for some cases when GP_1 (35 taxa, entomophilous taxa included) is used. Therefore, the choice of input data in terms of taxa selection, number of dates used for the chronology, and PPEs will not influence the rank orders of the GB REVEALS PFT-s, except in some cases with the inclusion of entomophilous taxa.

4.2. Effect of basin size and type on GB REVEALS PFT-s

To evaluate the effect of basin size and type, we chose to maximize the number of pollen records by selecting option d_3 (47 pollen records), to exclude all entomophilous taxa (option GP_2, 25 taxa) and to use the PPE dataset based on objective rules (option $PPE.st_2$), i.e. we used the combination $GP_2.PPE.st_2.d_3$.

We then selected grid cell N51:E15 that includes six pollen records from small bogs and one pollen record from a large bog (60 ha) for the time windows x–100 and 350–700 BP, and four pollen records from small bogs and one pollen record from a large bog for the time window 100–350 BP (Fig. 1, Table 2, Appendices 1A and 1B). Fig. 2 presents the GB REVEALS PFT-s for the individual small (4 and 6) and large (1) bogs, and the mean REVEALS estimates for small bogs (4 and 6) and all bogs (5 and 7 small and large bogs together). The results show that the site-to-site variation of the GB REVEALS PFT-s is large when small bogs are used. Moreover, the GB REVEALS PFT-s obtained with the pollen data from the large bog do not differ significantly from the mean estimates using GB REVEALS PFT-s from several small bogs, except for GB REVEALS Grassland (time window 350–700 BP). In that case, the mean of GB REVEALS PFT using pollen data from several small bogs is significantly higher than that from the large bog. To evaluate the effect of basin type, we selected six grid cells that include pollen data from both bog and lake sites (Fig. 1, Table 2). In two cases (grid cells N51:E13 and N51:E14) and one time window, the rank orders of GB REVEALS PFT-s for bogs are different from those for lakes (Table 4). Moreover, the results from grid cell N51:E14 illustrate the differences between GB REVEALS PFT-s when using pollen records from bogs and/or lakes applying two different sets of parameters $GP_2.PPE.st_2.d_3$ and $GP_1.PPE.st_2.d_3$. The Spearman correlation coefficients and their probability are listed for the three time windows 100–350, 350–700, and 2700–3200 BP in Table 5. The correlation between GB REVEALS PFT-s obtained using pollen data from a single type of basin (lake or bog) is not influenced by the parameter setting, i.e. H_0 is rejected. The same holds for the three time windows. The GB REVEALS PFT-s using pollen assemblages from bogs versus those from lakes are associated (H_0 is rejected) for the time windows 350–700 and 2700–3200 BP and not associated for time window 100–350 BP (H_0 is accepted) (Table 5). It should be noted that the number of lake pollen records available is low, i.e. two for 350–700 BP and only one for the two other time windows.

Fig. 3 illustrates the similarities between the GB REVEALS PFT-s using pollen data from bogs (left side) or lakes (right side) depending on the parameter settings, $GP_2.PPE.st_2.d_3$ or $GP_1.PPE.st_2.d_3$, for the three time windows. The two sets of parameters produced similar GB REVEALS PFT-s when standard errors are taken into account, irrespective of the basin type used. The GB REVEALS TBE1-s (shade-tolerant evergreen trees; Table 1) and their standard errors calculated with pollen data from bogs are similar to those obtained with pollen data from lakes for the three time windows. For the time window 350–700 BP (maximum number of pollen records available, i.e. 3 bogs and 2 lakes), in addition to the GB REVEALS TBE1-s, the GB REVEALS TBE2-s (shade-tolerant evergreen trees), -TBS-s (shade-tolerant summergreen trees), -LSE-s (low evergreen shrubs), -GL-s (grassland herbs), and -AL-s (agricultural land) are similar regardless of the basin type (i.e. H_0 is rejected).

Fig. 3 also demonstrates that in some cases, even when the GB REVEALS PFT-s have significant correlations (H_0 is rejected), the GB REVEALS PFT-s from one type of basin may not necessarily be similar to the GB REVEALS PFT-s from the other type of basin. In all time windows, the GB REVEALS IBE-s (shade-intolerant evergreen trees) show higher values when lakes rather than bogs are used. Further, the GB REVEALS LSD-s (low summergreen shrub) from bogs are higher than those from lakes. For the time-windows 100–350 BP (H_0 accepted) and 2700–3200 BP (H_0 rejected), the GB REVEALS GL-s and AL-s are higher when the lake data rather than bog data are used, while the GB REVEALS TBE2-s, IBS-s, TBS-s, LSE-s and LSD-s are higher when bog data rather than lake data are used.

5. Discussion

5.1. Use of the Spearman correlation test

The Spearman correlation measures the statistical association between two measurements using their rank orders. It is the non-

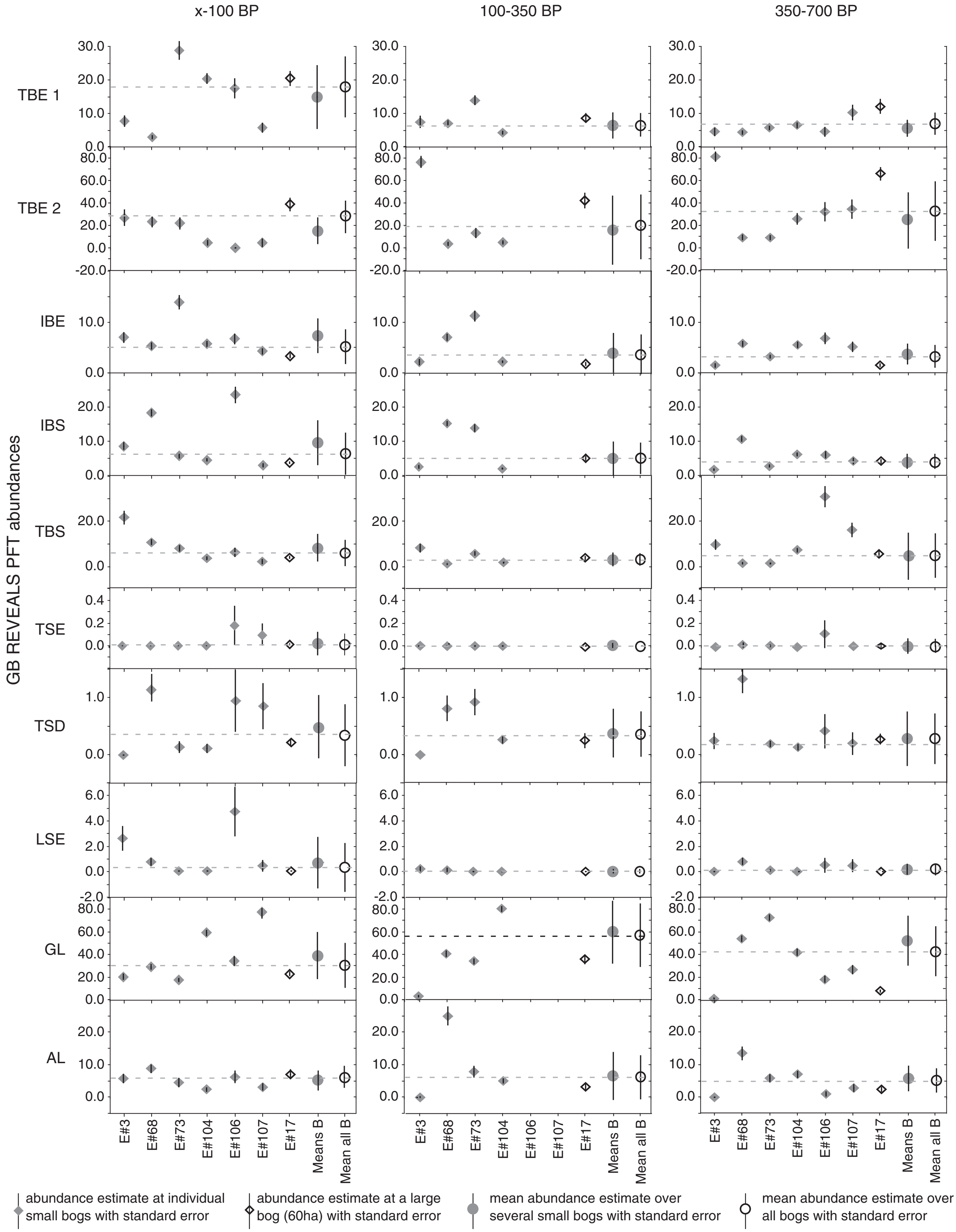

Fig. 2. Grid cell-based REVEALS estimates of PFTs (GB REVEALS PFT-s) in grid cell N51:E15 (Fig. 1, Table 2). Three time windows (x–100, 100–350 and 350–700 BP) are shown. Pollen assemblages from 4–6 small bogs and one large bog were used for the REVEALS reconstructions for each time window. The GB REVEALS PFT-s obtained with pollen data from a single small or large bog, and the mean GB REVEALS PFT-s obtained with pollen data from several small bogs (Means B) and from all bogs (Mean all) are shown with standard errors. The PFTs are: TBE1, shade-tolerant evergreen trees; TBE2, shade-tolerant evergreen trees; IBE, shade-intolerant evergreen; TSE, tall shrub, evergreen; IBS, shade-intolerant summergreen trees; TBS, shade-tolerant summergreen trees; TSD, tall shrub, summergreen; LSE, low evergreen shrub; LSD, low summergreen shrub; GL, grassland-all herbs; and AL, agricultural land-cereals. The label of the pollen record (e#) is also indicated (see Appendices 1A and 1B for more details).

Table 5
Spearman's rank-order correlation coefficients and their related two-sided significance level p (in bracket) for pairs of GB REVEALS PFT-s using different data and parameter inputs in grid cell N51:E14. Bold characters indicate good correlation (see text and legend Table 4).

Site type	Parameter setting	100–350 BP	350–700 BP	2700–3200 BP
Bogs vs lakes	$GP_2.PPE.st_2.d_3$	0.323 (p = 0.33)	**0.818 (p = 0.002)**	**0.634 (p = 0.03)**
Bogs vs lakes	$GP_1.PPE.st_1.d_3$	0.123 (p = 0.71)	**0.736 (p = 0.009)**	**0.669 (p = 0.02)**
Bogs	$GP_2.PPE.st_2.d_3$ vs $GP_1.PPE.st_2.d_3$	**0.855 ($p = 8 \times 10^{-4}$)**	**0.973 ($p = 5.14 \times 10^{-7}$)**	**0.984 ($p = 4.67 \times 10^{-8}$)**
Lakes	$GP_2.PPE.st_2.d_3$ vs $GP_1.PPE.st_2.d_3$	**1 (p = 0)**	**1 (p = 0)**	**1 (p = 0)**

parametric alternative to the Pearson product-moment correlation (Sokal and Rohlf, 1981). In this study, we use the Spearman correlation test to evaluate whether the rank orders of the GB REVEALS PFT estimates are comparable when two different sets of parameters are used for the REVEALS calculations (d_3 vs d_5, B vs L, etc.…). Because of the departures from the basic assumptions and necessary conditions for the REVEALS model (Sugita, 2007a), uncertainties on the results are expected to be large. In addition, the REVEALS results are not normally distributed, which violates one of the major assumptions of the Pearson correlation test. Nevertheless, we also applied the Pearson correlation test for comparison and found that the conclusions were mostly consistent to those based on the Spearman correlation test (results not shown); among 2826 test cases, the results (i.e. H_0 accepted or rejected) are the same for 93.6% of the cases. Therefore, we chose the non-parametric method as the most reasonable test to evaluate the correlations (or associations) between GB REVEALS PFT-s.

5.2. Effect of basin size and type on GB REVEALS PFT estimates

Previous studies have suggested that pollen data from 2–5 large lakes (>100-500 ha) provide reliable estimates of regional vegetation composition using the REVEALS model (Hellman et al., 2008a, 2008b). It was also shown by model simulations and empirical studies that REVEALS can also use pollen records from a number of smaller-sized sites for regional vegetation reconstruction by averaging out the between-site differences in pollen assemblages, although error estimates will usually be large (Sugita, 2007a). Our study is the first empirical test at a larger scale of the theoretical implications of the REVEALS model.

In the Czech Republic and its vicinities there are few pollen records from large sites, while those from small-sized bogs and mires are common, particularly in the mountainous areas. In grid cell N51:E15 (Fig. 2), the ranking orders of the GB REVEALS PFT-s based on the pollen record from the large bog and from several small bog pollen records together are not significantly different (Fig. 2). The same conclusion is drawn from the results in grid cell N50:E14 (three small bogs, one large bog and one large lake). These results suggest that the implication of the simulation outputs in Sugita (2007a) is also true for empirical pollen data. However, the approach of using several small sites instead of one or several large sites within a standard grid system – which is the case in the LANDCLIM project – may have obvious drawbacks. Grouping sites within a grid cell irrespective of their altitudes or locations in relation to vegetation zones (particularly in mountains) might provide non-relevant GB REVEALS PFT-s. In such cases, grouping sites according to topology would be sounder. Further tests of differences in GB REVEALS PFT-s across topologically simple and complex grids are underway within the LANDCLIM project for target grid cells with a sufficient number of sites (Trondman et al., in progress). Nevertheless, and in view of the LANDCLIM project's aims, the selection of a 1°×1° grid-cell system as the operational geographic basis appears to be the most practical, objective, and appropriate approach to produce a first-order approximation of the regional land-cover at a sub-continental spatial scale such as that of NW Europe.

The REVEALS model assumes that no source plants for pollen grow on the basin (Sugita, 2007a). Therefore, pollen records from bogs and fens could be problematic for REVEALS applications. Certain trees (e.g., *Pinus*, *Betula*) and herbs (e.g., *Calluna vulgaris*, Ericaceae, Poaceae and Cyperaceae) frequently grow on bog surfaces and produce pollen. Pollen of those taxa can also originate from other upland vegetation types. Thus, the GB REVEALS PFT-s from bog records might be biased and less representative of the regional vegetation than those from lake records, particularly for GB REVEALS LSE (*C. vulgaris*), LSD (Ericaceae), IBE (*Pinus*), and GL (Poaceae, Cyperaceae included). However, comparison of GB REVEALS PFT-s from bog and lake records in several grid cells does not indicate systematic differences in the results (Table 4). Even though the GB REVEALS PFT-s for LSE and LSD are clearly too high in grid cell N51:E14 (Fig. 3) when small bogs are used, this is not the case for IBE (*Pinus*) and GL (all herbs taxa). Meanwhile, one should keep in mind that Poaceae and Cyperaceae both important components in GL, can also be abundant in the littoral zones of lakes, which may contribute important quantities of these pollen taxa into lakes as well. Moreover, pollen grains from common reed (*Phragmites*) are seldom differentiated from other Poaceae pollen types by pollen analysts; this may explain the high values of Poaceae in the 100–350 and 2700–3200 BP windows (Fig. 3). Similar results are obtained in four other grid cells (i.e. N51:E13, N50:E20, N49:E13, N49:E17). If high pollen counts of Poaceae and Cyperaceae are associated with large lake or bog sites, the obtained high GB REVEALS GL-s might be biased by the site's local vegetation and thus represent unreliable estimates of the regional land cover and also, as a consequence, affect the other GB REVEALS PFT-s in the grid cell. However, it should be noted that the effect of such sites should decrease with the number of sites (bogs and lakes, small and large) available in a grid cell for a given time window, as the influence of pollen counts from a few sites will be smaller if the total pollen count for a grid cell is very large. Therefore, the more sites available, the better the reliability of GB REVEALS PFT-s, which is exemplified in Fig. 3 (350–700 BP). Further empirical tests will be useful for this specific issue.

5.3. Effect of number of pollen records and the inclusion of pollen records with three and four ^{14}C dates on GB REVEALS PFT-s

The choice of options d_3 or d_5 influences the number of pollen records per grid cell and time-window available for application of the REVEALS model. Moreover, the number of pollen records and time windows in each individual grid cell varies. Based on the validation of the REVEALS model using modern pollen samples and vegetation data, Hellman et al. (2008b) suggested that REVEALS estimates are more reliable when using one or two pollen samples per site from several sites in a region rather than several samples from one single site in the same region. Some results in this study also support this observation; the within-site variations of the REVEALS PFT-s can be larger than the between-sites variations (Fig. 2). The GB REVEALS TBE1-s, IBS-s and GL-s for x–100 BP obtained by using three pollen records from the same site show significant within-pollen record differences (Appendices 1A and 1B) while the between-site difference in the same grid cell and time window is small.

Our study also shows that the GB REVEALS PFT-s are not significantly different when the results using pollen records with chronologies based on three or four ^{14}C dates are compared with those using pollen records with five or more ^{14}C dates. This implies that, in the case of the Czech

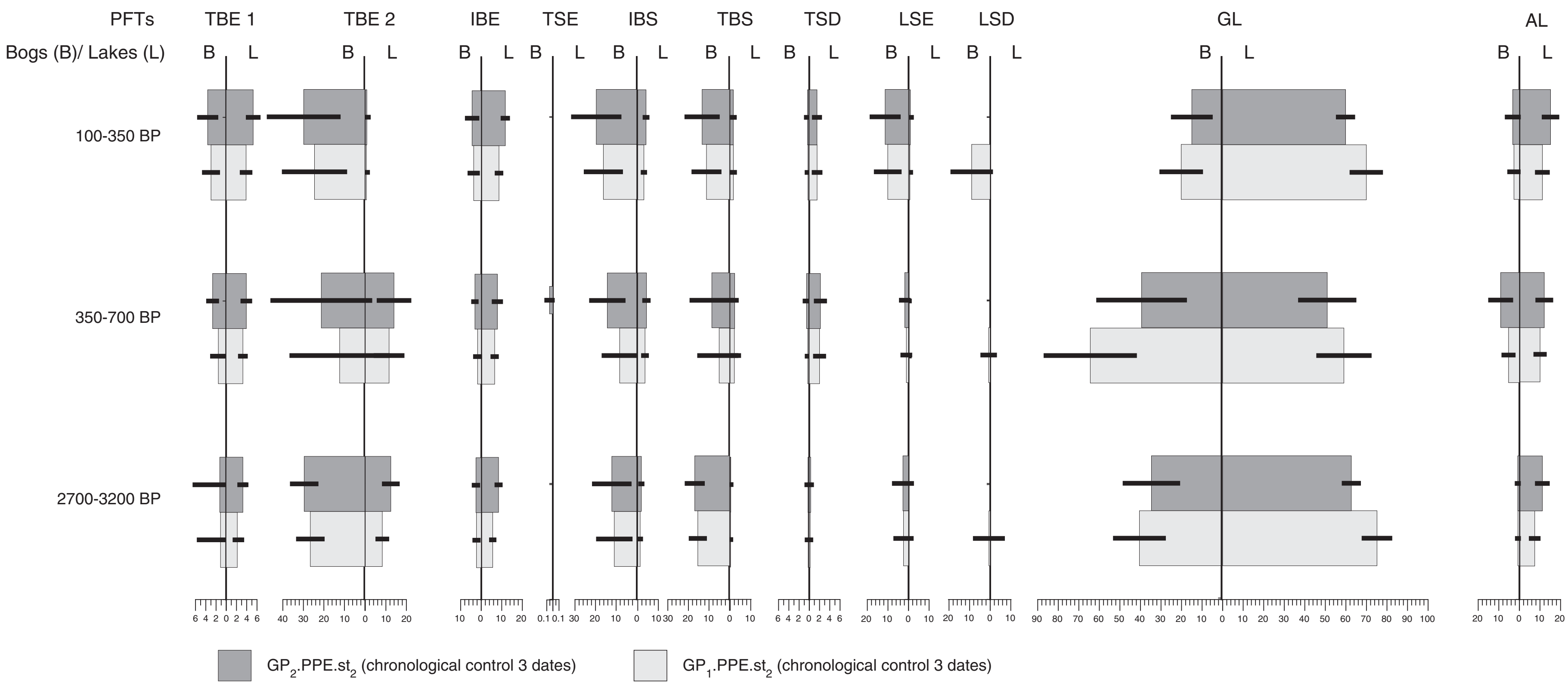

Fig. 3. Grid cell-based REVEALS PFT estimates (GB REVEALS PFT-s) in grid cell N51:E14 (Fig. 1) for which the Spearman-rank order correlation test indicates no significant correlation when comparing results using pollen data from bogs or lakes for the time window 100–350 BP (Table 5). Significant correlation is found for the time windows 350–700 and 2700–3200 BP. Here we compare the GB REVEALS PFT-s obtained with pollen data from bogs (left side) or lakes (right side) for two sets of combinations in terms of selection of taxa and PPEs, i.e. $GP_2.PPE.st_2$ (dark grey) and $GP_1.PPE.st_2$ (light grey) (see section text for details). All sites with a minimum of 3 ^{14}C dates per pollen record are used for this comparison. The PFTs are the same as in Fig. 2 and Table 1.

Republic, the uncertainties related to the chronologies are not severe enough to influence the GB REVEALS PFT-s calculated for time windows of 100 to 500 years. Although we acknowledge that robust chronologies for pollen records of several millennia ought to be based on a large number of ^{14}C dates, the research objectives of the LANDCLIM project justify that a precision of ± ca. 250 years for the selected time windows is sufficient. Moreover, we had to take a decision on whether to compromise on the number of dates or the number of sites. In the first stage of the project, we find it more relevant to work with the largest number of sites possible to achieve higher pollen counts for individual grid cells and higher geographical cover of the REVEALS reconstructions. Testing the effect of using option d_3 rather than d_5 for the GB REVEALS PFT-s in the entire study region of the LANDCLIM project is underway (Trondman et al., in progress).

5.4. Effect of pollen taxa and PPE datasets on GB REVEALS PFT-s

Pollen grains from entomophilous herbaceous taxa (i.e. Ericaceae, *Empetrum*, most of the herbs taxa in GL) are generally less frequent in lake sediments than in bog or mire deposits (peats). In our study entomophilous taxa tend to affect the GB REVEALS PFT-s (Table 4), particularly when pollen records from bogs are used. The differences in the GB REVEALS PFT-s between GP_2 and GP_1 are the most apparent of all tests performed in this study. Therefore, and also because the REVEALS model assumes that all pollen are airborne, it is justified to exclude as many strict entomophilous taxa (i.e. transported by wind only in very low quantities) as possible from REVEALS reconstructions.

The REVEALS model assumes that pollen productivity is constant in space and time. Differences in abiotic (e.g. geology, soil types, and climate), biotic (e.g. genetics, species' range limits, grazing pressure, demography, spatial structure of vegetation, and succession stages), and anthropogenic (e.g. land-use, different forms of agricultural practices, etc.) factors might – and in some cases were shown to – influence pollen productivity estimates (Hicks, 2001; Broström et al., 2008; Gaillard et al., 2008; Sugita et al., 2010b; Mazier et al., 2012). Differences in the methods and sampling designs can also affect the estimates (Bunting and Hjelle, 2010). Our study shows, however, that the rank orders of the GB REVEALS PFT-s are consistent in all cases regardless of the PPE dataset used. Therefore, we select $PPE.st_2$ for the first generation of LANDCLIM GB REVEALS PFT-s in NW Europe, because we consider the rules and assumptions applied for $PPE.st_2$ to be more objective and straightforward than those for $PPE.st_1$ and st_3.

Uncertainties associated with pollen productivity estimates and the size of pollen counts also affect SE estimates of the REVEALS reconstruction. With total pollen counts of 300–500 grains per sample, standard errors could be 10–20% larger than those with pollen counts of 1000 grains (Sugita, 2007a). In this paper, summed pollen counts in time intervals of 100–500 years are used; total pollen counts in a single time window and pollen record varies from 238 to 35,841 grains. Among all pollen records used, only ten have total pollen counts < 500 in one of the time windows tested (Appendices 1A and 1B). Pollen counts are in general high in the pollen records included in the PALYCZ database, which implies relatively small SE estimates for the obtained GB REVEALS PFT-s, which is not always the case for other parts of the LANDCLIM project study region.

6. Conclusions

Using the Spearman rank coefficient text, we have shown that different selections of data and parameter inputs applied on the Czech pollen database do not affect the rank orders of GB REVEALS PFT-s significantly, except for the use of entomophilous taxa. In general, the results suggest that there are no significant differences between rank orders of GB REVEALS PFT-s obtained with pollen data from lakes or bogs. The number of ^{14}C dates used for the chronology (i.e. ≥3 or ≥5) and the PPE dataset do not influence the results notably either. Therefore, we chose the following protocol for the LANDCLIM project's first generation of GB REVEALS PFT-s (Trondman et al., in prep.): we use options d_5, GP_2 and $PPE.st_2$, and all pollen records within a grid cell for a given time window, i.e. those from lakes and bogs, large and small. In this way, we increase the number of pollen records per grid cell and the number of grid cells with pollen records, which in turn should generally increase the quality of the REVEALS results per grid cell and maximise the geographical cover of the GB REVEALS PFT-s.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.revpalbo.2012.07.017.

Acknowledgments

We thank all data contributors of the Czech Quaternary Pollen Database. The LANDCLIM (LAND cover–CLIMate interactions in NW Europe during the Holocene) project and research network were sponsored by the Swedish Research Council (VR) and Nordic Council of Ministers (NordForsk), respectively, and coordinated by Marie-José Gaillard. The two initiatives are a contribution to the IGBP-PAGES-Focus 4 PHAROS programme (http://www.pages-igbp.org/science/focus4) on human impact on environmental changes in the past and to the Swedish Strategic Area "ModElling the Regional and Global Earth system, MERGE". Other funding sources include the Faculty of Natural Sciences and Technology of Linnaeus University (Kalmar-Växjö, Sweden) and MERGE for Gaillard and Trondman, the Estonian Science Foundation Mobilitas Programme (MTT3) and Estonian Ministry of Education grant (SF0280016 S07) for Sugita, the Czech Science Foundation (P504/12/0649) and Czech Academy of Sciences (IAAX00020701) for Kuneš, the VR Linnaeus grant "Lund Centre for studies of Carbon Cycle and Climate Interaction, LUCCI" (grant number: VR 349-2007-8705) and the French National Center of Research for Mazier.

References

Beug, H.-J., 2004. Leitfaden der Pollenbestimmung für Mitteleuropa und angrenzende Gebiete. Verlag Dr Friedrich Pfeil, Munich.

Blaauw, M., 2010. Methods and code for "classical" age-modelling of radiocarbon sequences. Quaternary Geochronology 5 (5), 512–518.

Boyle, J.F., Gaillard, M.-J., Kaplan, J.O., Dearing, J.A., 2011. Modelling prehistoric land use and carbon budgets: a critical review. The Holocene 21 (5), 715–722.

Broström, A., Sugita, S., Gaillard, M.-J., 2004. Pollen productivity estimates for the reconstruction of past vegetation cover in the cultural landscape of southern Sweden. The Holocene 14 (3), 368–381.

Broström, A., Nielsen, A.B., Gaillard, M.-J., Hjelle, K.L., Mazier, F., Binney, H., Bunting, M.J., Fyfe, R., Duffin, K., Meltsov, V., Poska, A., Räsänen, S., Soepboer, W., von Stedingk, H., Suutari, H., Sugita, S., 2008. Pollen productivity estimates of key European plant taxa for quantitative reconstruction of past vegetation — a review. Vegetation History and Archaeobotany 17, 461–478.

Bunting, J.M., Hjelle, K.L., 2010. Effect of vegetation data collection strategies on estimates of relevant source area of pollen (RSAP) and relative pollen productivity estimates (relative PPE) for non-arboreal taxa. Vegetation History and Archaeobotany 19, 365–374.

Bunting, M.J., Armitage, R., Binney, H.A., Waller, M., 2005. Estimates of 'relative pollen productivity' and 'relevant source area of pollen' for major tree taxa in two Norfolk (UK) woodlands. The Holocene 15 (3), 459–465.

Davis, M.B., 1963. On the theory of pollen analysis. American Journal of Science 261, 897–912.

Eisenhut, G., 1961. Untersuchungen über die Morphologie und Ökologie der Pollenkörner heimischer und fremdländischer Waldbaüme. Paul Parey, Hamburg.

Gaillard, M.-J., Sugita, S., Bunting, M.J., Middleton, R., Broström, A., Caseldine, C., Giesecke, T., Hellman, S., Hicks, S., Hjelle, K.L., Langdon, C., Nielsen, A.B., Poska, A., Von Stedingk, H., Veski, S., 2008. The use of modelling and simulation approach in reconstructing past landscapes from fossil pollen data: a review and results from the POLLANDCAL network. Pollandcal Members 17, 419–443.

Gaillard, M.J., Sugita, S., Mazier, F., Trondman, A.K., Broström, A., Hickler, T., Kaplan, J.O., Kjellström, E., Kokfelt, U., Kuneš, P., Lemmen, C., Miller, P., Olofsson, J., Poska, A., Rundgren, M., Smith, B., Strandberg, G., Fyfe, R., Nielsen, A.B., Alenius, T., Balakauskas, L., Barnekow, L., Birks, H.J.B., Bjune, A., Björkman, L., Giesecke, T., Hjelle, K.L., Kalnina, L., Kangur, M., van der Knaap, W.O., Koff, T., Lagerås, P., Latalowa, M., Leydet, M., Lechterbeck, J., Lindbladh, M., Odgaard, B., Peglar, S., Segerström, U., von Stedingk, H., Seppä, H., 2010. Holocene land-cover reconstructions for studies on land cover–climate feedbacks. Climate of the Past 6 (4), 483–499.

Gregory, P.H., 1973. The microbiology of the atmosphere. A Plant Science Monograph. Leonard Hill, Aylesbury.

Hellman, S., Gaillard, M.-J., Broström, A., Sugita, S., 2008a. 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. Vegetation History and Archaeobotany 17, 445–459.
Hellman, S., Gaillard, M.-J., Broström, A., Sugita, S., 2008b. The REVEALS model, a new tool to estimate past regional plant abundance from pollen data in large lakes — validation in southern Sweden. Journal of Quaternary Science 23 (1), 21–42.
Hicks, S., 2001. The use of annual arboreal pollen deposition values for delimiting tree-lines in the landscape and exploring models of pollen dispersal. Review of Palaeobotany and Palynology 117, 1–29.
Hjelle, K.L., 1998. Herb pollen representation in surface moss samples from mown meadows and pastures in Western Norway. Vegetation History and Archaeobotany 7 (2), 79–96.
Kaplan, J.O., Krumhardt, K., Zimmermann, N., 2009. The prehistoric and preindustrial deforestation of Europe. Quaternary Science Reviews 28, 3016–3034.
Kaplan, J.O., Krumhardt, K., Ellis, E.C., Ruddiman, W.F., Lemmen, C. Klein, Goldewijk, K., 2011. Holocene carbon emissions as a result of anthropogenic land cover change. The Holocene 21 (5), 775–791.
Klein Goldewijk, K., 2001. Estimating global land use change over the past 300 years: the HYDE database. Global Biochemical Cycles 15 (2), 417–434.
Klein Goldewijk, K., Beusen, A., Janssen, P., 2010. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. The Holocene 20 (4), 565–573.
Kuneš, P., Abraham, V., Kovářík, O., Kopecký, M., PALYCZcontributors, 2009. Czech Quaternary Palynological Database — PALYCZ: review and basic statistics of the data. Prelia 81, 209–238.
Lemmen, C., 2009. World distribution of land cover changes during Pre and Protohistoric Times and estimation of induced carbon releases. Géomorphologie : Relief, Processus, Géomorphogie 4, 303–312.
Mazier, F., Broström, A., Buttler, A., Gaillard, M.-J., Sugita, S., Vittoz, P., 2008. Pollen productivity estimates and relevant source area of pollen for selected plant taxa in a pasture woodland landscape of the Jura Mountains (Switzerland). Vegetation History and Archaeobotany 17, 479–495.
Mazier, F., Nielsen, A.B., Broström, A., Sugita, S., Hicks, S., 2012. Signals of tree volume and temperature in a high-resolution record of pollen accumulation rates in northern Finland. Journal of Quaternary Science. http://dx.doi.org/10.1002/jqs.2549.
Nielsen, A.B., 2004. Modelling pollen sedimentation in Danish lakes at C. AD 1800: an attempt to validate the POLLSCAPE model. Journal of Biogeography 31, 1693–1709.
Nielsen, A.B., 2005. Quantifying the relationship between pollen sedimentation in lakes and land cover using historical maps. Geological Survey of Denmark and Greenland Bulletin 7, 49–52.
Nielsen, A.B., Giesecke, T., Theuerkauf, M., Feeser, I., Behre, K.E., Beug, H.J., Chen, S.H., Christiansen, J., Dörfler, W., Endtmann, E., Jahns, S., de Klerk, P., Kühl, N., Latalowa, M., Vad Odgaard, B., Rasmussen, P., Stockholm, J.R., Voigt, R., Wiethold, J., Wolters, S., 2012. Quantitative reconstructions of changes in regional openness in north-central Europe reveal new insights into old questions. Quaternary Science Reviews 47, 131–149.
Olofsson, J., Hickler, T., 2008. Effects of human land-use on the global carbon cycle during the last 6,000 years. Vegetation History and Archaeobotany 17 (5), 605–615.
Pongratz, J., Reick, C., Raddatz, T., Claussen, M., 2008. A reconstruction of global agricultural areas and land cover for the last millennium. Global Biochemical Cycles 22 (3), GB3018. http://dx.doi.org/10.1029/2007GB003153.
Prentice, I.C., 1985. Pollen representation, source area, and basin size: toward a unified theory of pollen analysis. Quaternary Research 23 (1), 76–86.
Prentice, I.C., 1988. Records of vegetation in time and space: the principles of pollen analysis. In: Huntley, B., Webb, T. (Eds.), Vegetation history. : Handbook of Vegetation Science. Kluwer, Dordrecht, pp. 17–42.
Räsänen, S., Suutari, H., Nielsen, A.B., 2007. A step further towards quantitative reconstruction of past vegetation in Fennoscandian boreal forests: pollen productivity estimates for six dominant taxa. Review of Palaeobotany and Palynology 146, 208–220.
Samuelsson, P., Jones, C.G., Willén, U., Ullerstig, A., Gollvik, S., Hansson, U.L.F., Jansson, C., KjellströM, E., Nikulin, G., Wyser, K., 2011. The Rossby Centre Regional Climate model RCA3: model description and performance. Tellus A 63 (1), 4–23.
Siegel, S., Castellan, N.J.J., 1988. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill International Editions, Boston.
Smith, B., Prentice, I.C., Sykes, M.T., 2001. Representation of vegetation dynamics in modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. Global Ecology and Biogeography 10, 621–637.
Smith, B., Samuelsson, P., Wramneby, A., Rummukainen, M., 2011. A model of the coupled dynamics of climate, vegetation and terrestrial ecosystem biogeochemistry for regional applications. Tellus A 63 (1), 87–106.
Soepboer, W., Sugita, S., Lotter, A.F., 2010. Regional vegetation-cover changes on the Swiss Plateau during the past two millennia: a pollen-based reconstruction using the REVEALS model. Quaternary Science Reviews 29 (3–4), 472–483.
Sokal, R., Rohlf, F.J., 1981. Biometry, 2nd edition. W.H. Freeman and company, New York.
Stuart, A., Ord, J.K., 1994. Kendall's Advanced Theory of Statistics. Volume 1. Distribution Theory. Edward Arnold, London.
Sugita, S., 1993. A model of pollen source area for an entire lake surface. Quaternary Research 39, 239–244.
Sugita, S., 1994. Pollen representation of vegetation in quaternary sediments : theory and method in patchy vegetation. Journal of Ecology 82, 881–897.
Sugita, S., 2007a. Theory of quantitative reconstruction of vegetation. I: pollen from large lakes REVEALS regional vegetation composition. The Holocene 17 (2), 229–241.
Sugita, S., 2007b. Theory of quantitative reconstruction of vegetation II: all you need is LOVE. The Holocene 17 (2), 243–257.
Sugita, S., Gaillard, M.-J., Broström, A., 1999. Landscape openness and pollen records: a simulation approach. The Holocene 9, 409–421.
Sugita, S., Parshall, T., Calcote, R., Walker, K., 2010a. Testing the Landscape Reconstruction Algorithm for spatially explicit reconstruction of vegetation in northern Michigan and Wisconsin. Quaternary Research 74 (2), 289–300.
Sugita, S., Hicks, S., Sormunen, H., 2010b. Absolute pollen productivity and pollen–vegetation relationships in northern Finland. Journal of Quaternary Science 25 (5), 724–736.
von Stedingk, H., Fyfe, R., Allard, A., 2008. Pollen productivity estimates from the forest tundra ecotone in west-central Sweden: implications for vegetation reconstruction at the limits of the boreal forest. The Holocene 18 (2), 323–332.
Wolf, A., Callaghan, T., Larson, K., 2008. Future changes in vegetation and ecosystem function of the Barents Region. Climatic Change 87 (1), 51–73.