Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications

J. Whitmore[a], K. Gajewski[a,*], M. Sawada[a], J.W. Williams[b,1], B. Shuman[b], P.J. Bartlein[c], T. Minckley[d], A.E. Viau[a], T. Webb III[e], S. Shafer[f], P. Anderson[g], L. Brubaker[h]

[a]*Department of Geography, University of Ottawa, Ottawa, Ont., Canada K1N 6N5*
[b]*Department of Geology and Geophysical Sciences, Limnological Research Center, University of Minnesota, MN 55455-0219, USA*
[c]*Department of Geography, University of Oregon, Eugene, OR 97403-1251, USA*
[d]*Department of Geography, Geology, and Anthropology, Indiana State University, Terre Haute, IN 47809, USA*
[e]*Department of Geological Sciences, Brown University, Providence, RI 02912, USA*
[f]*US Geological Survey, 200 SW 35th St., Corvallis, OR 97333 USA*
[g]*Quaternary Science Center, University of Washington, Seattle, WA 98195-1360, USA*
[h]*College of Forest Resources, University of Washington, Seattle, WA 98195-1360, USA*

Abstract

The modern pollen network in North America and Greenland is presented as a database for use in quantitative calibration studies and paleoenvironmental reconstructions. The georeferenced database includes 4634 samples from all regions of the continent and 134 pollen taxa that range from ubiquitous to regionally diagnostic taxa. Climate data and vegetation characteristics were assigned to every site. Automated and manual procedures were used to verify the accuracy of geographic coordinates and identify duplicate records among datasets, incomplete pollen sums, and other potential errors. Data are currently available for almost all of North America, with variable density. Pollen taxonomic diversity, as measured by the Shannon–Weiner coefficient, varies as a function of location, as some vegetation regions are dominated by one or two major pollen producers, while other regions have a more even composition of pollen taxa. Squared-chord distances computed between samples show that most modern pollen samples find analogues within their own vegetation zone. Both temperature and precipitation inferred from best analogues are highly correlated with observed values but temperature exhibits the strongest relation. Maps of the contemporary distribution of several pollen types in relation to the range of the plant taxon illustrate the correspondence between plant and pollen ranges.
© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past 40 years, paleoecologists have accumulated sufficient paleoenvironmental data for North America and Greenland to enable the analysis of Quaternary environments at several spatial and temporal scales. The mapping and statistical analyses of these data in parallel with climate modelling studies have led to significant advances in our understanding of past environmental conditions (e.g. CLIMAP Members, 1976; COHMAP Members, 1988). Paleoecologists have established large databases of radiocarbon-dated fossil pollen data to facilitate such studies (www.ngdc.noaa.gov/paleo).

Modern pollen samples, cross-referenced with current distributions of climate and vegetation, provide the data required to infer temporal variations in vegetation and climate from fossil pollen data (e.g., Prentice et al., 1992; Webb et al., 1993, 1998; Williams et al., 2000; Sawada et al., 2004). Modern pollen databases in North America

---

*Corresponding author. Tel.: +1 613 562 5800x1057; fax: +1 613 562 5145.

E-mail address: gajewski@uottawa.ca (K. Gajewski).

[1]Present address: Department of Geography, University of Wisconsin Madison, Madison, WI 53706, USA.

have generally been assembled on a regional basis (e.g. Avizinis and Webb, 1985) and have received less attention than have fossil data. Recent studies have made initial attempts at building continental datasets for North America (Gajewski et al., 2000; Williams et al., 2000; Sawada, 2001), but there remains a need for a database that spans North America and ice-free Greenland.

In this paper, we present a database of modern-pollen samples with associated climatic and vegetational information useful for quantitative and qualitative interpretations at regional to continental scales in North America and Greenland. Databases serve two important functions: they are archives of the results of research projects and the source of data for new research programs. Because palynologists collect samples from a variety of depositional environments for differing research purposes, the assembled data are of variable quality.

The primary purpose of this paper is to describe an approach for database development that is relevant to other types of paleoenvironmental data. The general approach consists of automated and manual quality-control procedures that rely partly on the concepts and capabilities of geographic information systems (GIS). We discuss major issues, considerations, assumptions and impediments that are involved in the compilation of data from multiple sources.

Secondly, this study documents the history of the development of modern pollen databases in North America and Greenland, and reviews the current status of the available modern pollen network. Of all terrestrial paleoenvironmental data, pollen records are the most spatially extensive, widely accessible, and one of the few terrestrial records available for both temporal and spatial multi-scale applications in paleoecology and paleoclimatology. We describe the spatial scale and quality of the data in various regions of North America and thereby illustrate areas where pollen surface samples are still needed.

Thirdly, we discuss the considerations necessary when using these data in biogeographical and paleoenvironmental contexts. Unlike most other paleoclimate data sources (e.g. diatoms or chironomids), the biogeography of plant species and communities in relation to climate is well established (e.g. for North America: Fowells, 1965; Thompson et al., 1999). Although each plant taxon has particular pollen production and dispersal characteristics, studies have shown that vegetation patterns are adequately sensed by pollen (Webb, 1974; Davis and Webb, 1975; Bradshaw and Webb, 1985; Prentice, 1988; Jackson, 1990; Davis, 2000) and analyses of geographical networks of pollen samples (Jackson and Williams, 2004) may mitigate representation problems (Webb and McAndrews, 1976). The display of large-scale patterns allows over- and under-representation to be assessed through comparison of pollen and plant maps.

1.1. History and contributions of modern pollen databases in North America

The compilation of a modern pollen dataset for use in paleoclimate reconstructions began at the University of Wisconsin, during the late 1960s (Cole, 1969; Webb and Bryson, 1972). During the 1970s, Webb at Brown University collaborated with Davis and McAndrews to create a modern pollen database (BUPD) consisting of hundreds of samples from eastern North America (Davis and Webb, 1975; Webb and McAndrews, 1976) that were obtained from the literature and from contributions by many palynologists (e.g. Richard, 1976; Delcourt et al., 1983; Bradshaw and Webb, 1985). By the 1980s, the modern pollen network comprised several thousand samples with associated metadata and quality control information, largely focusing on eastern North America (Avizinis and Webb, 1985).

Studies resulting from the analysis of these modern and fossil datasets had a major impact on the development of Quaternary paleoecology and paleoclimatology (e.g. Bernabo and Webb, 1977; Webb, 1986; Wright et al., 1993; discussed in Gajewski, 1993). Parallel efforts by Huntley and Birks (1983) resulted in the development of a European atlas of pollen-percentage changes through time. Other regional databases were later established, including datasets from Québec (Richard), Canada (Gajewski and Ritchie), southeastern United States (Delcourt et al., 1983), Alaska (Anderson and Brubaker, 1986) and Southwestern United States (Davis, 1995).

In the late 1980s, European and North American Pollen Databases (EPD: Cheddadi, 2002; NAPD: Grimm, 2000a) were established as archives for pollen data from both continents, and these were gradually expanded into a Global Pollen Database (GPD: Grimm, 2000b). The emphasis of the GPD is to archive pollen core data; modern pollen data, for example, from the BUPD, are also included, but the collection is incomplete.

More recently, Williams built a continental modern pollen dataset of 57 taxa for biomization purposes by assembling samples from the BUPD, Davis (1995), and NAPD core-tops (Williams et al., 2000). This dataset was incorporated into a more comprehensive dataset independently assembled by Sawada and Gajewski (Sawada, 2001). However, this database, although suitable for continental-scale paleoclimate reconstructions, contained few data from the Arctic, Alaska and the southwestern United States. Sawada et al. (2004) also found that higher taxonomic resolution could improve paleoenvironmental inferences. The research database presented in this study is the result of efforts to develop and release to the scientific community an improved, high-quality modern-pollen dataset with associated environmental variables.

2. Data

The new modern pollen database includes 4634 sites from across North America (Fig. 1a). This research dataset is stored as a Microsoft Excel® file which contains four types of information for each record: (a) pollen counts for 134 pollen taxa that are either found across North America or are regionally important (Table 1); (b) site identification, geographic coordinates, source, depositional environment and auxiliary identification codes; (c) environmental data, including elevation-corrected climatic data based on the Climatic Research Unit gridded climatology (New et al., 2002) with improved lapse rate correction; and (d) information about natural and potential vegetation derived from cartographic sources and earth observation satellite sensors (AVHRR). Estimates of site positional errors are included to assess the suitability of individual sites for research at multiple spatial scales.

2.1. Pollen data

Our objective is to produce a data set with broad and uniform coverage and as consistent as possible in terms of pollen identification, spatial and temporal resolution. To be as consistent as possible, it would be preferable to use only pollen samples from similar depositional environments. Most contemporary pollen studies are currently undertaken using lake sediments; however, some important regions, such as the southwest United States, have few lakes. Thus, we included data from moss polsters, fens and bogs in some regions (e.g. western Oregon and Washington) where lake sediment samples were lacking.

The original data for the modern pollen database mainly came from two sources: (1) core-top samples originating from many sites where fossil data were analyzed during the past 40 years; and (2) local and regional surface sample datasets accumulated expressly for calibration studies. Data were obtained from over 450 studies and several unpublished sources; references are included in the metadata. Many of these data are included in several databases (Fig. 1a). A total of 64% of the pollen records are stored as the number of pollen grains, while the remaining 36% are identified as percentages or per mille abundance. Some of the latter samples were read from publications and then digitized, which could affect the quality of the data (see Section 3.2).

Surface pollen samples are generally collected using methods that enable a consistent comparison of samples across North America. Pollen processing follows a standardized procedure where a small volume of sediment (typically 1 cm³) is treated with a series of acids and bases to remove most of the sedimentary material, leaving the pollen grains intact for identification and counting (Faegri and Iversen, 1989).

The classification of pollen types is based on the taxonomic and morphological hierarchies defined in the NAPD (Grimm, 2000a), although case-by-case classification was required for certain pollen types. Because our goal was to produce a database ready for analysis, we did not include all taxa from the original sources, but rather combined some taxa into higher categories. The pollen grains represent 134 types from 65 families (Table 1). Major North American taxa and many less common taxa with a constrained geographic distribution and/or climatic optimum were included in the taxonomic list. *Pinus*, *Betula*, *Picea*, Poaceae, Cyperaceae and *Alnus* represent approximately 60% of all pollen grains in the dataset, and 99.25% of all pollen grains within the NAPD are represented under the taxonomy of our database. Identification of *Pinus*, *Picea*, *Acer* and *Alnus* was generally universal and consistent within the samples in the database, but distinctions within these genera were not uniformly made, especially in samples dating between 1960 and 1980. Where these distinctions were made, we have included them, but they are missing at many sites.

Factors such as depositional environment (e.g. lake, moss polster, peat bog, etc.), basin size and changes in sampling technique through time can bias pollen assemblages (Ritchie, 1974; Webb et al., 1978). These factors become more important with a finer spatial and temporal resolution, and stratifying modern pollen records according to sample properties (e.g. depositional environment, basin size) can minimize such bias-induced errors (Jackson and Williams, 2004). Therefore, the names of the principal investigator and the depositional environment for each sample are included as supplementary fields in the database. Original data contributors rarely provide other important information, such as basin size, sampling area or depth of water at sample location, but lake water depth is included when available.

2.2. Climate data

Modern climate and bioclimate variables were assigned to each record using a three-step procedure. First, long-term monthly mean (1961–1990) temperature (°C), precipitation (mm) and percent possible sunshine data (CRU; New et al., 2002) were used to calculate local lapse rates for these variables by fitting locally weighted third-degree trend-surface regression models (in latitude and longitude, with elevation as a covariate) to the data within the neighborhood of each grid point. The search radius was 500 km, and inverse-distance squared weighting was used. Second, for each modern pollen sample, the lapse rates at the surrounding 10-min grid points were used to adjust the CRU values at those points to
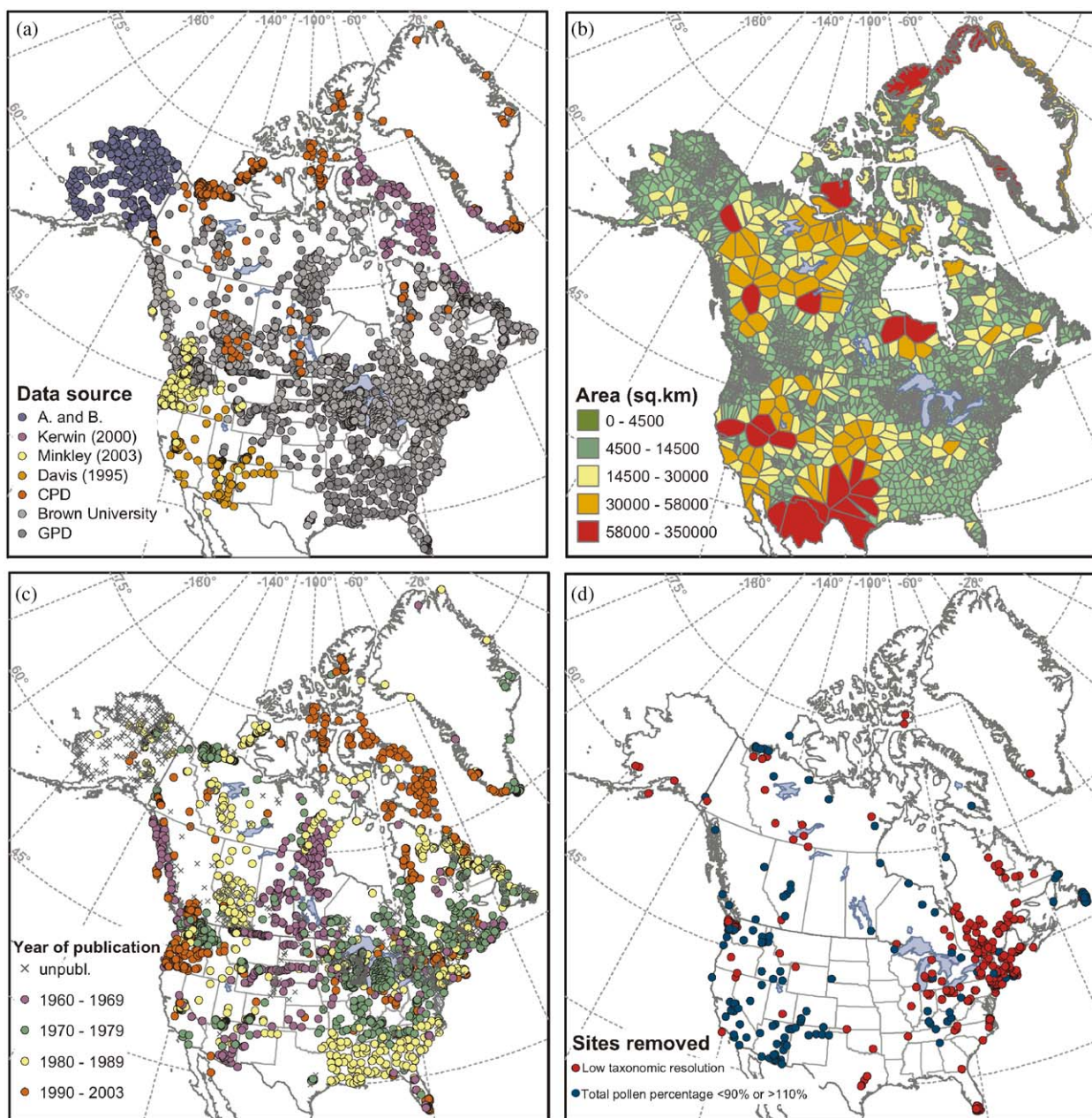
Fig. 1. (a) Distribution of modern pollen samples in North America classified by database origin. Abbreviations: A.&B. = Anderson and Brubaker (1986, unpublished), CPD = Canadian pollen database, GPD = global pollen database; (b) point density of modern pollen data in North America. The Voronoi polygon surrounding each point is classified by relative area; (c) year of publication for data in the North American modern pollen database; and (d) identification of "low quality" data site removed from database.

the elevation of the target point. Third, these adjusted values were interpolated at the location of the target point using geographically weighted bilinear interpolation.

Bioclimatic variables (Prentice et al., 1992) were calculated from the basic climate variables, including: growing-degree days with a 0 and 5 °C base; a "chilling period" (number of days when daily temperature was below 5 °C); mean temperature of the coldest and warmest months of the year; and the January/annual and July/annual precipitation ratios (Table 2). Moisture-balance variables, including actual annual evapotranspiration (AE), potential annual transpiration (PE) and the Priestley–Taylor moisture-index parameter ("alpha", annual AE/PE), were computed using a modification of the Cramer and Prentice (1988) approach and soil hydrologic properties derived from the International Geosphere–Biosphere Programme IGBP soils data set (Global Soil Data Task, 2000).

Table 1
List of pollen types in database

1.	*Abies*
2.	*Acer negundo*
3.	*Acer pensylvanicum*
4.	*Acer rubrum*
5.	*Acer saccharinum*
6.	*Acer saccharum*
7.	*Acer*
8.	*Alnus crispa*
9.	*Alnus rubra*
10.	*Alnus rugosa*
11.	*Alnus* undifferentiated
12.	*Ambrosia*
13.	*Amorpha*
14.	Anacardiaceae
15.	Apiaceae
16.	Aquifoliaceae
17.	*Arceuthobium*
18.	Arecaceae
19.	*Armeria*
20.	*Artemisia*
21.	Asteraceae
22.	Asteraceae subf. Cichorioideae
23.	*Betula*
24.	Boraginaceae
25.	*Botrychium*
26.	Brassicaceae
27.	Cactaceae
28.	Campanulaceae
29.	Caprifoliaceae
30.	*Carya*
31.	Caryophyllaceae
32.	*Castanea*
33.	*Ceanothus*
34.	*Celtis*
35.	*Cephalanthus*
36.	*Cercocarpus*
37.	Chenopodiaceae/Amaranthaceae
38.	*Chrysolepis/Lithocarpus*
39.	*Cornus*
40.	Corylus
41.	Cupressaceae
42.	Cyperaceae
43.	*Dodecatheon*
44.	*Dryas*
45.	Elaeagnaceae
46.	*Ephedra*
47.	*Equisetum*
48.	Ericaceae *Chamadaphne/Ledum*
49.	Ericaceae *Vaccinium*
50.	Ericaceae undifferentiated
51.	*Ericales*
52.	*Eriogonum*
53.	*Euphorbiaceae*
54.	Fabaceae
55.	*Fagus*
56.	*Fraxinus nigra*
57.	*Fraxinus pennsylvanica/americana*
58.	*Fraxinus* undifferentiated
59.	*Iva*
60.	*Juglans cinerea*
61.	*Juglans nigra*
62.	*Juglans* undifferentiated
63.	*Koenigia islandica*
64.	Lamiaceae
65.	*Larix/Pseudotsuga*
66.	*Larrea*
67.	Liliaceae

Table 1 (*continued*)

68.	*Liquidambar*
69.	*Liriodendron*
70.	*Lycopodium annotinum*
71.	*Lycopodium clavatum*
72.	*Lycopodium complanatum*
73.	*Lycopodium selago*
74.	*Lycopodium* undifferentiated
75.	Magnoliaceae
76.	Malvaceae
77.	Moraceae
78.	Myricaceae
79.	*Nyssa*
80.	Onagraceae
81.	Osmundaceae
82.	*Ostrya/Carpinus*
83.	*Oxyria*
84.	Papaveraceae
85.	*Pedicularis*
86.	*Picea glauca*
87.	*Picea mariana*
88.	*Picea* undifferentiated
89.	*Pinus* Diploxylon
90.	*Pinus* Haploxylon
91.	*Pinus* undifferentiated
92.	Plantaginaceae
93.	*Platanus*
94.	Poaceae
95.	Polemoniaceae
96.	Polygonaceae
97.	*Polygonum bistortoides*
98.	*Polygonum* undifferentiated
99.	*Polygonum viviparum*
100.	Polypodiaceae
101.	*Populus*
102.	*Potentilla*
103.	*Prosopis*
104.	*Pteridium*
105.	*Quercus*
106.	Ranunculaceae
107.	Rhamnaceae/Vitaceae
108.	Rosaceae
109.	Rubiaceae
110.	*Rubus*
111.	*Rumex*
112.	*Rumex/Oxyria*
113.	*Salix*
114.	*Sanguisorba*
115.	*Sarcobatus*
116.	*Saxifraga cernua*
117.	*Saxifraga hieracifolia*
118.	*Saxifraga oppositifolia*
119.	*Saxifraga tricuspidata*
120.	Saxifragaceae
121.	Scrophulariaceae
122.	*Selaginella*
123.	*Shepherdia canadensis*
124.	*Sphagnum*
125.	*Taxodium*
126.	*Taxus*
127.	*Thalictrum*
128.	*Tilia*
129.	*Tsuga heterophylla*
130.	*Tsuga mertensiana*
131.	*Tsuga* undifferentiated
132.	*Ulmus*
133.	Urticaceae
134.	*Xanthium*

Table 2
Climate and vegetation variables included in the pollen surface-sample database

Variable	Description
t [jan]…[dec]	Mean monthly temperature (°C)
p [jan]…[dec]	Mean total monthly precipitation (mm)
s [jan]…[dec]	Mean monthly percent possible sunshine (NOT clouds)
tmax, tmin	Absolute (all time over the period of record) maximum or minimum monthly temperature (not averaged across years of record)
gdd0, gdd5	Growing-degree days, 0 and 5 °C base
chill	Chilling requirement (number of days when pseudo-daily temperature < 5 °C)
mtwa, mtco	Mean temperature of the warmest or coldest month of the year (averaged across all years of record)
annp	Mean total annual precipitation
pjanpann, pjulpann	January/annual precipitation ratio, July/annual precipitation ratio
aaetpt, apetpt	Actual evapotranspiration (AE), potential evapotranspiration (PE)
mipt	Priestley–Taylor (alpha) parameter (AE/PE)
miptev	Alpha calculated for the evergreen plant functional type (pft) assimilation period (days when the pseudo-daily temperature > -4 °C)
aaetptev, apetptev	AE, PE for the evergreen pft assimilation period
miptev	Alpha calculated for the deciduous pft assimilation period (days when the pseudo-daily temperature $> +5$ °C)
aaetptev, apetptev	AE, PE calculated for the deciduous pft assimilation period
IGBP21	Modal vegetation type from DISCover dataset V2 within a $21 \times 21 \, km^2$ search window. IGBP classification scheme
BATS21	As IGBP21, BATS classification scheme
GLEC21	As IGBP21, Global Ecosystems classification scheme
NASLC21	As IGBP21, North American Seasonal Land Cover classification scheme
USGS21	As IGBP21, USGS Level II classification scheme
SIB2_21	As IGBP21, SIB2 classification scheme
RUNN21	As IGBP21, Running Life Forms classification scheme
IGBP03	As IGBP21, IGBP classification scheme
IGBP[11] [51] [101]	As IGBP21, for search windows of 11×11, 51×51, and $101 \times 101 \, km^2$
nland	Number of land pixels within $21 \times 21 \, km^2$ search window
npix	Total number of pixels within $21 \times 21 \, km^2$ search window
IGBP2[1_1]…[1_17]	Fraction land area within $21 \times 21 \, km^2$ window covered by 17 different biomes (IGBP classification scheme); see metadata
needavg [03] … [101]	Percent land surface covered by needleleaf canopy, for 3×3, 11×11, 51×51, and $101 \times 101 \, km^2$ search windows
broadavg [03] … [101]	Percent land surface covered by broadleaf canopy, same search windows as previous

2.3. Vegetation data

Three global land-cover datasets were used to characterize vegetation properties for all surface pollen samples (Williams, 2002). First, the potential vegetation (vegetation in the absence of significant anthropogenic land-use) was determined for each pollen site from the Fedorova et al. (1994) compilation of potential-vegetation maps. The IGBP DISCover database (Version 2) (http://edcdaac.usgs.gov/glcc/glcc.html) provided information about current land-cover type (Loveland et al., 2000). The pollen source radius can be at least several tens of kilometres for small lakes (Bradshaw and Webb, 1985), therefore, each surface-pollen sample was assigned the modal land-cover type within a search window centered on each site (Table 2). Land-cover types that consist of 'water' were not considered. In some cases, no modal vegetation type could be found, either because of a tie or because no terrestrial vegetation existed within a given search window. For these sites, the search window was repeatedly increased

by 2 km until a modal biome was identified or until the width of the search window was double the size of the starting width. Sites with no modal biome are flagged. Land cover assignments from seven land-cover classification schemes using $21 \times 21 \, km^2$ search windows are included. For the IGBP classification scheme, search windows of 3×3, 11×11, 51×51, and $101 \times 101 \, km^2$ were used.

Estimates of the percent land surface covered by needle-leaved tree canopy, broad-leaved tree canopy and open/unforested areas were obtained from the University of Maryland Continuous Fields Tree Cover dataset (DeFries et al., 2000). These data are based on AVHRR observations from 1992 to 1993. Coverage values surrounding each pollen site were averaged using the same search windows as the IGPB scheme (Table 2). Some sites in Greenland were outside the edge of the North American image and were not assigned tree-cover values. Analyses of surface sample and tree-cover data from eastern North America suggest that search-window dimensions of 50×50–$150 \times 150 \, km^2$ produced

the best agreement between the palynological and AVHRR sensors (Williams, 2002).

3. Methods and quality controls

The surface pollen samples originated from many studies done by different palynologists for different purposes, leading to possible inconsistencies. We developed quality-control procedures to identify and account for discrepancies, including the removal of duplicate and low-quality samples and the correction of positional and elevation errors.

3.1. Duplicate removal

Duplicate samples between and within the databases were removed using automated applications and on a case-by-case basis. Duplicates resulting from sites having similar geographical coordinates but different pollen assemblages (geographic duplicates) were discriminated, through filtering procedures, from real duplicates having either identical geographical coordinates and pollen assemblages or similar assemblages (assemblage duplicates).

3.1.1. Geographical and assemblage filters

Two automated applications were developed to identify geographical and assemblage duplicates (Sawada, 2001). If two sites shared identical geographical coordinates but had different pollen assemblages (as might occur in elevational transects of sites), both would be retained. However, if the two sites had identical geographic coordinates and assemblage values, then only one of the sites would be retained in the final dataset without preference as to database origin. To determine sites with duplicate coordinates, the latitude and longitude decimal degrees were rounded to two decimal places.

The first procedure retrieves identical geographical and assemblage duplicates and discards the first one found. However, imprecision in the geographical coordinates between databases can leave some duplicates unnoticed. Therefore, a second procedure identified assemblage duplicates by computing the squared chord distance (SCD) between all sites and extracting those sites having identical assemblages (SCD $< 2.220446e-16$). Sites within each group of assemblage duplicates were then manually verified, as described in the next procedure.

3.1.2. Case-by-case duplicate removal

Manual verification was required for assemblage duplicates and duplicates not recognized by the automated applications due to discrepancies in the taxonomic classification or resolution between different databases. Differences in taxonomic resolution among datasets depend on the degree to which pollen types were aggregated or split in the original data counts that were submitted to each database. Some records also contained decimal points in the pollen counts, while values were rounded in another database. The latter were retained to maintain whole pollen counts in the database, as the decimal places were due to rounding errors when converting percentages to counts. After removing duplicates using automated applications, the remaining sites were individually checked for hidden duplicates using associated information (e.g. site names, elevations).

Preference criteria were established to remove duplicates between datasets. When taxonomic resolution was greater, data from the Canadian Pollen Database, Kerwin (2000) and Anderson and Brubaker (1986, unpublished) were preferred over data from the GPD, BUPD, Davis (1995) or Minckley (2003), as small inconsistencies had been corrected in the former, although not always well documented. Sites in the GPD were preferred over those duplicated in the BUPD, Davis (1995) or Minckley (2003) for the same reason.

3.1.3. Duplicate removal using a Geographical Information System (GIS)

The remaining sites were imported into a GIS to identify spatial duplicates that could have been overlooked. Sites from one dataset that were within a 0.1-degree distance (approximately $9.5 \times 11 \, \text{km}^2$ at $30°\text{N}$ to $4.5 \times 11 \, \text{km}^2$ at $66.5°\text{N}$) of a site from another dataset were selected. Pollen assemblages for these selected records were then examined, as in step 2.

3.2. Low-quality data removal and sample averaging

In previous database compilations, older samples were retained simply because no other pollen data were available from those regions. While suitable for applications using small pollen sums, these data are not comparable to those produced in the past 20–30 years, as knowledge of pollen taxonomy has increased through time. Data produced before 1960 were therefore removed. Similarly, older samples that did not have, for example, standard Asteraceae breakdown or excluded some non-arboreal taxa (Davis and Webb, 1975; Avizinis and Webb, 1985), were also removed from the database. Moss polster samples from the Arctic that were collected from the same location were averaged due to the high between-sample variability. Averaging these samples renders them more comparable to others in the database while not altering their quality in terms of taxonomic resolution.

We also verified for poorly digitized samples. Sites with unusually low pollen sums, and/or percentage values totalling less than 90% or greater than 110%

were dropped. Although 731 problematic sites were removed, sufficient higher-quality data remained in these regions (SW US, NW US and Eastern North America) to compensate for the discarded samples. For archive purposes, the removed samples are available as a separate file.

3.3. Geographic coordinates

Site coordinates from source datasets are sometimes only approximate locations, sufficient to locate the region of collection, but not specific enough to locate the precise sampling location. Information on possible positional errors for these sites is required when integrating older site coordinates with more recent geographic positioning system (GPS) coordinates of greater accuracy. The problem of recorded location precision is often over-looked, therefore, we discuss this issue to provide guidelines for reporting the locations of future samples.

Longitude and latitude from the source datasets are stored as decimal degrees of varying precision. The majority of modern samples within our database were collected before 1990, therefore site coordinates were usually derived from printed maps. In the United States, these would be 1:24,000 scale 7.5-min quadrangles referenced to the geographic coordinate system known as the North American Datum of 1927 (NAD27). This series was completed in 1992 for the 48 lower states. In Canada, 1:50,000 scale National Topographic Series (NTS) maps utilized the NAD27 datum until 1990 when NAD83 was adopted. Misspecification of the geodetic datum can lead to errors in position ranging from a few meters in eastern North America to over 100 m in Alaska and northern Canada (Pinch, 1990). However, given that the majority of samples integrated into our database were collected before 1990, datum misspecification is of concern at a few sites only. The datum shift error is small when compared to errors that are possibly induced by the variable accuracy of coordinate precision for many sites.

Inspection of the data suggests that coordinates were commonly recorded in degrees and minutes (rarely seconds) with a resultant error of typically ±2 km depending on latitude. Coordinates in the source databases were stored in decimal degrees with precisions varying between 0 and 7 decimal places. However, the best precision available from a 1:24,000 paper map with the aid of mechanical rulers would be approximately 1 arcsec (4 decimal places).

Two complementary approaches were used to determine the variations in accuracy and precision of geographical coordinates as a function of topographic relief. First, site coordinates within high-relief regions were verified by comparing the elevations recorded by the author to a digital terrain model (GTOPO30; http://edcdaac.usgs.gov/gtopo30/gtopo30.asp). Large differ-ences between the author's and GTOPO30 elevations indicate either: (1) an erroneous sample elevation; (2) an erroneous sample location; or (3) a sample in a region of high-relief where the GTOPO30 dataset is smoothed compared to the actual topography. For sample locations with discrepant elevations, we checked the original site locations against the published site or topographic maps. Our default assumption was that the original recorded elevation was correct and that the site coordinate was in error, unless inspection of the topographic maps indicated an error in reported elevation. We attempted to confirm the location of all samples with elevation differences of greater than 300 m, except in the relatively flat eastern United States, where differences of greater than 100 m were evaluated. In addition, all sample locations west of 100° longitude and south of 60° latitude were checked against their published location. In some cases in western North America, when the original reference did not provide sample locations other than on a site map, the site map was digitized and spatially registered to confirm and update latitude and longitude values. Of the 937 sites identified as problematic, 365 within areas of high relief were corrected using topographic maps and the remainder were flagged.

The above approach was found to be insufficient to determine positional discrepancies in low-relief regions because of the level of generalization in the GTOPO30 elevation data. The original precision for all sites within our database was retained; however, we estimated the potential positional accuracy for each site according to the recorded precision of the coordinates from the source databases. Potential positional errors in units of meters in both the x and y directions were derived using a spherical model of the earth (radius = 6374 km). For a given coordinate precision (e.g., four decimal places), positional errors decrease towards the poles in the x-direction because of meridian convergence. Errors in the y-direction are considered constant for a given precision at all latitudes. The positional error for a site can be thought of as an ellipse with the semi-major axis oriented in y and minor in x. We assumed that the original coordinate values as entered into the database were rounded to the nearest nth decimal place, leading to precision-based error on the coordinate of at most $5 \times 10^{-(n+1)}$ decimal degrees.

3.4. Elevation data

Site elevations recorded by authors were retained but some sites with no elevation or erroneous values required correction. We retained the authors' elevations because these were most likely estimated directly from a topographic map; if the existing site coordinates in the database were utilized, then the positional errors could

lead to the extraction of elevation at points that did not correspond to the true site location.

High spatial resolution elevation data are available for all of Canada (http://www.geobase.ca) and the United States (http://seamless.usgs.gov/). For 189 sites with missing elevations within the lower 48 states and 5 sites in southern Greenland, elevations were extracted from the Shuttle Radar Topography Mission (SRTM) 3 arcsec dataset. For 52 sites in Canada, elevations were extracted from 3 arcsec Canadian Digital Elevation Data (CDED) Level 1. The Cartesian resolution of the SRTM and CDED Level 1 is 90 m or less depending on latitude and orientation. Seven sites in Alaska and the Mackenzie Delta region were assigned elevations from the Alaska 300 m Digital Elevation Model (ADEM). The difference between vertical datums in all cases is less than 2 m and within the inherent elevation error of the datasets.

3.5. Overall data quality and status of the modern pollen dataset

Although the various methods discussed above can identify problematic sites and outliers, errors are frequently identified by use of the database. Potential problems were identified using a number of graphical, mapping and analytical analyses. All pollen taxa were mapped to identify unusual values in some areas. The Shannon–Weiner Index, which takes into account species richness and proportion of each species (p_i) within the pollen assemblage ($H = -\Sigma p_i \ln p_i$) (Magurran, 1988) was computed for each sample as a general index of sample information context, as discussed below.

Finally, using the modern analogue technique (MAT), we evaluated the ability of the pollen dataset to discriminate between different vegetation formations in North America and to reconstruct modern climates (Gajewski et al., 2002). The MAT takes each pollen assemblage and compares it to all other assemblages to determine which is most similar. We used a "leave-one-out" cross-validation to estimate of the strength of the modern pollen–vegetation and pollen–climate relations. We used the SCD as a measure of dissimilarity (Overpeck et al., 1985) and all 134 taxa in our dataset. Including uncommon taxa in the pollen sum may increase the specificity of the selected best analogue for a site by excluding regionally inappropriate analogues (Sawada et al., 2004), although rare indicator taxa may not be identified in every sample, increasing the value of the SCD. If pollen assemblages in the dataset were consistently identified among contributing authors, then sites should find their best analogue in close geographical proximity. To determine the ability of the dataset to reconstruct modern conditions, the climate assigned to the site with the lowest SCD was compared to the observed climate at each pollen site. Small differences between estimated and observed climate values suggest that pollen assemblages select analogues from comparable climatic regions.

4. Results and discussion

The updated surface-sample dataset with its coverage across all of North America opens new possibilities for paleoecological and paleoenvironmental interpretation of fossil pollen data. We describe: (1) the spatial extent and uniformity of coverage of the dataset; (2) the taxonomic reliability, resolution and the information content of the surface-samples; and (3) positional errors due to precision rounding of geographic coordinates. We then provide two examples that illustrate the use of this database. First, we investigate the ability of modern samples to find analogues in their own biomes. Second, we present maps that provide examples of taxa that could be used in regional-scale studies, or to improve the resolution of continental-scale studies of pollen–vegetation or pollen–climate relations.

4.1. Description of the modern pollen data

The North America and Greenland database contains 4634 samples of modern pollen data compiled from seven source databases (Fig. 1a). Samples are available from across North America, although site density varies considerably. For those parts of North America with relatively little topographic relief, horizontal site density can be a measure of spatial resolution of the data points.

We measured site density by using Voronoi polygons around all sample-points (Boots and Getis, 1988) and classifying the relative area of these polygons (Fig. 1b). Large polygons correspond to regions having low density of points while small polygons represent regions with dense sample coverage. Site density is greatest along the axis from the Great Lakes through the St. Lawrence lowlands. Northern Alaska and the Pacific Northwest also have regular arrays of modern pollen data. In the latter region the Voroni polygons may be misleading about the site density needed to yield good coverage of the patterns in the vegetation because of high topographic relief. Texas is the region with by far the scarcest sample coverage. Other US regions with poor coverage include most of the southwestern and central states (e.g. California, New Mexico, Nevada and Colorado). These regions generally have few lakes and a semi-desert climate less favorable to pollen preservation and accumulation. Poorly covered regions in Canada include northern Ontario and much of the Northwest Territories, Nunavut and Yukon. Access is generally the greatest impediment to sample collecting in these regions. Surface samples from Mexico are not yet

available in the database due to ongoing development of regional databases.

The age of publication for the samples in the updated modern dataset spans over 40 years (Fig. 1c). The oldest samples are from the mid-continent and the British Columbian coast, whereas the most recent samples are from Alaska, the eastern Canadian Arctic, and Oregon. The age differences should little affect the interpretation of fossil data, but may be of importance when the data are compared to temporally highly resolved earth observation data, or when assessing whether a pollen sample may have been affected by recent land-use change. The year of publication does not necessarily correspond to the year of collection or analysis and some samples are older than the actual publication date. However, contributors rarely submit the collection date for each sample. Data published before 1960 were removed from the database because the level of taxonomic discrimination is frequently too low for current applications (Fig. 1d). Many of these samples originated from Potzger, Potter and other early workers in the 1940s and 1950s, which were digitized by Davis and Webb (1975). Some data in the database were digitized from diagrams, others were recorded from tables obtained from the original publications, but many came from unpublished tables supplied by the analyst, who only published summary figures of the data. Therefore complete counts are not always available.

4.2. Information content

Information content (H) depends on: (1) the botanical diversity of the region and the level of taxonomic resolution of the component taxa based on pollen morphology; (2) inter-regional differences in dominance by large pollen producers; and (3) the level of taxonomic detail reported by contributors. The first two factors should dominate in this database, since older samples with less taxonomic detail have been removed. We illustrate broad-scale trends in the information content of pollen spectra because H may have an impact on the results of some numerical methods like the MAT. In particular, the value of SCD that indicates a "non-analogue" sample is a function of H (Waelbroeck et al., 1998). Information-statistic indices can also be used to identify the variability of taxonomic quality between sites in any region. If modern pollen sites from comparable depositional environments are surrounded by similar vegetation, those with higher taxonomic resolution should have a higher H.

At the broad continental scale, low information content arises primarily because of low diversity of the flora, which affects the diversity of the pollen assemblage (Fig. 2a). This first factor is apparent in the region of high diversity in eastern North America, consistent with the high floristic diversity for this region and the many wind-pollinated taxa that rather evenly dominate samples in the mixed forest and deciduous forest regions. This pattern contrasts with western North America and the boreal forest where just two or three pollen types tend to dominate the pollen spectra, leading to lower H values in these regions. The effect of inter-regional differences in pollen production is apparent in comparing the palynological diversity in the boreal forest and tundra regions. *Picea*, *Alnus* and *Betula*, important pollen producers within the boreal forest, dominate the pollen spectra at the expense of other taxa, consequently reducing overall diversity. Towards the Arctic and the coastal tundra regions of Alaska, the values of H increase due to the reduced importance of large pollen producers. Another example is western North America, where *Pinus* and *Artemisia* dominate many pollen samples. Differences in reported taxonomic detail likely contribute to sub-regional variations in pollen diversity and may explain the low diversity values in Florida.

The taxonomic resolution of pollen assemblages has changed through time as the knowledge of pollen taxonomy increased. Major arboreal taxa have generally been consistently identified and reported for the past 40 years, but this is not necessarily the case for rare or non-arboreal taxa. Previous large-scale data syntheses have concentrated on major taxa of eastern and northern North America (e.g. Delcourt et al., 1984; Webb et al., 1993). Paleoclimate reconstructions were often based on a limited pollen sum, as this provided the most reliable results at the scale of eastern North America (e.g. Overpeck et al., 1985). In these circumstances, the use of older samples with lower taxonomic resolution (typically 15 pollen types) was acceptable for large-scale paleoclimate reconstructions; using a coarse taxonomic resolution reduced the uncertainties associated with taxonomic inconsistencies. However, with the data now available and increased interest in regional studies, higher taxonomic resolution, if applied carefully, may provide more precise paleoenvironmental reconstructions (Sawada et al., 2004).

4.3. Deposition environment

The nature of the deposition environment could also affect pollen assemblages (Fig. 2c), due to influences in the pollen trapping efficiency and preservation of different pollen types (Ritchie, 1974). Most samples in the database were collected from lake sediments (2261 sites), with approximately 75% of these sites having H values between 1.5 and 2.5 (Fig. 2d). Samples from peat deposits (338 sites) are comparable, with roughly 65% of sites between the same values. Samples originating from moss pollsters (860 sites) and terrestrial/soil
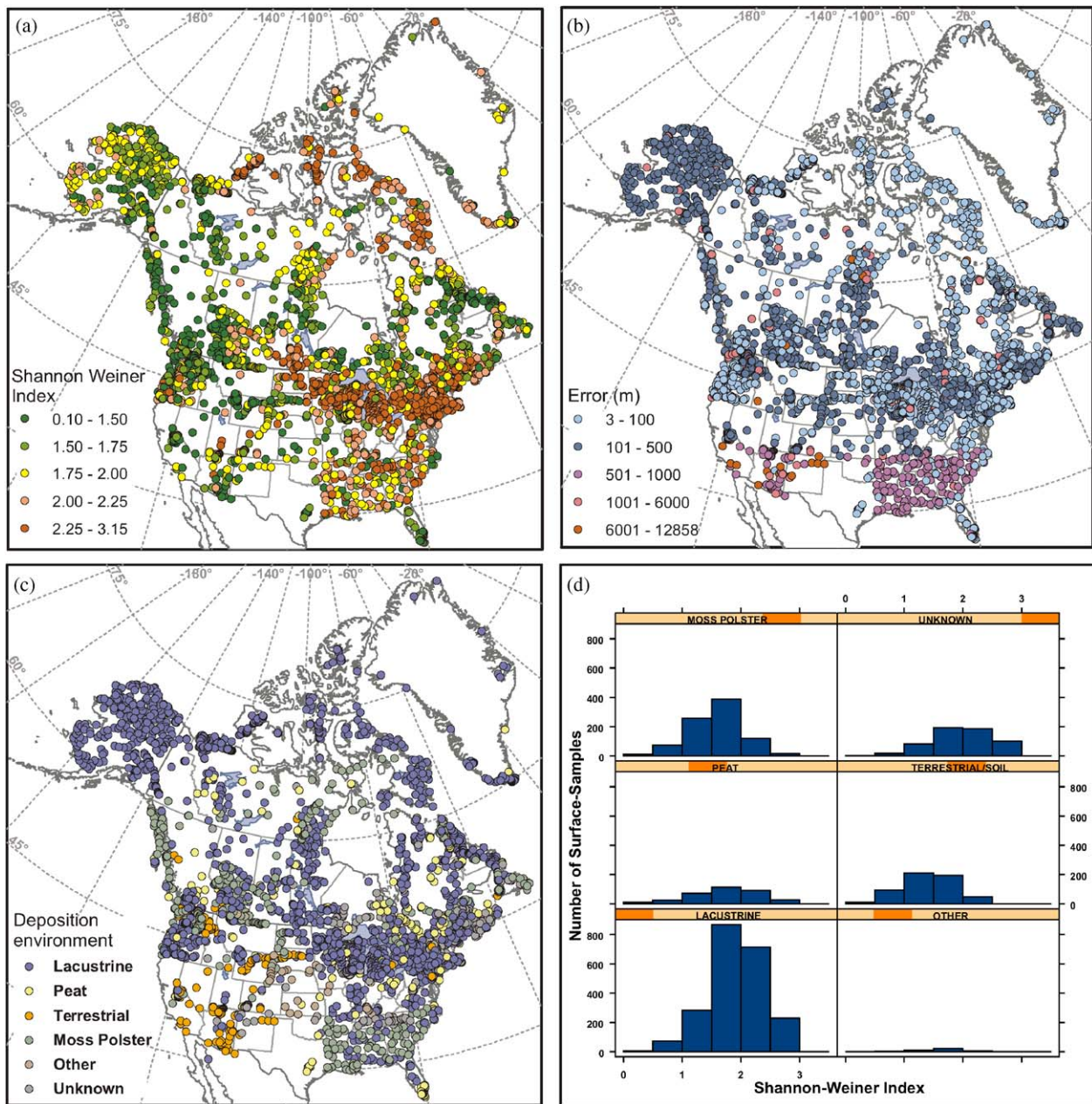
Fig. 2. (a) Information content (Shannon–Weiner index) of modern pollen samples in North America; (b) error in location (latitude/longitude) of modern pollen samples in North America as a result of rounding errors; (c) deposition environment of samples in the North American modern pollen database; (d) histogram conditioned by deposition environment of surface-samples showing the number of sites (counts) in relation to the Shannon–Weiner Index (information content).

(556 sites) deposits tend to have lower information content, with approximately 75% of samples between values of 1 and 2. Samples in which contributors did not specify sedimentary environment upon submission of their data (classified as "unknown") (580 sites) have high index values.

Where abundant lakes are found, such as across most of Canada, palynologists tend to sample lake sediments. Many pollen samples from the southeastern United

States, the northern Pacific Coast and the Central Plains are from moss polsters, whereas terrestrial/soil samples, which have lower information content (Fig. 2d) are restricted to the southwestern United States (Fig. 2c). Samples deposited in "other" depositional environments, such as stock ponds or pitcher plants, tend to be scattered across the central United States. In North America, there are fewer lakes south of the glacial limit, so other deposits have been sampled.

Fig. 2d includes data from all of North America, thereby confounding deposit type and spatial location. To attempt to separate these causes, we chose four smaller regions (southeastern Alberta, western Oregon, lower St. Laurent Valley and southeastern US excluding Florida) where there is a mixture of high and low information contents and samples from several deposition environments. In all cases, the mean Shannon–Weiner index was not statistically different between deposition environment (not shown) in part because of the high variability within any region and the low potential range of information content (see Fig. 2c). Thus, at a continental scale (Fig. 2b) deposition environment is a secondary influence on information content.

Although the information content among different depositional environments is generally comparable, some applications at a regional scale may need to stratify by sample depositional environment, age or data origin. Auxiliary information is available for each sample in the database describing the form of the data (i.e. digitized vs. raw counts or percentages), publication source and notes related to quality control issues.

4.4. Positional errors

Positional errors are of importance for local- to regional-scale studies, especially where data are related to vegetation surveys and where environmental properties are assigned to surface sample coordinates in areas of high relief. If site coordinates are accurate at the level of the recorded precision, then the potential positional error is less than 100 m for many modern pollen samples, but a more realistic estimate is probably greater as discussed above (Fig. 2b). Some sites with high positional errors within southwest California, for example, were referenced to corners of 7.5-min quadrangle maps and the sites could thus be anywhere within the quadrangle (Davis, 1995). In areas of high relief, where small changes in distance equate to large changes in elevation, low positional precision should positively associate with differences between authors' elevation and those extracted from a high-resolution DEM. We tested this in Alberta using the CDED Level 1 DEM data for Canada and found that while large elevational differences belong to sites with low levels of precision, not all sites with low positional precision have large elevational differences (not shown). Thus, the degree of positional precision may provide a first approximation of positional inaccuracies, but many of the low-precision sites are likely more accurate than evident within the database. Potential errors in location reported in the new surface sample database should be considered as approximations.

4.5. Climate range of the modern pollen dataset

Locations of all pollen samples were plotted on a coordinate system of precipitation and temperature; also plotted was the entire modern climate range of North America (New et al., 2002) classified by generalized biomes (Fedorova et al., 1994) (Fig. 3). The surface-samples are generally well distributed across most climatic zones of the continent. Areas lacking data include the very cold and dry arctic regions, as well as warm and dry climates within the desert and prairie biomes. A denser network of surface-samples is also required in the mountainous area of western North America.

4.6. Modern analogues

By comparing the modern pollen samples between themselves using the MAT, we demonstrate that our dataset is able to reconstruct current vegetation and climate patterns with a high degree of accuracy. In this discussion, we use the term "best analogue" for that site with the lowest SCD, irrespective of whether it is a "good", "possible" or non-analogue (sensu Anderson et al., 1989). For more than 93% of sites from the mediterranean, desert and arctic biomes, the sample with the lowest SCD is found within their own vegetation zone (Fig. 4). A further 4–5% of the samples find best analogues in adjacent biomes. Characteristic taxa from these biomes include *Dryas* or *Papaver* for the Arctic, *Chrysolepis/Lithocarpus* or *Ceanothus* for the mediterranean and Cactaceae for the desert (Fig. 5). There are only 19 samples in the mediterranean biome, so a detailed analysis is not possible.

Pollen spectra from the coastal, boreal, mountain and deciduous vegetation zones find 80–88% of samples with lowest SCD within their own biomes. Sites that do not find analogues in their own biome typically find them in adjacent biomes, with few outliers. For instance, one sample within the deciduous forest finds its most similar site in Greenland, while six others find their best analogues in the desert, boreal or mountain regions. The sample that finds its best analogue in Greenland is from a bog with approximately 40% Ericales and 17% *Betula* pollen, which resembles the tundra sample. Two samples from the deciduous forest that find samples with lowest SCD in the mountain and boreal biomes are dominated by two pollen types (*Pinus* and *Quercus* and *Pinus* and *Picea*, respectively), so there is little information in these particular spectra to discriminate among the various biomes. Not surprisingly, samples from the mountain biome found samples with lowest SCD in a variety of biomes, as this region actually contains a variety of vegetation types stratified along the altitudinal and directional gradients.
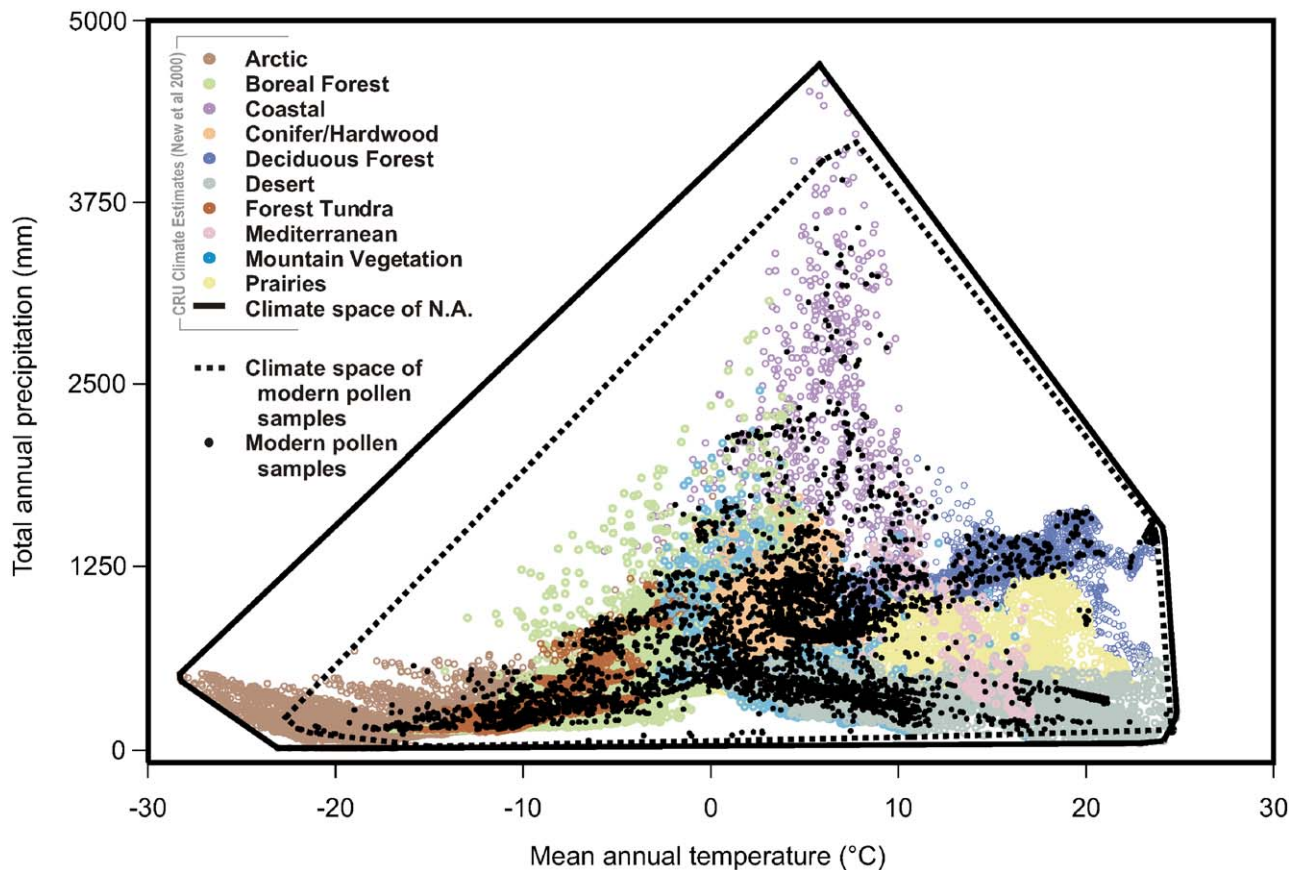
Fig. 3. North American and Greenland surface-sample climate range (black points) and modern climate range across North America (New et al., 2000) classified by biome (colored points; Fedorova et al., 1994).
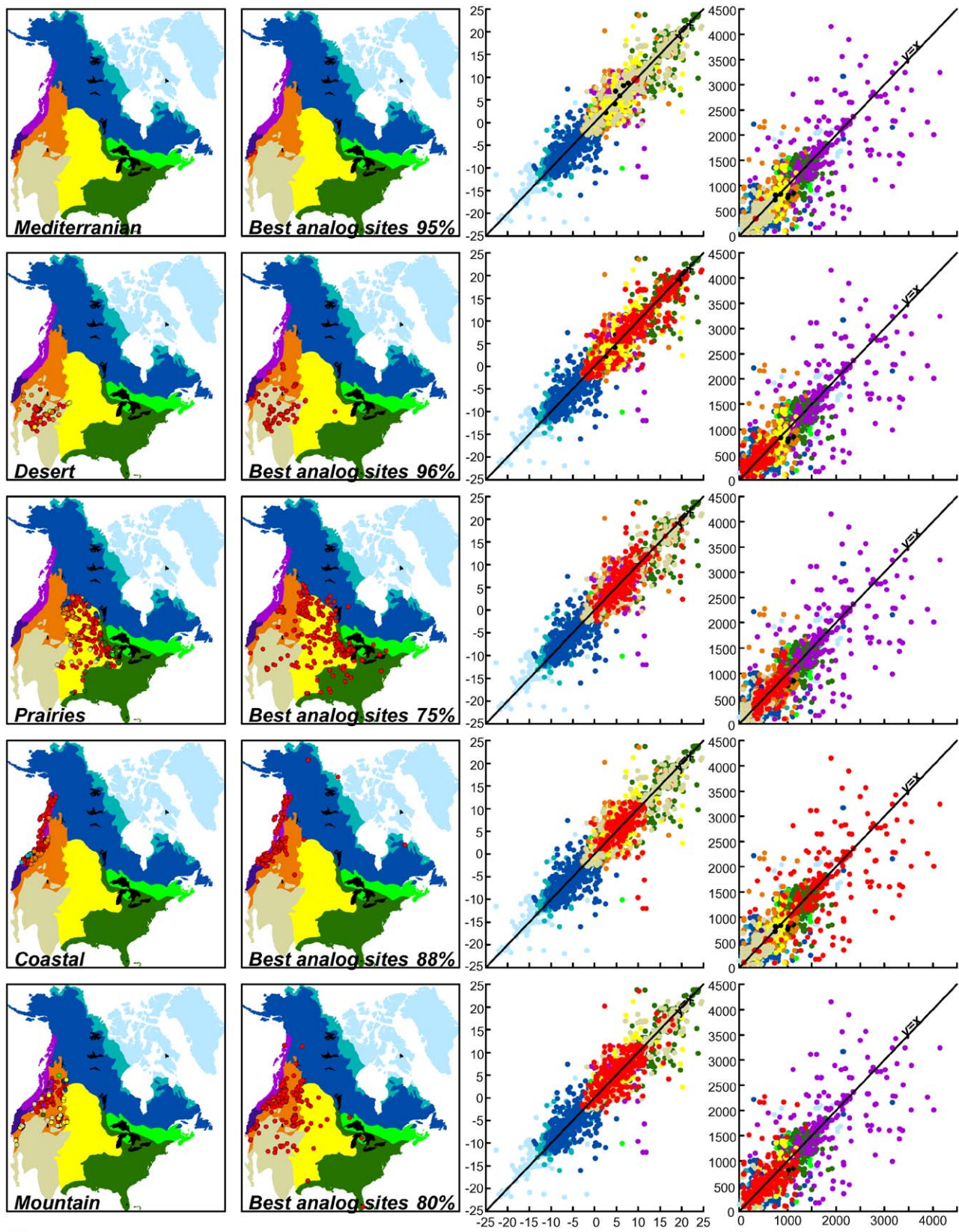
Pollen samples from the prairies, conifer/hardwood and forest–tundra biomes find their samples with lowest SCD within their own boundaries 70–75% of the time. Comparable assemblages to sites in the prairies are found across North America, probably due to the large impact of agriculture on the vegetation, which has increased the non-arboreal pollen percentages in formerly forested regions. Although sites in the conifer/hardwood and forest–tundra biomes have fewer analogues within their own vegetation zones, most similar sites are generally found in adjacent biomes. These two biomes are relatively small and are considered ecotones in other vegetation classifications. Therefore it is not unsurprising that pollen samples from these biomes have lower success in finding analogues in their own

vegetation zone. Nevertheless, for modern pollen samples from the conifer-hardwood forest, only 1.4% of best analogues are not found in either the conifer-hardwood forest or two adjacent biomes (boreal and deciduous forest). Those modern pollen spectra from the forest–tundra that do not find best analogues in the forest–tundra find them in either the arctic or boreal regions.

4.7. Climate reconstruction

The ability of the surface samples to produce accurate climate reconstructions was evaluated by comparing actual climate values to the climate inferred from the best analogue. Overall, inferred climate values are highly

Fig. 4. Modern analogues for sites in North America. From left to right, the first column shows the location of modern pollen sites in each biome. A point is red if that site finds its modern analogue in the same biome. A colored point indicates that a site's best analog is from another biome. The second column shows the location of the best analogue sites for each biome in column 1. Percentage value corresponds to percentage of sites within biome that found their best analogue within the same biome. The third and fourth columns show a comparison of the temperature (third) and precipitation (fourth) of each of the sites (sites within the corresponding biome are in red) with that estimated by the best analogue. Measured values are on the x-axis (°C mean annual temperature and mm for precipitation) and estimated values for both variables from the best modern analog are on the y-axis. The r^2 values: temperature (0.93; $p < 0.001$) and precipitation (0.79; $p < 0.001$).
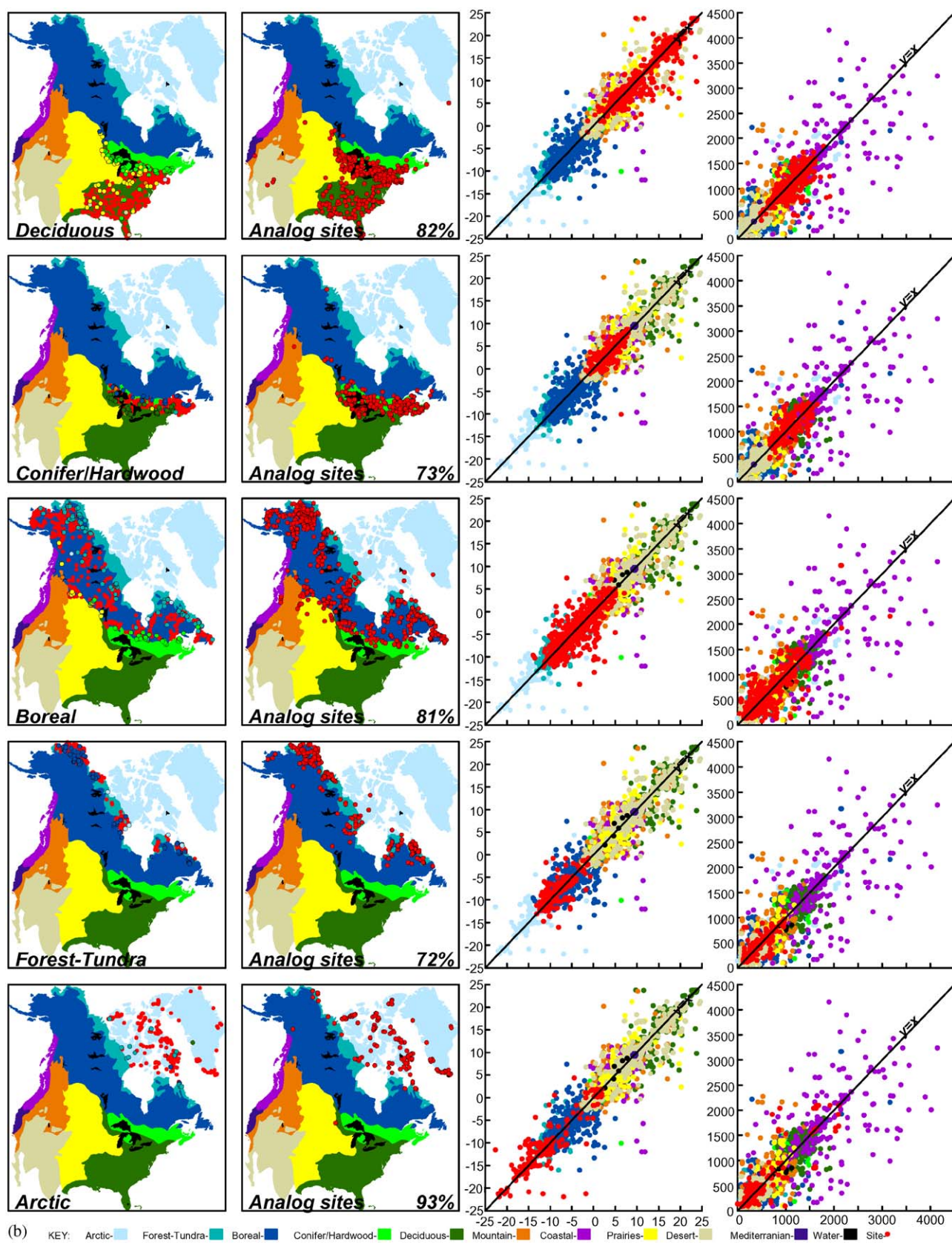
(a) KEY: Arctic- Forest-Tundra- Boreal- Conifer/Hardwood- Deciduous- Mountain- Coastal- Prairies- Desert- Mediterranian- Water- Site-
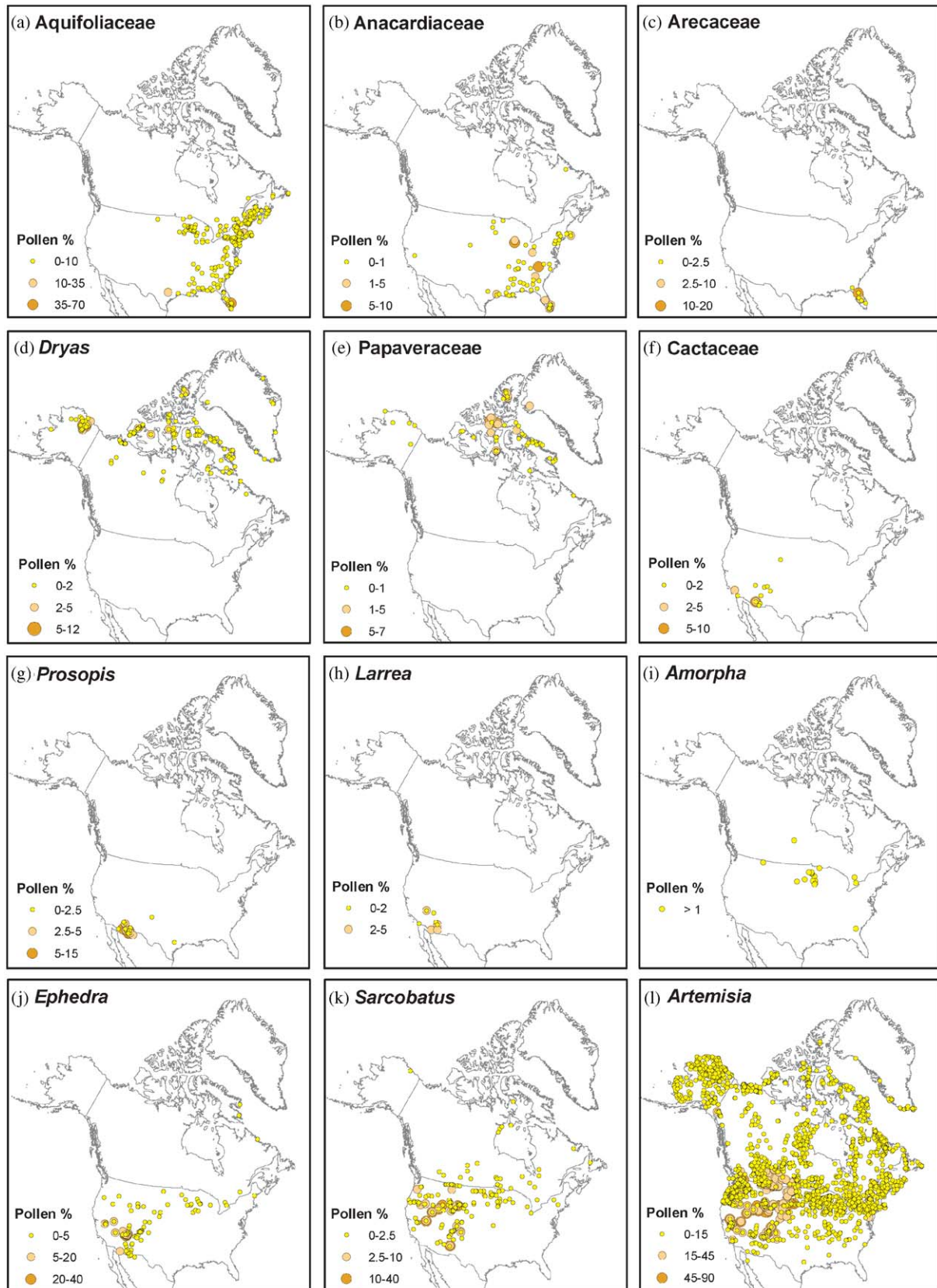
Fig. 4. (*Continued*)

Fig. 5. Maps of the distribution of selected pollen of taxa in North America. Range of the plant taxon (Thompson et al., 1999) in gray, when available.
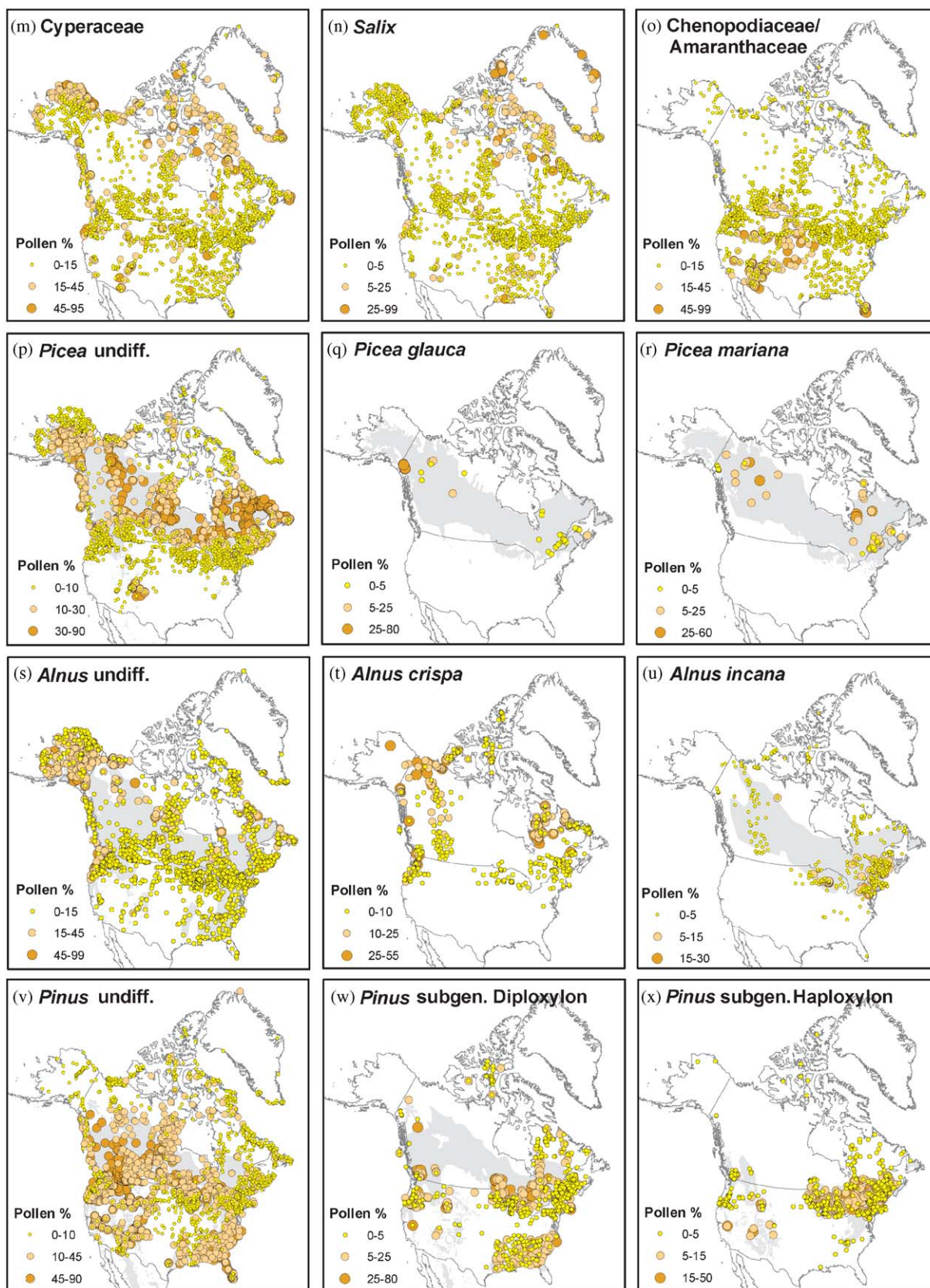
Fig. 5. (*Continued*)

correlated with observed ones, with some scatter (Fig. 4). The r^2 value between temperature inferred from the best analogue and that observed at each modern pollen site is 0.93 ($p < 0.001$), while the r^2 value for precipitation is 0.79 ($p < 0.001$).

Temperature is generally well estimated by the best analogue for all biomes. A few samples from the Arctic and the coastal regions find "analogues" in other vegetation zones and consequently underestimate observed temperatures by up to 20 °C. Precipitation is also reasonably estimated in most biomes, with the exception of the coastal region of the Pacific Northwest, British Columbia and Alaska. The coastal region has the highest scatter because the wide difference in altitude can cause estimates of the precipitation to be unreliable. In other mountain regions, the difference between the climate at a sample and the climate at its best analogue is lower due in part to the lower precipitation amounts, which thus limit the size of the discrepancy.

4.8. Distribution of selected taxa

Maps illustrate the distribution of the taxa and help with quality control, because outlying points identified on these maps may be based on a coding error or misidentification. Inspection of the geographical distribution of modern pollen types indicates the potential for further study of many taxa, if the limits of taxonomy are accounted for. Maps of these data allow the identification of well-constrained and well-represented taxa as well as those that are problematic. A selection of maps (Fig. 5) illustrates several spatial patterns in the pollen taxa:

(1) For regional-scale studies, many pollen types have restricted ranges that lie within the range of the plant taxon. Although the major arboreal taxa, mainly from eastern North America, have been mapped (e.g. Webb and McAndrews, 1976; Delcourt et al., 1984), many other pollen types have not been systematically investigated. For example, Aquifoliaceae and Anacardiaceae pollen are restricted to eastern North America, although individual species of plants from these families are found as far away as Mexico (Gleason and Cronquist, 1963). Less common pollen types are diagnostic of certain regions: Arecaceae (Palmae) pollen is restricted to Florida, *Dryas* and *Papaveraceae* to the Arctic, and Cactaceae, *Prosopis* and *Larrea* to the deserts of the south-western United States. Conversely, *Amorpha* pollen is found in only a small area of the prairie, although the plant is known to be more widely distributed along the Mississippi Valley and in dry prairies from Ohio westward to Colorado and Mexico and northward to Manitoba (Gleason and Cronquist, 1963; McDougall, 1973; Weber, 1976; Scoggan, 1978).

(2) Some pollen types have ranges that extend beyond the range of their plants, yet in a consistent way that may be interpreted from knowledge of wind patterns. *Ephedra* and *Sarcobatus* may be examples of these (Fig. 5; Maher, 1964). Within the range of these two taxa, pollen percentages are greater than 5% (Fig. 5), but they are found to the east and north in low amounts. Maher (1964) discussed the long-distance transport of these pollen taxa but based his analysis on a few sites from the Midwest. The patterns are seen more strongly in the more complete data now available.

(3) Although some widely distributed pollen types cannot be used as indicator taxa, spatially consistent patterns in abundance are apparent and informative and help constrain analogue selection. *Artemisia* pollen is ubiquitous (Fig. 5), yet its highest values are found in the Sagebrush steppe of the west. Cyperaceae and *Salix* have clear maxima in the Arctic, especially along the North Slope of Alaska (Cyperaceae) or High Arctic (*Salix*), and Chenopodiaceae–Amaranthaceae have maxima in the western mountains and prairies.

(4) Distinctions within some types must be used with caution. Four examples include *Picea*, *Pinus*, *Alnus* and Ericales pollen taxa (Fig. 5). At the generic level (order for Ericales) these pollen types are routinely distinguished in the database and provide reliable distribution maps, but at finer taxonomic distinctions, the maps show that the data are often too scattered for reliable mapping at the continental scale. Although *Alnus crispa* and *A. incana* pollen are easy to distinguish, species level identifications were not routinely recorded within the dataset, and *A. rubra* pollen is rarely distinguished (not shown). *Picea* pollen is rarely distinguished into species within the dataset (Lindbladh et al., 2003). At a continental scale, the two major species of eastern spruce (*P. glauca* and *P. mariana*) overlap over most of their range, but there are regional differences: *P. glauca* is absent from northwestern Québec; *P. mariana* is absent from the southwest Yukon but extends a bit further north in Keewatin. Unfortunately, the database does not yet contain many pollen samples in which *Picea* species have been distinguished.

Pinus grains are frequently distinguished into Haploxylon and Diploxylon types. In eastern North America, *Pinus strobus* is the only Haploxylon pollen taxon and high abundance in the pollen corresponds to the range of the tree. Diploxylon pollen and tree ranges also match closely in the east, but not in the west due to the lack of distinction of the pollen in that region. In western North America, *Pinus* pollen is often undifferentiated because this type can be easily identified despite being torn, crumpled or otherwise deteriorated. Therefore, in the east, where enough samples with the distinctions exist, *Pinus* subgenus Haploxylon and Diploxylon could be used in analogue studies and reconstruction work. However,

only the subset of the modern pollen samples in which these types are distinguished should be selected and used.

5. Conclusion

This study documents the modern pollen database for North America and Greenland, and describes its use for multi-scale paleoecological studies. Maps show some areas with few samples, for example, the southwestern United States and northwestern Canada, and these regions offer the opportunity for new data collection and regional syntheses.

The availability of this set of modern pollen samples should encourage more paleoenvironmental reconstructions. Although many of the samples are in the Global Pollen Database, this archive requires checking and quality control before it can be used for specific research goals. Our new synthesis of several previous data sets of modern pollen samples from North America now makes the data readily available as an Excel spreadsheet with marker fields that allow for selection of samples using simple operations, such as sorting or "cut-and-paste". We removed samples that contained no non-arboreal pollen or did not include distinctions among non-arboreal taxa. We have also reconciled many site location discrepancies.

The mapped summary of the pollen distribution for selected taxa illustrates the potential of these data for regional-scale studies. Several examples of indicator taxa are illustrated (Fig. 5), and there are others among the 134 taxa in the database. These pollen taxa are not necessarily distinguished in all samples within the range of the plant, but there are now sufficient data in most regions to attempt to map their distributions. The deposition of rare taxa is partly a matter of chance, and use of these data should account for this fact. Caution must also be exercised for major taxa like *Pinus, Alnus* and *Picea* in which species distinctions are not always consistent. Another potential consideration when using the dataset was illustrated by the distribution of *Amorpha* (Fig. 5). This pollen type appears to have a limited range, but may not have been identified in other samples within its range, or may be too rare unless the plant is particularly abundant. Although many palynologists distinguish important types, pollen counts submitted to the GPD may not always be complete. Nevertheless, some rare as well as some higher-level taxa may be used for landscape- to regional-scale studies.

The data file and associated metadata are available on the Internet at the National Geophysical Data Center (www.ngdc.noaa.gov/paleo), as well as the web site of the Laboratory for Paleoclimatology and Climatology (www.lpc.uottawa.ca) and the University of Wisconsin Geography Department (http://www.geography.wisc.edu/faculty/williams/data/data.html).

References

Anderson, P.M., Brubaker, L.B., 1986. Modern pollen assemblages from northern Alaska. Review of Palaeobotany and Palynology 46, 273–291.

Anderson, P., Bartlein, P., Brubaker, L., Gajewski, K., Ritchie, J., 1989. Modern analogues of Late-Quaternary pollen spectra from the western interior of North America. Journal of Biogeography 16, 573–596.

Avizinis, J., Webb III, T., 1985. The computer file of modern pollen and climatic data at Brown University. Department of Geological Sciences, Brown University, Providence.

Bernabo, J.C., Webb III, T., 1977. Changing patterns in the Holocene pollen record of northeastern North America: a mapped summary. Quaternary Research 8, 64–96.

Boots, B.N., Getis, A., 1988. Point Pattern Analysis. Sage Publications, Newbury Park, CA 93pp.

Bradshaw, R.H., Webb III, T., 1985. Relationship between contemporary pollen and vegetation data from Wisconsin and Michigan, USA. Ecology 69, 941–955.

Cheddadi, E., 2002. European pollen database. IGBP PAGES/World Data Centre for Paleoclimatology, NOAA/NGDC Paleoclimatology Program, Boulder, Co, USA http://www.ngdc.noaa.gov/paleo/pollen.html.

CLIMAP Members, 1976. The surface of the ice-age earth. Science 191, 1138–1144.

COHMAP Members, 1988. Climatic changes of the last 18,000 years: observations and model simulations. Science 241, 1043–1052.

Cole, H., 1969. Objective reconstruction of the paleoclimate record through application of eigenvectors of present-day pollen spectra and climate to the late Quaternary pollen stratigraphy. Ph.D. Dissertation, University of Wisconsin, Madison.

Cramer, W., Prentice, I.C., 1988. Simulation of regional soil moisture deficits on a European scale. Norsk Geografisk Tidsskrift 42, 149–151.

Davis, O.K., 1995. Climate and vegetation patterns in surface samples from arid western USA: application to Holocene climatic reconstructions. Palynology 19, 95–117.

Davis, M.B., 2000. Palynology after Y2K-Understanding the source area of pollen in sediments. Annual Review of Earth and Planetary Sciences 28, 1–18.

Davis, R.B., Webb III, T., 1975. The contemporary distribution of pollen in eastern North America: a comparison with the vegetation. Quaternary Research 5, 395–434.

DeFries, R.S., Hansen, M.C., Townshend, J.R.G., 2000. Global continuous fields of vegetation characteristics: a linear mixture model applied to multi-year 8 km AVHRR data. International Journal of Remote Sensing 21, 1389–1414.

Delcourt, P.A., Delcourt, H.R., Davidson, J.L., 1983. Mapping and calibration of modern pollen–vegetation relationships in the southeastern United States. Review of Palaeobotany and Palynology 39, 1–45.

Delcourt, P.A., Delcourt, H.R., Webb III, T., 1984. Atlas of mapped distributions of dominance and modern pollen percentages for important tree taxa of eastern North America. AASP Contributions Series No. 14, AASP Foundation, Dallas, 131pp.

Faegri, K., Iversen, J., 1989. Textbook of Pollen Analysis, fourth ed. Blackburn Press, New Jersey 328pp.

Fedorova, I.T., Volkova, Y.A., Varlyguin, E., 1994. World vegetation cover. Digital raster data on a 30-minute cartesian orthonormal geodetic (lat/long) 1080×2160 grid. In: Global Ecosystems Database Version 2.0. USDOC/NOAA National Geophysical Data Center, Boulder, CO.

Fowells, H.A., 1965. Silvics of forest trees of the United States. US Department of Agriculture Forest Service Agriculture Handbook No. 271. Superintendent of Documents, Washington, DC.

Gajewski, K., 1993. The role of paleoecology in the study of global climatic change. Review of Palaeobotany and Palynology 79, 141–151.

Gajewski, K., Vance, R., Sawada, M., Fung, I., Gignac, L.D., Halsey, L., John, J., Maisongrande, P., Mandell, P., Mudie, P., Richard, P., Sherrin, A., Soroko, J., Vitt, D., 2000. The climate of North America and adjacent ocean waters ca. 6ka. Canadian Journal of Earth Sciences 37, 661–681.

Gajewski, K., Lézine, A.-M., Vincens, A., Delestan, A., Sawada, M., African Pollen Database contributors, 2002. Modern climate–vegetation–pollen relations in Africa and adjacent areas. Quaternary Science Reviews 21, 1611–1631.

Gleason, H.A., Cronquist, A., 1963. Manual of Vascular Plants of Northeastern United States and Adjacent Canada. D. van Nostrand, Co., New York 810pp.

Global Soil Data Task, 2000. Global Soil Data Products CD-ROM (IGBP-DIS). CD-ROM. International Geosphere–Biosphere Programme, Data and Information System, Potsdam, Germany. Available from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN, USA. http://www.daac.ornl.gov.

Grimm, E., 2000a. North American Pollen Database. IGBP PAGES/World Data Centre for Paleoclimatology. NOAA/NGDC Paleoclimatology Program, Boulder, CO, USA http://www.ngdc.noaa.gov/paleo/pollen.html.

Grimm, E., 2000b. Global Pollen Database. IGBP PAGES/World Data Centre for Paleoclimatology. NOAA/NGDC Paleoclimatology Program, Boulder, CO, USA http://www.ngdc.noaa.gov/paleo/pollen.html.

Huntley, B., Birks, H.J.B., 1983. An Atlas of Past and Present Maps for Europe: 0–13,000 years ago. Cambridge University Press, Cambridge.

Jackson, S.T., 1990. Pollen source area and representation in small lakes of northeastern United States. Review of Palaeobotany and Palynology 63, 53–76.

Jackson, S.T., Williams, J.W., 2004. Modern analogs in Quaternary paleoecology: here today, gone yesterday, gone tomorrow? Annual Review of Earth and Planetary Sciences 32, 495–537.

Kerwin, M.W., 2000. Quantifying and modeling Holocene climate variability based on modern and fossil pollen records from the eastern Canadian Arctic and Subarctic. Ph.D. Dissertation, Department of Geology, University of Colorado, 307pp.

Lindbladh, M., Jacobson Jr., G.L., Schauffler, M., 2003. The postglacial history of three *Picea* species in New England, USA. Quaternary Research 59, 61–69.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, J., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. International Journal of Remote Sensing 21, 1303–1330.

Magurran, A.E., 1988. Ecological Diversity and its Measurement. Princeton University Press, Princeton 179pp.

Maher, L., 1964. *Ephedra* pollen in sediments of the Great Lakes Region. Ecology 45, 391–395.

McDougall, W.B., 1973. Seed Plants of Northern Arizona. Museum of Northern Arizona, Flagstaff 594pp.

Minckley, T.A., 2003. Holocene environmental history of the northwestern Great Basin and the analysis of modern pollen analogues in western North America. Ph.D. Dissertation, Department of Geography, University of Oregon, Eugene, 310pp.

New, M., Lister, D., Hulme, M., Makin, I., 2002. A high-resolution dataset of surface climate over global land areas. Climate Research 21, 1–25.

Overpeck, J.T., Webb III, T., Prentice, I.C., 1985. Quantitative interpretation of fossil spectra: dissimilarity coefficients and the method of modern analogs. Quaternary Research 23, 87–108.

Pinch, M.C., 1990. Differences between NAD27 and NAD83. In: Barnes, D.C. (Ed.), Moving to NAD'83 the New Address for Georeferenced Data in Canada. The Canadian Institute of Surveying and Mapping, Ottawa, pp. 1–15.

Prentice, I.C., 1988. Records of vegetation in time and space: the principles of pollen analysis. In: Huntley, B., Webb, III, T. (Eds.), Vegetation History. Kluwer Academic Publisher, Dordrecht, pp. 17–42.

Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., Solomon, A.M., 1992. A global biome model based on plant physiology and dominance, soil properties and climate. Journal of Biogeography 19, 117–134.

Richard, P., 1976. Relations entre la végétation actuelle et le spectre pollinique au Québec. Naturaliste Canadien 103, 53–66.

Ritchie, J.C., 1974. Modern pollen assemblages near the arctic tree line, Mackenzie Delta region, Northwest Territories. Canadian Journal of Botany 52, 381–396.

Sawada, M.C., 2001. Late Quaternary Paleoclimates and Biogeography of North America. Ph.D. Dissertation, Department of Geography, University of Ottawa, Ottawa, 253pp.

Sawada, M., Viau, A.E., Vettoretti, G., Peltier, W.R., Gajewski, K., 2004. Comparison of North American pollen-based temperature and global lake-status with CCCma AGCM2 output at 6ka. Quaternary Science Reviews 23, 225–244.

Scoggan, H.J., 1978. The Flora of Canada. Part 3 Dicotyledoneae (Saururaceae to Violaceae) National Museum of Natural Sciences. Publications in Botany No. 7(3), Ottawa, pp. 547–1115.

Thompson, R.S., Anderson, K.H., Bartlein, P.J., 1999. Atlas of relations between climatic parameters and distributions of important trees and shrubs in North America. Volume A: Conifers. Volume B: Hardwoods. US Geological Survey Professional Paper 1650A & B, Denver, CO, 269pp., 423pp.

Waelbroeck, C., Labeyrie, L., Duplessy, J.-C., Guiot, J., Labracherie, M., Leclaire, H., Duprat, J., 1998. Improving past sea surface temperature estimates based on planktonic faunas. Paleoceanography 13, 272–283.

Webb III, T., 1974. Corresponding patterns of pollen and vegetation in Lower Michigan: a comparison of quantitative data. Ecology 55, 17–28.

Webb III, T., 1986. Is vegetation in equilibrium with climate? How to interpret Late-Quaternary pollen data. Vegetatio 67, 75–91.

Webb III, T., Bryson, R.A., 1972. Late and post-glacial climate change in Northern Mid-West USA: quantitative estimates derived from fossil pollen spectra by multivariate statistical analysis. Quaternary Research 2, 70–115.

Webb III, T., McAndrews, J.H., 1976. Corresponding patterns of contemporary pollen and vegetation in central North America. Geological Society of America Memoir 145, 267–299.

Webb III, T., Laseski, R.A., Bernabo, J.C., 1978. Sensing vegetation with pollen data: Control of the signal-to-noise ratio. Ecology 59, 1151–1163.

Webb III, T., Bartlein, P.J., Harrison, S.P., Anderson, K.H., 1993. Vegetation, lake levels, and climate in eastern North America for the past 18000 years. In: Wright, H.E., Kutzbach, J.E., Webb, III, T., Ruddiman, W.F., Street-Perrott, F.A., Bartlein, P.J. (Eds.), Global Climates Since the Last Glacial Maximum. University of Minnesota Press, Minneapolis 569pp.

Webb III, T., Anderson, K.H., Bartlein, P.J., Webb, R.S., 1998. Late Quaternary climate change in eastern North America: a comparison of pollen-derived estimates with climate model results. Quaternary Science Review 17, 587–606.

Weber, WA., 1976. Rocky Mountain Flora. Colorado Associated University Press, Boulder 479pp.

Williams, J.W., 2002. Variation in tree cover in North America since the last glacial maximum. Global and Planetary Change 35, 1–23.

Williams, J.W., Bartlein, P.J., Webb III, T., 2000. Data-model comparisons for eastern North America—inferred biomes and climate values from pollen data. In: Braconnot, P. (Ed.), Proceedings of the Third Paleoclimate Modeling Intercomparison Project Workshop, October 4–8, 1999. Montreal, Canada WCRP-111, WMO/TD-No. 1007.

Wright, H.E., Kutzbach, J.E., Webb, III, T., Ruddiman, W.F., Street-Perrott, F.A., Bartlein, P.J. (Eds.), 1993. Global climates since the Last Glacial Maximum. University of Minnesota Press, Minneapolis 560pp.