Université d'Ottawa · University of Ottawa

Université d'Ottawa • University of Ottawa

FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES

FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES

Johanne WHITMORE
AUTEUR DE LA THÈSE - AUTHOR OF THESIS

M. Sc.(Geography)
GRADE - DEGREE

Department of Geography
FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

North American and Greenland Modern Pollen Data for Multi-Scale
Paleoecological and Paleaclimatic Applications

K. Gajewski
DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

CO-DIRECTEUR DE LA THÈSE - THESIS CO-SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

B. Lauriol

M. Sawada

J.-M. De Koninck, Ph.D

LE DOYEN DE LA FACULTÉ DES ÉTUDES
SUPÉRIEURES ET POSTDOCTORALES

DEAN OF THE FACULTY OF GRADUATE
AND POSTDOCTORAL STUDIES

NORTH AMERICAN AND GREENLAND MODERN POLLEN DATA FOR MULTI-SCALE PALEOECOLOGICAL AND PALEOCLIMATIC APPLICATIONS

JOHANNE WHITMORE

DEPARTMENT OF GEOGRAPHY
UNIVERSITY OF OTTAWA

A THESIS SUBMITTED TO
THE SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES
IN PARTIAL FULFILLMENT FOR THE DEGREE OF
MASTER OF SCIENCE IN GEOGRAPHY
APRIL 7, 2004

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-01642-6
Our file Notre référence
ISBN: 0-494-01642-6

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

The status of the modern pollen network in North America and Greenland is presented by assembling a database for use in quantitative calibration studies and paleoenvironmental reconstructions. The geo-referenced database includes 4569 samples from all regions of the continent. The database includes 134 pollen taxa that range from common taxa to those that may be used for regional-scale reconstructions. Climate data and vegetation characteristics are assigned to every site. A series of procedures, both automated and manual, were used to check the quality of the pollen data. Data are currently available for almost all of North America, with variable density. The squared-chord distance computed between samples shows that most modern pollen samples find analogues in the same vegetation zone and the temperature and precipitation computed from the best analogue are highly correlated with actual value at the site. Analysis of the contemporary distribution of pollen taxa in relation to their vegetation range illustrates factors that must be considered when using these data as well as the potential for more detailed analysis of understudied taxa.

Résumé

Cette étude présente l'état de l'ensemble des échantillons polliniques contemporains disponible en Amérique du Nord et au Groenland et présente une nouvelle base de données polliniques pour la recherche quantitative de calibrage et de reconstruction paléoenvironnementale. La base de données geo-référencée inclut 4569 échantillons polliniques de toutes les régions du continent et comprend 134 taxa. Des données climatiques et des caractéristiques de végétation sont associées à chaque échantillons. Une série de procédures, automatisées et manuelles, a été employée pour vérifier la qualité des données. Les données polliniques sont présentement disponibles pour presque toute l'Amérique du Nord, avec une densité variable. Une analyse par méthode analogique a montré que la plupart des échantillons polliniques modernes trouvent des analogues dans la même zone de végétation, et que la température et la précipitation estimées par le meilleur analogue sont fortement corrélées avec les observations climatiques modernes à l'emplacement des échantillons. Une analyse de la distribution contemporaine des taxa polliniques en relation à leur limite de végétation révèle qu'il y a des facteurs qui doivent être considérés losqu'on utilise ces données et que certains taxa non étudiées antérieurement ont un potentiel pour les reconstructions paléoenvironnementales.

Acknowledgements

My deepest gratitude and respects are dedicated to my supervisor, Dr. Konrad Gajewski. His guidance and profound cynicism have been highly intellectually stimulating throughout my graduate studies, and have taught me that "*something as timid as word* " can change the course of a life. Many of his ideas are reflected in this work.

Foremost, this project was possible thanks to the data contribution of hundreds researchers. Through their generosity and willingness to make available their data, palynologists have set a precedent in the field of environmental sciences. Only through transparency and openness will science maintain its integrity and remain accessible to all researchers, regardless of their institutional patronage, reputation or financial subsidies.

As noted by Thomas S. Kuhn, the scientific development of a discipline is a fragmented process through time, where individuals or groups add elements to the scientific knowledge and techniques of the field. The compilation of this database is one of those elements to be added to the field of paleoenvironment studies, and could not have been realized without the efforts of several collaborators. The third chapter of this dissertation was written for publication in peer-reviewed scientific journals; although I am the primary author, the chapter was written with the co-authors, including my thesis supervisor and research partners.

The framework of this project is an extension of the previous work done by Dr. Mike Sawada (2001), at the University of Ottawa. Dr. Sawada (a.k.a Computer Genius) was central in the programming aspect of this project and in the preparation of several figures (MAT, positional error, Thessien density maps and the conditional plots). Dr. E. Viau, at the University of Ottawa, was key in database management programming. Dr. J. Williams at the University of Minnesota facilitated much of the communication between collaborators and provided the AVHRR data; Dr. P. Bartlein, at the University of Oregon provided the improved CRU Climate data; Drs

B. Schuman, M. Sawada and T. Minckley were responsible for the positional correction of the sites. I would also like to thank the following unpublished data contributors: P. Anderson, L. Brubaker, K. Gajewski, E. Kerwin, and T. Minckley.

À l'idée de mon professeur
et celle de ces microrganismes que sont les pollens;
À l'impression qu'ils m'auront laissé
au moment de mon passage vers l'ailleurs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NORTH AMERICAN AND GREENLAND MODERN POLLEN DATA FOR MULTI-SCALE PALEOECOLOGICAL AND PALEOCLIMATIC APPLICATIONS

J. Whitmore

1. INTRODUCTION

In recent years, sufficient environmental proxy data have accumulated over North America and Greenland to permit the mapping and synthesis of Quaternary environments at several spatial and temporal scales. Analyses of these data have led to significant advances in our understanding of past environmental conditions (e.g. CLIMAP members 1976; COHMAP members 1988; Wright et al 1993). The need to perform multi-scale applications, such as the verification of paleoclimate simulations (e.g. Jaussaume and Taylor 2000), has motivated several research groups to develop paleoenvironmental databases that enable large-scale data syntheses.

Databases serve two important functions: 1) they are archives of the results of research projects, and 2) are the source of data for new research programs. Because they derive from many sources, the data may be of variable quality, as techniques and knowledge accumulate through greater experience. It is only through continual use and analysis that errors can be removed and the limits of the data identified. Databases can also be used to develop hypotheses that can be tested by new data collection and further analysis.

Of all paleoenvironmental data, pollen is the most widely accessible to the scientific community and currently the only terrestrial proxy record available for multi-scale applications in paleoecology and paleoclimatology. Plants produce abundant identifiable micro and macrofossils (Jackson et al 1997) that record changes in population distributions through time. Although each plant taxon has particular pollen production and dispersal characteristics, studies have shown that vegetation patterns are adequately sensed by pollen (Webb 1974; Bradshaw and Webb 1985;

Prentice 1988; Sugita 1994; Davis 2000) and that an analysis of geographical networks of pollen samples may mitigate representation problems (Webb et al 1993; Jackson et al 1995; Jackson and Williams 2004).

Modern values of environmental variables can be associated with contemporaneous proxy data at one site (ter Braak 1987) and calibration function used to estimate the value of an environmental variable in the past from fossil assemblages. Pollen data have been used extensively for paleoclimate reconstructions and calibrations (Webb and Bryson 1972; Gajewski 1988; Guiot et al 1993; Sawada et al 1999; Fauquette et al 1998; Peyron et al 1998; Gajewski et al 2000; Kühn et al 2002), ecological studies (Graumlich and Davis 1993), reconstructions of past terrestrial carbon storage (Peng et al 1998; Gajewski et al 2001), verification of global climate model simulations (Webb et al 1998; Gajewski et al 2000; Sawada et al 2004), analysis of biogeographic dynamics in plant communities and biome modeling (Prentice and Webb 1988; Anderson et al 1989; Prentice et al 1992; Huntley and Birks 1983; Jackson et al 2000; Shuman et al 2002; Williams 2002). Contrary to many 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, Little 1971, 1976, 1977; Thompson et al 1998; 2000).

However, to have precise and accurate plaeoenvironmental or paleoclimate reconstructions, it is fundamental that modern pollen calibration samples span the entire range of the present-day distribution of the constituent taxa (Birks and Gordon 1985, Birks 1995). Otherwise, reconstructions for some regions could be under-represented by the modern pollen samples or lead to the existence of fossil pollen spectra without any present-day analog, known as "non-analogs" (Overpeck et al 1985; Sawada 2001; Jackson and Williams 2004). Unique fossil pollen assemblages make it difficult to reconstruct past climatic conditions at one site because there are no modern equivalents from which to infer past environmental variables. Thus, past environments may not be adequately reconstructed in the past.

Since the 1960s, several modern pollen databases have been developed. These include the Brown University Database (Avizinis and Webb 1985), the Canadian Pollen Database (Gajewski, unpubl.) and the Global Pollen Database (Grimm 2000). The Global Pollen Database (GPD) is now the leading archive of surface-sample results from research projects and includes many of the data from the two previously mentioned databases. However, because the GPD is an archive, the data is not available in a readily usable format. Furthermore, updates are not methodically implemented and authors may not routinely submit data. This can partly explain the lack of data represented in several North American regions, such as the Arctic, Alaska, western North America and southwestern United States. Although suitable for continental-scale paleoenvironmental reconstructions, for regional-scale projects a quality controlled and detailed database may be useful.

This study presents and makes available a high quality database of modern pollen and associated environmental variables applicable for quantitative and qualitative applications at regional to continental scales in North America and Greenland. Discussed is the database development methodology that is relevant to other proxy-data collection and compilation endeavors. This general methodological approach is composed of automated and manual quality-control procedures that rely partly on palynological consistency and partly on the concepts and capabilities of Geographic Information Systems (GIS). Major issues, considerations, assumptions and impediments that are involved in the compilation of data from multiple sources are reviewed, with emphasis on those that are ultimately necessary to ensure product usability at different spatial and temporal scales.

A second purpose of this study is to document and review the current status of the modern pollen network available in North America and Greenland. The evolution of the current modern pollen network for this region is reviewed through examination of the database history and major contributors. Areas where surface samples are still needed have been identified, as well as the spatial scale and quality of the data in the various regions of North America. The fundamental considerations necessary when using these data in biogeographical and paleoenvironmental contexts are discussed. Finally, spatial distribution maps of all taxa included in the surface-sample database are analyzed and, when possible compared to their corresponding vegetation extent.

This study is divided into three parts. The first section provides a general review of the assumptions and problems underlining vegetation, pollen and climate relations, different applications used in pollen based paleoenvironmental reconstructions and the historical development of modern pollen databases. The second section describes the data and quality-control procedures used to compile the database. Finally, the last section summarizes the modern pollen distribution across North America and Greenland in relation to the corresponding vegetation range of the taxa, followed by an overview of future requirements needed to improve and maintain the modern pollen database.

2. VEGETATION-POLLEN-CLIMATE RELATION

It has long been acknowledged that the broad-scale distribution of vegetation communities (i.e. biomes) is controlled by climatic factors (e.g. Seward 1892: *In* Chaloner 1994; Woodward 1987). Vegetation responds continuously to changes in the environmental variables controlling its growth (Prentice 1986). Climate variability can therefore alter the distribution and associations of species within these communities as well as their pollen production.

Over broad-scale regions, the distribution of plant taxa can be represented by the variation in the percentage of their pollen. Recorded spatio-temporal changes in fossil pollen abundance from sediment cores can be used to estimate past climate variability (Prentice 1988). Variations in pollen abundance through time therefore reflect changes in climate variables such as precipitation and temperature, which influence the plants producing the pollen. By combining both spatial and temporal variations, pollen records can be used to reconstruct past changes in the vegetation composition, and atmospheric circulation patterns can be inferred from these reconstructions (Guiot 1990; Prentice et al 1992). Methods used for past climate interpretation include qualitative analysis of pollen diagrams (e.g. Colinvaux, 1964; Cwynar 1982), response surfaces and isopoll maps (Webb III 1986; Prentice et al 1991), weighted averaging regression and calibration (ter Braak 1987; Birks 1994), density plots (Kühl et al 2002) and the modern analog technique (MAT) (Overpeck et al 1985; Sawada et al 2004; Jackson and Williams 2004).

However, these techniques rely on several assumptions, without which their method would render meaningless results. This section will begin by underlining the fundamental premises relating pollen, vegetation and climate that form the basis of these techniques. A brief overview of the North American climate history since the last glacial maximum will follow, to put in perspective the boundary conditions and controlling factors which have shaped the climate, hence the biome development through time and space. The development and different applications used in paleoevironmental research will then be discussed. The last section reviews the history of modern pollen databases in North America.

2.1 Pollen-Vegetation Relationship

Pollen grains deposited in lake or peat sediment serve as records of past vegetation, which in turn provides information about the climate (Davis 2000). Numerous studies have examined the relationship between vegetation and pollen and between vegetation and climate. The general finding from studies on surface sample sediments (the uppermost centimeters of a sediment core and is defined as pollen data less than 30-50 yr BP), indicates that vegetation patterns are adequately sensed by pollen (Davis and Webb 1975; Prentice 1988; Davis 2000). However, some considerations need to be taken. The transport and deposition of pollen is also a function of the differential production and dispersal characteristics of each species and the source area of each lake (Janssen 1973; Davis and Webb 1975; Jacobson and Bradshaw 1981; Sugita 1993; Graumlich and Davis 1993; Davis 2000, Sawada 2001). Therefore, pollen proportions in lacustrine sediment are not always linearly correlated with the proportion of plant abundance on the landscape (Jackson et al

1995; Sawada 2001). This non-linearity between pollen and vegetation percentages, also know as the "Fagerlind effect", results from the interdependency of percentages within a closed sum (Jackson et al 1995).

Most of the pollen grains released by a plant are deposited nearby, but some can be carried a considerable distance before settling. Each vegetation formation therefore produces a characteristic association of pollen grains, which depends on the constituent taxon abundance per unit area (Janssen 1973; Prentice 1988). However, the distribution limit of any plant taxon is associated with some error. The differences in dispersal and production properties among taxa causes bias in pollen assemblages towards high pollen producers (Prentice and Parsons 1983; Prentice 1988). Arboreal and wind-blown taxa, such as *Pinus*, are generally over-represented in pollen samples, while rare or insect pollinated taxa (e.g. *Acer, Larix*) are under-represented.

Pollen source area increases with basin size (Janssen 1973; Bradshaw and Webb 1985; Sugita 1993) meaning that the larger the basin (or the sparser the vegetation) the more exotic pollen will be introduced in the sediment. Exotic pollen is a term used for pollen transported far from the source, that is, pollen found in areas where the plant is not. Other factors such as input of pollen to the lake by stream inflow (Bonny 1976) and sediment focusing (Lehman 1975; Davis et al 1984) can influence pollen representation of the vegetation. However, these factors are more significant when using absolute pollen frequency measurements (see Davis et al 1973), which renders pollen percentages as influx (gains cm^{-2} yr^{-1}) and provides more reliable quantitative information if sedimentation factors are accounted for.

Biases are mitigated when a dense geographical network of pollen sites is analyzed or when distance-weighted techniques (e.g. maximum likelihood linear calibration, extended R-Value, geometric mean regression) account for the pollen source, productivity and background pollen input (Prentice and Parsons 1983; Sugita 1993; Jackson et al 1995; Sawada 2001). Pollen assemblages that are averaged over a region can spatially smooth out local vegetation variation and override the differences in source areas of different pollen types (Prentice 1988). Furthermore, for vegetation with moderately high diversity, the Fagerlind effect is negligible (Jackson et al 1995). According to Prentice (1988), there is a generally coherent quantitative correspondence between pollen samples collected from typical basin sizes (1-1000 ha) and regional vegetation composition.

Sawada (2001) compared several pollen/plant distance-weighting schemes for 17 plant taxa in southern Quebec. Results are in general agreement with a previous study conducted by Bradshaw and Webb (1985) in the Midwestern United States, which demonstrated that background pollen component (i.e. variance) decreases when sampling radius is increased. Sawada (2001) found that no specific weighing method rendered better-explained variance estimates for all taxa. Major arboreal taxa showed a stronger correlation with their pollen at a regional scale, while other types, such as *Larix* and *Tilia* showed a reduced correlation and produced a greater variance when included in paleoclimatic analysis. Thus, regional scale vegetation tends to be a better indicator of macroclimate variability (Sawada 2001).

Davis and Webb (1975) compared vegetation and pollen relationship at a continental-scale. Overall, they found that vegetation was generally well represented by modern pollen. The range limits of certain pollen types corresponded to that of the vegetation. However, the recurrent problem of exotic pollen input exaggerated values of certain species in pollen diagrams. Pollen samples from sparsely vegetated landscapes (e.g. tundra) often contain significant amounts of well-dispersed tree pollen (e.g. *Pinus, Picea*), decreasing with distance from the treeline (Gajewski 1995; Campbell et al 1999; Hicks 2001; Rousseau et al 2003). Long-distance pollen can increase the variance within pollen spectra sampled at sites where these taxa are absent from the surrounding landscape. Consequently, pollen representation is not proportional within the pollen spectra, which can affect the dissimilarity coefficient and the choice of best modern analog (Sawada 2001).

2.2 Vegetation-Climate Relationship

Biomes are large scale grouping of the vegetation based on the floristic and structural characteristics of plants. The groupings are generally composed of species having similar ecological tolerances, which renders the characteristic properties of each biome. Köppen (1936) and later, Holdridge (1947) are examples of efforts to represent the geographical distribution of major biomes in relation to the modern climate range. The primary variables explaining the distribution of plant formation are precipitation and temperature (Prentice et al 1992). Therefore, biome boundaries are determined by the major climatic air masses (Bryson 1966).

Presently, in North America, the northern regions are generally too cold to support trees. The southern limits of the resulting tundra vegetation coincides with the mean summer position of the polar front, characterized by mean annual temperature less than 0°C, low precipitation and the presence of permafrost. The axis center of the upper-air Rossby wave makes the treeline dip southward over the Hudson Bay region. South of the tundra lies the boreal forest, a circumpolar ecosystem dominated by conifers (*Picea, Pinus, Abies, Larix*). The arctic air also influences the boreal forest in the winter. However, during the summer, the biome is under the influence of other air mass. The influence of the mid-latitude or tropical air brings warmer temperature and more precipitation during the growing season. These more complacent conditions allow for more nutrients and energy to become available to the vegetation, which accounts for greater biomass productivity than in the tundra.

In temperate latitudes, in eastern United-States and southeastern Canada, more favorable climate conditions are associated with the eastern deciduous forest biome. The predominantly hardwood (e.g. *Quercus, Acer, Fagus*) and conifer (e.g. *Pinus strobus, Tsuga*) forest is under the influence of the subtropical high-pressure cell in the summer, which brings hot and moist air, with high convective activity. The biome is delimited by winter cold in the north and increasing dryness to the west.

The xerophytic vegetation in southwestern United-States results because it is under the influence of the descending branch of the Hadley Cell and rain shadow effect in the mountains. The prevalent high-pressure system over the region accounts for the low annual precipitation, which results in a desert environment. In the interior of the continent, to the east of the Rockies, dry conditions also prevail. Moisture from the Pacific air masses is lost in the mountains as the air is forced to rise, producing dry conditions on the lee side. Predominantly herbaceous vegetation lie in the treeless central interior of the continent. The prairie biome results from continental conditions associated with low precipitation, cold winters and summer droughts.

Finally, the temperate alpine biomes have a more complex vegetation composition. Vegetation tends to follow an altitudinal zonation, as a function of the local temperature and moisture conditions, with gradual replacement by different tree species. Conifers generally dominate the biome.

2.2.1 North American Climate History Since The Last Glacial Maximum

Climate is a combination of changing properties of the atmosphere interacting at different spatial and temporal scales. Climate variations are not homogeneous across the atmosphere, even at a regional-scale. Both external and internal forcings are responsible for broad-scale climate alterations ($>10^4$ year). At large time-scales, the Milankovitch orbital parameters and variations in the intensity of the solar radiation are the primary external controls. The principal internal forcing consist of the lithosphere plate movements (e.g. continental drift, orogenesis), the presence of large land masses near the polar regions, which enables ice to accumulate and facilitates cooling of the atmosphere by the albedo effect, and changes in the thermohaline circulation (Lowe and Walker 1997). However, periodic variations in the seasonal and latitudinal distribution of the solar radiation are established as being the driving forces behind long-term climate changes (CLIPMAP members 1976; COHMAP members 1988; Whitlock and Bartlein 1997).

Three orbital parameters control the aspects of the earth-sun geometry (Ruddiman 2001): the precession of the equinox, the eccentricity of the earth's orbit and changes in the tilt of the earth's axis. The 100 ka year eccentricity cycle regulates the shape of the earth's orbit around the sun. The obliquity of the axis varies with a periodicity of 41ka year, where an increase in the angle of the planet's tilt amplifies the seasonal cycle. Finally, the 22 ka year precession cycle determines the position of the summer and winter solstice and of the spring and autumn equinox along the annual orbit, thus affecting the earth's distance from the sun at various seasons.

Global Climate models (GCM) revealed the important influence of the Laurentide Ice Sheet and seasonal insolation variation on regional climate (Whitlock and Bartlein 1997). In North America, during the last glacial maximum, the ice sheet would have caused a cooling of the northern mid-latitudes, increased the latitudinal temperature gradient and diverted the jet stream south of the present-day position. Although the seasonal and latitudinal distributions of solar radiation at the last glacial maximum were similar to today, the anticyclonic circulation over the Laurentide Ice induced cold and dry conditions south of the ice margin. Conversely, during the late-glacial and early Holocene (15-9 ka), the seasonality of climate in the Northern Hemisphere was amplified by an increase in the axial tilt and the shorter earth-sun distance during the northern summer. Therefore, the smaller ice-volume during the late-glacial period caused cool humid conditions due to the northern retreat of the jet stream and the latitudinal temperature depression. In the Early Holocene, model simulations indicate that greater summer insolation and temperature caused lower effective precipitation and more droughts. Seasonal insolation extremes decreased towards modern values after 9 ka.

The spatial and temporal variations of the boundary conditions affect temperature and precipitation patterns at a regional scale. Regional climate alterations are recorded by changes in the composition, abundance and extent of biotic and abiotic systems. Simulated climatic patterns were generally consistent with variations in the observed paleoclimate indicators.

However, models convey climatic conditions unlike any today from 18 to 12 ka (COHMAP 1988; Overpeck 1992; Jackson 2000). Unique pollen assemblages occurring between 18 and 12ka coincide with the abnormal climatic conditions previously described (i.e. glacial anticyclone, strong easterlies, increasing insolation) (Webb 1986; Prentice 1986; COHMAP 1988; Overpeck et al 1992; Jackson et al 2000). Withlock and Bartlein (1997) also concluded that unique combination of large-scale climate patterns led to unique vegetation assemblages that do not have modern analogs.

2.2.2 Vegetation response to climate changes

Although climate explains the modern broad-scale vegetation distribution, do time lags exist between the vegetation response and the climate change? The maximum resolution of a past climate reconstruction is equal to the maximum resolution of vegetation response to climate changes. According to Prentice (1986), vegetation is in "dynamic equilibrium with climate if its response time is sufficiently fast in relation to the rate of climatic change to which it is observed to be responding". Therefore, the relative importance of causal processes responsible for long-term vegetation dynamics changes as a function of time and space resolution (Graumlich and Davis 1993; Prentice 1986; Webb 1986; Ritchie 1986).

At broad temporal (>10^4 yr) and regional scales, macroclimate factors are predominant influences on vegetation patterns. According to Ritchie (1986) it is at this scale that paleoecology provides the most secure records because fossil indicators show a clear, qualitative and quantitative response in both terrestrial and marine records between glacial and non-glacial and humid and arid regimes. Conversely, at finer temporal and spatial scales (1ha, 10-100 year) biotic factors such as migration lags (Davis et al 1986), soil development, topographic differentiation (Pennington 1986), and vegetation succession (Brubaker 1986) are important processes influencing vegetation patterns (Graumlich and Davis 1993). Williams et al (2001) concluded that "the extent to which the vegetation is in equilibrium with climate is scale dependent, with the equilibrium hypothesis clearly applicable to temporal scales $>10^4$ year and the disequilibrium hypothesis relevant at timescales $\leq 10^2$ year". Defining the relative importance of abiotic and climatic factors on vegetation dynamics at the intermediate spatial (10^8-10^{12} m^2) and temporal scales (10^3 year) remains problematic.

However, Gajewski (1987), and later Williams et al (2002) demonstrated that vegetation could respond rapidly to abrupt climate changes. In Williams et al (2002) the analysis was based on indicators of paleovegetation assemblages (pollen) and of climate ($\delta^{18}O$ and chironomids) from 11 high-resolution lacustrine records in North America and Europe. The results show a fast response of plants to abrupt climate changes across sites and continents; the shortest time lag at all site being <100 year and 76 % of the significant time lags < 200 year. The results confirm what Gajewski (1987) had shown earlier.

Prentice et al (1991) conducted a response surface analysis on several major pollen taxa in eastern North America to test the equilibrium hypothesis (Prentice 1986; Ritchie 1986; Webb 1986). Continental-scale isopoll maps depicting the migration of taxa at a 3000 year interval since the last glacial maximum indicate that vegetation response was in dynamic equilibrium with orbitally-induced climate changes. The findings were consistent with results from previous studies (COHMAP members 1988). Webb (1986) concluded that although non-climatic processes influence the climatic response in vegetation, the response times are short compared to the time scale of the climatic changes. Whitlock and Bartlein (1997) also showed that individualistic and non-climatic plant responses were of secondary importance in vegetation changes on a millennial timescale. Interglacial and interstadial vegetation assemblages are essentially controlled by global climate changes.

The vegetation-climate dynamic is complex and will remain a widely discussed subject in the field of Quaternary research. Davis (1989) summarized four insights in paleoecology, important to research in global climate changes: 1) species respond individualistically to climate change, 2) migration lags are generally associated with biological response to climate change, 3) disturbance regimes alter with changes in the climate, and 4) as we go back in time, various climatic and non-climatic combination can produces environmental conditions not occurring today, which translates to having plant assemblages without modern analogs.

2.3 Methods for Climate Calibration Using Pollen Assemblages

The objective of paleoenvironmental sciences is to reconstruct past environments as a function of biological data assemblages deposited in lacustrine, terrestrial and marine sediments (Birks 1995). Pollen records are the most comprehensive independent source of paleoclimatic data. Therefore, pollen-based climate reconstructions have been widely used to estimate historical climate and vegetation changes during the Quaternary. Moreover, the dense geographical distribution of pollen data allows for robust statistical analysis and quantitative estimates of environmental variables, such as temperature and precipitation (Birks and Gordon 1985).

Theoretically, paleoclimate reconstruction is achieved by transforming pollen data into climate estimates using a quantitative transfer function (Webb and Clark 1977). Surface pollen studies have shown that a quantitative relationship exists between pollen percentage of major plant taxa and their relative abundance in the vegetation (see section 2.1). Therefore, past climate estimates can be derived directly from pollen data based on the assumption that pollen distribution corresponds to that of the vegetation and also to the climatic limits (e.g. equilibrium hypothesis Webb and Clark 1977; Prentice 1986; Webb 1986; Prentice 1988).

Pollen records are generally represented by a time series diagram depicting variations in the relative abundance of a taxon through time. Abundance values for a taxon are estimated as a percentage of the total pollen grains counted at the sampled depth. The chronology of the sample is based on interpolated radiocarbon dates of several sections within the sediment core. Analyzing changes in pollen assemblages down core enables paleovegetation reconstructions at one site through time. When a network of pollen records from various geographical locations is analyzed, spatial vegetation reconstructions over an area are obtained.

Pollen deposited by modern vegetation, found at the top of a core, or surface-samples, are required for calibration. Calibration is used to estimate values of climatic variables at a fossil site based on modern data and environmental parameters (ter Braak 1987). Climate can be inferred from pollen data using response surface analyses. Response surfaces are non-linear functions that define existing relationships between pollen taxa (response variable) and the combined effect of multiple environmental gradients (predictor variables) (Bartlein et al 1986; Webb et al 1993). Plant species generally have a unimodal relationship with environmental variables, for each species thrives best at a particular optimal range of climate conditions and cannot survive when values are either too low or too high (ter Braak 1987).

Alternatively, the modern analogue technique (MAT) can be used to estimate paleoclimate by finding best modern analogs for the fossil assemblages. This method quantifies the compositional similarities between fossil and modern pollen spectra. Past pollen assemblages are then assigned environmental parameters of the modern site having the most similar pollen association. The modern analog method is based on the assumption that the functional relationship between modern climate and pollen data also existed in the past, and that both modern and past vegetation have been in equilibrium with the climate (Ritchie 1986; Prentice 1986; Webb 1986).

The modern analogue technique has been widely used in paleoclimate reconstructions because of its conceptual simplicity and ease of application (Overpeck et al 1985). The MAT has been used to reconstruct past vegetation (Overpeck et al 1985; Anderson et al 1989; Williams 2002) and paleoclimates (Guiot et al 1993; Cheddadi et al 1998; Fauquette et al 1998; Peyron et al 1998; Gajewski et al 2000) using fossil pollen. In conjunction with macrofossils it has been used to study past climates (Jackson et al 2000), and with lake level data used to estimate precipitation patterns (Guiot et al 1993). The method has been used to validate climate models (Gajewski et al 2000; Williams et al 2001; Sawada et al 2004) and to test the importance of edaphic controls on tundra vegetation (Oswald 2002).

Finally, verification of changes in past climate can be obtained by simulating the spatial changes of pollen percentages on an isopoll map (Prentice et al 1991; Anderson et al 1991; Jackson et al 2000). Global Climate Model simulations can be used in conjunction with response surfaces to generate distribution maps of taxa abundance in the past (Webb et al 1987).

2.4 History and Contribution of Pollen Databases

Modern pollen samples have long been used to aid in interpreting of pollen diagrams. By the 1960s, several studies attempted to compile pollen surface-samples in North America to aid in the interpretation of pollen diagrams (e.g. Potter and Rowley 1960; Maher 1963; Lichti-Federovich and Ritchie 1968; Heusser 1969). As emphasis moved from the descriptive comparison of modern and fossil assemblages to quantitative comparisons, extensive datasets were needed.

The compilation of a modern pollen dataset for use in paleoclimate reconstructions began in the late 1960s at the University of Wisconsin (Webb 1971). Between 1971 and the present, T. Webb at Brown University created a modern pollen database (BUPD) by accumulating hundreds of samples from Eastern North America that were obtained from the literature and from contributions by many palynologists (Webb and Bryson 1972; Webb and Clark 1977). In addition, field campaigns dedicated to accumulating new modern pollen data were mounted, especially at the Center for Climatic Research (CCR) at the University of Wisconsin. By the 1980s, the modern pollen network comprised several thousand samples, largely focusing on eastern North America (Avizinis and Webb 1985). Studies resulting from the analysis of these modern samples and cores archived in the database had a major impact on the development of Quaternary paleoecology and paleoclimatology (e.g. Cooperative Holocene Mapping Project: COHMAP Members 1988; 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. These efforts demonstrated the importance and the need for accessible, standardized and extensive pollen databases for the advancement of paleoenvironmental and paleoclimatic research.

In the 1980s, a Quebec database was mounted by P. Richard and associates (http://www.geog.umontreal.ca/palyno/umont.html), and a Canadian Pollen Database (CPD) by K. Gajewski and J.C. Ritchie (http://www.uottawa.ca/academic/arts/geographie/lpcweb). Other regional databases were established in the 1980s and 1990s, including Alaska (Anderson and Brubaker 1986) and Southwestern United States (Davis 1995). By 1989, a European and North American Pollen Database (EPD and NAPD; Cheddadi 2002; Grimm 2000) were established as archives for all pollen data from both continents and these were gradually expanded, through other continental efforts, into a Global Pollen Database (Grimm 2000). In the late-1990s both the Brown University database and the CPD were incorporated into the North American Pollen Database. In North America, the development of extensive data centers was stimulated by the Paleoclimatology Program created at the National Oceanic and Atmospheric Administration (NOAA), as part of the Global Climate Change Programs at the National Geophysical Data Center (NGDC). The Global Pollen Database (GPD) is freely available over the Internet at the NOAA/NGDC web site (http://www.ngdc.noaa.gov/paleo/gpd.html). The GPD remains an ongoing project, where pollen data is regularly added and updated.

The emphasis of the GPD is to archive pollen core data. Modern pollen data are also included, but the collection is incomplete. In the 1970s and 1980s, the Brown University Modern Database was extensively used for paleoclimatic reconstructions. Between 1997 and 1999, M. Sawada and K. Gajewski assembled a comprehensive modern database consisting of 4590 continental surface pollen samples and 89 pollen taxa (Sawada 2001). The modern pollen dataset was designed for North American paleoclimate reconstructions and included sites from the Brown University database, the CPD and the GPD. Duplicates and low-quality sites were removed in collaboration with J. Williams and A. Viau and data from Davis (1995) was added. Climate data from Leemans and Cramer (1996) were associated to the modern pollen records for paleoclimate reconstructions (Sawada et al 1999; Gajewski et al 2000; Viau et al 2002; Sawada et al 2004). Williams (2002) extracted AVHRR data at the locations of the pollen sites to reconstruct changes in vegetation cover through time. This modern pollen dataset was suitable for continental-scale paleoclimate reconstructions, yet contained few data from several regions of North America such as the Arctic, Alaska and the South-western United States. Research with this dataset also indicated that higher taxonomic resolution could improve paleoenvironmental inferences (Sawada et al 2004; Jackson and Williams, 2004). The research database presented in this study is the result of recent efforts by many collaborators to develop and release to the scientific community an improved, high-quality modern pollen dataset with associated environmental variables

3. NORTH AMERICAN MODERN POLLEN DATABASE

The new modern pollen database includes 4569 sites from across North America (Figure 1a). This research dataset is stored as a Microsoft Excel file which contains three types of information for each record (Appendix I): 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, elevation, 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 present-day tree-cover estimates derived from satellite sensors (Advanced Very High Resolution Radiometer - AVHRR). Estimates of site positional errors and the spatial scale of representation for each site are also derived in order to allow usability of the data at multiple spatial scales.

3.1 Data

3.1.1 Pollen data

The original data for the modern pollen database came from two general sources. First are core tops and samples from many sites across North America, analyzed during the past 40 years. The second source is local and regional surface sample datasets accumulated expressly for calibration studies, including Maher (1963), Lichti-Federovich and Ritchie (1965, 1968), Heusser (1969, 1973, 1978), McAndrews and Wright (1969), Fredskild (1973), Ritchie (1974), Davis and Webb (1975), Richard (1976, 1981), Peterson (1978), Davis (1995), Delcourt et al (1983), Funder and Abrahamsen (1988), Gajewski (1988, 1991, 1995, 2002), Lamb (1984), MacDonald and Ritchie (1986), Minckley and Whitlock (2000), Anderson and Brubaker (1986), Short et al (1986), Ritchie et al (1987), Anderson and Davis (1988), Willard and Weimer (1997) and several unpublished datasets (eg. McAndrews, Swain). Many other smaller collection or individual samples are included (e.g. Kerwin 2000). Data were obtained from over 350 studies and several unpublished sources. Many of these data had been previously included in several databases including the Brown University Modern Pollen Database (Avizinas and Webb 1985), the Global Pollen Database (GPD; Contributors to the GPD 2000; Grimm 2000), the Canadian/Arctic Pollen Database (Gajewski unpubl), a southwest USA dataset (Davis 1995) and an Alaskan pollen database (Anderson and Brubaker 1986; unpubl.) (Figure 1a). The pollen composition of each record is typically stored as the number of pollen grains, although some samples are stored as percent or per mille abundance (Data form information available in the database [*DataForm*]). Transforming all pollen data to percentage values renders samples comparable to each other.

Surface pollen data are generally collected using a standardized methodology, which enables the comparison of samples across North America. For instance, most lake sediment samples are collected using a surface corer or dredge from lakes 1-5 ha in size, while peat samples are typically collected by hand. Pollen processing follows a standardized procedure where a small volume of sediment (typically 1 cm^3)

is treated with a series of acids and bases to remove most of the sedimentary material, leaving the pollen grains intact (Faegri and Iversen, 1989).

The classification of pollen types is based on the taxonomic and morphological hierarchies defined in the NAPD (Grimm 2000), although case-by-case classification was required for certain pollen types. To produce a dataset that could be easily manipulated in a spreadsheet, some pollen taxa were excluded or combined into lower taxonomic levels. The pollen grains can be grouped into 65 botanical families and represent 134 taxa (Table 1). Major North American taxa and many less common types 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 the total pollen grains. Taxa that were identified at only a few locations with percentages <0.05% were excluded. The taxa incorporated into the final dataset include 99.25% of the total pollen grains present in the NAPD.

Factors such as depositional environment (e.g. lake, moss polsters, peat bogs, etc.), basin size, and changes in sampling technique through time can bias pollen assemblages (Webb et al 1978). Generally, the sensitivity of pollen assemblages to these factors increase with the geographical and temporal resolution of the study and stratifying modern pollen records according to sample properties (e.g. depositional environment, basin size) can minimize errors (Jackson and Williams 2004). Therefore, original publication references, the name of the principal investigator and the depositional environments for each samples are included as supplementary fields in the database. However, other important information, such as basin size of sampling area or depth of sample at location, was rarely provided by the original authors.

3.1.2 Geographic reference coordinates and elevation

Latitude and Longitude are stored as decimal degrees and assumed to be referenced to the geographic system of the North American Datum of 1927

(NAD27). Because the majority of modern samples within our database were collected before 1990, the site coordinates are generally derived from the most commonly available and accurate analog maps. In the United States, these analog maps would be 1:24,000 scale 7.5-minute quadrangles all referenced to the NAD27 with the series completion 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 meters in Alaska and the northern Canada (Pinch 1990). However, given the dates of collection this would be of concern only at a few sites. The datum shift error is small when compared to errors that are possibly induced by the variable coordinate precision stored for many sites in the database.

The precision of coordinate storage ranges from 0 to 7 decimal places. However, precision to 7 decimal places is not accurate to that level if extracted from visual inspection of paper maps, as seven decimal places of precision represents approximately 3 cm of longitude at the equator and 1.5 cm at 60° N. This kind of precision is only possible with highly accurate differential geographic positioning system (DGPS) and/or survey equipment, or when converting Latitude and Longitude from degrees, minutes and seconds to decimal degrees. The best precision available from a 1:24,000 paper map with the aid of mechanical rulers would be approximately 1 arc-second (4 decimal places). Because we do not know the true precision of data obtained from older databases, the original precision for all sites within our database was retained.

Table 1. List of pollen taxa variables in modern database

1. *Abies* - Fir
2. *Acer negundo* - Box Elder
3. *Acer pensylvanicum* - Striped Maple
4. *Acer rubrum* - Red Maple
5. *Acer saccharinum* - Silver Maple
6. *Acer saccharum* - Sugar Maple
7. *Acer* - Maple
8. *Alnus crispa* - Mountain Alder
9. *Alnus rubra* - Red Alder
10. *Alnus rugosa* - Speckled Alder
11. *Alnus* - Alder
12. *Ambrosia* - Ragweed
13. *Amorpha* - Leadplant
14. Anacardiaceae - Sumac Family
15. Apiaceae - Celery Family
16. Aquifoliaceae - Holly Family
17. *Arceuthobium* - Mistletoe
18. Arecaceae - Palm Family
19. *Armeria* - Thrift
20. *Artemisia* - Sagebrush, Wormwood
21. Asteraceae - Daisy Tribe
22. Asteraceae subf. Cichorioideae - Dandelion Tribe
23. *Betula* - Birch
24. Boraginaceae - Borage Family
25. *Botrychium* - Grape Fern (spore)
26. Brassicaceae - Mustard Family
27. Cactaceae - Cactus Family
28. Campanulaceae - Bluebells
29. Caprifoliaceae - Honeysuckle Family
30. *Carya* - Hickory
31. Caryophyllaceae - Chickweed Family
32. Castanea - Chestnut
33. *Ceanothus* - New Jersey Tea
34. *Celtis* - Hackberry
35. *Cephalanthus* - Button bush
36. *Cercocarpus* - Mahogany
37. Chenopodiaceae/Amaranthaceae – Goosefoot/Pigweed Families
38. *Chrysolepis/Lethocarpus* - Chinquapin
39. *Cornus* - Dogwood
40. *Corylus* - Hazel
41. Cupressaceae - Cedar Family
42. Cyperaceae – Sedge Family
43. *Dodecatheon* – Shooting Star
44. *Dryas* - Mountain Avens
45. Elaeagnaceae - Oleaster Family
46. *Ephedra* - Mexican Tea
47. *Equisetum* - Horsetails (spore)
48. Ericaceae Chamadaphne/Ledum – Leather Leaf/Labrador Tea
49. Ericaceae Vaccinium - Blueberry
50. Ericaceae – Heath Family
51. Ericales
52. *Eriogonum*
53. Euphorbiaceae - Spurge Family
54. Fabaceae - Pea Family
55. *Fagus* - Beech
56. *Fraxinus nigra* - Black Ash
57. *Fraxinus pennsylvanica/americana* – White/Green Ash
58. *Fraxinus* - Ash
59. *Iva* - Marsh Elder
60. *Juglans cinerea* - Butternut
61. *Juglans nigra* - Balck Walnut
62. Juglans - Walnut Family
63. *Koenigia islandica* - Polygonaceae
64. Lamiaceae - Mint Family
65. *Larix/Pseudotsuga* - Larch/Douglas Fir
66. *Larrea* - Creosote bush
67. Liliaceae - Lily Family
68. *Liquidambar* - Sweet Gum

69. *Liriodendron* - Tulip Tree
70. *Lycopodium annotinum* – Club moss
71. *Lycopodium clavatum* – Club moss
72. *Lycopodium complanatum* – Club moss
73. *Lycopodium selago* – Club moss
74. *Lycopodium* undifferentiated – Club moss
75. Magnoliaceae - Umbrella tree Family
76. Malvaceae – Mallow Family
77. Moraceae - Mulberry Family
78. Myricaceae - Myrtle Family
79. *Nyssa* - Sour Gum
80. Onagraceae - Evening Primerose Family
81. Osmundaceae - Royal Fern Family
82. *Ostrya/Carpinus* – Hop Hornbeam/ Ironwood
83. *Oxyria* - Mountain Sorrel
84. Papaveraceae - Poppy Family
85. *Pedicularis* – Louse Wort
86. *Picea glauca* - Black Spruce
87. *Picea mariana* - White Spruce
88. *Picea* - Spruce
89. *Pinus* diploxylon - Hard Pines
90. *Pinus* haploxylon – Soft Pines
91. *Pinus* - Pine
92. Plantaginaceae - Plantain Family
93. *Platanus* - Sycamore
94. Poaceae - Grass
95. Polemoniaceae - Phlox Family
96. Polygonaceae – Buckwheat Family
97. *Polygonum bistortoides*
98. *Polygonum* undifferentiated - Knotweed
99. *Polygonum viviparum* - Bistort
100. Polypodiaceae - Fern Family
101. *Populus* - Poplar, Aspen
102. *Potentilla* - Cinquefoil
103. *Prosopis* - Mesquite
104. *Pteridium* - Bracken (Fern)
105. *Quercus* - Oak
106. Ranunculaceae - Buttercup Family
107. Rhamnaceae/Vitaceae – Buckthorn/Grape
108. Rosaceae - Rose Family
109. Rubiaceae - Bedstraw Family
110. *Rubus* – Raspberry, Cloudberry
111. *Rumex* - Sorrel
112. *Rumex/Oxyria* - Sorrel
113. *Salix* - Willow
114. *Sanguisorba* - American Burnet
115. *Sarcobatus* - Greasewood
116. *Saxifraga cernua* - Nodding Saxifrage
117. *Saxifraga hieracifolia*
118. *Saxifraga oppositifolia* - Purple Saxifrage
119. *Saxifraga tricuspidata* – Prickled Saxifrage
120. Saxifragaceae - Saxifrage Family
121. Scrophulariaceae - Figwort Family
122. *Selaginella* -Club moss
123. *Shepherdia canadensis* - Soapberry
124. *Sphagnum* - Peat moss
125. *Taxodium* - Cypress
126. *Taxus* - Yew
127. *Thalictrum* - Meadow Rue
128. *Tilia* - Basswood
129. *Tsuga heterophylla* - Western Hemlock
130. *Tsuga mertensiana* – Mountain Hemlock
131. *Tsuga* - Hemlock
132. *Ulmus* - Elm
133. *Urticaceae* - Nettle Family
134. *Xanthium* - Cocklebur

Site elevations recorded by authors are retained but some sites with no elevation or erroneous values required correction. We choose to retain author's elevation measures 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 would lead to the extraction of elevation at points that did not correspond to the actual site location and hence render inaccurate elevation measurements for many sites. However, for 249 sites, author's elevations were missing and site elevations were extracted from high-resolution digital elevation models (DEM) recently made available for North America (www.geobase.ca).

3.1.3 Climate data

Modern climate and bioclimate variables were assigned to each record using a two-step procedure (see Appendix I). First, the Climate Research Unit (CRU) data that contain long-term monthly means (1961-1990) for temperature, precipitation and percent possible sunshine were used to calculate local lapse rates for these variables by fitting a locally-weighted trend surface regression model to the data within the neighborhood of each grid point. The model used was a third-degree polynomial in latitude and longitude, with elevation as a covariate. The search radius was 500 km, and inverse-distance squared weighting was used. Secondly, the lapse rates at each 10-minute grid point were used to adjust the CRU values to the elevation of target points. The adjusted values were then interpolated to the target points using geographically weighted bilinear interpolation.

Bioclimatic variables were calculated from the basic climate variables. The following variables were calculated following Bartlein (pers. com): growing-degree days with a 0°C and 5°C base; the "chilling requirement" (number of days when daily temperature was below 5°C); mean temperature of the coldest and warmest month of the year; January and July/annual precipitation ratio. The Priestley-Taylor (alpha) parameter (annual AE/PE) was computed using soil properties derived from the IGBP soils data set.

3.1.4 Vegetation data

Two land-cover datasets, both based upon Advanced Very High Resolution Radiometer (AVHRR) observations, provide a consistent characterization of vegetation properties for all surface pollen samples (see Appendix I). Firstly, the International Geosphere-Biosphere Programme (IGBP) DISCover database (Version 2), archived at the Land Processes Distributed Active Archive Center (http://edcdaac.usgs.gov/glcc/glcc.html) provided information about land cover type (Loveland et al 2000). Three different sets of information were extracted from the DISCover data. The modal vegetation type for each of seven land cover classification schemes (all provided by the DISCover database) within a 21km^2 search window around each site was extracted. The classification schemes included here are Biosphere-Atmosphere Transfer Scheme (BATS), Global Ecosystems, IGBP, North American Seasonal Land Cover Regions (NASLC), U.S. Geological Survey (USGS) Level 2, SIB2, and Running Vegetation Lifeforms (see http://edcdaac.usgs.gov/glcc/nadoc2_0.html for description of these classification schemes; references for each of these schemes).

The fraction of each IGBP land cover type within a 21km^2 search window was extracted. Each data column records the proportion of each land cover type within the search window, relative to total land area within the search window. Finally, the modal vegetation type, using the IGBP classification scheme, for the following search windows of 3 km^2, 11 km^2, 21 km^2, 51 km^2 and 101 km^2 was obtained. In all three cases, land cover types that consist of 'water' were not considered. The use of search windows reflects the fact that the pollen source radius can be at least several tens of kilometers for small lakes (Bradshaw and Webb 1985). In some cases, no modal biome could be found for a given search window, either because of a tie or because no terrestrial vegetation existed within the 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 identified are flagged with a –9999 in the data table.

Estimates of the percent land surface covered by needle-leaved tree canopy, broadleaved tree canopy, and open/unforested areas was obtained from the University of Maryland Continuous Fields Tree Cover dataset (DeFries et al 1999). These data rely on phenological AVHRR observations from 1992 to 1993. Search windows of dimensions 3 km^2, 11 km^2, 51 km^2 and 101 km^2 were placed around each pollen sample and the coverage values for all pixels within this window were averaged. Several site and search window combinations have no tree cover assignments, because no terrestrial vegetation existed within the search window. In addition, some sites in Greenland were outside the edge of the North American image and were not assigned tree cover values. Analyses of surface-samples and tree cover data from eastern North America suggest that search-window dimensions of 50 to 150 km produced the best agreement between the palynological and AVHRR sensors (Williams 2002).

3.2 Methodology and Quality Controls

The 4569 surface pollen samples originated from many data sources, therefore leading to possible errors and inconsistencies. Various quality control applications were developed to identify and account for possible discrepancies. Source of errors that were addressed include the removal of duplicate and low-quality samples and correction of positional and elevation errors. Other applications were performed to measure the overall quality of the final surface-sample dataset.

3.2.1 Duplicate removal

Replicate samples between and within the several databases used in the compilation of the present surface-sample dataset were removed using automated applications and on a case-by-case basis. Duplicates resulting from sites having similar geographical coordinates but different pollen assemblage (geographic duplicates) need to be discriminated through filtering procedures from real duplicates having either identical geographical coordinated and pollen assemblages or only similar assemblages (spectral duplicates).

1) Geographical and spectral filters: Two automated applications were developed to identify geographical and spectral duplicates (Appendix II; Sawada 2001). If sites A and B shared identical geographical coordinates but had different pollen spectra, both sites would be retained in the dataset. However, if the two sites had the same geographical and spectral values, then only one of the sites would be retained in the final modern dataset without any preference as to database origin. For this, the latitude and longitude decimal degrees were rounded to two decimal precisions to determine sites common between datasets.

The first application retrieves identical geographical and spectral duplicates and discards the first one found. However, imprecision in the geographical coordinates between databases can leave some duplicates unnoticed. Therefore, a second application identified spectral duplicates by computing the squared chord distance between all sites and extracted those sites having identical assemblages. Sites within each group of replicates were then manually verified, as described in the next procedure.

2) Case-by-case duplicate removal: Spectral duplicates and those that were not recognized by the automated applications due to discrepancies in the taxonomic classification or resolution between different databases required manual verification. In other instances, one database had a greater taxonomic resolution than the other. Different taxonomic resolution between datasets depends on the original data counts that were submitted by the contributors to each database. Some records also contained decimal points in the pollen counts, while another database had rounded the values; the latter were retained. Therefore, after removing duplicates using automated application the remaining sites were individually examined.

Preference criteria were established to remove remaining replicate sites between datasets. When taxonomic resolution was greater, data from the Canadian Pollen Database, Kerwin (2000), Anderson and Brubaker (1986, unpubl.) were preferred over data from the GPD, BUPD, Davis (1995) or Minkley (unpubl.), 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 Minkley (unpubl.) datasets for the same reason.

3) Duplicate removal using Geographical Information System: The remaining sites were imported into a Geographical Information System (GIS) to identify spatial duplicates record that could have been overlooked. Sites from one dataset that were within a 0.1-degree distance of a site from another dataset were selected. Pollen assemblages for these selected records were then examined, as in step 2.

3.2.2 Low-quality data removal

In previous work, some older samples were retained in the database simply because no other pollen data were available from those regions. While suitable for applications using small pollen sums, these data are simply not comparable to those produced in the past 20-30 years, as knowledge of pollen taxonomy has increased through time. Therefore, data produced before 1960 were removed. Similarly, samples clearly lacking particular taxa (such as non-arboreal taxa) were also removed from the database. A number of moss polster samples from the Arctic that had been collected from a similar location were averaged due to the high between-site variability of moss polster samples. Sites with unusually low pollen sums, or percentage values less than 90% and greater than 110% were dropped, after verifying the publication source of samples. In total, 731 problematic sites were removed after filtering all duplicate samples. For archive purposes, these low-quality samples are available as a separate file from the dataset.

3.2.3 Geographic coordinates

With the availability of Geographic Positioning System (GPS) technology, accurate location data (+/- 10 m) are presently obtainable. However, information on possible positional errors for each site within the present database is required when relating our sample site locations to existing paper maps or when integrating our data with newly obtained data of greater accuracy. Because this problem is often over-looked, we here discuss it, as understanding the sources of such errors also provides guidelines to palaeoecologists for reporting the locations of future samples.

Site coordinates within the surface-sample database have been collected from other databases and coordinates entered by original authors are often only approximate locations, sufficient to locate the region of sample collection but often not specific enough to locate the exact site itself. Site coordinates are sometimes reported to the nearest arc second, minute or in some cases only the nearest degree. While any of these coordinates may be accurate, it is not possible to know whether the value recorded was some approximation or exact in nature. Therefore, potential positional accuracy was quantified as a function of the precision with which the site coordinates were recorded, reported or entered into databases by contributors and compilers.

Two complimentary approaches were used to determine the accuracy of geographical coordinates. First, a sub-sample of site coordinates was verified within high-relief regions by comparing the elevations recorded by the author to a digital terrain model. If there are large discrepancies then it is likely that the site coordinate is in error, under the assumption that the author's elevation record is correct. Author's elevations were compared to those obtained from GTOPO30 (http://edcdaac.usgs.gov/gtopo30/gtopo30.asp) to identify problematic sites.

Large differences 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. Locations of samples in regions of high relief were targeted for validation because they require the greatest spatial accuracy; in mountainous areas, small location errors could yield inaccurate climate and vegetation assignments because of steep climatic and ecological gradients. The location of all samples with elevation differences of 300 m were identiied, except in the relatively flat eastern United States, where differences of 100 m were evaluated. All sample locations west of 100° longitude and south of 60° latitude (the high-relief Cordillera of the western United States and Canada) were checked against their originally 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 oriented to confirm and update latitude and longitude values. A total of 937 sites were identified as problematic. Of those, 365 of the sites within areas of high relief were corrected using topographic maps and the remainders were flagged.

The above approach is not appropriate for low-relief regions, because there is not enough differences in elevation between two points at a given distance. Therefore, a first approximation to site positional error can be based on the storage precision for all the longitude and latitude coordinates (number of decimal places). 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., 4 decimal places), with varying latitude, potential errors decrease towards the poles due to the convergence of meridians. Errors in the y-direction are constant for a given precision at all latitudes. For each coordinate pair, we defined the measured precision as the maximum number of decimal places recorded for either the longitude or latitude in the database. This assumes that original coordinate values as entered into the database were rounded to the nearest n^{th} decimal place, leading to precision-based error on the coordinate of at most $5 \times 10^{-(n+1)}$. Although we did not do so, one could add an additional 25 m error for sites in the United States and 50 m for sites in Canada. These latter errors embody positional errors due to national map accuracy standards for both countries assuming that coordinates were originally taken from NTS 1:50,000 maps in Canada and 7.5-minute quadrangle maps in the United States.

3.2.4 Elevation data

Similarly to geographic coordinates, accurate site elevations are important to this database for reasons of yielding accurate climate and vegetation assignments. A high-resolution digital elevation model (DEM) data within a GIS were utilized to retrieve elevations missing at 249 sites. High spatial resolution elevation data is 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 arc-second dataset. For 52 sites in Canada, elevations were extracted from recently released 3 arc-second 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 less than 2 m and well within the inherent elevation error of the datasets.

3.2.5 Overall data quality

Although the various methods discussed above can identify problematic sites and outliers, errors are frequently identified by extensive use of the database. Potential problems were also identified using a number of graphical, mapping and analyses through the use of Geographic Information System (GIS). All pollen taxa were mapped to identify unusual values in some areas.

Taxonomic resolution of pollen assemblages has changed through time as the knowledge of pollen taxonomy increased. Previous large-scale data syntheses have concentrated on major arboreal taxa of eastern and northern North America (e.g. Delcourt and Delcourt 1987; Delcourt et al 1984; Webb 1987; Webb et al 1993). Major arboreal taxa have generally been more consistently identified and reported for the past 40 years, but this is not necessarily the case for more rare or non-arboreal taxa. Therefore, paleoclimate reconstructions were often based on limited pollen sum, as this seemed to provide more reliable results (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. However, while using a coarse taxonomic resolution reduced the uncertainties associated with taxonomic inconsistencies, it also may have had the adverse result of generating low-resolution paleoclimate reconstructions.

Information statistic indices can be used to identify the variability of taxonomic quality between sites. Thus, the information content of the samples was computed using the Shannon-Weiner Index to identify sites with low pollen taxonomic resolution. The Shannon-Weiner Index (H) takes into account species richness and proportion of each species (P_i) within the pollen assemblage ($H = -\sum p_i \ln p_i$) (Maugurran 1988). The index comes from information science and measures the order observed within a particular system. In this circumstance, this order is characterized by the percentage of pollen observed for each plant taxa in the pollen assemblage. Sites with high values suggest that they have a higher taxonomic resolution than those with lower values.

Finally, the pollen assemblages of all surface-samples were compared between each other, using the modern analogue technique (MAT), to determine the ability of the dataset to discriminate between different vegetation formations in North America or to reconstruct modern climate (Gajewski et al 2002). The modern analog method is a multivariate procedure that compares each pollen assemblage to all other pollen assemblages to determine the pollen sites with the most statistically similar value to another site. If pollen assemblages in the dataset were consistently identified between contributing authors, then sites should find their best analog within the same vegetation zone, or at least within close geographical proximity. Conversely, low-quality sites (e.g. low taxonomic resolution, poorly digitized samples) would tend to find their best analogue outside of their expected biome boundaries. To determine the ability of the dataset to reconstruct modern total annual precipitation and mean temperatures, the climate assigned to the best analogue was compared to the observed climate at each pollen site. Smaller differences between the estimated and observed climate values suggest that pollen assemblages can adequately select analogues from comparable climatic regions.

The dissimilarity between modern pollen spectra was tested using squared chord distance (SCD) as a measure of dissimilarity (Overpeck et al 1985). SCD emphasizes the large-scale patterns of the differences between spectra while reducing local variability. The coefficient is computed as the summed differences in the percentages between each pollen type at one site with those of all other sites. Although including more taxa in the pollen sum can introduce more noise, it also increases precision as to the specificity of the selected best analogue for a site (Sawada et al 2004). The value ranges between 0 and 2, where the best analogs for a site are is the site having the least dissimilar pollen assemblage (i.e. lowest SCD value):

$$D_{ij} = \sum_k \left(\sqrt{p_{ik}} - \sqrt{p_{jk}}\right)^2$$

where, D_{ij} is the dissimilarity coefficient between two spectra i and j whose proportions at a fossil sample site i , p_{ik} are compared to the proportions at the modern sample site j , p_{jk}, for pollen type k. A dissimilarity coefficient, d_{ij}, is computed for each modern site for each sample spectrum (Overpeck et al 1985).

3.3 State of the Modern Pollen Dataset

The surface-sample dataset holds great potential for applications in the fields of biogeography and environmental reconstructions. However, it is important to understand the limitations and "state" of the pollen dataset to ensure that future studies provide accurate results. Several aspects of the database should be considered when used in research. These include the size and spatial extent of the dataset, the taxonomic resolution and the information content of the surface-samples, positional errors due to precision rounding of geographic coordinates, the depositional environment of samples (Jackson and Williams 2004).

3.3.1 Description of the modern pollen data

At the present time, 4569 high-quality modern pollen samples are available in North America and Greenland database (Figure 1a). Data are available from across North America, although site density varies considerably. The highest resolution available for regional-scale analyses has been determined by quantifying the site density using Voronoi polygons around all sample-points (Boots 1986) and classifying the relative area of these polygons (Figure 1b). Large areas correspond to regions having low density of points (red), while small areas represent regions with high sample coverage (dark green). 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. However, in the latter region this may be misleading, as the density map does not account for the high topographic variability. Texas is by far the region with the scarcest sample coverage. Other American regions with poor coverage include most of the southwestern and central states (California, New Mexico, Nevada and Colorado). These regions generally have few lakes and a semi-desert climate not favorable to pollen preservation and accumulation. Poorly covered regions in Canada included 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 the lack of studies conducted in the region. The site density map suggests that high spatial resolution analyses may be considered in Eastern North America, Oregon and Northern Alaska, but not elsewhere unless new data are obtained. These results demonstrate the need and opportunity for new sampling to increase areas with low site density for future work.

3.3.2 Information content and deposition environment

The taxonomic resolution between samples can vary. In general, older samples generally have lower information content as knowledge of pollen taxonomy has increased through time. Figure 1c depicts the year of publication for each data point included in the present modern pollen database. The year of publication does not necessarily correspond to the year of analysis and some samples are older than the actual publication date. However, while publication references are more readily available, contributors rarely submit the analysis date for each sample. Moreover, older studies may not have been routinely included. Data published before 1960 were removed from the database, as the level of taxonomic discrimination is frequently too low for current applications (Figure 1d). Apart from indicating potential taxonomic quality issues, this map illustrates the development of palynological knowledge in North America, which has generally progressed from the central part of the continent towards the periphery.

Other problematic samples were removed from the database (Figure 1d). Some sites were only available as percentage values. However, many samples, particularly in the southwestern United States region, had pollen sums that did not equal to 100%. Therefore, samples having a pollen sum less than 90% or greater than 110% were dropped. Other sites, located mainly in Eastern North America had low taxonomic resolution. These were typically older samples that either did not have normal Asteraceae breakdown or excluded some non-arboreal taxa (Avizinis and Webb 1985). Although 731 problematic sites were removed, sufficient higher-quality data remained in these regions to compensate for the discarded samples.

Samples in Eastern North American (excluding Florida), St Lawrence region and the Arctic generally have higher Information content (Figure 2a). Sites with lower information content are generally found across Western North America and the Boreal Forest region. The Central Prairie regions have higher between-sample variability. Information content depends on two factors: 1) the plant and pollen diversity of the region and; 2) the level of taxonomic detail reported by the contributors. Low information content could therefore arise due to diversity patterns of the flora as well as that of the pollen assemblages. However other factors, such as large pollen producers dominating the assemblage, can affect the results. That is likely the case for low-valued sites across the Boreal Forest region spanning from Alaska, through mid Canada and up to Northern Quebec/Labrador. *Picea*, *Alnus* and *Betula*, important pollen producers within the Boreal Forest, dominate the pollen spectra at the expense of other taxa, consequently reducing the overall richness. Towards the Arctic and the coastal regions of Alaska, characterized by Tundra vegetation, the values increase due to the reduced importance of large pollen producers. On the other hand, higher information content in eastern North American samples coincides with increase taxonomic resolution within the Temperate Deciduous Forest. In the case of Florida, low values could result from several sources, including the overabundance of the overrepresentation of local types in the swamp.

The nature of the deposition environment could also influence the level of information content (Figures 2c and 2d), due to influences in the pollen trapping efficiency and preservation of different pollen types (Ritchie 1974). Most sites in the database are collected from lacustrine deposits (2196 sites), with approximately 75% of sites having Shannon-Weiner index values between 1.5 and 2.5 (Figure 2d). Samples from peat deposits (338 sites) are comparable with roughly 65% of sites between the same values. Samples originating from moss polsters (860 sites) and terrestrial/soil (556 sites) deposits, tend to have lower information content, with approximately 75% of samples between values of 1 and 2. Samples that contributors did not specify sedimentary environment upon submission of their data (classified as "unknown") (580 sites) have high index value, and likely originate from lakes.

Many samples from the southeastern United States, the northern Pacific Coast and the Central Plains are moss polsters, while terrestrial/soil samples having lower information content are restricted to the southwestern United States region (Figure 2c). However, there is no formal way to determine whether samples with low information content result solely from low regional plant diversity or from low reported taxonomic resolution. It should also be noted that the apparent pollen information content might not necessarily reflect that of the data contributor, because complete counts are not always published. Some surface-samples in the database were digitized from diagrams, while others were recorded from tables obtained from the original publications.

Samples deposited in “other” depositional environment, such as stock ponds, tend to be scattered across central United States. Although, the information content between different sediment types is generally comparable, the results suggest that it may be necessary for some studies to stratify by sample depositional environment. Auxiliary information describing the form of the data (i.e. digitized vs. raw counts or percentages), publication source and notes, such as updates or corrections brought to a record, is available for each sample in the database.

3.3.3 Positional errors

Positional errors are important, given that some modern samples may originate from a very precise location, such as moss polsters, while others may be more general, such as data originating from a lake or a bog of several hectares. At a continental scale, positional errors are of minor considerations, as most analysts have located sample positions to at least the nearest minute. However, for regional scale studies, especially where data are related to vegetation surveys, positional errors could have greater importance.

Assuming that most sites are accurate at the level of the recorded precision, then the potential positional error is less than 100 m for many of the modern pollen samples, but can be over 1 km, as is the case in the southwestern United States (Figure 2b). A number of the sites with high positional errors within southwest California for example were found to be referenced to corners of 7.5 minute quadrangle maps and the sites could thus be anywhere within the quadrangle. In areas of high relief where small changes in distance equate to large changes in elevation, it stands to reason that positional errors should be positively associated with differences between authors' elevation and those extracted from a high resolution DEM. This was tested in Alberta using the CDED Level 1 DEM data for Canada. Results suggest that while all of the large differences belong to sites with low-levels of precision, not all sites with low-levels of precision have large differences (not shown). As such, while the estimates of potential errors provide a first approximation, many of the sites with low precision are likely more accurate than evident within the database. Therefore, potential errors reported in the new surface sample database should be interpreted with caution for any given site.

3.3.4 Climate range of the modern pollen dataset

Location of all pollen samples was plotted on a coordinate system of precipitation and temperature; also plotted was the entire modern climate range of North America (New et al 2000) classified by biomes (Fedorova et al 1994) (appendix III). The surface-samples are generally well distributed across most climatic zones of the continent. Area lacking data coverage include the very cold and dry Arctic regions, as well as warm and dry climates within the Desert and Prairie biomes. A better network of surface-samples is also required in the mountainous area of western North America. Results suggest that past climates comparable to those under-represented by the modern dataset might not adequately be reconstructed.

3.3.5 Modern analogues

By comparing the modern pollen samples between themselves using the modern analog technique (MAT), it is possible to estimate the extent that increased pollen information content has a practical impact on the use of these data. Similar surface-samples should find their best analogue within their own vegetation zone, or in close proximity. This is generally the case (Figure 4). More than 90% of sites from the Mediterranean, Desert and Arctic biomes find best analogues within their own vegetation zone. Characteristic taxa from these assemblages, such as *Dryas* or *Papaveraceae* for the Arctic, *Chrysolepsis/Lithocarpus* or *Ceanothus* for the Mediterranean and Cactaceae in the Desert most likely constrained the modern analogue selection of these sites to their respective geographic and climatic regions.

Samples from the Coastal, Boreal, Mountainous and Deciduous vegetation zones find 80-88% of modern analogues within their own biomes. Sites that do not find analogues in their own biome typically find analogues in adjacent biomes, with some outliers. For instance, one sample within the Deciduous forest finds its most similar site in Greenland, while several others find their best analogues in the Desert and Mountainous regions. However, these outliers only represent five samples and may be due to human disturbance increasing the NAP percentages. Samples found in the southern half of eastern United States have lower information content and different deposition environments than those in the northern-half of the Deciduous biome, which could also explain why their best analog is found outside of the biome range. While the altitudinal gradient in western North America could contribute to some discrepancies in the analogue selection for sites in the Coastal and Mountainous regions.

The Prairies, Conifer/Hardwood and Forest-Tundra biomes find their best analogues 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 and therefore the pollen rain. 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. The biome classification was generalized from Fedorova et al (1994) World Vegetation Cover. These biomes are quite small and considered ecotones in many vegetation classifications. Ecotones are transition vegetation zone between two biomes, and therefore share characteristic of their adjacent biomes. Therefore, it is not surprising that samples within ecotones have lower success in finding analogues within their own vegetation zone.

The ability of the surface-samples to find best analogue from similar climate regions was evaluated by comparing actual climate estimates to predicted climate values from the best analogue. Overall, predicted climate estimates highly correlate with observed climate values, with some scatter (Figure 4). The r^2 of the temperature recorded by the best analogues and those at the modern pollen site is 0.93 ($p < 0.001$), while the r^2 of the precipitation is 0.79 ($p < 0.001$), suggesting that temperature is more accurately estimated than precipitation.

Temperature is generally well estimated by the best analogues for all biomes. A few cases from the Arctic and the Coastal regions finds analogues in other vegetation zones that underestimate observed temperatures by approximately 20°C. However, the greatest divergences occur in the precipitation estimates. The Coastal region has the most scattered results because of the wide differences in ecological and climate gradients, which can cause estimates of the precipitation to be unreliable.

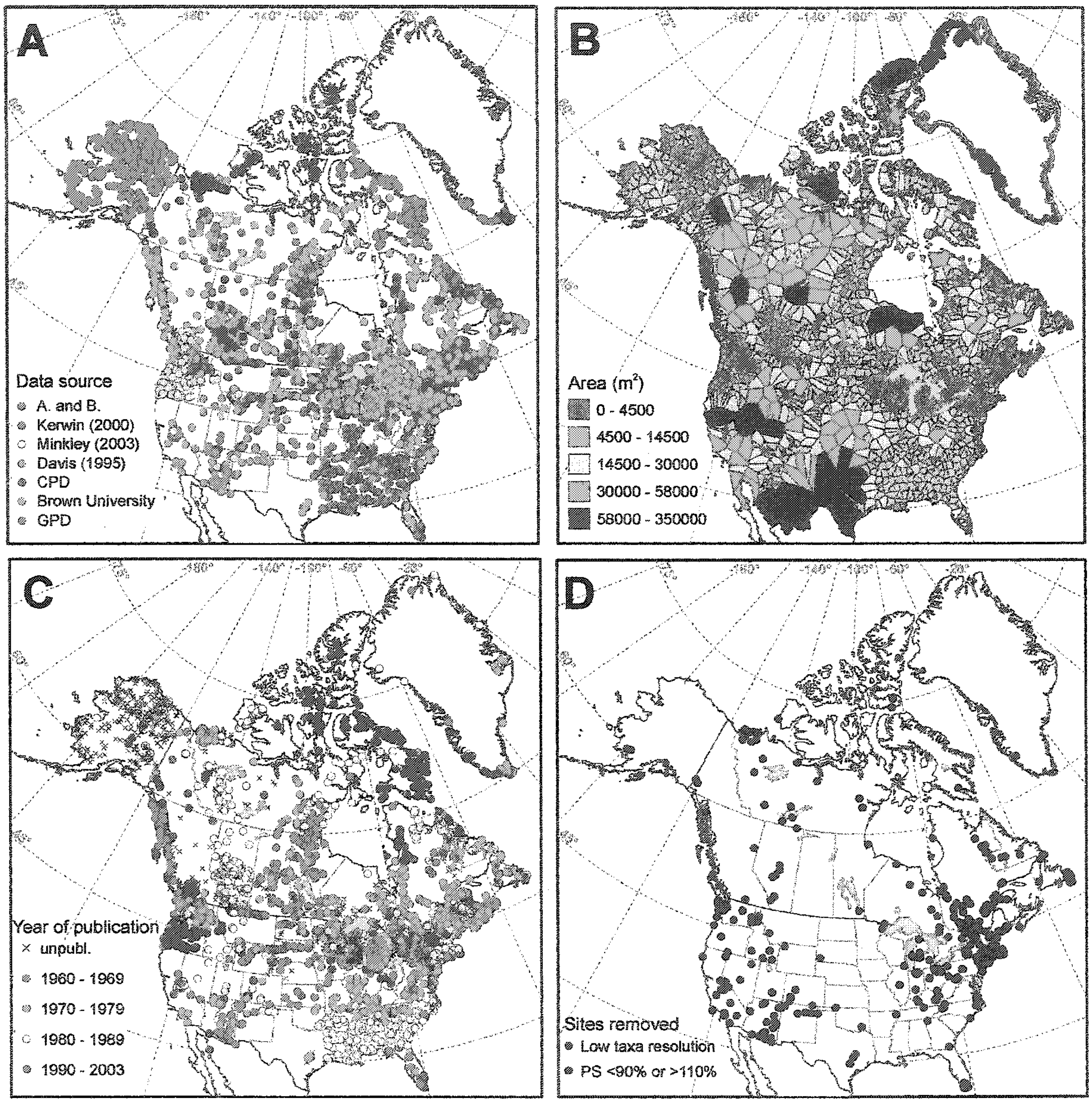

Figure 1. a) Distribution of modern pollen samples in North America classified by database origin. A. and B.=Anderson and Brubaker (1986, unpublished), CPD=Canadian Pollen Database, GPD=Global Pollen Database; b) Point density of modern pollen data in North America. The Voronoi polygon surrounding each point is classified by relative area; c) Year of publication for data in the North American modern pollen database; d) Identification of "low quality" data site removed from database.

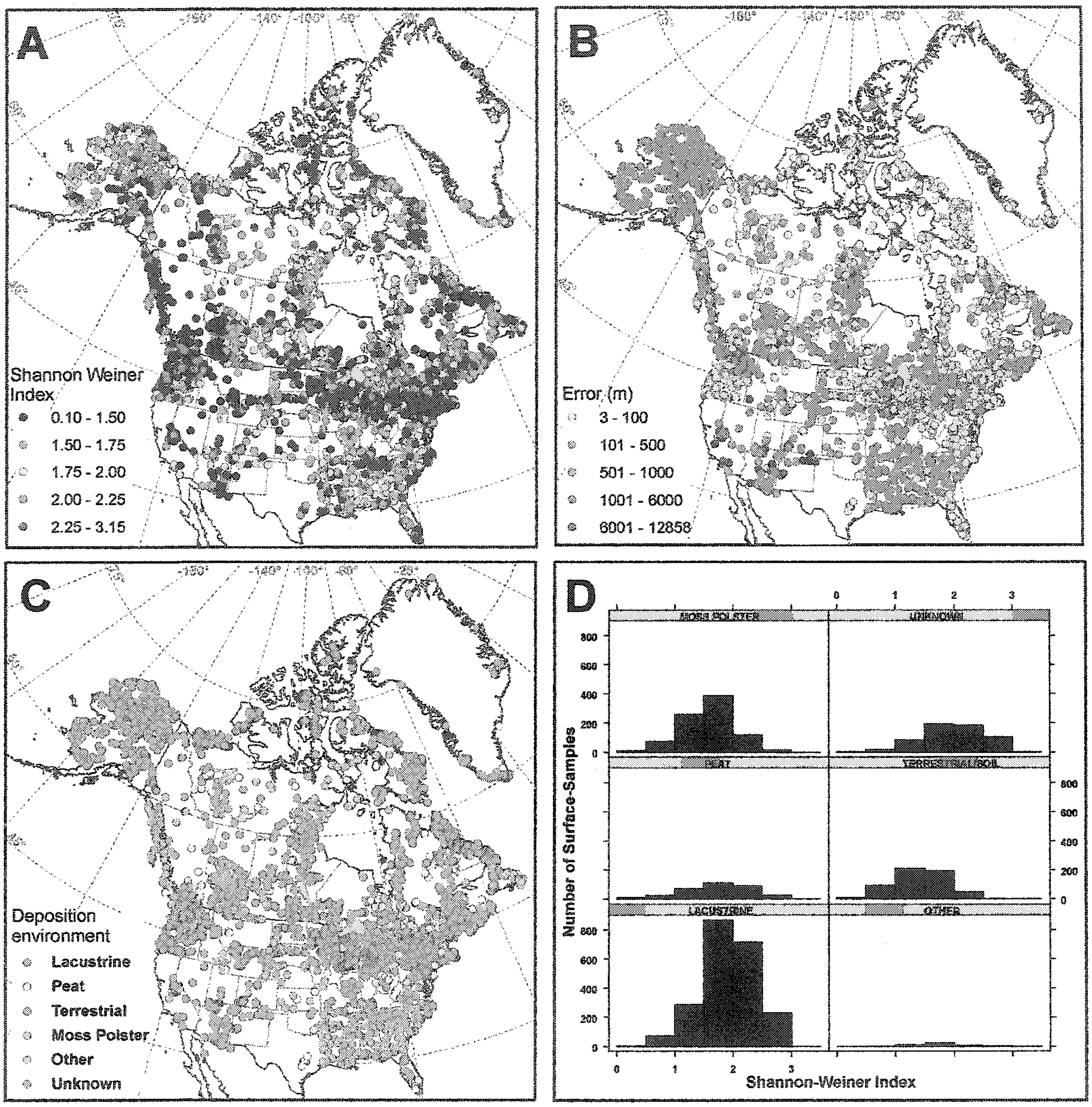

Figure 2. a) Information content (Shannon-Weiner index) of modern pollen samples in North America; b) Error in location (latitude/longitude) of modern pollen samples in North America due rounding error; c) Deposition environment of samples in the North American modern pollen database; d) Histogram conditioned by deposition environment of surface-samples showing the number of sites (counts) in relation to the Shannon-Weiner index.

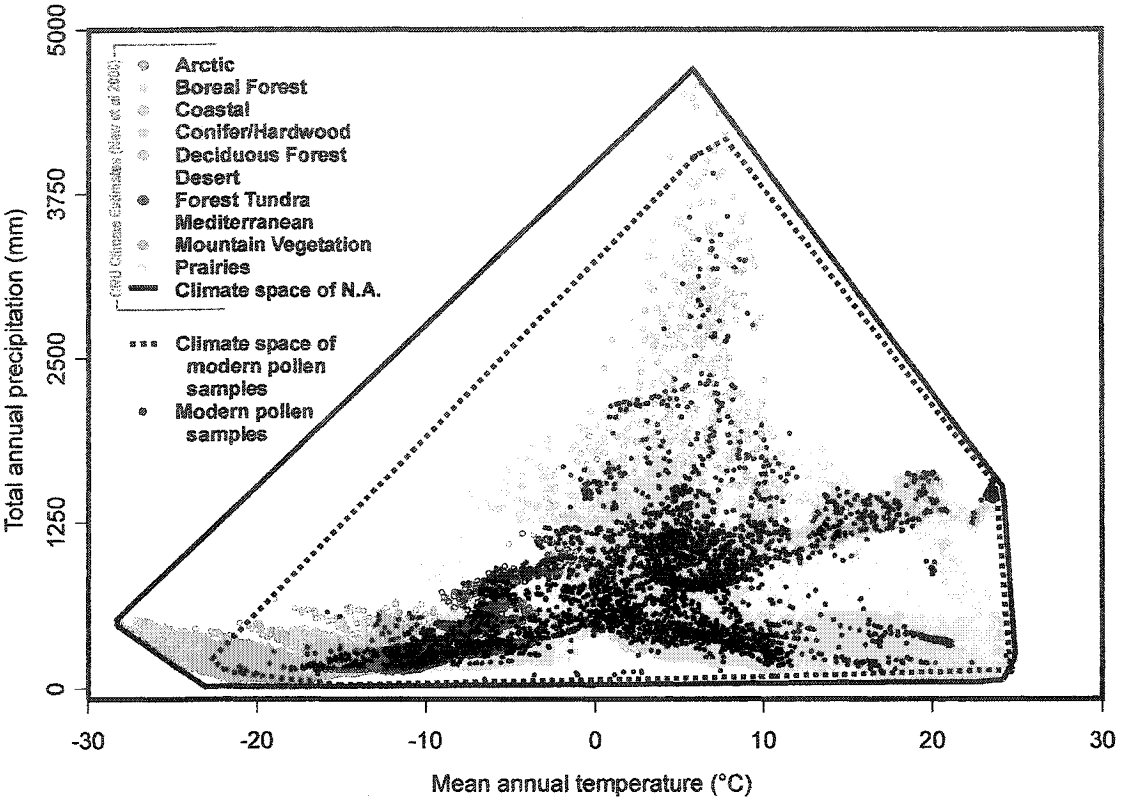

Figure 3. North American and Greenland surface-sample climate range (black points) against the modern climate range across North America (New et al 2000) classified by biomes (coloured points).

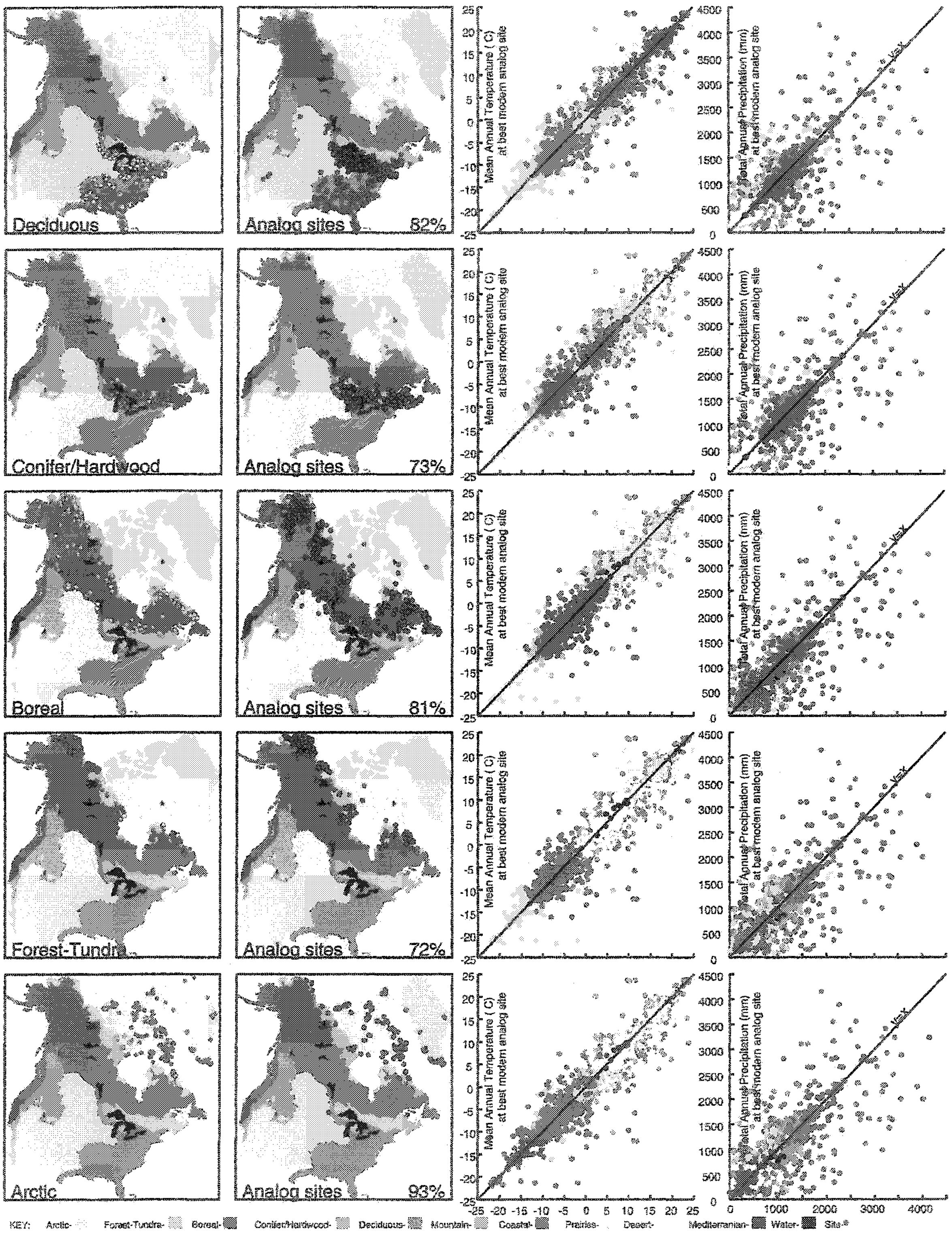

Figure 4. Modern analogues for sites in North America. From left to right, the first column shows the location of modern pollen sites (in red if modern analogue is from the same biome; coloured according to scheme in map if best analogue is from a different biome), classified by biome. The second column shows the location of sites of the best analogue for each of the sites in column 1. Percentage value corresponds to percentage of sites within biome that found their best analogue with in the same biome. The 3rd and 4th columns show a comparison of the temperature (3rd) and precipitation (4th) of each of the sites (sites within the corresponding biome are in red) with that estimated by the best analogue. The r^2 values: temperature (0.93; $p < 0.001$) and precipitation (0.79; $p < 0.001$).

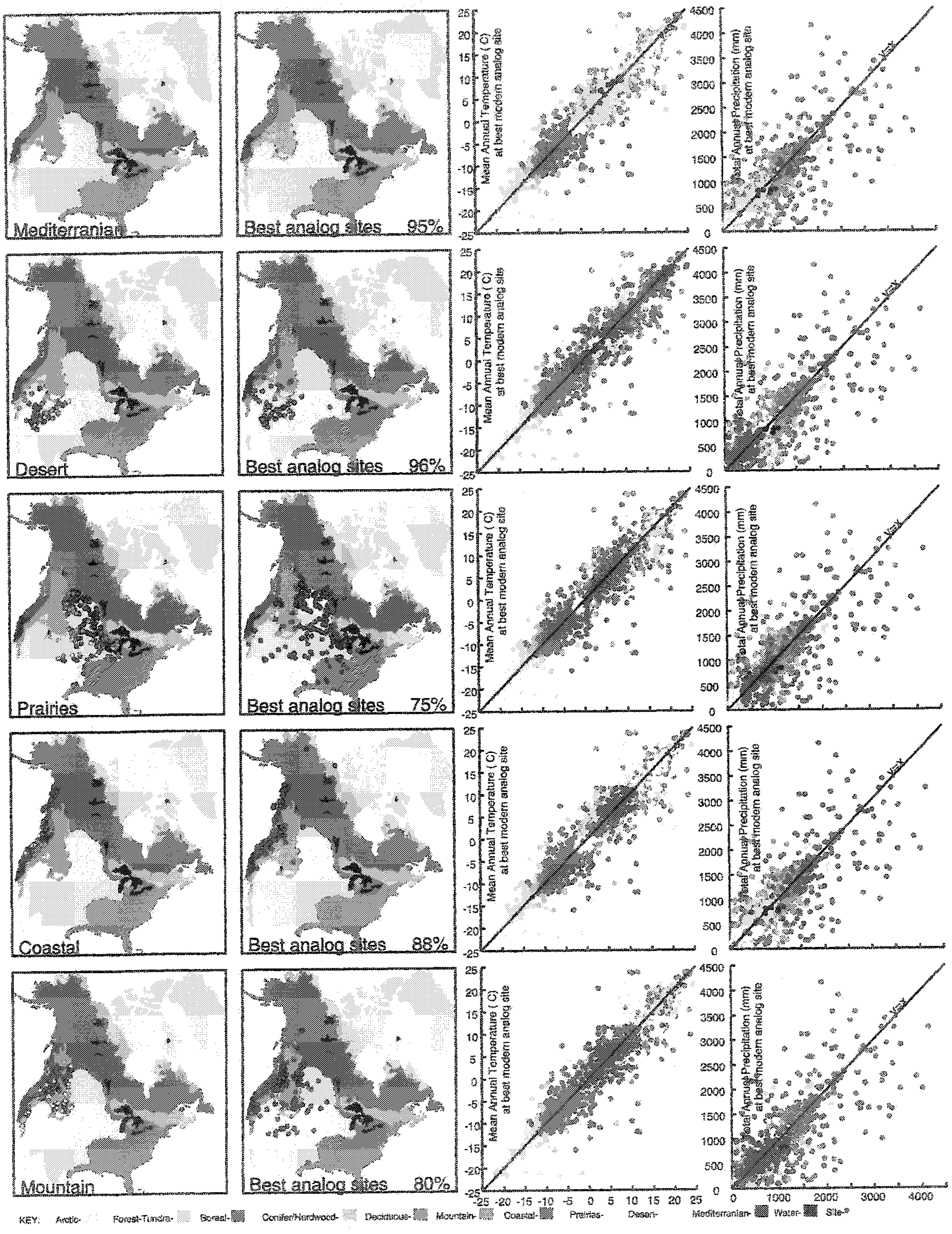

Figure 4. Modern analogues for sites in North America (cont.).

4. SPATIAL DISTRIBUTION OF MODERN POLLEN TAXA

Maps of relative abundance of the modern pollen have been prepared for all taxa (Figure 5). These serve two purposes. First, they were used as a quality control measure to identify outlying points that were investigated to ensure they are not simply based on a coding error. Secondly, 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. Previous work has concentrated on only the major taxa (e.g. Delcourt et al 1984) and only certain region of North America (e.g. Webb and Davis 1975; Bradshaw and Webb 1985; Bartlein et al 1986; Anderson et al 1989). Maps of these data allow for the identification of well constrained and represented taxa as well as those that are problematic.

For some family or genera, there is no objective basis for differentiating between species. The inability to identify pollen at lower taxonomic levels can decrease the resolution of environmental reconstructions because species from the same family or genus can have different ecological tolerances, like the Boreal/cool-temperate *Pinus resinosa* and the warm-Temperate/Subtropical *Pinus taeda* found in south United States (Jackson et al 2000). Conversely, taxa with a modern pollen distribution that coincides with their vegetation boundaries offer more reliable and accurate paleoenvironmnental reconstructions (e.g. Cactaceae or *Dryas*). As the climatic and geographical range of a taxon is reduced, the more precision it will add to paleoenvironmental reconstructions.

Therefore, to determine whether the pollen rain of a taxon concords with its geographic and climate boundaries, pollen distributions were mapped against the vegetation extent of the corresponding taxa, when available. Digitized maps from Thompson et al (1998; 2000) were used to depict the distribution of arboreal taxa. For reasons of practicality, pollen distribution maps in Figure 5 are classified in alphabetical order of taxa. The following section describes general patterns of pollen types by regional or taxa distribution.

4.1 Cosmopolitan Taxon

Many taxa, particularly forbs and grasses, are widely distributed across North America and Greenland. Some cosmopolitan pollen types nevertheless have relatively high values in some regions, as is the case for types in the Betulacaeae family. *Alnus* is transcontinental, but maximum values are found in Alaska. Individual species have been differentiated (*A. crispa, A. rubra, A. rugosa*), but only since the 1970s, as seen by the lack of distinction in older samples from the central Boreal region (see Figure 1c). Similarly, *Betula* pollen is dominant in the Alaska and Boreal regions (35-99%), but regional maximum correspond to individual species distribution. As discussed in Anderson et al (1991), high *Betula* percentages in the northern Boreal vegetation range coincide with Shrub Birch (*B. nana*), while high values in the Alaskan region correspond to Paper Birch (*B. papyrifera*). A third area of high values in the St Lawrence region corresponds to Birch (*B. lenta*). Although Anderson et al (1991) was limited Canada and Alaska, adding the southern United States does not add more information, as pollen could be long distance transported. It has been suggested that *Betula* types can be distinguished by size, but this is still controversial (Anderson et al 1989). Nevertheless, how this can be used in paleoenvironmental reconstruction is not clear.

Similarly to *Alnus* and *Betula* types, Cupressaceae is transcontinental with conspicuous maxima that coincide with individual species. Many Cupressaceae species are found in the east (Parent and Richard 1990), but maximum values in southwestern United States correspond to *Juniperus occidentalis, Sequoia* and *Cupressus* species. The specific climate requirements and constrained geographical range of these southwestern types could offer great precision in reconstructions. However, palynologists have not routinely segregated them (e.g. only three samples in the NAPD distinguish *Cupressus*).

In the Asteraceae family, *Artemisia* is ubiquitous, yet the highest values (35-90%) are found in the Sagebrush steppe of the west. *Ambrosia* is often associated with human disturbance over the past two centuries, and therefore excluded from environmental reconstructions (Jackson et al 2000; Sawada 2001). However, a definite pattern progressing westward is noticeable, with maximum values found in the eastern interior and southwest United States. The closely related genera of *Iva* and *Xanthinium* are more spatially restricted to eastern and Central United States. Asteraceae subf. Asteroideae is well distributed across North America, with percentages clearly higher in the United States (25-99%) and a scarcer representation in the Boreal region, where higher-pollen producers, such as *Picea*, dominate the pollen spectra. Asteraceae subf. Cichorioideae (dandelion subfamily) has a comparable distribution, but at considerably lower abundance. Maxima (5-15%) are generally found in the United States, East of the Cordilleras.

Apiaceae, Brassicaceae and Caryophyllaceae are ubiquitous herbs which include species that grow best in the temperate to warmer parts of the United States (especially in the west), but also include Arctic species. These taxa typically have low relative pollen abundances. Lower percentages in the Boreal region may be a function of the high percentages of *Pinus, Picea, Betula* and *Alnus* pollen that characterizes this zone. These higher pollen producers may act to reduce the percentage of the herbaceous taxa in that region, followed by an increase in their relative abundances at higher latitudes, where trees are not present. Alternatively, these herbaceous types may be less abundant in the relatively closed parts of the Boreal forest. Chenopodiaceae-Amaranthaceae has a maximum (35-98%) in the western mountains and prairies, where they are significant component of the the Prairies and Steppe ecosystems. Other species of this taxon are important weeds in agricultural area. Cyperaceae and *Salix* have a clear maxima in the Arctic, especially the North Slope of Alaska (Cyperaceae) or High Arctic (*Salix arctica*). The Alaskan Tundra is a relatively moist environment where Cyperaceae is a dominant plant, compared to the drier Tundra towards the East.

Pinus is a wind-borne over-represented taxon that is found everywhere, with maximum in the southeastern and west mountainous regions of North America. The effects of high-pollen producers, such as Pine, on the relative abundance of other taxa are noticeable within the pollen distribution of Poaceae. Although the pollen distribution of Poaceae and *Pinus* (undifferentiated) mostly reflect their relative abundance on the landscape, percentages of *Pinus* are generally the opposite of those of Poaceae; that is, where Poaceae values are high, those of *Pinus* are low, and vice versa. Nevertheless, Poaceae maxima are characteristic of the High Arctic, whereas Cyperaceae better define the Mid-Arctic (Gajewski 2002). Poaceae is also highly represented in the Coastal Tundra of Alaska, the Central Plains and southeastern United States, with lower percentages in the western United States and Boreal areas.

There are many transcontinental types with overall low percentage values. Generally, these types are low pollen producers. Plataginaceae has low percentages everywhere, with some higher concentrations in the warm-temperate and southeastern United States, while Ranunculaceae is more important in the Arctic and Greenland. High values within Polypodiacea and *Thalictrum*, on the other hand, show no definite patterns. In the case of *Populus*, pollen percentages is low more or less everywhere because the grains do not preserve well. However, a maxima is apparent in the southern Boreal forest, which coincides with the Poplar woodland transition zone between the Prairies and Boreal forests. *Populus* is also a significant a succession species in Alaska (Anderson et al 1989).

Other types are cosmopolitan at the family-level, although individual plant species of these families are regionally constrained. For example, Rosaceae is widely distributed across the continent, with higher values in the Arctic and St Lawrence regions. Some species within the Rosaceae family, such as *Dryas* and *Potentilla*, are more abundant in Arctic/alpine regions. Although *Dryas* is an Arctic plant and therefore the pollen relatively well represented its vegetation range, *Potentilla* is more widely distributed and its presence in other region may be due to under-representation or because it was not distinguished. Conversely, *Rubus* and *Sanguisorba* yield ambiguous pollen patterns, with no distinct regional maximums and disjunct distributions. Although pollen grains in both cases are quite distinctive, these types may not have been routinely distinguished, which can lead to discrepancies, or the pollen of these taxa are simply rare. Continental-scale studies should combine both *Rubus* and *Sanguisorba* counts into those of Rosaceae, whereas regional-scale studies restricted to northern Alaska or southeastern United States could use *Rubus* as a distinct type. Similarly, Saxifragaceae is ubiquitous, with maximum in the Arctic, but species could be diagnostic of the Arctic region, as shown by *S. cernua, S. hieracifolia* and *S. oppositifolia*. However, many palynologists have not always distinguished these types.

In some cases, pollen grains of some taxa are morphologically indistinguishable, such as *Rumex/Oxyria* types. Pollen distributions between the two types appear to be well discriminated; *Oxyria* is restricted to the Arctic, while *Rumex*, a weed associated with disturbance, is more wide-spread with higher concentrations west of the Hudson Bay and within the St Lawrence regions. However, because *Oxyria* is known to grow in the Arctic, while *Rumex* is more ubiquitous within the temperate region, palynologists might have distinguished these types based on their knowledge of the plant distribution, rather than the actual physical differences between the pollen grains. Fredskild (1983) claims they can be discriminated, but this has not been universally accepted. Therefore, some southern samples classified under *Oxyria* could in fact correspond to *Rumex*. In some studies conducted in Alaska and Baffin Island, either type was routinely classified under *Rumex/Oxyria* indicating that the analysts used the pollen morphology rather than the geographical distribution.

To summarize, cosmopolitan pollen types include many widely distributed arboreal and no-arboreal taxa, but they are typically at the family level. Therefore, it is not surprising that they are found across the continent. Many times, these types have maximums in the Arctic, perhaps due to their relative importance in that region, but also due to the lack of presence of other types. Some types have high percentages where expected, such as *Artemisia*, Poaceae and *Alnus*. Maximum relative abundance for these types could be used with more reliability in paleoenvrionmnental interpretation, but palynologists should remain cautious.

4.2 Eastern North America

Taxa from eastern North America have been extensively studied (e.g. Bradshaw and Webb 1985; Delcourt et al 1984). This analysis therefore builds on the previous regional mapped summaries by examining these taxa within a continental context. Here, other types not previously studied will be emphasized. Including the species distribution of three major genera: *Acer, Fraxinus* and *Juglans*. Although the distribution of the species within these genera may overlap, discrimination between species may increase the resolution of paleoclimate inferences. Most pollen grain of *Acer* can be separated using the surface sculpturing or structure (McAndrews et al 1973); the two eastern North American *Juglans* species (*J. cinera* and *J. nigra*) are distinguishable by the number and the size of the pores (McAndrews et al 1973); and species of *Fraxinus* can be divided into two types, *F. nigra* and *F. pensylvanica/Americana*, based on the number of colpae (McAndrews et al 1973). There are more species in these genera, but few are consistently distinguished. Finally, some families that have not systematically been analyzed in previous studies will be discussed.

Ten species of *Acer* are found in eastern North America and several others are found in the West coast (e.g. A. *macrophyllum, A. glabrum*) and through central North America (*A. negundo*). The overall pollen distribution of *Acer* often falls within the vegetation boundaries, particularly in the southeastern and central United States. The highest diversity of *Acer* pollen taxa is concentrated in the St Lawrence region; with some long-distance transported pollen grains, north of the vegetation range. The pollen distribution of *Acer rubrum* best corresponds to its vegetation range, with maximum values along the east coast. *Acer saccharum* has high concentrations across the St Lawrence region. There is a noticeable gap, where no pollen of *A. saccharum* are identified, within the southern half of its vegetation boundary. There are however many sites with low abundance located beyond the southern range. This gap is also visible in the distribution *A. rubrum* pollen, suggesting that the *Acer* pollen identified from sites in that region were classified under undifferentiated *Acer*. The three remaining species of *Acer* are found with percentages less than 7%. *A. negundo* has a much broader vegetation range within the continent, extending to the Canadian Prairies with some presence in the southwestern United-States (not shown in map). The vegetation range is generally well sensed by the pollen, although perhaps under-represented in the eastern United States. Similarly, *A. pensylvanicum* grows in the southern half of the St Lawrence region with an extension of its range in the eastern United States interior (not shown). Pollen of *A. pensylvanicum* are present in eastern Canada and adjacent United States, but does not extent to its southern vegetation boundary. The vegetation range of *A. saccharinum* overlaps considerably with the other types (not shown). However, the pollen distribution falls well within the plant distribution boundaries, with higher values south of the Great Lakes.

Fraxinus is also found mainly in southeastern North America, with its range extending into the Central and West Coast regions. Compared to *Acer*, *Fraxinus* has more long-distance transported pollen. Nevertheless, maximum values are found within its plant range. However, this is not always the case for the two taxa, *F. nigra* and *F. pensylvanica/americana*. While numerous pollen samples in southeastern United States contain *F. nigra*, many samples with the pollen taxon fall outside of its Great Lakes/St Lawrence vegetation range. The range of the two trees, *F. pensylvanica* and *F. americana* is better sensed by its pollen distribution, but is under-represented in the western and Prairies regions. It should be noted that this pollen taxon consists of two species, thereby reducing its use as a paleoenvironmental indicator. As with the *Acer* pollen distribution, a noticeable gap in the eastern mid-latitude United States suggests that *Fraxinus* species were not discrimminated; and there is long-distance transport of the pollen to the south leading to a high frequency of pollen surface-samples containing *F. nigra*, in southeastern United States.

Juglans cinera and *J. nigra* both have comparable vegetation extents in eastern North America, but with boundaries displaced more to the North in the first case, while more to the South in the second (vegetation range not available for *J. cinera*). Both of these types are under-represented. Although the difference in pollen distribution between these taxa appears negligible, some dissimilarity is noticeable. Samples with *J. cinera* pollen are more frequent in the northern St Lawrence region, whereas *J. nigra* is slightly more abundant in southern United States. Again, the absence of those types in the eastern mid-latitude United States, as seen in the previously discussed taxa, suggest that these species were not routinely discriminated, and should therefore be used with caution in local- and global-scale studies.

At a regional-scale, many eastern North American pollen taxa have restricted distribution that lie within the range of the plant taxon. But at larger scale, the distribution of common eastern types, such as *Carya, Fagus, Ostrya/Carpinus, Quercus* and *Ulmus,* although well sensed by the pollen data, are affected by long-distance transport. Pollen from these taxa can be found as far as Greenland and the High Arctic.

Maximum abundances (30-90%) of *Quercus* are found throughout eastern United States. However, *Quercus* is over-represented and its pollen can be transported well beyond the vegetation range, although at low abundance. In the West, several species of *Quercus* found in Mexico and southwestern United States correspond to slightly higher values of pollen. Optimal values for *Fagus* are concentrated along the St Lawrence region, with low percentages (0-5%) across southeastern United States. *Castanea* has also traditionally been associated with the eastern Deciduous forest, with the bulk of its pollen distribution found across New England. However, its southern range is under-represented. Several grains identified in Alaska are probably a coding error and will be corrected in the future version of the database. Pollen from *Liquidambar* and *Nyssa* are confined to the southeast and are less influenced by long-distance transport. Likewise, *Liriodendron* is well constrained in the southeastern Unites States, but has an overall low pollen frequency and abundance (>2%). These types could be better indicators of the warm-temperate and moist climate of the Deciduous forest. In the east, *Tilia* and *Tsuga* generally have lower percentages and are restricted to the St Lawrence/Great Lakes region (Davis et al 1986). Although there is only one species of *Tsuga* found in the east (*T. canadensis*), there are several others in the western Mountains, notably *T. heterophylla* and *T. mertensiana*. The eastern and western *Tsuga* species clearly have distinct geographical and climate preferences, but have not routinely been distinguished.

Although major arboreal taxa from eastern North America have been extensively analyzed, many other pollen taxa have not been systematically investigated. For example, Aquifoliaceae, Anacardiaceae, and *Cephalanthus,* pollen are restricted to eastern North America, although individual species of plant from these families are found as far as Mexico. These three types generally have low percentage (<10%), but Aquifoliaceae has maximum abundance along the eastern shorelines, while the two others show no conspicuous pattern within their region of high values. Good regional delimitation of pollen from Moraceae, Osmundaceae and to a lesser extent, *Platanus* could also be useful indicators eastern North American environments.

At a continental-scale, previously studied eastern North American taxa are influenced by long-distance transport. Therefore, pollen of these types found in a core may not be unequivocal proof of local presence of these taxa. Analysis of species within the *Acer, Fraxinus* and *Juglans* genera suggests that they have not consistently been distinguished and should therefore be use with caution at large-scale studies. Furthermore, the persistent lack of species distinction in the eastern mid-latitude United States indicates a need for more sampling in that region. Less common types not previously used (e.g. Aquifoliaceae, Anacardiaceae, *Cephalanthus*) were shown to be well constrained and could increase precision in paleoenvironmental inferences. Although some of these taxa may be used as indicators, others are more widespread (e.g. *Platanus*) and perhaps not routinely identified. Therefore, it is not clear how useful these would be in environmental reconstructions.

4.3 Regional Taxa

As shown in the previous section, pollen types may be well constrained by a geographic or climatic boundaries. The following section examines other taxa that are diagnostic of different regions across North America. These include areas from the Arctic, Boreal, Mountain, Prairies, Western, Florida and Southeast, Desert and Californian regions.

a) Arctic/Boreal

Taxa restricted to the Arctic and Boreal forests have previously been discussed by Anderson et al (1991) and Gajewski (2002). With the present modern pollen dataset, new Arctic samples and many more from the Boreal forest have been added. Taxa from these regions are generally indicators of cool temperature and moderate to low precipitation. Important types in the Boreal forest include *Abies, Alnus, Betula, Corylus*, Elaeagnaceae, *Equisetum*, Ericaceae and Lycopodiaceae, *Picea, Populus* and *Sphagnum*. In the Arctic, *Dryas*, *Oxyria, Papaver, Pedicularis*, *Potentilla* and *Polygonum* species are important. Note that some types important in the Boreal forest (e.g. *Alnus, Betula*) and in the Arctic (e.g. *Dryas* and *Potentilla*) were discussed above.

Abies is mostly found within its range, but with some species in the Mountainous regions having high values (25-65%). Several species of *Abies* are found in the west, although there is only one (*A. balsamea*) in the east. *Corylus* has low abundance (>15%), but is mainly concentrated within the southern transition zone between the Boreal and the Deciduous forests. However, *Corylus* pollen is transported as far north as the High Arctic and Greenland. Elaeagnaceae *and Equisetum* are also common across the Boreal forest. *Equisetum* is more widespread across the northern half of the continent, with maximum values south of the Great Lakes and in the Tundra, whereas Elaeagnaceae has consistent low pollen values (>3%).

The Ericaceae and Lycopodiaceae are widespread taxa which also have maxima in the Boreal region. Ericaceae pollen can be distinguished at least to genus (Cwynar 1982; Comtois and Larouche 1981) but this is not routinely done. Maximum values of Ericaceae pollen, as well as *Chamadaphne/Ledum* and *Vaccinium* are found in the Boreal forest and Tundra. In this region, heaths are found across the landscape and are more important than to the south, where the taxa may be more restricted. The distribution of the two genus-level taxa investigated here - *Chamadaphne/Ledum* and *Vaccinium* - do not correspond to their distribution, suggesting a lack of discrimination. In addition, some palynologists have lumped Ericaceae and Pyrolaceae pollen in the Ericales group, and might require regrouping all Ericaceae in this group when using these taxa for paleoenvironmental reconstructions. Although species identification of Lycopodium spores is possible (McAndrews et al 1973), it also appears that it was not consistently done.

Picea pollen is also rarely distinguished in pollen analyses (Lindbladh et al 2002). At a continental scale, the two major species of eastern spruce (*P. glauca* and *P. mariana*) overlap over most of their range, so there is probably little ecological information to be gained from their distinction. However, there are regional differences; *Picea glauca* is absent from northwestern Quebec, while *Picea mariana* is absent from the southwest Yukon and extends a bit further north in Keewatin. Unfortunately, it is impossible at the present time to use these species due to their lack of distinction in the database. It should be noted, that in some cases, researchers have distinguished these grains, but these have not been entered into the database. *Sphagnum* is a spore, where distribution is highly correlated with that of the boreal forest range. Some grains are transported north, but maximum percentages (30-95%) are well within its vegetation boundaries (Gajewski et al 2000).

Dryas, Papaveraceae and *Pedicularis* are restricted to the Arctic, with little to no exception. Although pollen percentages for these types are generally low, their geographic and climatic constraint make them sound proxy indicators of arctic climate. *Oxyria* pollen could also be a valuable indicator, but is often mixed with *Rumex*, as discussed previously.

Koenigia islandica is a rare type located only in the Arctic region. Presence of types such as this in pollen assemblages increases the precision of reconstructions, but it may also have been classified at the higher taxonomic level of its ubiquitous family, Polygonaceae. Although *Polygonum* should be cosmopolitan, its distribution is predominantly within the United States, probably due to representation problem discussed above. Several species are distinguished and show more definite patterns, with *P. viviparum* mainly concentrated in the Arctic and *P. Bistortoides*, in Alaska. However, it is not clear how routinely these were discriminated.

b) Mountainous and southwestern North America

Coverage of the western mountainous region is very sparse. Because the local climate changes across altitudal gradient differences, the plant communities can change rapidly over small area. However, some types characterize these regions, for example *Alnus rubra*, which has a very localized distribution on the West coast of the United States, near the Canadian borders, and *Pinus*, with its conspicuous maximum values (45-90%) across the Cordilleras.

Other types such as *Arceuthobium, Cercocarpus, Erigonum* and Malvaceae are generally well constrained within the southwestern North America region and can be used as indicators for a warm and dryer climates. Other pollen types from the Midwest have distributions that extend beyond their vegetation range, yet in a consistent way that may be interpreted from wind patterns. *Ephedra* and *Sarcobatus* may be examples of these. Maher (1964) discussed the long-distance transport of these pollen taxa, but based his analysis on a few sites from the Midwest. Patterns of decreasing pollen percentages from the source in the southwest to the northeast are seen more clearly in the current dataset. For the Desert region, Cactaceae, *Larrea* and *Prosopis* are restricted to the southwestern United States, while *Ceanothus* and *Chrysolepsis/Lethocarpus* are found in maximum abundance in the region of the Mediterranean climate along the Western United States coast, as expected. *Agave* and *Fouqueria* are also characteristic of Deserts and should be useful to define these environments using analogues. However, there are no pollen grains in the modern dataset, although several are identified in fossil assemblages archived in the NAPD.

d) Prairies

The Prairie climate consists of both winter cold and summer drought, and plants must be adapted to this environment. Many grasses and forbs are found in this region. However, most of these are ubiquitous. *Amorpha* pollen is found in only a small area of the Prairies. However, this plant taxon 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; MacDougall, 1973; Scoggan, 1978).

e) Florida and Southeastern North America

Arecaceae (Palmae) and Magnoliaceae are diagnostic types of the extreme southeastern United States and Florida regions. The climate of this region is sub-tropical, and therefore warm and moist all year round. *Taxodium* pollen is also exceptionally well distributed within its localized vegetation range, in the coastal swamps of the southeast. Maximum values (5-15%) are found along the Atlantic coast and Florida, with less frequency within the Mississippi region.

Results from this analysis indicate that most biomes have a few typical taxa, which are representative of the climate dominating the vegetation zone. Analysis of the pollen distribution for important types in the Arctic/Boreal forests confirms much of the findings from the previous works done in these regions (Anderson et al 1989; 1991). Results show that diagnostic types may be use for paleoclimate inferences, but that important types are still not consistently distinguished (e.g. *Picea*, Ericaceae). More data is still needed, particularly in the Mountainous and Prairies regions.

4.4 Rare and Problematic types

The poor pollen representation of the taxa at a continental-scale can be misleading. Less-common types, such as Campanulaceae, Liliaceae and Onagraceae, are cosmopolitan, but are poorly represented across the continent because they are insect or animal pollinated or because they are low pollen producers, such as *Bothrychium* and Lamiaceae. For example, the pollen distribution of Euphorbiaceae suggests that the taxa is regionally constrained in southern North America, when it is actually more widespread. The pollen range of *Celtis*, on the other hand, does not correspond with its vegetation range. Other types show no definite patterns or are found sporadically across the entire continent, such as *Armeria*, Boraginaceae and *Dodecathon*. *Sangisorba* and *Sherperdia* pollen grains are found in the Boreal forest and the pollen grains are very distinctive. However, these pollen types have either not routinely been identified or are under-represented in pollen assemblages. Therefore they would not be adequate to infer past distribution of the taxa.

As with *Rumex/Oxyria*, some pollen types are indistinguishable. For example, *Larix* and *Pseudotsuga*, both from the Pinaceae family, have distinct climate tolerances, but the pollen cannot be distinguished. Both pollen types coincide well with their vegetation range. However, *Larix* grows in the cooler boreal region, while *Pseudotsuga* is dominant in western North America. Consequently, this pollen type could decrease the precision of environmental reconstructions at fossil sites with a low taxonomic resolution, unless the analysis was regionally constrained. On the other hand, Rhamnaceae and Vitaceae are both low pollen producers found in the tropical and warm climates of the United States. Some species extend to the temperate zone. Although perhaps useful for continental-scale reconstructions, the pollen taxon is too general to supply ecological information at a regional scale.

The use of some individual taxa must be done with caution. Four major examples, some already discussed above, include *Picea, Pinus, Alnus* and *Ericales* pollen taxa. Although *Alnus crispa* and *A. incana* pollen are easy to distinguish, it appears that this is not routinely done, and *A. rubra* pollen are rarely distinguished. *Pinus* grains are frequently distinguished. In eastern North America, *Pinus strobus* is the only haploxylon pollen taxon, and the ranges of tree and pollen are highly correlated. 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. Probably in western Canada, most diploxylon pollen are simply lumped in undifferentiated *Pinus*. The various taxa in the *Ericales* are not routinely distinguished and the range of the pollen does not correspond well to the range of the plants. However, because workers in some regions may distinguish them, this gives the appearance of a regional distinction, and may falsely be used to find a narrow range of analogues. Some palynologists have routinely classified Ericaceaea pollen grains under Ericales, while other only use the former class, which can lead to confusion.

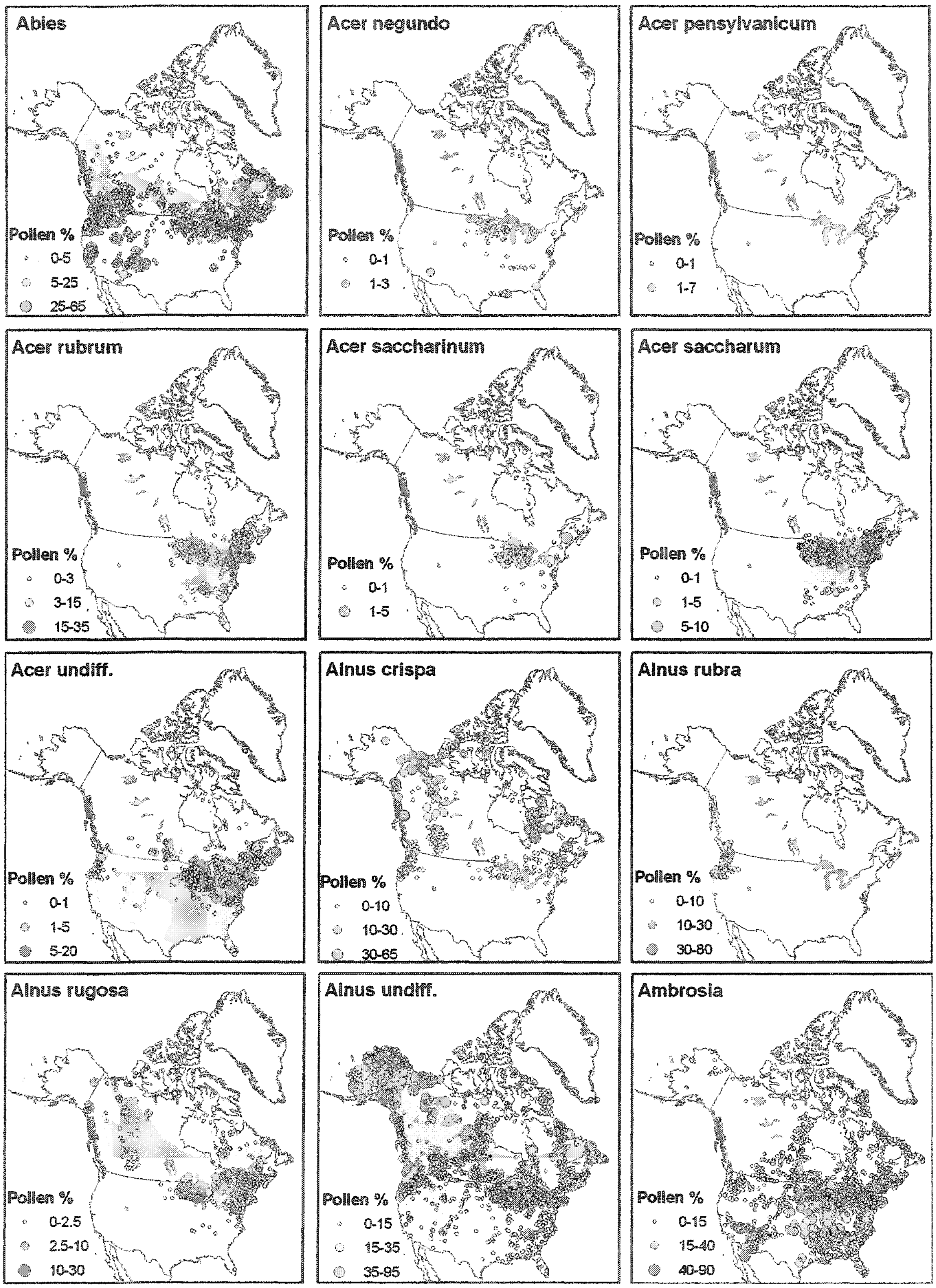

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available

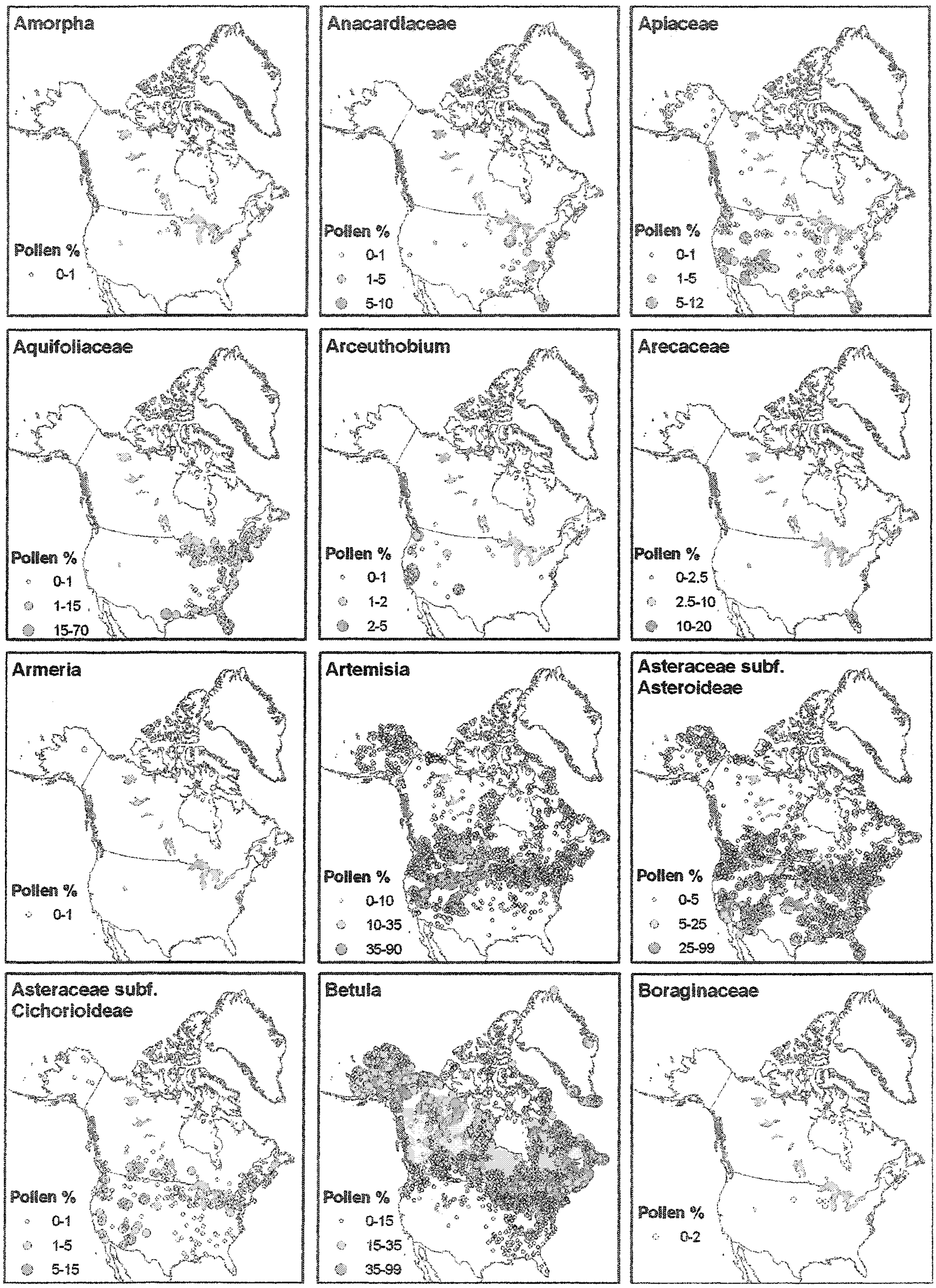

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available(cont.).

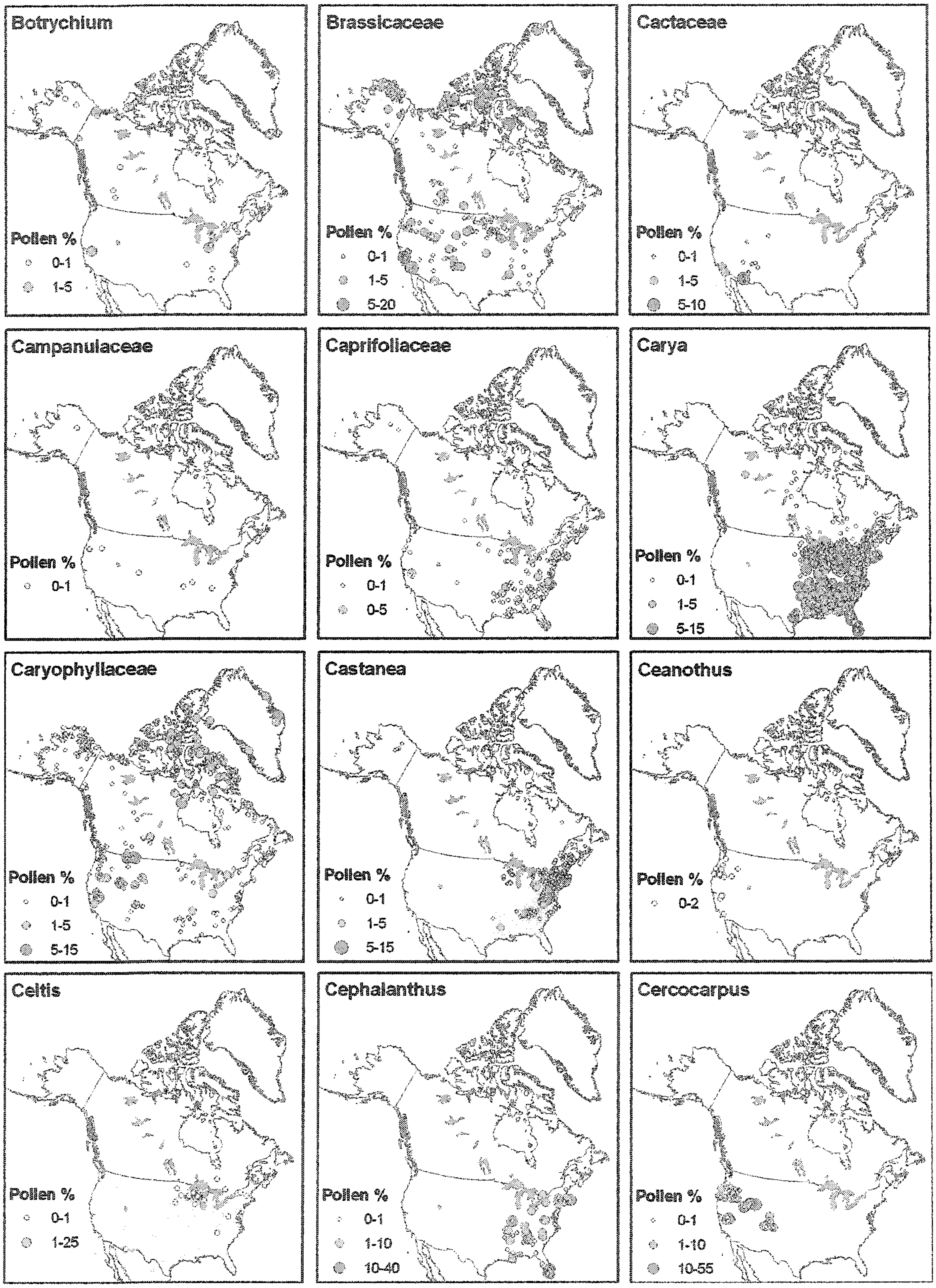

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available

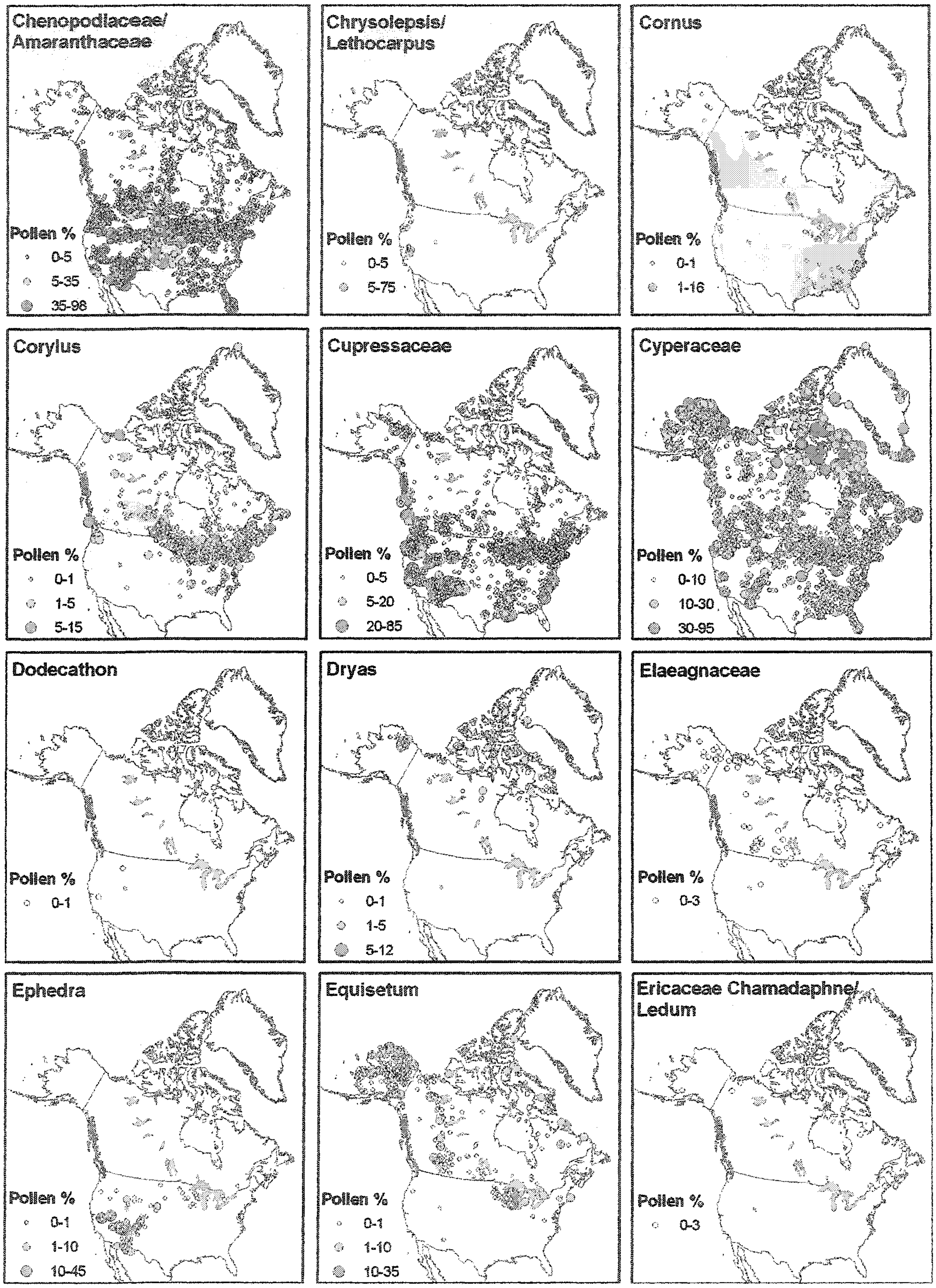

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available

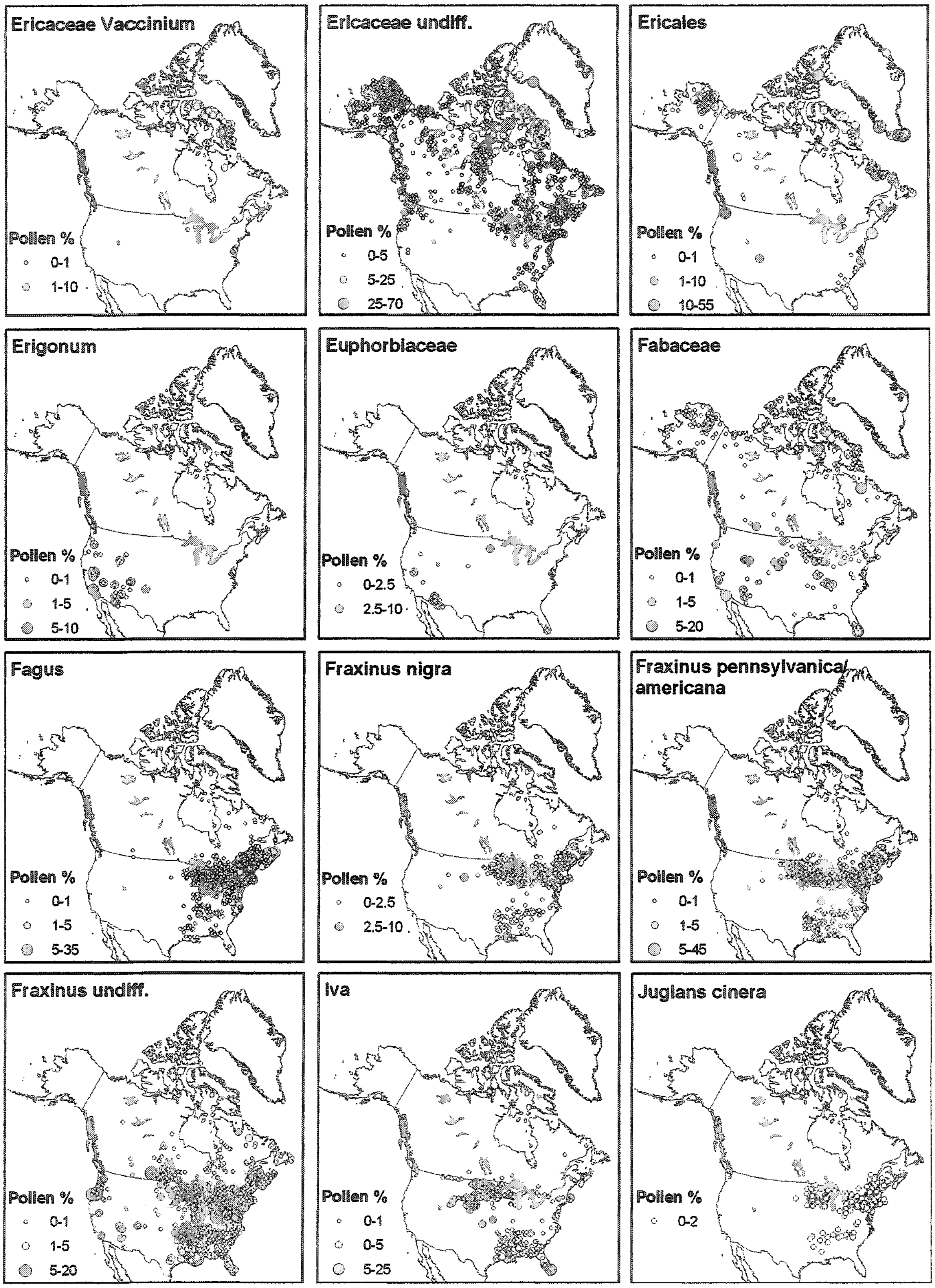

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available (cont.).

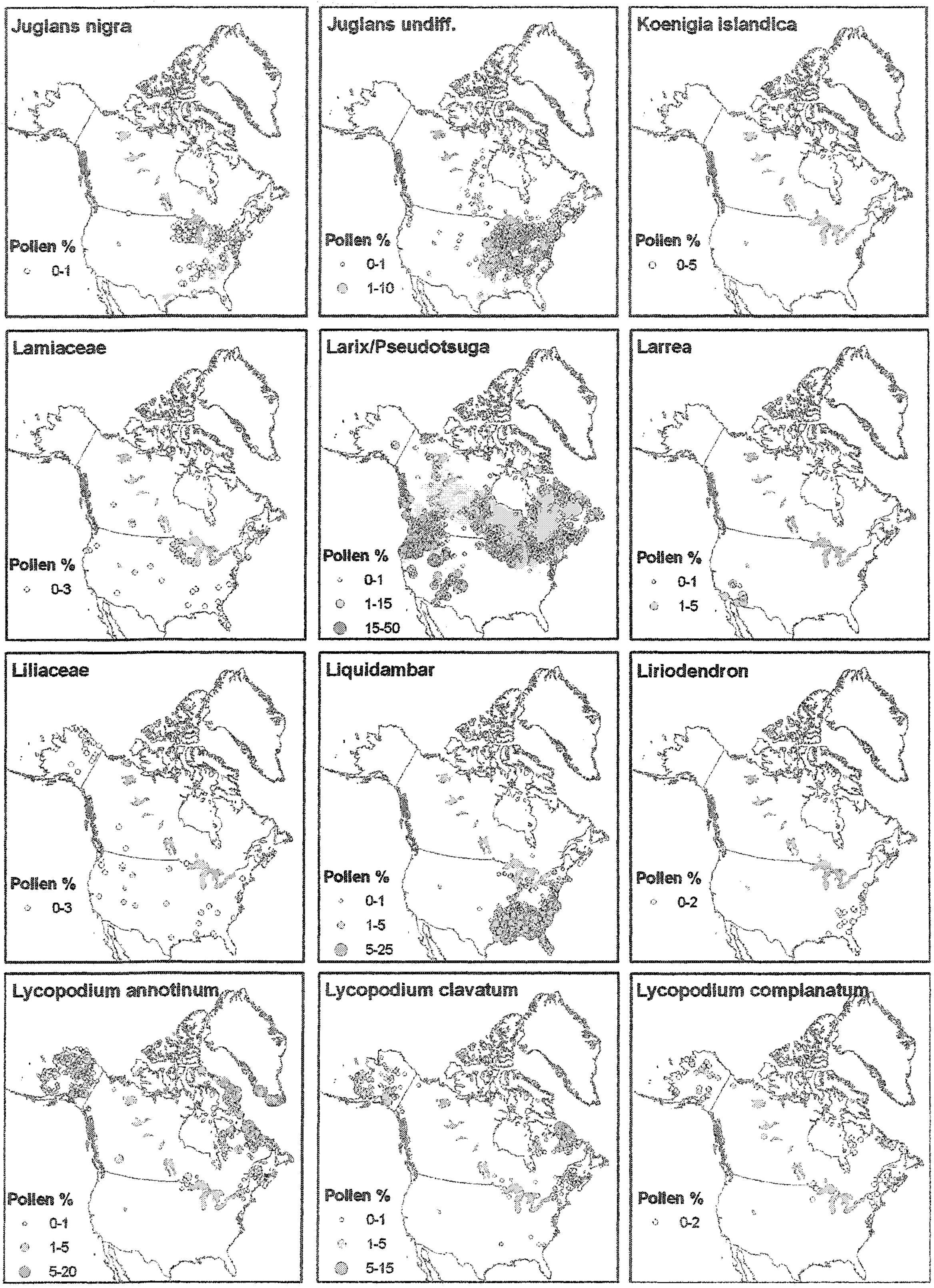

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available (cont.).

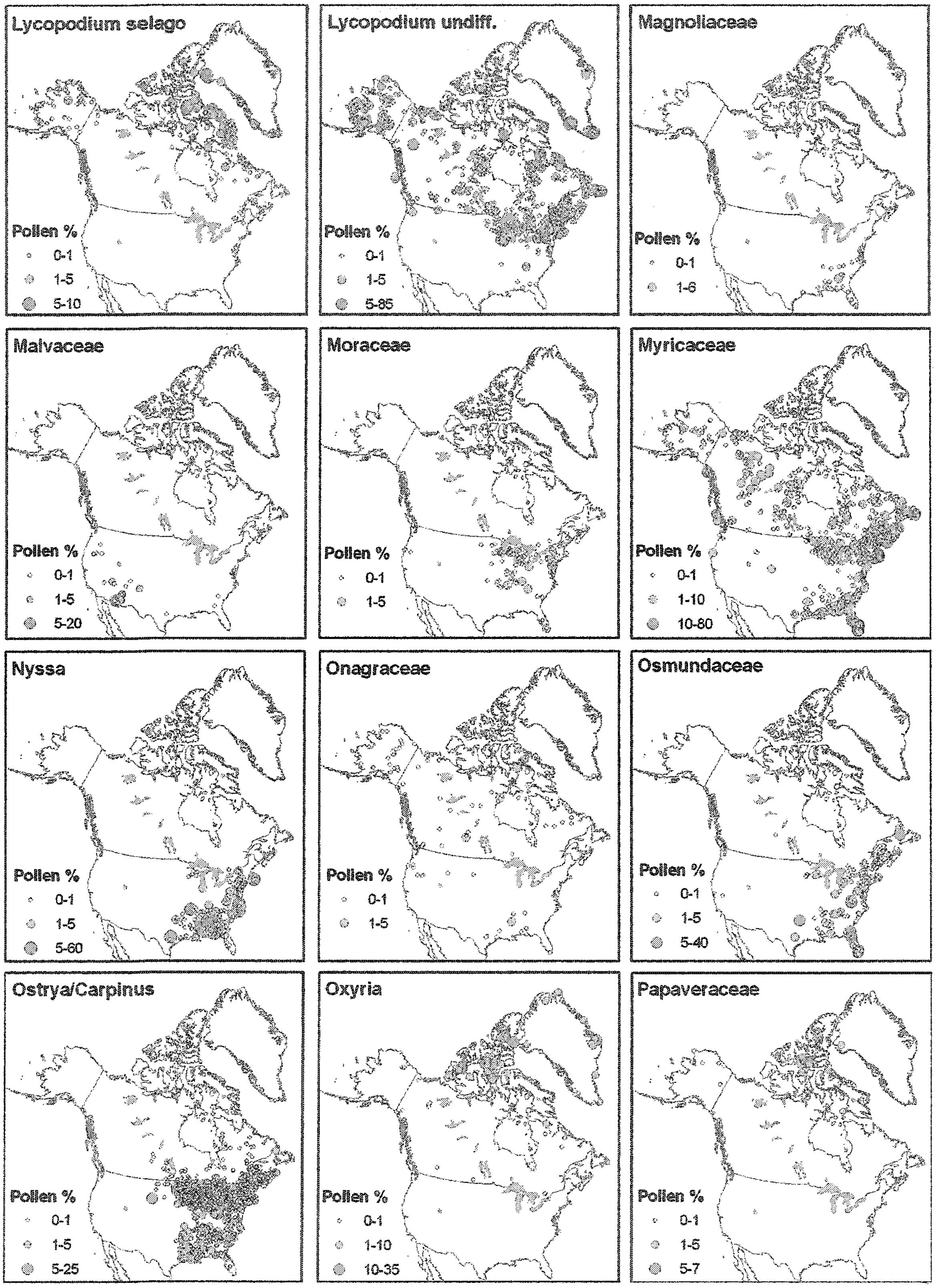

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available (cont.).

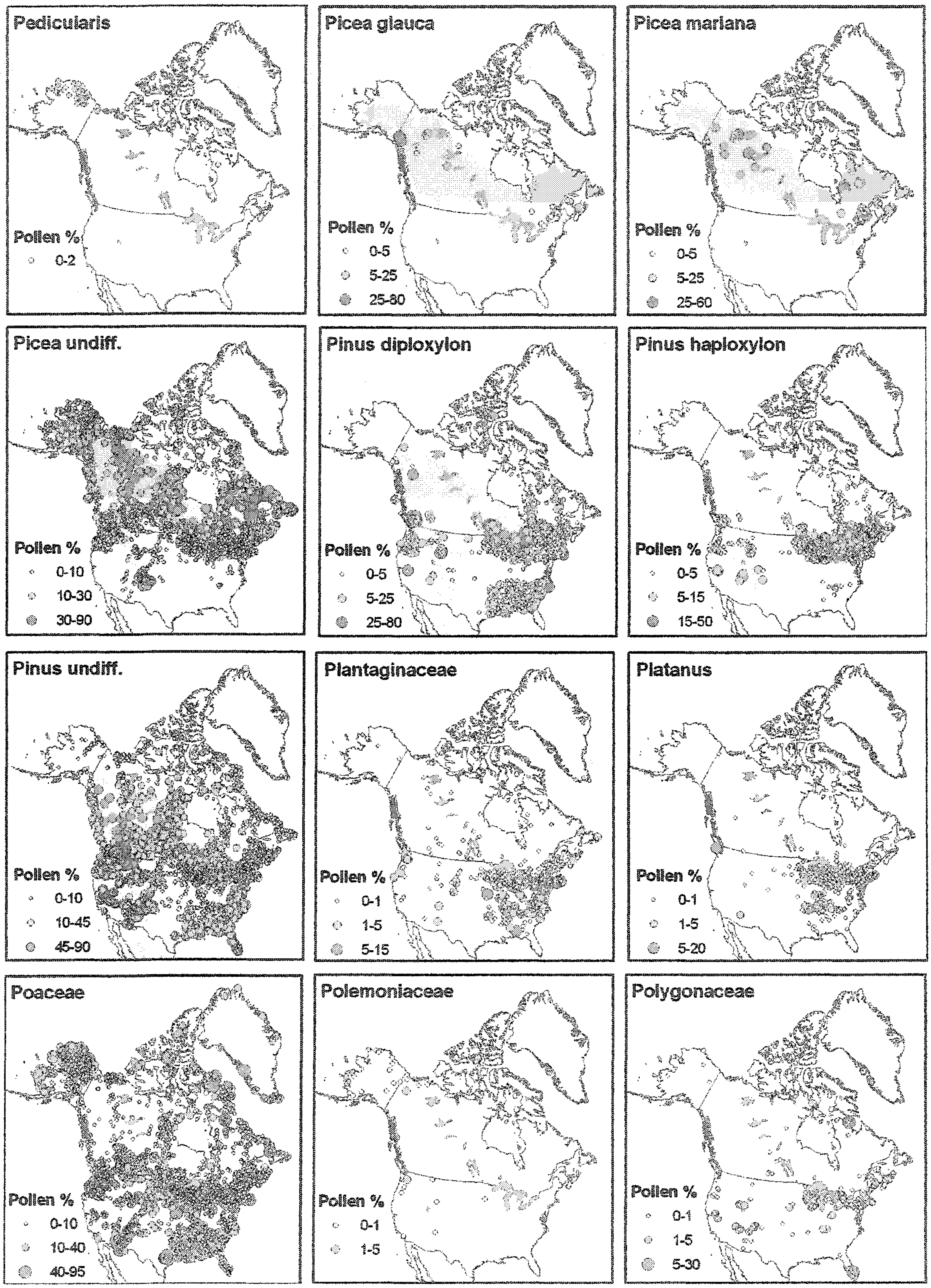

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available (cont.).

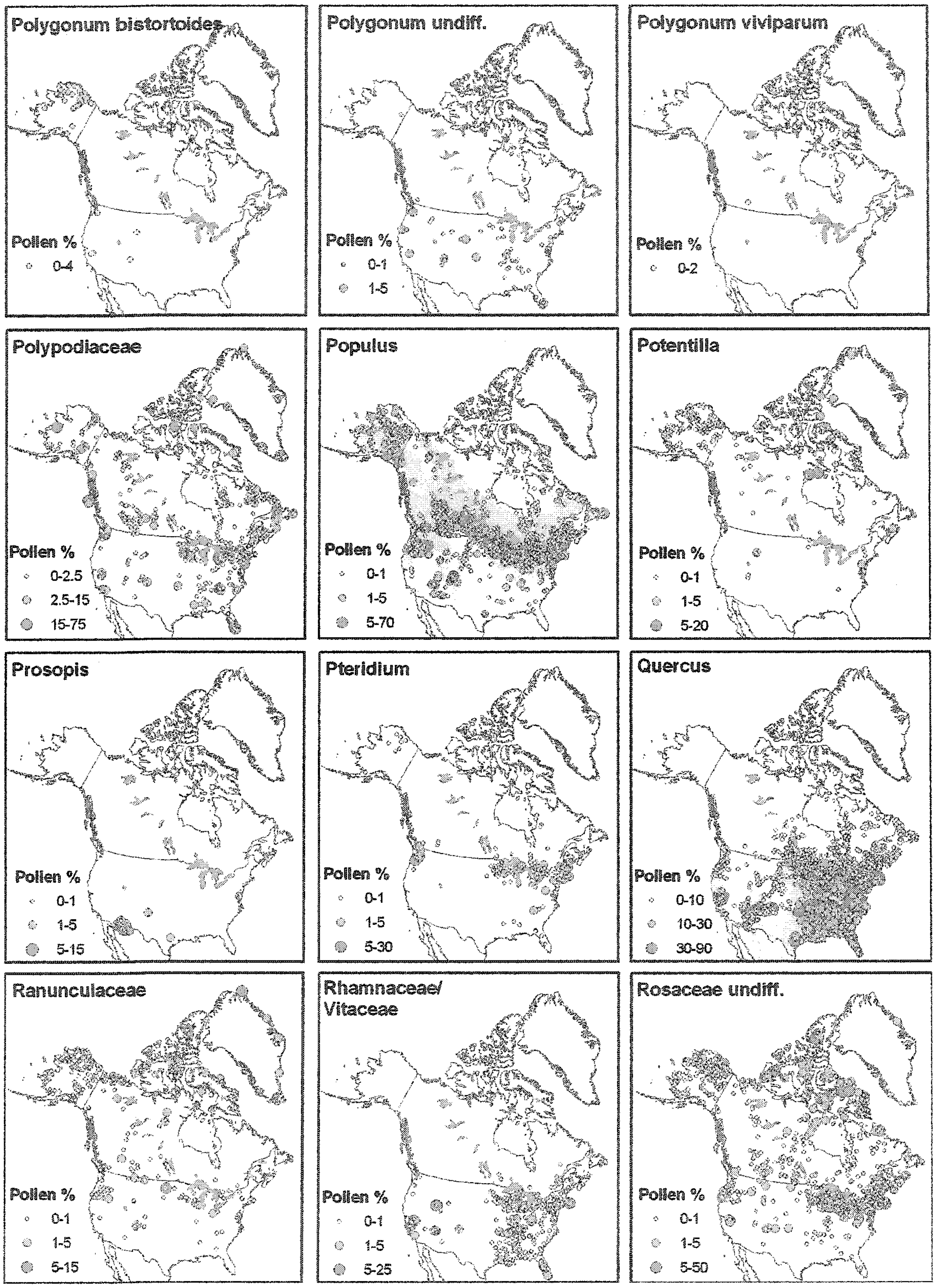

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available (cont.).

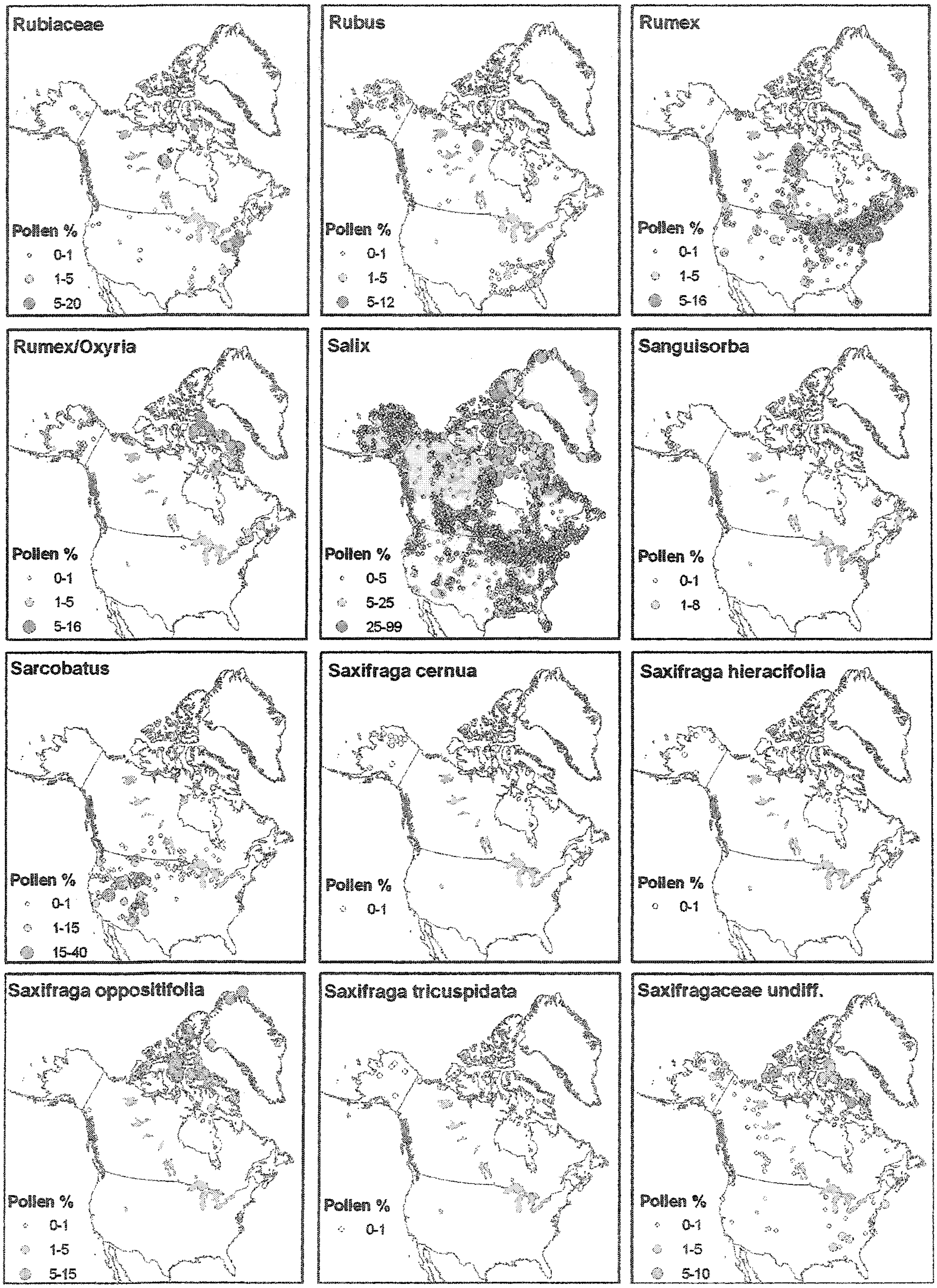

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available (cont.).

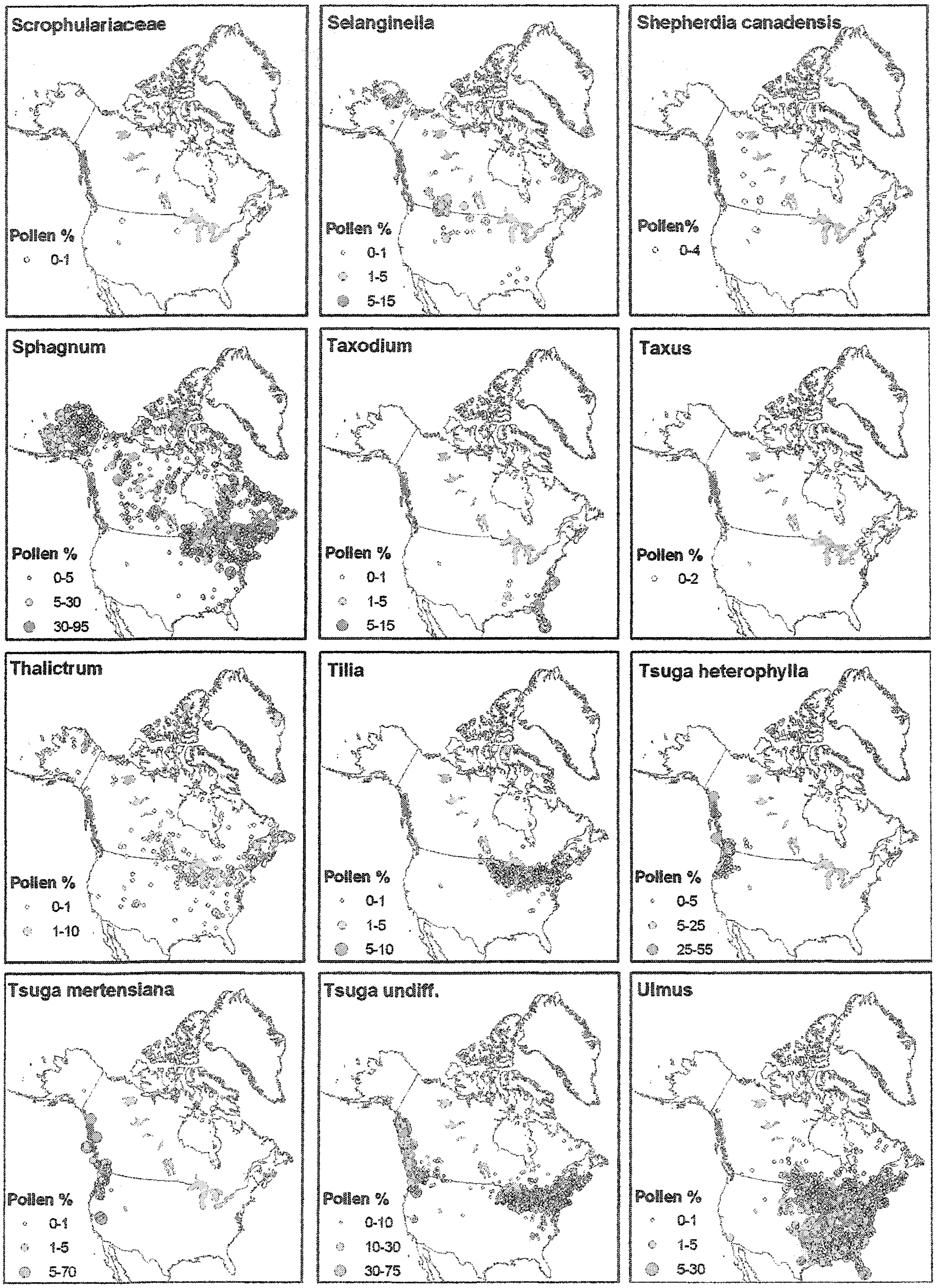

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available (cont.).

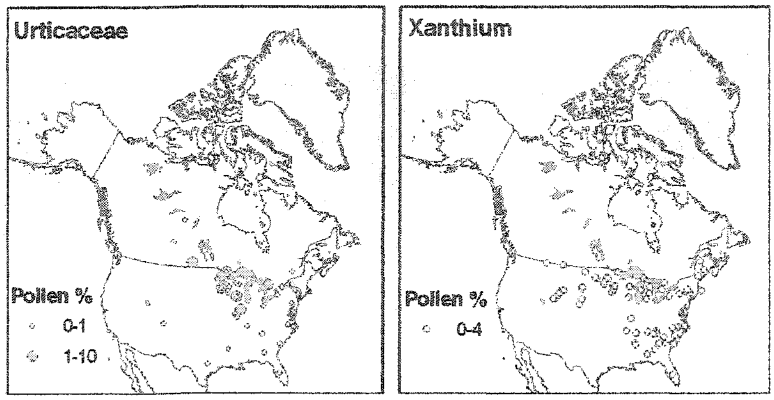

Figure 5: Maps of the distribution of the pollen of taxa in North America. Extent of vegetation for arboreal taxa (Thompson et al 1996) in gray, when available (cont.).

4. CONCLUSION

This study documents the status of the North American and Greenland modern pollen database, and indicates its use for multi-scale paleoenvironmental studies. The dataset includes 4569 samples and 134 pollen taxa and associated environmental data. This dataset will enable the quantitative reconstruction of past vegetation and climate from individual fossil pollen diagrams or regional networks of sites. A series of procedures, both automated and manual, were used to ensure the quality of the dataset. Samples older than 1960 or that did not include non-arboreal pollen or distinguish non-arboreal taxa were excluded from the database. Duplicated data between datasets compiled by different researchers at different times were also removed, and many location and elevation discrepancies were corrected. The verification and database development methodology described in this thesis is applicable to other database projects in the paleoenvironmental sciences.

Data is currently available for almost all of North America, with variable density. Areas with low data coverage (High Arctic, Texas, Northern Ontario, North West Territories, South West United States and Yukon) offer the opportunity for new data collection and regional syntheses. Pollen information content index (H) was shown to vary as a function of location, which can be a factor of regional plant diversity and the level of taxonomic detail reported by the contributors. Other factors to consider when interpreting H include the nature of the depositional environment and date of sample analysis, as taxonomic knowledge has increased through time. For studies involving high spatial precision, it may be required to stratify records by different depositional environments. The squared-chord distance computed between samples shows that most modern pollen samples find analogues in the same vegetation zone and the temperature and precipitation computed from the best analogue are highly correlated with the actual value at the site.

The mapped summary indicates that coherent geographic patterns exist in the pollen distribution of several taxa that have not previously been studied, such as

Aquifoliaceae, *Acer* and *Juglans* species. Many pollen types are well distributed within their modern vegetation range. Regional analyses have identifies several taxa (e.g. Aracaceae, Cactaceae, *Dryas*, Papaveraceae) which can be used as regional climate indicators, in paleoclimate inferences. However, some caution must be applied when using the database because of inconsistencies in the identification of certain species (e.g. *Picea*, Ericaceae) and differences in pollen representation amongst taxa. Some species, such as *Oxyria* and *Rumex*, or *Pinus* subgroups may be used in landscape- to regional-scale studies, but not in continental scale. Because the knowledge of pollen taxonomy has increased through time, it might be necessary to conduct new sampling campaigns in areas previously sampled regions (central and eastern North America) to perform high-resolution environmental reconstructions at a continental-scale. Increased sample coverage is also needed in the western Cordilleras, Desert and Prairies regions.

The availability of these data should enable better paleoenvironmental reconstruction than previously and encourage more synoptic studies. Although many of the samples are in the Global Pollen Database, the GPD is an archive of pollen data and its data require checking and quality control before they can be used for specific research goals. Data extraction can be cumbersome for users with no programming experience, and many modern data are not included. The new synthesis of the 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 among the different types of samples. Needed data may be easily extracted by using simple operations such as sorting, cut-and-paste, etc. Complete metadata is included with the dataset file, which is available on the Internet at the National Geophysical Data Centre (http://www.ngdc.noaa.gov/ngdc.html), as well as the web sites of the Laboratory for Paleoclimatology and Climatology (LPC).

REFERENCES

Anderson, P.M., Bartlein, P.J., Brubaker, L.B., Gajewski, K. and Ritchie, J.C. (1991) Vegetation-pollen-climate relationships for the arcto-boreal region of North America and Greenland. *Journal of Biogeography*, 18: 565-582.

Anderson, P.M., Bartlein, P.J., Brubaker, L.B., Gajewski, K. and Ritchie, J.C. (1989) Modern analogues of Late-Quaternary pollen spectra from the western interior of North America. *Journal of Biogeography*, 16: 573-596.

Anderson, P.M. and Brubaker, L.B. (1986) Modern pollen assemblages from northern Alaska. *Review of Palaeobotany and Palynology*, 46: 273-291.

Anderson, R.S and Davis, O.K. (1988) Contemporary pollen rain across the central Sierra Nevada, California: the relationship to modern vegetation types. *Arctic and Alpine Research*, 20:448-460.

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

Bartlein, P.J., Prentice, I.C. and Webb III, T. (1986) Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography*, 13: 35-57.

Birks, H.J.B. (1995) Quantitative paleoenvironmental reconstructions. In Maddy, D. & Brew J.S. (Eds.), *Statistical Modelling of Quaternary Science Data*, 161-254. Technical Guide 5, Quaternary Research Association, Cambridge.

Birks, H.J.B. (1994) The importance of pollen and diatom taxonomic precision in quantitative palaeoenvironmental reconstruction. *Review of Palaeobotany and Palynology*, 83: 107-117.

Birks, H.J.B. and Gordon, A.D. (1985). *Numerical Methods in Quaternary Pollen Analysis.* Academic Press, London, 317 pp.

Bonny, A.P. (1976) Recruitment of pollen to the seston and sediment of some lake district lakes. *Journal of Ecoogy*, 64, 859-887.

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

Brubaker, L.B. (1986) Response of tree populations to climatic change. *Vegetatio*, 67: 119-130.

Bryson, R. A. (1966) Airmasses, Streamlines and the Boreal Forest. *Geographical Bulletin*, 8(3): 228-269.

Boots, B.N. (1986) *Voroni (Thiessen) Polygons*. Geo Books, Norwich.

Campbell, I.D., McDonald, K., Flannigan, M.D. and Kringayark, J. (1999) Long-distance transport of pollen into the Arctic. *Nature*, 399: 29-30.

Chaloner, W.G. (1994) Fossil plants as palaeoenvironmental indicators. *In* Boulter, M.C. and Fisher, H.C. (eds) *Cenozoic Plants and Climates of Arctic*, NATO ASI Series, 27:13-21. Springer-Verlag, Berlin.

Cheddadi, R. (Ed.) (2000) *European Pollen Database*. Data archived at IGPB PAGES/World Data Center-A for Paleoclimatology, NOAA/NGDC Paleoclimatology Program. Boulder, Co. (www.ngdc.noa.gov/paleo)

Cheddadi, R., Lamb, H.F., Guiot, J., van der Kaars, S. (1998) Holocene climate change in Morocco: a quantitative reconstruction from pollen data. *Climate Dynamics*, 14: 883-890.

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*. PhD Dissertation, University of Wisconsin, Madison.

Colinvaux, P.A. (1964) The environment of the Bering Land Bridge. *Ecological Monographs*, 34: 297-329.

Comtois, P. and Larouche, A. (1981) Morphologie pollinique des Éricales du Québec. *Naturaliste Canadien*, 108 : 245-262.

Cwynar, L.C. (1982) A late-Quaternary vegetation history from Hanging Lake, Northern Yukon. *Ecological Monographs*, 52(1): 1-24.

Davis, M.B. (2000) Palynology after Y2K-Understanding the source area of pollen in sediments. *Ann. Rev. Earth Planet Sci.*, 28: 1-18.

Davis, M.B., Brubaker, L.B. and Webb III, T. (1973) Calibration of absolute pollen influx. *In* Birks, H.J.B. and West, R.G. (eds), *Quaternary Plant Ecology.* Blackwell Scientific Publication, Oxford, 9-25.

Davis, M.B., Moeller, R.E. and Ford, J. (1984) Sediment focusing and pollen influx. *In* Haworth, E.Y. and Lund, J.W.G. (eds), *Lake Sediments and Environmental History.* University of Minnesota Press, Minneapolis, 261-293.

Davis, M.B. and Webb III, T. (1975) The contemporary distribution of pollen in eastern North America: a comparison with the vegetation. *Quaternary Research,* 5: 395-434.

Davis, M.B., Woods, S.L., Webb, S.L. and Futyma, R.P. (1986) Dispersal versus climate: Expansion of *Fagus* and *Tsuga* into the Upper Great Lakes region. *Vegetatio*, 93-103.

Davis, O.K. (1995) Climate and vegetation pattern in surface samples from arid western U.S.A.: application to Holocene climatic reconstruction. *Palynology* 19: 95-119.

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

Delcourt, P.A. and Delcourt, H.R. (1987) *Long-Term Forest Dynamics of the Temperate Zone*. Ecological Studies vol. 63, Springer-Verlag, New York, 439 pp.

Delcourt, P.A., Delcourt, H.R. and 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. and Webb III, T. (1984) Atlas of mapped distributions of dominance and Moden percentages for important tree taxa of Eastern North America. *AASP*, Contributions Series: 14.

Faegri, K., and Inversen, J. (1989) *Textbook of pollen analysis*. 4th Edition. Blackburn Press, Caldwell, 328 pp.

Fauquette, S., Guiot, J. and Suc, J.-P. (1998) A method for climatic reconstruction of the Mediterranean Pliocene using pollen data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 144: 183-201.

Fedorova, I.T., Volkova, Y.A. and Varlyguin, E. (1994) *World Vegetation Cover*. Digital Raster Data on a 30-minute Cartesian Orthonormal Geodetic (lat/long) 1080x2160 grid. *In* Global Ecosystems Database Version 2.0. USDOC/NOAA National Geophysical Data Center, Boulder, Co. Three independent single-attribute spatial layers and one tabular data file.

Fredskild, B. (1983) The Holocene vegetational development of Godthabsfjord area, West Greenland. Meddelelser om Gronland, Geoscience, 10: 28 pp.

Fredskild, B. (1973) Studies of the vegetational history of Greenland, Palaeobotanical investigation of some Holocene lake and bog deposits. Meddelelser om Gronland, 198(4): 245 pp.

Funder, S. and Abrahamsen, N. (1988) Palynology in the polar desert, eastern North Greenland. *BOREAS*, 17: 195-208.

Gajewski (2002) Modern pollen assemblages in lake sediments from the Canadian Arctic. *Arctic, Antarctic and Alpine Research*, 34(1): 26-32.

Gajewski, K. (1995) Modern and Holocene pollen assemblages from some small Arctic lakes on Somerset Island, NWT, Canada. *Quaternary Research*, 44: 228-236.

Gajewski, K. (1993) The role of paleoecology in the study of global climatic change. *Review of palaeobotany and palynology* 79:141-151.

Gajewski, K. (1991) Representation pollinique actuelle a la limite des arbres au Nouveau-Quebec. *Can J. Earth Sci.,* 28:643-648.

Gajewski, K. (1988) Late-Holocene climates of eastern North America estimated from pollen data. *Quaternary Research*, 29:255-262.

Gajewski, K. (1987) Climatic impacts on the vegetation of eastern North America during the past 2000 years. *Vegetatio*, 68: 179-190.

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

Gajewski, K., Vance, R., Sawada, M., Fung, I., Gignac, L.D., Halsey, L., John, J., Maisongrande, P., Mandell, P., Mudie, P.J., Richard, P.J.H., Sherin, A.G., Soroko, J., Vitt, D.H. (2000) The climate of North America and adjacent ocean waters ca. 6 Ka. *Can. J. Earth Sci.*, 37: 661-681.

Gajewski, K., Viau, A., Sawada, M.C., Atkinson, D. and Wilson, S. (2001) *Sphagnum* peatland distribution in North America and Eurasia during the past 21,000 years. *Global Biochemical Cycles*. 15: 297-310.

Gleason, H.A. and Cronquist, A. (1963) *Manual of Vascular Plants of Northeastern United States and Adjacent Canada*. D. Van Nostrand Company, New York, pp. 810.

Graumlich, L., J. and Davis, M., B. (1993) Holocene variation in spatial scales of vegetation patterns in the upper Great Lakes. *Ecology*, 74(3): 826-839.

Grimm, E. (2000) *North American Pollen Database*. IGBP PAGES/World Data Center-A for Paleoclimatology, NOAA/NGDC Paleoclimatology Program, Boulder, Co., (http://www.ngdc.noaa.gov/paleo/pollen.html).

Guiot, J. (1990) Methodology of the last climatic cycle reconstruction in France from pollen data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 80: 49-69.

Guiot, J., Harrison, S., Prentice, C.I. (1993) Reconstruction of the Holocene precipitation patterns in Europe using pollen and lake-level data. *Quaternary Research*, 40: 139-149.

Heusser, C.J. (1978). Modern pollen spectra from western Oregon. *Bulletin of the Torrey Botanical Club*, 105:14-17.

Heusser, C.J. (1973) Modern Pollen Spectra from Mount Rainer, Washington. *Northwest Science*, 47(1): 1-8.

Heusser, C.J. (1969) Modern pollen spectra from the Olympic Peninsula, Washington. *Bulletin of the Torrey Botanical Club*, 96:407-417.

Hicks, S. (2001) The use of annual arboreal pollen deposition values for delimiting tree-line in the landscape and exploring models of pollen dispersal, *Review of Palaeobotany and Palynology*, 117: 1-29.

Holdridge, L.R. (1947) Determination of world formations from simple climatic data. *Science*, 105: 367-368.

Huntley, B. and Birks, H. J. B. (1983) *An atlas of past and present maps for Europe: 0-13000 years ago*. Cambridge University Press, Cambridge.

Jacobson, G.L. and Bradshaw, R.H.W. (1981) The selection of sites for paleovegetational studies. *Quaternary Research*, 16: 80-96.

Jackson, S.T., Overpeck, J.T., Webb III, T. Keattch, S.E. and Anderson, K.H. (1997) Mapped plant-micro-fossil and pollen records of Late-Quaternary vegetation change in eastern North America. *Quaternary Science Reviews*, 84:65-374

Jackson, S.T., Webb, R.S., Anderson, K.H., Overpeck, J.T., Webb III, T., Williams, J.W. and Hansen, B.C.S. (2000) Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews*, 19: 489-508.

Jackson, S.T., Webb III, T., Prentice, I.C. and Hansen, J.E. (1995) Exploration and calibration of pollen/vegetation relationships: a PC program for the extended R-value models. *Review of Palaeobotany and Palynology*, 84: 365-374.

Jackson, S.T. and Williams, J.W. (2004) Modern analogs in Quaternary paleocology: here today, gone yesterday, gone tomorrow? *Annual Review of Earth and Planetary Sciences*, 32. (in press).

Janssen, C.R. (1973) Local and regional pollen disposition. *In* Birks, H.J.B. and West, R.G. (eds.) *Quaternary Plant Ecology*, p.31-42, Blackwell, Oxford.

Joussaume, S. and Taylor, K.E. (2000) The Paleoclimate modeling intercomparison Project. In Braconnot, P. (Ed.) *Paleoclimate of the Third PMIP Workshop*. WCRP-111, WMO/TD-Np 1007, ICPO Publication Series No.34: 9-24.

Kerwin, M.W. (2000) *Quantifying and modelling Holocene climate variability based on modern and fossil pollen records from Eastern Canadian and Subarctic*. Ph.D. Dissertation, Dept. of Geological Sciences, University of Colorado, Boulder.

Koppen, W. (1936) Das geographisched system der klimate. *In* Koppen, W. and Geiger, R. (Eds), *Handbuch der Klimatologie I(C)*. Gebruder Borntraeger, Berlin.

Kühl, N., Gehardt, C., Litt, T. and Hense, A. (2002) Probability density functions as botanical-climatologicak transfer functions for climate reconstruction. *Quaternary Research*, 58: 381-392.

Lamb, H.F. (1984) Modern pollen spectra from Labrador and their use in reconstructing Holocene vegetational history. *Journal of Ecology*, 72:37-59.

Lehman, J.T. (1975) Reconstructing the rate of accumulation of lake sediment: the effect of sedimentation focusing. *Quaternary Research*, 5, 541-50.

Leemans, R. and Cramer, W. (1996) The IISA Database for Mean Monthly Values of Temperature, Precipitation and Cloudiness on Global Terrestrial Grid. RR-91-18, IISA, Laxenburg.

Lichti-Federovich, S., and Ritchie, J.C.. (1968) Recent pollen assemblages from the western interior of Canada. *Review of Palaeobotany and Palynology*, 7:297-344.

Lichti-Federovich, S., and Ritchie, J.C. (1965) Contemporary pollen spectra in Central Canada, II. The forest-grassland transition in Manitoba. *Pollen & Spores*, 7: 63-87.

Lindbladh, M., O'Connor, R., and Jacobson, G.L., Jr. (2002) Morphometric analysis of pollen grains for paleoecological studies: classification of Picea from eastern North America. *American Journal of Botany*, 89:1459-1467.

Little, E.L. Jr. (1977) *Atlas of United States trees minor Western hardwoods.* Vol.4, U.S. Department of Agriculture Miscellaneous Publication, v. 1342, 17 pp., 230 maps.

Little, E.L. Jr. (1976) *Atlas of United States trees minor Western hardwoods.* Vol.3, U.S. Department of Agriculture Miscellaneous Publication, v. 1314, 13 pp., 290 maps.

Little, E.L. Jr. (1971) *Atlas of United States trees, conifers and important hardwoods*. Vol. 1, U.S. Department of Agriculture Miscellaneous Publication, v. 1146, 9 pp., 200 maps.

Lowe, J.J. and Walker, M.J.C (1997) Reconstructing Quaternary Environments. Prentice Hall, Pearson Education Limited, Essex, 446 pp.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, J., Yang, L. and 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.

MacDonald, G.M., and Ritchie, J.C. (1986) Modern pollen spectra from the western interior of Canada and the interpretation of late Quaternary vegetation development. *New Phytologist*, 103:245-268.

Maher, Jr., L.J. (1981) Statistics for microfossil concentration measurements employing samples spiked with marker grains. *Review of Palaeobotany and Palynology*, 32, 153-191.

Maher, L.J. (1964) Ephedra pollen in sediments of the Great Lakes region. *Ecology*, 45:391-395.

Maher, L.J. (1963) Pollen analyses of surface materials from the southern San Juan Mountains, CO. *GSA Bulletin*, 74: 1485-1504.

Maugurran, A.E. (1988) *Ecological diversity and its measurement*. Princeton University Press, Princeton, 179 pp.

McAndrews, J.H., Berti, A.A. and Norris, G. (1973) *Key to the Quaternary Pollen ansd Spores of the Great Lake Region.* Life Sci., Misc Publ., Royal Ontario Museum, pp. 61.

McAndrews, J.H., and Wright, H.E., Jr. (1969) Modern pollen rain across the Wyoming basins and the northern Great Plains (U.S.A.). *Review of Palaeobotany and Palynology*, 9:17-43.

McDougall, W.B. (1973) *Seed Plants of Northern Arizona.* The Museum of Northern Arizona, Flagstaff, pp. 594.

Minckley, T.M and Whitlock, C. (2000) Spatial variation of modern pollen in Oregon and southern Washington, USA.. *Review of Palaeobotany and Palynology* 112, 97-123.

New, M., Lister, D., Hulme, M., Makin, I. (2000) A high-resolution data set of surface climate over global land areas. *Climate Research*, 21: 1-25.

Oswald, W.W. (2002) *Holocene vegetational history of the central Arctic Foothills, northern Alaska: pollen representation of tundra and edaphic controls on the response of tundra to climate change*. PhD thesis, University of Washington, pp.152.

Overpeck, J.T., Webb, R.S. and Webb III, T. (1992) Mapping eastern North American vegetation change of the past 18 Ka: No-analogs and the future. *Geology*, 20: 1071-1074.

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

Parent, J. and Richard, P.H. (1990) Morphologie pollinique des Cupressaceae de l'est du Canada et du nord-est des Étas-Unis appliquée à l'Étude des sédiments quaternaires. *Canadian Journal of Botany*, 68(1): 79-89.

Peng, C. Guiot, J. and van Campo, E. (1998) Estimating changes in terrestrial vegetation and carbon storage: using paleoecological data and models. *Quaternary Science Reviews*, 17: 719-736.

Pennington, W. (1986) Lags in adjusment of vegetation to climate caused by the pace of soil development: evidence from Britain. *Vegetatio*, 67: 105-118.

Peterson, G.M. (1978) Pollen spectra from surface sediments of lakes and ponds in Kentucky, Illinois, and Missouri. *American Midland Naturalist*, 100: 333-340.

Peyron, O., Guiot, J., Cheddadi, R., Tarasov, P., Reille, M., Beaulieau, J.-L., Bottema, S. and Andrieu, V. (1998) Climatic reconstruction in Europe for 18,000 yr BP from pollen data. *Quaternary Research*, 49: 183-196.

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*, p. 1-15. The Canadian Institute of Surveying and Mapping, Ottawa.

Potter, L.D. and Rowley, J.S. (1960) Pollen rain and vegetation, San Augustin Plains, NM. *Botanical Gazette*, 122: 1-25.

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

Prentice, I.C. (1986) Vegetation responses to past climatic variation. *Vegetatio*, 67:131-141.

Prentice, I.C., Bartlein, P.J., Webb T., III (1991) Vegetation and climate change in eastern North America since the Last Glacial Maximum. *Ecology*, 72(6): 2038-2056.

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.

Prentice, I.C. and Parson, R.W. (1983) Maximum likelihood linear calibration of pollen spectra in terms of forest composition. *Biometric*, 39: 1051-1057.

Prentice, I.C. and Webb III, T. (1988) BIOME 6000: reconstructing global mid-Holocene vegetation patterns from palaeoeological records. *Journal of Biogeography*, 25: 997-1005.

Richard, P.J.H. (1981) Paléophytogéographie postglaciaire en Ungava par l'analyse pollinique. *Collection Paléo-Québec*, No. 13, 154 pp.

Richard, P.J.H. (1976) Relations entre la vegetation actuelle et le spectre pollinique au Quebec. *Le Naturaliste Canadien*, 103: 53-66.

Ritchie, J.C. (1986) Climate change and vegetation response. *Vegetatio*, 67: 65-74.

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

Ritchie, J.C., Hadden, K.A. and Gajewski, K. (1987) Modern pollen spectra from lakes in arctic western Canada. *Can. J. Bot.*, 65: 1605:1613.

Ruddiman, W.F. (2001) *Earth's Climate, Past and Future*. W.H. Freeman and Company, New York, 465 pp.

Rousseau, D.-D., Duzer, D., Cambon, G., Jolly, D., Poulsen, U., Ferrier, J., Schevin, P. and Gros, R. (2003) Long distance transport of pollen in Greenland. *Geophysical Research Letters*, 30(14): 1765-1768.

Sawada, M.C. (2001) *Late Quaternary Paleoclimates and Biogeography of North America.* PhD. Dissertation, Department of Geography, University of Ottawa, 253 pp.

Sawada, M.C., Gajewski, K., de Vernal, A. and Richard, P. (1999) Comparison of marine and terrestrial Holocene climatic reconstruction from northeastern North America. *The Holocene*, 9(3): 267-277.

Sawada, M., Viau, A.E., Vettoretti, G., Peltier, W.R., and 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*. National Museum of Natural Sciences Publication in Botany, No.7, Vol.1-4.

Short, S.K., Andrews, J.T. and Webber, P.J. (1986) Pollen, vegetation, and climate relationships along the Dalton highway, Alaska, U.S.A.: A basis for Holocene paleoecological and paleoclimatic studies. *Arctic and Alpine Research*, 18: 57-72.

Shuman, B., Bartlein, P., Logar, N. Newby, P. and Webb III, T. (2002) Parallel climate and vegetation responses to the early Holocene collapse of the Laurentide Ice Sheet. *Quaternary Science Reviews*, 21: 1793-1805.

Sugita, S. (1994) Pollen representation of vegetation in Quaternary sediments: theory and method in patchy vegetation. *Journal of Ecology*, 82: 881-897.

Sugita, S. (1993) A model of pollen source area for an entire lake surface. *Quaternary Research*, 39: 239-244.

Thompson, R.S., Anderson, Bartlein, P.J. and Smith, S.A. (2000) *Atlas of Relations Between Climatic Parameters and Distributions of Important Trees and Shrubs in North America – Additional Conifers, Harwoods, and Monocots*. USGS, Professional Paper 1650-C, Denver, Co.

Thompson, R.S., Anderson and Bartlein, P.J. (1998) *Atlas of Relations Between Climatic Parameters and Distributions of Important Trees and Shrubs in North America*. USGS, Professional Paper 1650-A&B, Denver, Co.

ter Braak, C.F.J. (1987) *Unimodal Models to Relate Species to Environment*. Pudoc, Wageningen, pp.152.

Viau, A.E, Gajewski, K., Fines, P., Atkinson, D.E., Sawada, M.C. (2002). Widespread evidence of 1500 yr climate variability in North America during the past 14 000 yr. *Geology*, 30: 455-458.

Webb, R.S., Anderson, K.H. and Webb III, T. (1993) Pollen response-surface estimates of Late-Quaternary changes in the moisture balance of the northeastern United States. *Quaternary Research*, 40: 213-227.

Webb, III, T. (1970) The late- and post-glacial sequence of climatic events in Wisconsin and east-central Minnesota: quantitative estimates derived from fossil pollen spectra by multivariate statistical analysis. PhD Dissertation. University of Wisconsin, Madison, USA.

Webb III, T. (1987) The appearance and disappearance of makor vegetational assemblages: long-term vegetational dynamics in eastern North America. *Vegetatio*, 69:177-187.

Webb III, T. (1986) Is vegetation in equilibrium with climate? How to inerpret late-Qyaternary data. *Vegetatio*, 67:75-91.

Webb III, T. (1974) Corresponding patterns of pollen and vegetation in Lower Michigan: a comparison of quantitative data. *Ecology*, 55(1): 17-28.

Webb III, T., Anderson, K.H., Bartlein, P.J. and 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.

Webb III, T., Bartlein, P.J., Harrison, S.P. and Anderson, K.H. (1993) Vegetation, lake levels and climate climate in the eastern North America during the Holocene. In Wright, H.E., Jr., Kutzbach, J.E., Webb, T. III, Ruddiman, W.F., Street-Perrott, F.A. and Bartlein, J.P. (Eds), *Global Climates since the Last Glacial Maximum*. 415-467. University of Minnesota Press, Minneapolis.

Webb III, T., Bartlein, P.J., Kutzback, J.E. (1987) Climate change in Eastern North America during the past 18 000 years: Comparisons of pollen data with model results. In Ruddiman, W.F. and Wright, H.E. (Eds) *North America and the Adjacent Oceans during the Last Glaciation. The Geology of North America*, vol. K-3, pp.447-462. The geological Society of America.

Webb III, T. and 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. and Clark, D.R. (1977) Calibrating micropaleontological data in climate terms: a critical review. *Ann. N.Y. Acad. Sci*, 288: 93-118.

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

Webb, III, T. (1970) *The late- and post-glacial sequence of climatic events in Wisconsin and east-central Minnesota: quantitative estimates derived from fossil pollen spectra by multivariate statistical analysis*. PhD Dissertation. University of Wisconsin, Madison, USA.

Whitlock, C. and Bartlein, P.J. (1997) Vegetation and climate change in northwest America during the past 125 kyr. *Nature*, 388: 57-61.

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., Post, D.M., Cwynar, L.C. and Lesvesque, A.J. (2002) Rapid and widespread vegetation responses to past climate change in the North Atlantic. *Geology*, 30(11): 971-974.

Williams, J.W., Shuman, B.N. and Webb, T., III (2001) Dissimilarity analysis of Late-Quaternary vegetation and climate in eastern North America. *Ecology*, 82(12): 3346-3362.

Willard, D.A. and Weimer, L.M. (1997) *Palynological census data from surface samples in south Florida*. USGS Open-File Report 97-867, 12 pp.

Woodward, F.I. (1987) *Climate & Plant Distribution*. Cambridge University Press, Cambridge, pp.174.

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

APPENDIX I

Fields in the new modern pollen database, containing site information for each record:

Variable	Data type	Description
ID	[numeric]	Unique key identifier (1-5171)
E	[numeric]	GPD E-number for GPD sites, -9999 for sites from other databases
SITE	[numeric]	Original Site number from GPD, Original MOD5SEQ number for sites from the Brown University Pollen Database, -9999 for sites from other databases
DBCODE	[string]	Original database source and sequence code for each record: The first characters specify the database from which a record originated. The numbers correspond to the unique sequential identifier of a record within the corresponding database origin. (ABD=Anderson and Brubaker 1986, unpublished;BUPD=Brown University Pollen Database (Avizinis and Webb 1985); CPD=Canadian Pollen Database (Gajewski unpubl.); GPD=Global Pollen Database (Grimm 2000); KER=Kerwin 2000; MIN=Minckley 2002; SWUS=Davis 1995)
SITECODE	[string]	Original site code: First characters specify which dataset a record came from. The numbers correspond to the original sites code number of the record prior to duplicate removal to provide continuity. (AB=Anderson and Brubaker 1986, unpublished; BU=Brown University Pollen Dataset; G=Global Pollen Database; K=Kerwin 2000; SW=Davis 1995. Data from the CPD were categorised by subdatasets within the CPD database: FRE=Fredskild 1973; GRE=Greenland dataset; LAC=Lacourse 1998; NQ=Northern Quebec dataset; WC=West Canadian Dataset; C=for all other Canadian Pollen Database samples; BON=Gajewski unpublished)
SITENAME	[string]	Site names from original datasets. Site name from Webb is [Name1]+[Name2] fields
LONG	[numeric]	Longitude in full precision from original databases
LAT	[numeric]	Latitude in full precision from original databases
ELEVATION	[numeric]	Elevation values from original databases. In meters.
POLLSUM	[numeric]	pollen sum as calculated from [Abies]+…+[Xanthium]
DEPENV	[string]	Unique four character identifier for depositional environment for modern sample.
DESCRENV	[string]	Description of depositional environment for the modern entity site (e.g. "Bog", "Lacustrine"...)
DataForm	[string]	Two character code indicating the form of the data: RC=raw counts, RP=raw percentages, DC=digitized counts, DP=digitized percentages
YrOfPubl	[string]	Year of first publication.
AUTHYR	[string]	Identifier for authors and year of publication.
CITATION	[string]	Complete bibliographic citation for publication.
Collector	[string]	Sample collector, when available.
PrincInvest	[string]	Principal Investigator, when available.
SamDATE	[string]	Sampling date, when available.
DepthLoc	[numeric]	Depth of sample at location, when available.
Notes	[string]	Supplemental information (e.g. updates, problems) specific to record.

Fields in the new modern pollen database, for climate and bioclimate data:

CLIMATE + BIOCLIMATE CODE	Data Type	Description
t *[jan]...[dec]*	[numeric]	mean monthly temperature (°C)
p *[jan]...[dec]*	[numeric]	mean monthly precipitation (mm)
s *[jan]...[dec]*	[numeric]	mean monthly percent possible sunshine (NOT clouds)
tmax	[numeric]	absolute (all time over the period of record) maximum temperature (NOT average maximum temperature)
tMin	[numeric]	absolute (all time over the period of record) minimum temperature (NOT average minimum temperature)
gdd0	[numeric]	growing-degree days, 0C base
gdd5	[numeric]	growing-degree days, 5C base
chill	[numeric]	chilling requirement (number of days when pseudo-daily temperature was below 5C)
mtco	[numeric]	mean temperature of the coldest month of the year (NOT absolute minimum temperature)
mtwa	[numeric]	mean temperature of the warmest month of the year (NOT absolute maximum temperature)
annp	[numeric]	mean annual precipitation
pjanpann	[numeric]	January/Annual precipitation ratio
pjulpann	[numeric]	July/Annual precipitation ratio
mipt	[numeric]	Priestley-Taylor (alpha) parameter (AE/PE)
aaetpt	[numeric]	actual evapotranspiration (AE)
apetpt	[numeric]	potential evapotranspiration (PE)
miptev	[numeric]	alpha calculated for the evergreen pft assimilation period (days when the pseudo-daily temperature > -4 C0
aaetptev	[numeric]	AE calculated for the evergreen pft assimilation period
apetptev	[numeric]	PE calculated for the evergreen pft assimilation period
miptev	[numeric]	alpha calculated for the deciduous pft assimilation period (days when the pseudo-daily temperature > +5 C0
aaetptev	[numeric]	AE calculated for the deciduous pft assimilation period
apetptev	[numeric]	PE calculated for the deciduous pft assimilation period

Fields in the new modern pollen database, for AVHRR data:

AVHRR CODE	Description
IGBP21	Modal vegetation type from DISCover dataset V2, 21 x 21 km search window. IGBP classification scheme
BATS21	Modal vegetation type from DISCover dataset V2, 21 x 21 km search window. BATS classification scheme
GLEC21	Modal vegetation type from DISCover dataset V2, 21 x 21 km search window. Global Ecosystems classification scheme
NASLC21	Modal vegetation type from DISCover dataset V2, 21 x 21 km search window. North American Seasonal Land Cover classification scheme
USGS21	Modal vegetation type from DISCover dataset V2, 21 x 21 km search window. USGS Level II classification scheme
SIB2_21	Modal vegetation type from DISCover dataset V2, 21 x 21 km search window. SIB2 classification scheme
RUNN21	Modal vegetation type from DISCover dataset V2, 21 x 21 km search window. Running Life Forms classification scheme
IGBP03	Modal vegetation type from DISCover dataset V2, 3 x 3 km search window. IGBP classification scheme
IGBP11	Modal vegetation type from DISCover dataset V2, 11 x 11 km search window. IGBP classification scheme
IGBP51	Modal vegetation type from DISCover dataset V2, 51 x 51 km search window. IGBP classification scheme
IGBP101	Modal vegetation type from DISCover dataset V2, 101 x 101 km search window. IGBP classification scheme
NLAND	Number of land pixels within search window
NPIX	Number of pixels within search window
IGBP21_1	Fraction land surface within search window covered by Evergreen Needleleaf Forest (IGBP classification scheme)
IGBP21_2	Fraction land surface within search window covered by Evergreen Broadleaf Forest (IGBP classification scheme)
IGBP21_3	Fraction land surface within search window covered by Deciduous Needleleaf Forest (IGBP classification scheme)
IGBP21_4	Fraction land surface within search window covered by Deciduous Broadleaf Forest (IGBP classification scheme)
IGBP21_5	Fraction land surface within search window covered by Mixed Forest (IGBP classification scheme)
IGBP21_6	Fraction land surface within search window covered by Closed Shrublands (IGBP classification scheme)
IGBP21_7	Fraction land surface within search window covered by Open Shrublands (IGBP classification scheme)
IGBP21_8	Fraction land surface within search window covered by Woody Savannas (IGBP classification scheme)
IGBP21_9	Fraction land surface within search window covered by Savannas (IGBP classification scheme)

Fields in the new modern pollen database, for AVHRR data (cont.):

AVHRR CODE	Description
IGBP21_21	Fraction land surface within search window covered by Grasslands (IGBP classification scheme)
IGBP21_11	Fraction land surface within search window covered by Permanent Wetlands (IGBP classification scheme)
IGBP21_12	Fraction land surface within search window covered by Croplands (IGBP classification scheme)
IGBP21_13	Fraction land surface within search window covered by Urban and Built-Up (IGBP classification scheme)
IGBP21_14	Fraction land surface within search window covered by Cropland/Natural Vegetation Mosaic (IGBP classification scheme)
IGBP21_15	Fraction land surface within search window covered by Snow and Ice (IGBP classification scheme)
IGBP21_16	Fraction land surface within search window covered by Barren or Sparsely Vegetated (IGBP classification scheme)
IGBP21_17	Fraction land surface within search window covered by Water Bodies (IGBP classification scheme) Not included in sum so may exceed 1
NeedAvg03	Percent of land surface covered by needleleaf canopy, averaged across all pixels in 3 x 3 km search window
NeedAvg11	Percent of land surface covered by needleleaf canopy, averaged across all pixels in 11 x 11 km search window
NeedAvg51	Percent of land surface covered by needleleaf canopy, averaged across all pixels in 51 x 51 km search window
NeedAvg101	Percent of land surface covered by needleleaf canopy, averaged across all pixels in 101 x 101 km search window
BroadAvg03	Percent of land surface covered by broadleaf canopy, averaged across all pixels in 3 x 3 km search window
BroadAvg11	Percent of land surface covered by broadleaf canopy, averaged across all pixels in 11 x 11 km search window
BroadAvg51	Percent of land surface covered by broadleaf canopy, averaged across all pixels in 51 x 51 km search window
BroadAvg101	Percent of land surface covered by broadleaf canopy, averaged across all pixels in 101 x 101 km search window

APPENDIX II

DUPLICATE REMOVAL PROCEDURE in S-PLUS:

1) Get proportion value for pollen counts for each taxon (columns 13 to 145)

```
modpercent_cbind(mod[,1:12],mod[,13:145]/rowSums(mod[,13:145]))
```

2) Verification procedure: summarize proportion for all taxa by site row. Should get the value of 1 for each site.

```
rowSums(modpercent[,13:145])
```

3) Verification procedure: summarize all rows. (Should get the value of the total number od rows in the data base, before duplicate removal.)

```
sum(rowSums(modpercent[,13:145]))
[1] 8211
```

4) Verification procedure: Determine length of first column. (Value should be equal to the sum of rowSums)

```
length(mod[,1])
[1] 8211
```

5) Verification procedure: Subtract Sum of pollen taxa from original pollen sum [,12]. Discrepancy, in this case is delectable. Difference probably due to the rounding of decimal point.

```
sum(mod[,12]-rowSums(mod[,13:145]))
[1] 5.82645e-013
```

6) Apply elimduplicates2 function to proportioned dataset. Function returns a vector with the row number of sites having geographical and spectral duplicates. 1329 sites had geographical and spectral duplicates

```
dupmod_elimduplicates2(modpercent,13,145,9,10)
dupmod
```

#Because some duplicate sites between database categorized pinus differently (some database distinguished between Pine hap and dip, while other put them all in pine undiff), it was necessary not to include those taxon in the function. The following statement compares spectrum between column 13 and 97, which doesn't include Pinus and Picea types.

```
elm2_elimduplicates2(mod,13,139,9,10)
```

ELIM DUPLICATES 2 F

```
INPUTS
      inModern - name of modern pollen dataframe
      t1 - column with first taxon
      t2 - column with last taxon
      long - column with longitude - should be a projected coordinate such
as from an equidistant since this function uses euclidean distance to
calculate distance between sites
      lat - column with latitude

function(inModern, t1, tn, long, lat)
{
      duplist       <- vector("numeric") # create the vector to hold the duplicates
      data1         <- as.matrix(inModern[, t1:tn]) # create a matrix of just the
                    spectra
      dimnames(data1) <- NULL # get rid of matrix names - speeds things up
      data2         <- data1       # create a second matrix from the first
      data1         <- t(data1) # transpose the first matrix
```

#for each row in data2 e.g., each sample

```
      for(i in 1:length(data2[, 1])) {
```

#this next (if) statement just ensures that if a row has already been tagged as a duplicate of some previous row then skip checking that sample for further duplicates - e.g., don't look for duplicates since they are already tagged.

```
      if(length(which(duplist == i)) == 0) {
```

#This gives the squared chord distances for all sites {data1} vs. the current site being considered {data2[i,]}

```
x <- colSums((data1^0.5 - data2[i,  ]^0.5)^2)
```

gets the row numbers that have 0 SQD with the current test row

```
dups <- which(x == 0)[which(x == 0) != i] #
```

Next, if there are spectral duplicates for this particular test row then

```
                if(length(dups) > 0) {
```

get the euclidean distance between the test row and each of the duplicates to see if they are geographic duplicates as well

```
dists <- apply(inModern[dups, c(long, lat)], 1, +euclidean.distance.f, unlist(inModern[i, c(long, lat)]))
```

get the row numbers of those spectral duplicates that are also geographic duplicates. You could modify this to dists <= SomeNumber if you want to get sites separated by say 1000 meters or less as possible geographic duplicates

```
distZero <- which(dists == 0)
```

now get only the spectral duplicates that are also geographic duplicates and put the row numbers in this next vector

```
dups1 <- dups[distZero]
```

If there are spectral & geographic duplicates then add the row numbers to the duplist before testing the next row

```
if(length(dups1) > 0) {
```

This "duplist" vector contains the rows tagged as both geographic and spectral duplicates

```
duplist <- c(duplist, dups1)   #
```

#Here these two statements write out the duplicates vector in the first and the actual duplicates in the second. You need to modify the column numbers to get what you might want from the second in a textfile

```
    write(paste(c(i, dups1)),"c:\\rdupsIdenticalLongLatSpectra.txt",
     ncolumns = length(dups1) + 1, append = T)
    write(paste(i, inModern[i, 4], inModern[i, 5], inModern[i, 6],
     inModern[i, 7], inModern[i, 10], inModern[dups1, 4],
     inModern[dups1, 5], inModern[dups1, 6], inModern[dups1, 10]),
     "c:\\rdupssiteinfo.txt", ncolumns = length(dups1) * 10, append =
     T)
}
}
```

```
}
}
```

here only the vector of duplicate row numbers is output from the function

```
	duplist
}
```

7) Return matrix which sites having spectral and geographical duplicates have been removed.

```
	mod2_mod[-c(elm2),]
```

8) Apply elimduplicates3 function to subsequent dataset. Function tags sequentially spectral duplicates in a second vector.

```
	mod3_elimduplicates3(mod2,13,139,9,10)
```

Sawada

a

```
elim.duplicates.3a.f<-function(inModern, t1, tn, long, lat)
{
# I have modified this function to tag sequentially (e.g., 1, 2, 3
etc..) spectral duplicates in a second vector
# called idlist.  This way you can determine the original spectra and
its duplicates, e.g., all idlist
# This function only looks at spectral duplicates only.
#Inputs: Same as for elim.duplicates.2.f
#
#
	duplist <- vector("numeric")
	idlist <- vector("numeric")
	id <- 0
	data1 <- as.matrix(inModern[, t1:tn])
	dimnames(data1) <- NULL
	data2 <- data1
	data1 <- t(data1)
	for(i in 1:length(data2[, 1])) {
		if(length(which(duplist == i)) == 0) {
	x <- colSums((data1^0.5 - data2[i,  ]^0.5)^2)
	dups <- which(x == 0)[which(x == 0) != i]
	if(length(dups) > 0) {
	write(paste(c(i,dups)),"c:\\rdupsIdenticalLongLatSpectra.txt",ncolum
ns = length(dups) + 1,append = T)
	write(paste(i, inModern[i, 4],inModern[i, 5], inModern[i, 6],
inModern[i, 7], inModern[i, 10],inModern[dups, 4], inModern[dups, 5],
```

```
inModern[dups, 6],inModern[dups,10]),"c:\\rdupssiteinfo.txt", ncolumns
=length(dups) * 10, append = T)

      duplist <- c(duplist, c(i, dups))
#
      id <- id + 1
      idlist <- c(idlist, rep(id, length(dups)+ 1))
}
}
}
      cbind(duplist, idlist)
}
```

9) Extract spectral duplicates from database with geographical and spectral duplicates removed.

```
      #create numeric vector of duplist column
            mod33_as.numeric(mod3[,1])

      #Extract rows from mod2 which are listed in mod33
            mod4_(mod2[c(mod33),])
```

10) After having manually examined (in excel) the spectral duplicates, and having chosen the sites to be kept into the database according to the predefined criteria,

```
      # create numeric vector of duplist column from sites to remove
            dup.remove_as.numeric(duptoremove[,1])

      #Extract all rows from newmod2 which are listed in dup.remove
      LPCMOD2003_(mod2[-c(dup.remove),])
```

APPENDIX III

CLIMATE SPACE PLOT

Modified CRU climate data (Bartlein, P. pers. com.) (annual precipitation and mean temperature) were assigned to modern pollen sites. A biome class was assigned to grid points of CRU climate data (new et al 2000) and modern pollen sites that were within the boundaries of the biome, in ArcView. A new table containing the sites and the newly assigned climate and biome information could the be plotted in S-PLUS. A convex hull delimiting the climate space of the modern pollen sites and CRU estimates was generated in S-plus (Venables and Ripley 1994).

The biome map is a simplification of Fedorova et al.'s (1994) World Vegetation Map based on remote sensing information, where new classes were created by merging some of the biome categories form the original map.

Within ArcView 3.2, biomes that needed to be merged were selected and converted into separate layers. This procedure was repeated until there were individual layers of each generalized biome. The boundaries within each separate layer were then dissolved (using the geoprocessing extension in ArcView 3.2). The new generalized biome layers were then merged to constitute a new map.

- *S-plus functions to returning a convex hull that delimits the climate space of the modern pollen sites and Leemans and Cramer grid estimates.*

```
x_mydata[seq(0,length(mydata[,1]),2),]

h_chull(modernpollen$TMEAN,  modernpollen$TANNP)#creates  convex  hull
   for climate variables from modern pollen sites

plot(modernpollen$TMEAN, modernpollen$TANNP)# Plot climate variables
  for modern pollen sites.

polygon(modernpollen$TMEAN[h],modernpollen$TANNP[h],dens=0)      #draw
  convex hull for climate variable of modern pollen data on plot
```

- *In S-plus interface, the data points symbols were varied as a function of the biome class.*