Review of Palaeobotany and Palynology 162 (2010) 203–212

Contents lists available at ScienceDirect

Review of Palaeobotany and Palynology

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

Are pollen-based climate models improved by combining surface samples from soil and lacustrine substrates?

Simon Goring [a,*], Terri Lacourse [b], Marlow G. Pellatt [c,d], Ian R. Walker [e], Rolf W. Mathewes [a]

[a] Department of Biological Sciences, Simon Fraser University, Burnaby, Canada V5A 1 S6
[b] Department of Geography, University of Victoria, Victoria, British Columbia, Canada V8W 3R4
[c] Parks Canada, Western and Northern Service Centre, 300-300 West Georgia Street, Vancouver, British Columbia, Canada V6B 6B4
[d] School of Resource and Environmental Management, Simon Fraser University, Burnaby, Canada V5A 1 S6
[e] Biology, and Earth & Environmental Sciences, University of British Columbia Okanagan, Kelowna, BC, Canada V1V 1 V7

ARTICLE INFO

Article history:
Received 11 December 2009
Received in revised form 29 June 2010
Accepted 30 June 2010
Available online 7 July 2010

Keywords:
pollen databases
depositional environment
lacustrine
terrestrial
palynology
climate models
weighted averaging
randomForest
modern analogue technique
partial least squares
non-metric-multidimensional-scaling

ABSTRACT

Differences between pollen assemblages obtained from lacustrine and terrestrial surface sediments may affect the ability to obtain reliable pollen-based climate reconstructions. We test the effect of combining modern pollen samples from multiple depositional environments on various pollen-based climate reconstruction methods using modern pollen samples from British Columbia, Canada and adjacent Washington, Montana, Idaho and Oregon states. This dataset includes samples from a number of depositional environments including soil and lacustrine sediments.

Combining lacustrine and terrestrial (soil) samples increases root mean squared error of prediction (RMSEP) for reconstructions of summer growing degree days when weighted-averaging partial-least-squares (WAPLS), weighted-averaging (WA) and the non-metric-multidimensional-scaling/generalized-additive-models (NMDS/GAM) are used but reduces RMSEP for randomForest, the modern analogue technique (MAT) and the Mixed method, although a slight increase occurs for MAT at the highest sample size. Summer precipitation reconstructions using MAT, randomForest and NMDS/GAM suffer from increased RMSEP when both lacustrine and terrestrial samples are used, but WA, WAPLS and the Mixed method show declines in RMSEP.

These results indicate that researchers interested in using pollen databases to reconstruct climate variables need to consider the depositional environments of samples within the analytical dataset since pooled datasets can increase model error for some climate variables. However, since the effects of the pooled datasets will vary between climate variables and between pollen-based climate reconstruction methods we do not reject the use of mixed samples altogether. We finish by proposing steps to test whether significant reductions in model error can be obtained by splitting or combining samples from multiple substrates.

1. Introduction

Large surface-sample databases allow researchers to relate modern pollen distribution to regional and continental-scale climates (North America: Whitmore et al., 2005; Africa: Gajewski et al., 2002; Europe: Dormoy et al., 2009; China: Members of China Quaternary Pollen Data Base, 2001). The size of the dataset used in an analysis plays an important role in model quality. For chironomids it has been shown that more precise and accurate reconstructions of past environments from sedimentary archives become possible as dataset breadth across a climate gradient increases (Walker et al., 1997). Chironomid datasets are obtained only from lacustrine-type environments and so may only be extended using other lacustrine samples.

In contrast to chironomid datasets, pollen datasets can extend coverage of a region or climate gradient by including modern pollen assemblages obtained from several depositional environments. These depositional environments may include peat, soil or lacustrine sediments. If a climate reconstruction uses pollen assemblages obtained from lacustrine sediments, it may be possible to extend coverage along a climate or vegetation gradient by including pollen obtained from both lacustrine and soil samples. This may be of particular interest in dry regions such as grasslands where lacustrine depositional environments may be absent or limited. The literature makes it clear that the effects of mixed depositional environments have been of concern for some time, and most large databases contain information about depositional environment (Davis and Webb, 1975). Although extending the dataset in this way may seem desirable, the effect of mixing samples from multiple depositional environments on model error for pollen–climate reconstructions has yet to be examined in detail.

* Corresponding author. Tel.: +1 778 782 4458.
E-mail address: sgoring@sfu.ca (S. Goring).

0034-6667/$ – see front matter
doi:10.1016/j.revpalbo.2010.06.014

Pollen-based climate reconstructions with relatively broad coverage of individual climate gradients are subject to error resulting from sample selection (Xu et al., 2010) and may be subject to error resulting specifically from the use of modern pollen samples from multiple depositional environments. It is well known that there are differences in source area (Jacobson and Bradshaw, 1981; Davis, 2000) and preservation (Sangster and Dale, 1961; Havinga, 1964; Havinga, 1984) between depositional environments, and it is clear by the inclusion of depositional data in early pollen databases that this has been of concern for some time. Researchers have already shown significant differences in pollen representation between depositional environments in a number of regions (lacustrine and terrestrial: Zhao et al., 2009; lacustrine, soil and moss: Wilmshurst and McGlone, 2005; lacustrine and peat: Webb et al., 1978). Thin-walled pollen taxa may be under-represented in soil samples as a result of oxidation (Havinga, 1964), while prolific pollen producers (e.g.: *Pinus*) may be more abundant in lacustrine samples because of their over-representation in the regional component of the pollen rain (Davis, 1984). These factors will affect the pollen–vegetation relationship for any site, and will also affect the way in which the pollen assemblage represents the vegetation–climate relationship of the site.

The North American Modern Pollen Database (NAMPD: Whitmore et al., 2005) defines six depositional environments (that may have further sub-categories), with varying numbers of pollen assemblages in each group (Fig. 1). The use of samples from multiple depositional environments varies among researchers. Gonzales et al. (2009) propose removing all terrestrial samples from the NAMPD to improve pollen–climate relationships in eastern North America but do not explicitly test the effects of sample site removal on pollen-based climate reconstructions. Among pollen-based climate reconstructions, some use pooled datasets where samples from all depositional environments are used (Viau et al., 2006; Dormoy et al., 2009) while others use only lacustrine samples (Seppä et al., 2004; Goring et al., 2009). Datasets using all samples must rely on the assumption that differences in pollen–climate representation between depositional environments are not sufficient to reduce model performance within large datasets. Datasets using pollen samples from one depositional environment are likely to ascribe to the assumption that complex relationships between depositional environment, pollen and climate representation may increase model error regardless of dataset size.

Since pollen-based climate reconstructions are often carried out on samples obtained from lacustrine sediments we are interested in testing the effect of adding pollen assemblages from terrestrial (soil) samples to a dataset composed of lacustrine samples. Increasing the dataset size by adding terrestrial samples to a lacustrine dataset may improve model performance by reducing the variability in the pollen–climate relationship. Predictions from pollen-based climate models may also be improved when the additional samples extend coverage of a climatic or spatial gradient, however, the dataset we present does not allow us to test this facet of mixed depositional environments in detail. In this study we compare the effect of adding terrestrial samples to the effect of adding an equivalent number of lacustrine samples. In this way we can examine how several pollen-based climate models perform with respect to mixed depositional environments. We also present insight into the various pollen-based climate models based on our results.

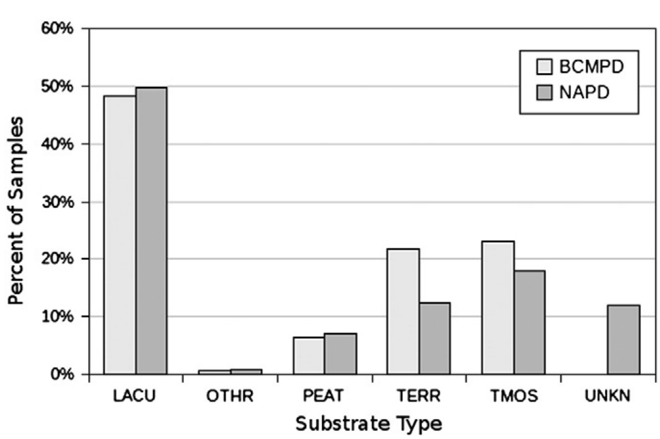

Fig. 1. Depositional environments for pollen assemblages in the expanded British Columbia Modern Database (BCMPD, $n=990$) and the North American Pollen Database (NAMPD, $n=4815$). Depositional environments include lacustrine (LACU), other (OTHR), peat (PEAT), terrestrial/soil (TERR), moss polsters (TMOS) and unknown (UNKN).

2. Regional setting

This study focuses on western North America, with samples spread from the southern border of Oregon to northern British Columbia (BC) and from the Pacific coast to Montana (Fig. 2). The region has an area of approximately 1 450 000 km^2 and spans a broad range of climate and vegetation zones. Boreal forest in north-eastern BC is typified by long, very cold winters and short growing seasons. Temperate rain forest is found along the Pacific coast, interspersed with dry Garry oak savanna stretching southward from southern Vancouver Island, through the Puget Trough in Washington State to Oregon. A second belt of interior rain forest can be found further inland in BC along the western slope of the Rocky Mountains. Each of the mountain ranges in the region has alpine tundra environments at high elevations. A belt of interior desert spreads from south central BC and continues to the Great Basin in southern Oregon.

Typically, summer temperatures increase from north to south, and precipitation and winter temperatures decrease from west to east. Sharp vertical gradients for precipitation and temperature also exist among the many mountain ranges that occur in western North America.

3. Methods

3.1. Pollen samples and climate data

The British Columbia Modern Pollen Database (BCMPD: Goring et al., 2009) contains 284 surface pollen samples (145 lacustrine samples and 139 terrestrial samples). To this dataset we have added 331 lacustrine and 77 terrestrial samples obtained from the NAMPD and from Gavin et al. (2005) to produce a larger dataset (Figs. 1 and 2). Because of the complexity of the problem, we use two extremes from the continuum of natural depositional environments, lacustrine and terrestrial substrates. Although samples obtained from moss polsters and peat are included in both the BCMPD and the NAMPD, we exclude them from analysis because they have characteristics that are intermediate between terrestrial and lacustrine samples. To generalize, pollen assemblages from a moss polster or from peat have lower rates of abiotic decomposition than terrestrial samples because of potential saturation, have a greater regional component than terrestrial samples because of larger canopy openings, and have less secondary reworking than lacustrine samples once pollen is deposited because of a lack of sediment focusing and a shallower reworking depth. It is possible that under certain conditions one or all of these assumptions may be violated but for the purpose of this analysis we will allow them to guide the selection of depositional environments.

Downscaled climate data for each site was obtained using the ANUSPLIN model (McKenney et al., 2007) at a 10×10 km scale. Preliminary results from spatially weighted principal components analysis (sPCA: Dray and Dufour, 2007) using only lacustrine samples showed that summer growing degree days (sGDD) was most heavily weighted on the first sPCA axis. This indicates that, when the spatial distribution of sites is taken into account, sGDD is responsible for the largest proportion of variance in the climate data set so this variable was chosen to be the basis of our modelling.