Quaternary International 311 (2013) 3–11

Contents lists available at SciVerse ScienceDirect

Quaternary International

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

# Pollen–climate transfer functions intended for temperate eastern Asia

Ruilin Wen [a,*], Jule Xiao [a], Yuzhen Ma [b], Zhaodong Feng [c], Yuecong Li [d], Qinghai Xu [d]

[a] Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China
[b] MOE Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Beijing 100875, China
[c] Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China
[d] College of Resources and Environment, Hebei Normal University, Shijiazhuang 050016, China

ARTICLE INFO

*Article history:*
Available online 29 April 2013

ABSTRACT

Pollen data of 646 surface samples from northern China and Mongolia and climatic data from the relevant meteorological stations were collected in this study to develop more reliable pollen–climate transfer functions for temperate eastern Asia. Canonical correspondence analysis (CCA) was used to examine the pollen–climate relationships, and mean summer precipitation (MSP) and mean January temperature (MJaT) were inferred to be the first and second important factors controlling the spatial distribution of the surface pollen in the study area. The original dataset was screened with CCA for MSP and MJaT separately to detect anomalous samples that show the extreme values. The first screened dataset was established after excluding those anomalous samples, and the initial transfer function was generated using the weighted averaging partial least squares (WAPLS) method. The jackknife test was then applied to the WAPLS model for determining the optimum transfer function and for detecting large-residual samples, and the final transfer function was generated after removing the large-residual samples from the first screened dataset. The final dataset for MSP and MJaT consists of 428 and 419 samples, respectively. The root mean square errors of prediction for both WAPLS models are 34 mm and 2.7 °C, and the coefficients of determination are 0.85 and 0.73. This study suggests that the values of climatic parameters could be better estimated and the reliability of pollen–climate transfer functions would be significantly improved through removing anomalous and large-residual samples from the original dataset with mathematical methods.



## 1. Introduction

Since Imbrie and Kipp (1971) introduced the algorithm to reconstruct environmental variables on biological proxies, pollen–climate transfer functions have been widely attempted for quantitatively reconstructing climatic parameters of the geological past (Bartlein et al., 1984; Guiot, 1987; ter Braak and Juggins, 1993; Birks, 1995). Pollen–climate transfer functions are normally developed based on the relationships between the surface pollen and the modern climate of sampling sites along certain climatic gradients. However, the developed transfer functions are often less reliable because many factors might have complicated or even distorted the expected relationship between the pollen and the climate. Among these factors, two are outstanding in most cases: (1) locally-resulted deviation of plant growth from climate equilibrium-resulted plant growth and (2) inadequate bioclimatic coverage of sampling sites. The local vegetation inferred from pollen assemblages of some surface samples may not closely related to the regionalization of natural vegetation under modern climatic conditions due to possible human disturbance or/and the effect of local topography. Such samples must be recognized and removed while the pollen–climate transfer functions are developed to improve the reliability of the transfer functions. In addition, the directional length or/and the areal coverage of ecological and climatic gradients along which the surface samples were obtained is also important. For example, a small length or/and a small coverage of bioclimatic gradients of surface pollen sampling is absolutely inadequate in developing pollen–climate transfer functions if the transfer functions are used to reconstruct paleoclimate of the past during which large shifts in climatic and associated vegetation zones occurred. A large coverage of surface pollen samples is therefore required to enhance the adaptability of the transfer functions.

Temperate eastern Asia is currently under the combined influence of the mid-latitude westerlies and the low-latitude monsoons (Chinese Academy of Sciences, 1984; Zhang and Lin, 1985). It covers

* Corresponding author.
*E-mail address:* rlwen@mail.iggcas.ac.cn (R. Wen).

1040-6182/$ – see front matter © 2013 Elsevier Ltd and INQUA. All rights reserved.
http://dx.doi.org/10.1016/j.quaint.2013.04.025

a large gradient of climatic conditions and associated vegetation landscapes and is thus an ideal region for developing pollen–climate transfer functions. In this study, we collected both pollen data of 646 surface samples from northern China and Mongolia and climatic data from the relevant meteorological stations. With the aid of statistical methods of canonical correspondence analysis and weighted averaging partial least squares, we attempt to develop reliable pollen–climate transfer functions for the quantitative reconstruction of paleoclimatic conditions in temperate eastern Asia.

## 2. Study area

The 646 surface samples used in this study are distributed in northern China and Mongolia, and display a latitudinal coverage of ~2000 km (33.6°–51.5°N) and a longitudinal coverage of ~3000 km (88.3°–128.1°E) (Fig. 1). The elevation of the sampling sites ranges from 228 to 3600 m above sea level. A total of 323 meteorological stations were collected for modern climatic data from the sampled region and the surrounding area (Fig. 1).

In the sampled area, the modern natural vegetation spans from deciduous broadleaved forests in the warm temperate zone to mixed conifer-broadleaved forests, steppes and deserts in the temperate zone, and conifer forests in the cold temperate zone (Fig. 1). Mean annual precipitation varies from 0 to 900 mm with a summer (June, July and August) average ranging from 0 to 500 mm and a winter (December, January and February) average from 0 to 40 mm. Precipitation decreases northwestwards in northern China and then increases northwards in Mongolia. Mean annual temperature ranges from −8 to 13 °C with a July average from 7 to 26 °C and a January average from −32 to −1 °C. Temperature declines with increasing latitudes and altitudes of the sampling sites.

## 3. Data and methods

### 3.1. Pollen data

Pollen data used in the present study are from 646 surface samples collected in northern China and Mongolia in the years 1998–2009 (Fig. 1). The sites of the surface samples were selected from localities more than 500 m far away from road. To ensure the even representation of surface samples in northern China, 3–5 subsamples of moss or surface litter leaves or soil were randomly collected from a plot of 10 m × 10 m and then mixed into one sample. In Mongolia, each sample was a composite of 10–20 subsamples of surface soil obtained randomly from an area of 20 $m^2$. Some samples from Mongolia contained a small amount of moss and litter material. The latitude, longitude and elevation of the sampling sites were determined with a receiver of Global Positioning System. Pollen was extracted from the surface samples following the HCl–NaOH–HF-sieving-acetolysis procedure described by Fægri et al. (1989) and more than 300 pollen grains were identified and counted for each sample. Aquatic pollen in the samples was not considered in this study because the growth of aquatic plants is closely related to local environments rather than the regional climate.

In order to eliminate undesirable effects of pollen types with occasional occurrences in the samples and/or with divergent identifications among different researchers, 25 major pollen taxa were selected from all the surface samples as representative pollen types in the present study (Table 1). These pollen types occur frequently and have higher percentages in different samples. They also show typical features in morphology and unambiguous implications in ecology. These pollen types include coniferous trees of *Pinus*, *Picea*, *Abies* and *Larix*, broadleaved trees of *Betula*, *Alnus*, *Quercus* and *Ulmus*, shrubs of *Corylus*, *Ostryopsis*, *Ephedra*, *Nitraria*, Rosaceae, Elaeagnaceae and Tamaricaceae, and herbs of *Artemisia*, Chenopodiaceae, Poaceae, Asteraceae, Cyperaceae, *Thalictrum*, Caryophyllaceae, Asclepiadaceae, Labiatea and Liliaceae. The pollen percentages were based on the sum of the 25 pollen taxa in a sample.

**Table 1**
Statistics of 25 representative pollen types in 646 surface samples from northern China and Mongolia. The percentages of pollen types are based on the sum of the total terrestrial pollen in a sample.

| Pollen type | Occurrence | Maximum (%) | Mean (%) | Standard deviation (%) |
|---|---|---|---|---|
| *Pinus* | 622 | 96.56 | 16.07 | 21.52 |
| *Picea* | 222 | 78.62 | 2.47 | 9.72 |
| *Abies* | 93 | 54.92 | 0.60 | 4.14 |
| *Larix* | 173 | 26.42 | 0.63 | 2.44 |
| *Betula* | 547 | 90.12 | 6.90 | 14.01 |
| *Alnus* | 194 | 4.67 | 0.19 | 0.48 |
| *Quercus* | 338 | 74.05 | 3.12 | 9.38 |
| *Ulmus* | 258 | 19.83 | 0.34 | 1.17 |
| *Corylus* | 134 | 20.74 | 0.21 | 1.08 |
| *Ostryopsis* | 245 | 63.71 | 0.86 | 4.16 |
| *Ephedra* | 417 | 63.16 | 1.25 | 4.79 |
| *Nitraria* | 312 | 39.34 | 0.90 | 3.16 |
| Rosaceae | 427 | 30.79 | 1.14 | 2.67 |
| Elaeagnaceae | 243 | 88.70 | 0.82 | 4.68 |
| Tamaricaceae | 58 | 37.66 | 0.30 | 2.16 |
| *Artemisia* | 623 | 94.42 | 28.55 | 21.94 |
| Chenopodiaceae | 621 | 94.11 | 16.87 | 20.61 |
| Poaceae | 609 | 41.06 | 4.67 | 5.89 |
| Asteraceae | 569 | 27.29 | 2.35 | 3.80 |
| Cyperaceae | 442 | 78.08 | 3.06 | 8.68 |
| *Thalictrum* | 201 | 12.31 | 0.26 | 1.01 |
| Caryophyllaceae | 229 | 53.21 | 0.68 | 2.68 |
| Asclepiadaceae | 102 | 8.72 | 0.20 | 0.79 |
| Labiatae | 305 | 17.76 | 0.53 | 1.36 |
| Liliaceae | 259 | 27.61 | 0.40 | 1.43 |

### 3.2. Climate data

Climate data used in this study include mean summer (June, July and August) precipitation (MSP), mean winter (December, January and February) precipitation (MWP), mean annual precipitation (MAP), mean July temperature (MJuT), mean January temperature (MJaT), and mean annual temperature (MAT). These data are based on the 30-year averages of instrumental observations of the years 1961–1990 at 323 meteorological stations (data source: http://cdc.cma.gov.cn/) (Fig. 1). For each sampling site, the six climate parameters were yielded from observed data of the surrounding stations within a radius scope of three latitudinal or longitudinal degrees by inverse distance squared interpolation using the program Polation version 1.0 (T. Nakagawa, unpublished software). Altitudinal temperature corrections were conducted according to the universal temperature lapse rate of 0.6 °C per 100 m.

The accuracy of the interpolation method was tested with the leave-one-out method. The correlation coefficients between estimated and observed values of the six climate parameters for the 323 meteorological stations all are higher than 0.95. The estimated values of MSP, MWP, MAP, MJuT, MJaT and MAT for the 646 sampling sites range from 28 to 485 mm, 1 to 38 mm, 45 to 828 mm, 7.0 to 26.0 °C, −31.7 to −0.1 °C and −7.3 to 12.6 °C, respectively.

### 3.3. Canonical correspondence analysis

The collinearity among climatic variables was assessed through examining the variance inflation factors (VIFs). One variable would be assumed to be almost perfectly correlated with the other variables in the dataset if it has a VIF value higher than 20 (ter Braak, 1988). Our results show that MAP, MJuT, MJaT and MAT have VIF

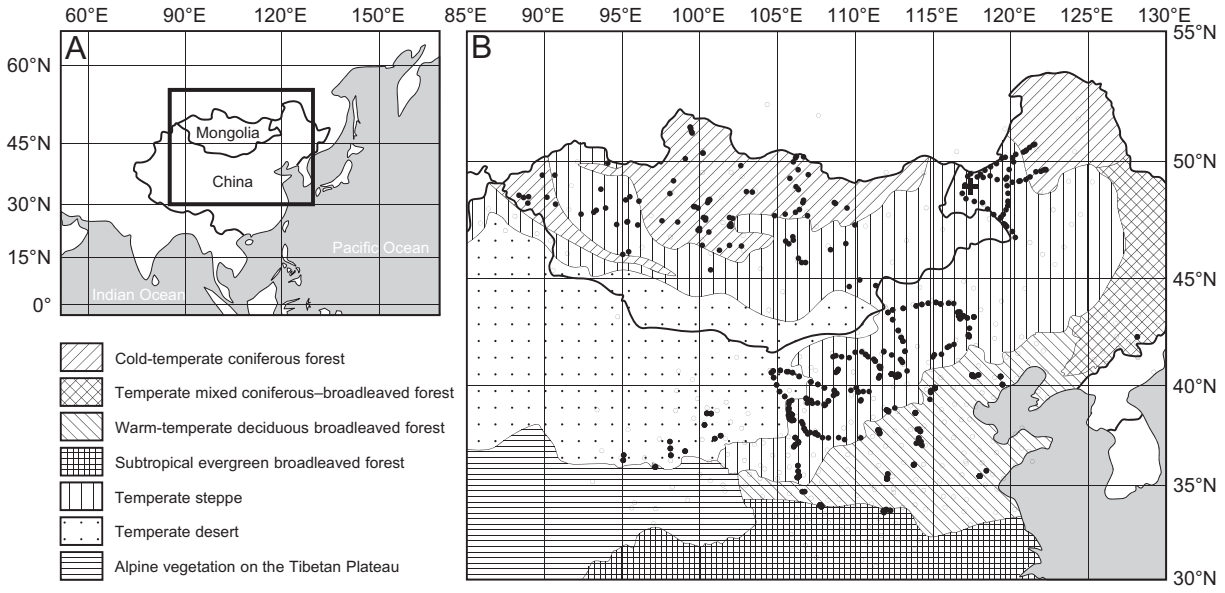


**Fig. 1.** (A) Map of eastern Asia showing the study area. (B) Map of northern China and Mongolia showing vegetation regionalization (Chinese Academy of Sciences, 2007; Hilbig, 1995) and the locations of 646 surface samples (dots) as well as 323 meteorological stations (circles) used in this study. Note that some of the sampling sites are too close to be distinguished from each other on the scale of the map. The location of Hulun Lake (cross) discussed in the text is shown.

values higher than 20, implying a significant collinearity among these variables. In view of the fact that the growth of plants is more closely associated with seasonal variations in precipitation and temperature, MAP and MAT that represent average climatic conditions within a year were not considered here. In this way, the remaining four variables (MSP, MWP, MJuT and MJaT) all have VIF values lower than 20 and should be thought to have independent impacts on the surface pollen data.

To relate the surface pollen data to climatic variables, canonical ordination techniques were performed using the program CANOCO version 4.5 (ter Braak, 1988; ter Braak and Smilauer, 2002). In CANOCO, a square-root transformation was applied to the pollen data in addition to the default settings. Detrended correspondence analysis showed that the gradient length of the first axis is 2.6, indicating a unimodal response of the pollen assemblages to climatic variables (ter Braak and Prentice, 1988; Birks, 1995). Canonical correspondence analysis (CCA) was therefore chosen for examining the relationship between pollen and climate (Jongman et al., 1995). The CCA results of the relationship between 25 representative pollen types in the surface samples and 4 climatic variables (MSP, MWP, MJuT and MJaT) are shown in Fig. 2 and Table 2.

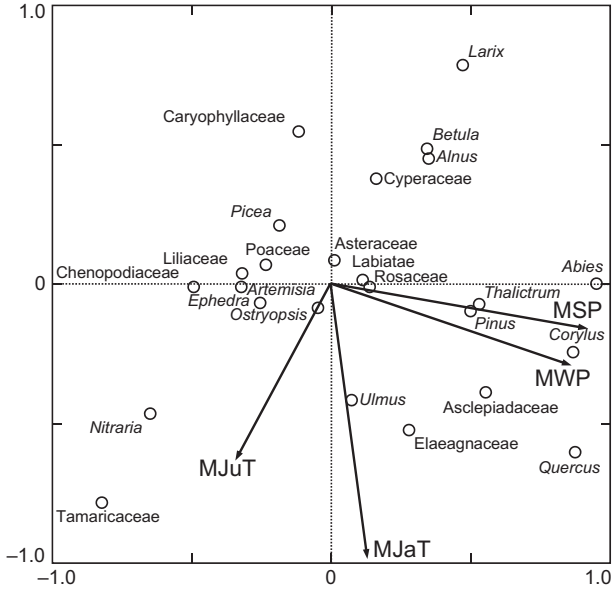


**Fig. 2.** Canonical correspondence analysis (CCA) of the relationship between 25 representative pollen types in the surface samples from northern China as well as Mongolia and 4 climatic variables (MSP, mean summer precipitation; MWP, mean winter precipitation; MJuT, mean July temperature; MJaT, mean January temperature).

**Table 2**
Results of the first four axes of canonical correspondence analysis of 25 representative pollen types in the surface samples and 4 climatic variables (MSP, MWP, MJuT and MJaT).

| | Axis 1 | Axis 2 | Axis 3 | Axis 4 | Total inertia |
|---|---|---|---|---|---|
| Eigenvalues | 0.165 | 0.073 | 0.038 | 0.013 | |
| Species–environment correlations | 0.850 | 0.798 | 0.604 | 0.446 | |
| Cumulative percentage variance | | | | | |
| Of species data | 11.4 | 16.4 | 19.1 | 20.0 | |
| Of species–environment relation | 57.1 | 82.3 | 95.5 | 100.0 | |
| Sum of all eigenvalues | | | | | 1.450 |
| Sum of all canonical eigenvalues | | | | | 0.290 |
| Inter set correlations of environmental variables with axes | | | | | |
| Mean summer precipitation (MSP) | 0.78 | −0.13 | −0.11 | −0.14 | |
| Mean January temperature (MJaT) | 0.11 | −0.78 | 0.09 | −0.02 | |
| Mean July temperature (MJuT) | −0.29 | −0.50 | −0.42 | −0.02 | |
| Mean winter precipitation (MWP) | 0.73 | −0.23 | −0.08 | 0.18 | |

In order to estimate the explanatory power of each of the 4 climatic variables individually with respect to the pollen data, a series of CCA analyses were carried out using each variable as a sole constraining variable. The variance in the pollen data explained by MSP, MWP, MJuT and MJaT shows percentages of 9.9%, 9.0%, 4.6% and 5.1%, respectively; and the statistical significance of each variable is at the

99% level ($p < 0.01$) as assessed with Monte Carlo permutation tests involving 499 unrestricted permutations. Then CCA with forward selection was conducted to obtain the minimal subset of climatic variables that best explained the pollen data and to identify the order of importance of each variable; and the statistical significance of each added variable was assessed with a Monte Carlo permutation test with a Bonferroni-type adjustment for significant levels (see Lotter et al., 1997; Finsinger et al., 2007). The results of the forward selection procedure indicate that MSP, MJaT, MJuT and MWP ($\lambda = 0.14$, 0.08, 0.04 and 0.03, respectively) are all statistically significant in relation to the variance in the pollen data, allowing quantitative inference models to be established for them.

The relationship between the pollen assemblages of some sampling sites and one specific climatic variable might be skewed due to possible disturbances by human activities and/or local topographic settings. On one hand, the occurrence of synanthropic plants (e.g., *Artemisia*, Chenopodiaceae) nearby some sampling sites could be attributed at least partly to human activities rather than climatic changes. Thus the pollen data from these sites should be used with caution when quantitative inference models are developed. On the other hand, plants growing in some geomorphic units might respond to some climate parameters more sensitively but to the others less sensitively. For example, the growth of plants in the intermountain basins would be more closely associated with temperature rather than precipitation because of relatively high moisture in the mountain areas resulted from lowered temperature. It is therefore necessary to establish a dataset separately for each single climatic variable by screening all the surface samples and selecting those samples that are sensitive to the climatic variable.

Although MAP and MAT are collinear with the other four climatic variables representing seasonal climatic conditions, MSP, MWP, MJaT and MJuT, we also screened the datasets and then established the relevant transfer functions for both variables in the present study. This attempt is motivated by the considerations that (1) MAP and MAT are two proxies widely used in paleoclimatic and paleoenvironmental reconstructions and (2) the statistical relation between one single climatic variable and the pollen data would not be affected by the other variables even though these "other variables" are more or less correlated with each other. Consequently the relationship between the pollen assemblages of 646 surface samples and each climatic variable (including MSP, MJaT, MJuT, MWP, MAP and MAT) was analyzed separately with CCA to detect anomalous samples that show extreme values in the specific climatic variable (Jongman et al., 1995). The extreme values of anomalous samples, in our opinion, might result from the occurrence of synanthropic and/or locally growing plants near the sampling sites. These samples must be excluded from the original dataset of 646 samples to improve the estimate of pollen–climate relationships. After the anomalous samples were detected with CCA and excluded from the original dataset, the first screened datasets for each climatic variable were then established.

### 3.4. Weighted averaging partial least squares

The method of weighted averaging partial least squares (WAPLS; ter Braak and Juggins, 1993) was used in this study to develop the initial pollen–climate transfer function for each one of the climatic variables according to its first screened dataset. Based on the concept of niche-space partitioning and ecological optima of species, the WAPLS model contains additional components that utilize the residual structure in the species data so as to improve the species optima in the final weighted averaging predictor (ter Braak and Juggins, 1993; Birks, 1998). This method performs well with noisy and species-rich compositional data that cover a large ecological gradient (ter Braak and Juggins, 1993; Jongman et al., 1995; Birks, 1995, 1998) and thus is applicable to the development of pollen–climate transfer functions in the present study.

With respect to each climatic variable, the jackknife test was applied to the WAPLS model using the first $x$ ($x = 1$–6) components to determine the optimum transfer function and to detect the samples with large residuals. The WAPLS model that gave the lowest root mean square error of prediction (RMSEP) and the highest coefficient of determination ($R^2$) was selected. Samples with large residuals would be analogous to the anomalous samples detected with CCA and are excluded from the pollen dataset in development of transfer functions.

In the present study, samples that show an absolute residual (difference between predicted and observed values) larger than an equal value of the RMSEP were identified as the outlying samples. A new transfer function was generated for each climatic variable after the outlying samples were removed from the dataset. In order to assess the improvement of the quantitative inference models through the two-step screening (i.e., CCA and WAPLS), statistics of the RMSEP and $R^2$ derived from the original and final datasets for each climatic variable were performed (Table 3). The improved pollen–climate transfer function was finally developed for each climatic variable based on the second screened dataset (Table 4).

**Table 3A**
Jackknifed root mean square errors of prediction (RMSEP) and coefficients of determination ($R^2$) of the weighted averaging partial least squares models using the first $x$ ($x = 1$–6) components to reconstruct MSP and MJaT based on the original dataset and the final datasets.

| WAPLS model | MSP | | MJaT | |
|---|---|---|---|---|
| | RMSEP | $R^2$ | RMSEP | $R^2$ |
| Component 1[a] | 66.7 | 0.60 | 5.03 | 0.38 |
| Component 2[a] | 63.8 | 0.63 | 4.89 | 0.42 |
| Component 3[a] | 63.0[c] | 0.64[c] | 4.84[c] | 0.43[c] |
| Component 4[a] | 63.1 | 0.64 | 4.85 | 0.43 |
| Component 5[a] | 63.4 | 0.64 | 4.88 | 0.42 |
| Component 6[a] | 63.7 | 0.64 | 4.90 | 0.42 |
| Component 1[b] | 42.5 | 0.76 | 3.18 | 0.61 |
| Component 2[b] | 35.7 | 0.83 | 2.91 | 0.67 |
| Component 3[b] | 34.0 | 0.85 | 2.76 | 0.71 |
| Component 4[b] | 33.8 | 0.85 | 2.69 | 0.72 |
| Component 5[b] | 33.8 | 0.85 | 2.67 | 0.73 |
| Component 6[b] | 33.8[c] | 0.85[c] | 2.66[c] | 0.73[c] |

[a] Original dataset, $n = 646$ for both MSP and MJaT.
[b] Screened datasets, $n = 428$ for MSP, and $n = 419$ for MJaT.
[c] Models with the lowest RMSEP and highest $R^2$ used for this study.

**Table 3B**
Jackknifed root mean square errors of prediction (RMSEP) and coefficients of determination ($R^2$) of the weighted averaging partial least squares models using the first $x$ ($x = 1$–6) components to reconstruct MAP, MWP, MAT and MJuT based on the original dataset and the final datasets.

| WAPLS model | MAP | | MWP | | MAT | | MJuT | |
|---|---|---|---|---|---|---|---|---|
| | RMSEP | $R^2$ | RMSEP | $R^2$ | RMSEP | $R^2$ | RMSEP | $R^2$ |
| Component 1[a] | 109.3 | 0.67 | 4.9 | 0.60 | 3.29 | 0.39 | 3.00 | 0.33 |
| Component 2[a] | 106.0 | 0.69 | 4.9[c] | 0.61[c] | 3.21[c] | 0.42[c] | 2.93[c] | 0.36[c] |
| Component 3[a] | 105.2[c] | 0.69[c] | 4.9 | 0.61 | 3.22 | 0.42 | 2.94 | 0.36 |
| Component 4[a] | 105.5 | 0.69 | 4.9 | 0.61 | 3.23 | 0.42 | 2.96 | 0.35 |
| Component 5[a] | 106.3 | 0.69 | 4.9 | 0.60 | 3.26 | 0.41 | 2.99 | 0.34 |
| Component 6[a] | 106.9 | 0.69 | 5.0 | 0.60 | 3.28 | 0.41 | 3.02 | 0.33 |
| Component 1[b] | 63.3 | 0.83 | 1.6 | 0.85 | 1.96 | 0.63 | 1.59 | 0.63 |
| Component 2[b] | 55.6 | 0.87 | 1.6 | 0.86 | 1.80 | 0.69 | 1.49 | 0.67 |
| Component 3[b] | 54.1 | 0.88 | 1.6 | 0.86 | 1.71 | 0.72 | 1.46 | 0.68 |
| Component 4[b] | 53.9 | 0.88 | 1.6 | 0.86 | 1.68 | 0.73 | 1.46[c] | 0.69[c] |
| Component 5[b] | 53.9[c] | 0.88[c] | 1.6[c] | 0.86[c] | 1.67 | 0.73 | 1.46 | 0.68 |
| Component 6[b] | 54.1 | 0.88 | 1.6 | 0.86 | 1.66[c] | 0.73[c] | 1.46 | 0.69 |

[a] Original dataset, $n = 646$ for all the four parameters.
[b] Screened datasets, $n = 448$, 467, 430 and 447 for MAP, MWP, MAT and MJuT, respectively.
[c] Models with the lowest RMSEP and highest $R^2$ used for this study.

**Table 4A**
Coefficients (optima) of 25 representative pollen types for the final prediction models of MSP and MJaT. The predicted value is equal to the sum of the products of coefficient and percentage of each pollen type.

| Pollen type | MSP | MJaT |
|---|---|---|
| *Pinus* | 399.2 | −9.81 |
| *Picea* | 158.0 | −15.99 |
| *Abies* | 349.8 | −10.79 |
| *Larix* | 127.1 | −29.76 |
| *Betula* | 327.4 | −30.25 |
| *Alnus* | 550.0 | −50.11 |
| *Quercus* | 346.7 | 2.94 |
| *Ulmus* | 160.0 | 20.78 |
| *Corylus* | 1329.4 | −8.88 |
| *Ostryopsis* | 373.2 | −21.27 |
| *Ephedra* | −72.8 | −2.91 |
| *Nitraria* | −229.2 | 26.18 |
| Rosaceae | 297.2 | −16.49 |
| Elaeagnaceae | 336.5 | −4.54 |
| Tamaricaceae | 7.1 | 16.77 |
| *Artemisia* | 195.8 | −11.72 |
| Chenopodiaceae | 115.2 | −19.31 |
| Poaceae | 42.3 | −20.61 |
| Asteraceae | 315.1 | −25.00 |
| Cyperaceae | 255.1 | −20.77 |
| *Thalictrum* | 1184.5 | 2.40 |
| Caryophyllaceae | −207.8 | −97.73 |
| Asclepiadaceae | 1165.6 | −20.13 |
| Labiatae | 359.9 | −20.02 |
| Liliaceae | −56.9 | −22.52 |

**Table 4B**
Coefficients (optima) of 25 representative pollen types for the final prediction models of MAP, MWP, MAT and MJuT. The predicted value is equal to the sum of the products of coefficient and percentage of each pollen type.

| Pollen type | MAP | MWP | MAT | MJuT |
|---|---|---|---|---|
| *Pinus* | 671.7 | 18.4 | 5.29 | 20.49 |
| *Picea* | 243.9 | 3.9 | −0.47 | 10.71 |
| *Abies* | 1221.1 | 30.4 | −4.75 | 4.76 |
| *Larix* | −221.3 | 1.8 | −23.18 | 4.22 |
| *Betula* | 480.4 | 14.6 | −6.30 | 14.94 |
| *Alnus* | 555.5 | 14.1 | −16.34 | 24.68 |
| *Quercus* | 839.3 | 32.6 | 9.51 | 18.67 |
| *Ulmus* | 304.7 | 3.8 | 29.56 | 28.86 |
| *Corylus* | 1343.8 | −1.4 | −6.05 | 22.12 |
| *Ostryopsis* | 491.2 | 9.4 | −0.03 | 15.16 |
| *Ephedra* | −77.3 | 0.3 | 4.99 | 11.15 |
| *Nitraria* | −55.2 | −0.9 | 24.70 | 31.42 |
| Rosaceae | 512.8 | 12.3 | −2.22 | 13.41 |
| Elaeagnaceae | 625.4 | 12.3 | 10.22 | 19.58 |
| Tamaricaceae | 177.9 | 3.5 | 28.62 | 30.42 |
| *Artemisia* | 305.9 | 6.5 | 5.53 | 20.87 |
| Chenopodiaceae | 151.1 | 3.5 | 1.17 | 21.37 |
| Poaceae | 47.1 | 3.2 | −1.83 | 16.53 |
| Asteraceae | 505.0 | 8.2 | −7.22 | 16.02 |
| Cyperaceae | 380.6 | 4.6 | −3.41 | 8.47 |
| *Thalictrum* | 1408.8 | 15.3 | 16.31 | 11.64 |
| Caryophyllaceae | −375.5 | −7.7 | −46.10 | −1.45 |
| Asclepiadaceae | 1411.0 | 14.7 | 9.08 | 10.15 |
| Labiatae | 647.5 | 7.3 | −2.00 | 13.29 |
| Liliaceae | −136.5 | 0.7 | 18.42 | 20.31 |

## 4. Results and discussion

### 4.1. Explanatory power of climatic variables for pollen data

The pattern of variations in the surface pollen in response to changes in modern climatic variables was illustrated by the CCA results (Fig. 2, Table 2). The first two CCA axes account for 16.4% of the variance of species data. The first axis shows an eigenvalue of 0.165 and the second axis an eigenvalue of 0.073. Moreover, MSP and MWP are obviously correlated with the first axis (correlation coefficient: 0.78, 0.73, respectively), implying that precipitation is the dominant factor controlling the spatial distribution of surface pollen in the study area. Whereas MJaT and MJuT exhibit significant correlations with the second axis (correlation coefficient: −0.78 and −0.50, respectively), suggesting that temperature is the second important factor. As shown in Fig. 2, the CCA results also reflect the climatic optima of different pollen types. Pollen types on the right (such as *Abies*, *Quercus*, *Corylus*, Asclepiadaceae, *Thalictrum* and *Pinus*) prefer humid climate while pollen types on the left (such as Tamaricaceae, *Nitraria*, Chenopodiaceae, *Ephedra* and Liliaceae) can tolerate arid climate. Pollen types on the lower part (such as Tamaricaceae, *Quercus*, Elaeagnaceae, *Nitraria*, *Ulmus* and Asclepiadaceae) prefer warm climate while pollen types on the upper part (such as *Larix*, Caryophyllaceae, *Betula*, *Alnus* and Cyperaceae) can tolerate cold climate.

The CCA with forward selection and Monte Carlo permutation tests indicate that MSP and MJaT are the two most important climatic variables with respect to the variance in the surface pollen data and capture 48% and 28% of the total variance, respectively. Percentage diagrams of 25 representative pollen types in 646 surface samples from northern China and Mongolia plotted against MSP and MJaT are shown in Fig. 3. As expected, MSP exerts a great influence on the spatial distribution of surface pollen in the study area because precipitation in summer (MSP) is usually the limiting factor for the growth of most plants in northern China and Mongolia. But, it is beyond our expectation that January temperature (MJaT) can explain the variance in the pollen data better than July temperature (MJuT). We speculate that extremely low winter temperature might have considerably severer damages or disturbances than extremely high summer temperature on the plant community. The inference about the dominant control of precipitation over the distribution of the surface pollen is in agreement with the result of analyses of the surface pollen samples collected from arid northern and western China (Luo et al., 2010). However, in the Japanese Islands, the variance of surface pollen assemblages was mainly attributed to spatial changes in temperature (Nakagawa et al., 2002). Such a difference might have resulted from the difference of climatic conditions between Japanese Islands and inland Asia. In Japanese Islands, precipitation is always abundant and has never been a limiting factor from the north to the south for the growth of the plants; whereas temperature often dictates the growth of the plants simply because it varies greatly along both latitudinal and altitudinal gradients. On the contrary, in inland Asia where precipitation varies greatly from the north to the south and from the east to the west, it is understandable that precipitation becomes the most important limiting factor for the growth of the plants.

### 4.2. Climatic optima and tolerances for pollen taxa

Based on the final datasets through two-step screenings, the optima and tolerances of 25 representative pollen types with respect to MSP and MJaT were estimated with the weighted averaging method (Fig. 4). It was suggested that the estimates of climatic optimum and tolerance of a taxon would be perfect if the taxon could display a unimodal distribution along the climatic gradient (c.f., ter Braak and Juggins, 1993). In our datasets, all the pollen taxa assume unimodal variations in response to changes in both MSP and MJaT except for *Pinus* and *Betula*. *Pinus* pollen shows high percentages almost along the entire gradients of both MSP and MJaT, whereas *Betula* pollen exhibits a bimodal distribution along the MJaT gradient (Fig. 3). The non-unimodal distributions of *Pinus* and *Betula* along the climatic gradient would be attributed partly to the fact that both genera comprise several species differing in the preference and tolerance for moisture and/or temperature conditions and partly to the fact that the both genera have an over-

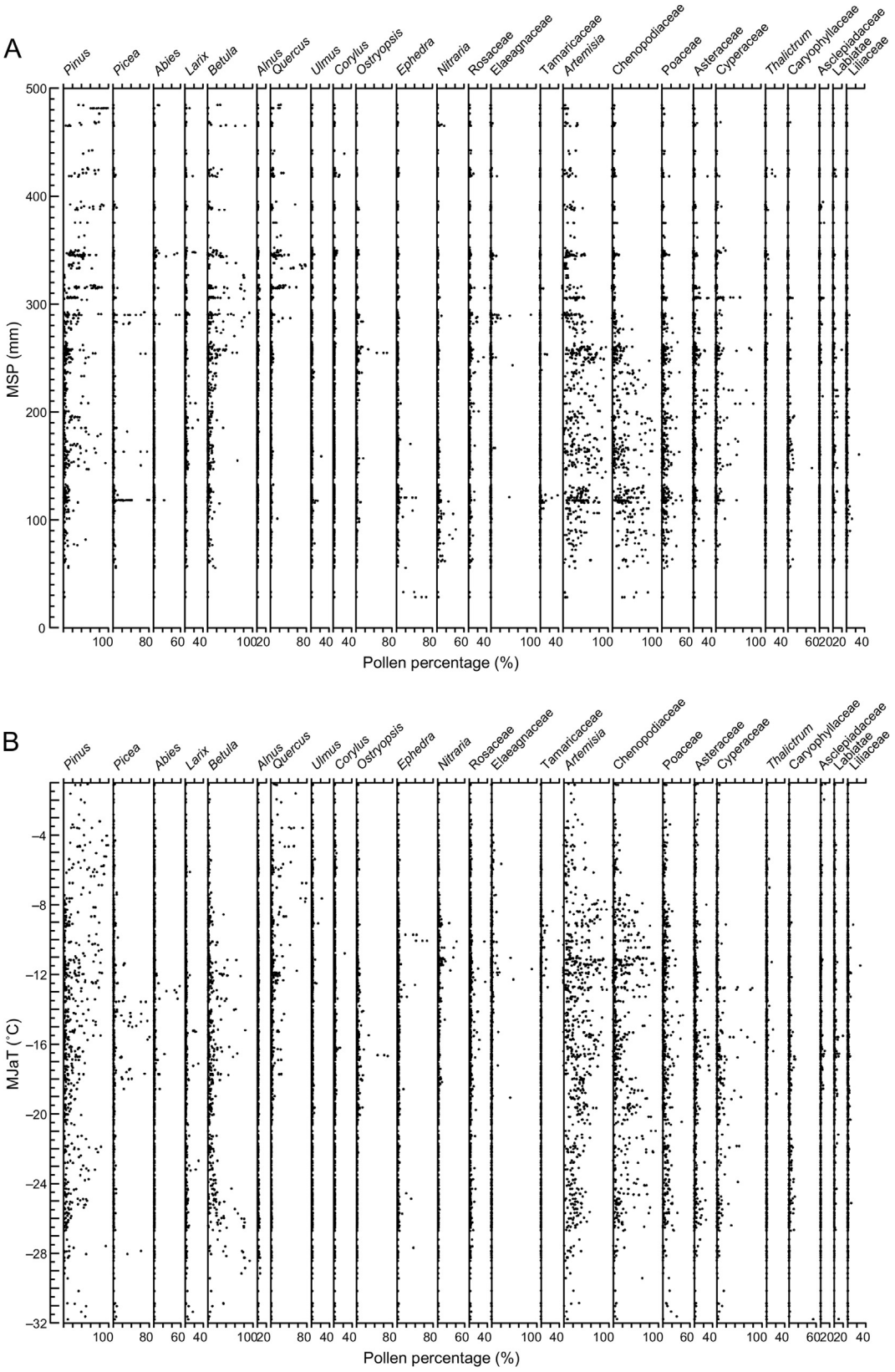


**Fig. 3.** (A) Percentage diagram of 25 representative pollen types in 646 surface samples from northern China and Mongolia, plotted against mean summer precipitation (MSP, mm). (B) Percentage diagram of 25 representative pollen types in 646 surface samples from northern China and Mongolia, plotted against mean January temperature (MJaT, °C). The percentages of pollen types are based on the sum of 25 representative pollen types in a sample.

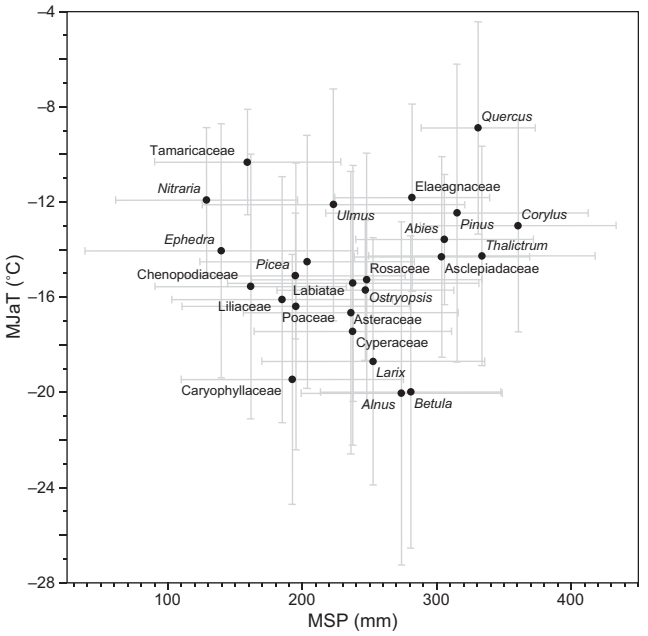


**Fig. 4.** Optima and tolerances of 25 representative pollen types in 646 surface samples from northern China and Mongolia with respect to mean summer precipitation (MSP, mm) and mean January temperature (MJaT, °C) based on the weighted averaging method.

representation in pollen spectrum because of the high production and long-distance transport of pollen grains. Nevertheless, *Pinus* and *Betula* were still adopted in the present study considering the fact that both genera dominate different types of the forest vegetation and display high pollen percentages in most of the surface samples.

As shown in Fig. 4, *Nitraria*, *Ephedra*, Tamaricaceae and Chenopodiaceae, which dominate the desert vegetations of western China, adapted to dry summer and warm winter and display MSP optima between 125 and 165 mm and MJaT optima between −16 °C and −10 °C. *Corylus*, *Thalictrum*, *Quercus*, *Pinus*, *Abies* and Asclepiadaceae that are mainly distributed in the alluvial plains of eastern China prefer humid summer and warm winter and exhibit MSP optima higher than 300 mm and MJaT optima between −15 °C and −8 °C. *Alnus*, *Betula*, Caryophyllaceae, *Larix* and Cyperaceae that grow in the high-latitude mountains are resistant to cold winters and display MJaT optima between −20 °C and −17 °C. Plants of Liliaceae, *Picea*, Poaceae, *Artemisia*, *Ulmus*, Asteraceae, Labiatea, *Ostryopsis*, Rosaceae and Elaeagnaceae occur in the conditions of intermediate humidity and moderate temperature and have MSP optima between 180 and 285 mm and MJaT optima between −17 °C and −11 °C.

### 4.3. Development of transfer functions

The climatic optima of 25 representative taxa based on the weighted averaging method are in agreement with the climatic information reflected by the pattern of the geographical and ecological distributions of these plants. In order to define accurate coefficients for each species in the quantitative inference models, however, these data of the climatic optima require further optimizations because of the differences in predictive power. In the present study, the optima of each species (i.e., species coefficients) were repeatedly improved by extracting further components from the residual structure in the pollen data with the WAPLS method. This procedure would effectively enhance the statistical performance of the inference models, despite the fact that the discrepancy between the calculated optima and the actual climatic preferences of the taxa may occur. Based on the final screened datasets for each climatic variable, the WAPLS models with the lowest RMSEP and the highest $R^2$ were selected (Table 3). The formula of each transfer function was expressed as the sum of the products of the coefficient and percentage of each pollen type (Table 4).

For each transfer function, the RMSEP decreased and the $R^2$ increased after two-step screening of the original datasets, indicating that the predictive power of each quantitative inference model based on the final dataset for each climatic variable was significantly improved (Table 3). In addition, the WAPLS models derived from the original and the final datasets differ in the distribution of the residuals on the gradient of each climatic variable (Fig. 5). The WAPLS models derived from the original datasets tend to overestimate the climatic optima at the low-value end of the gradient but to underestimate the optima at the high-value end. In other words, there exists a negative correlation between the residuals and the interpolated values for each climatic variable based on the original datasets. These phenomena can also be seen in the transfer functions previously established with the WAPLS method (c.f., Lu et al., 2006; Shen et al., 2006; Xu et al., 2010). Birks (1998) pointed out that all the regression methods based on weighted averaging would inevitably suffer from the problems of 'edge effects'. Differences in the distribution of the residuals based on the original datasets, in our view, might have resulted from the so-called 'edge effects'. The WAPLS models derived from the final datasets through the two-step screenings of the original datasets would be less affected by the 'edge effects' because the resultant residuals varied regularly within a limited range throughout the gradient of each climatic variable. These results suggest that it is necessary to screen the raw pollen data properly in order to establish robust pollen–climate transfer functions.

As mentioned in Section 4.2, *Pinus* pollen is strongly overrepresented in the modern pollen rain because of the pollen's high production as well as long-distance transport (Andersen, 1970). Previous studies suggested that *Pinus* pollen can account for over 30% in the percentage of surface pollen samples collected from localities where no pine trees grow (Li and Yao, 1990; Wang et al., 1996; Li et al., 2005; Ma et al., 2008). Such an overrepresentation implies that the presence of *Pinus* pollen may probably have an unfavorable effect on the pollen–climate transfer functions. In order to examine the possible interference of *Pinus* in the transfer functions, we arbitrarily excluded *Pinus* pollen from the datasets and established the pollen–climate transfer functions without *Pinus* with the same methods and procedures described above. The CCA with forward selection and Monte Carlo permutation tests indicate that MSP and MJaT remain being the two most important seasonal climatic variables and capture 46% and 29% of the total variance, respectively. The final WAPLS models for MSP and MJaT after two-step screening of the original datasets show the RMSEP of 36 mm and 2.6 °C, respectively, and the $R^2$ of 0.82 and 0.73, respectively. These data suggest that the pollen–climate transfer functions did not perform better when *Pinus* pollen was excluded. Judging from these data together with the significantly high percentages of *Pinus* pollen in the surface samples, it is not necessary to exclude *Pinus* from the original datasets in generating the pollen–climate transfer functions.

### 4.4. Application of the pollen–climate transfer functions

The pollen–climate transfer functions based on the original and the final datasets were applied to the pollen profile of a sediment core recovered at Hulun Lake in northeastern Inner Mongolia,

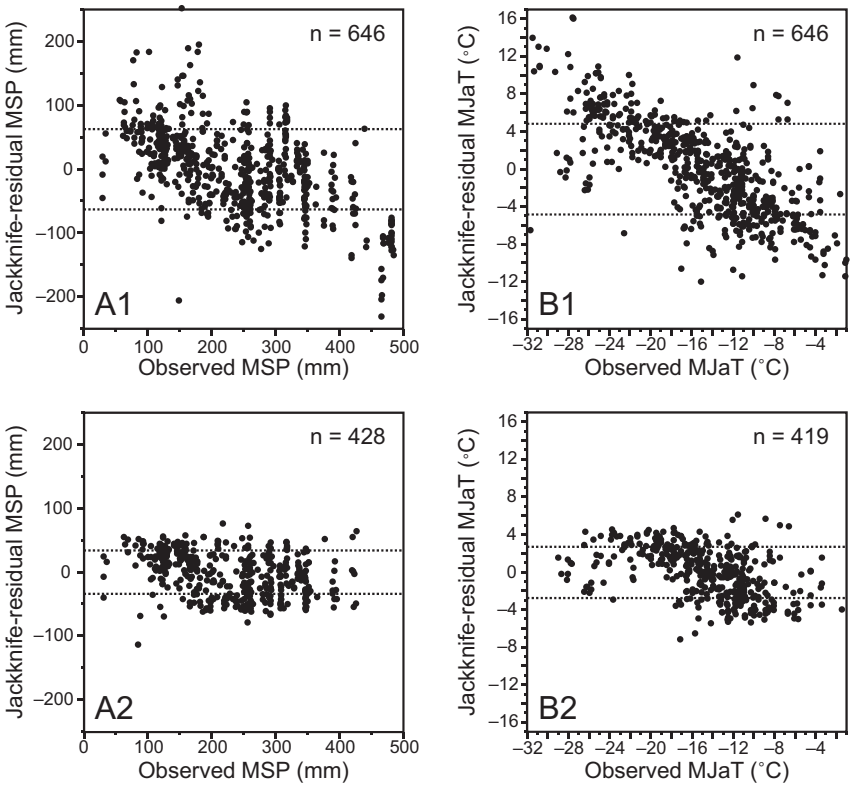


**Fig. 5.** (A1) Error estimates for mean summer precipitation (MSP, mm) based on the original dataset. (A2) Error estimates for mean summer precipitation (MSP, mm) based on the final dataset. (B1) Error estimates for mean January temperature (MJaT, °C) based on the original dataset. (B2) Error estimates for mean January temperature (MJaT, °C) based on the final dataset. The dashed lines denote residuals equal to the jackknifed root mean square errors of prediction (RMSEP). The number of the surface samples included in each dataset is shown.

China (Wen et al., 2010) in order to quantitatively reconstruct the Holocene history of changes in precipitation and temperature and to evaluate the validity of the transfer functions established in this study. As shown in Fig. 6, both MSP and MJaT reconstructed on the final datasets show a trend of variations similar to that on the original datasets. During the early Holocene between 11,000 and 8000 cal. BP, MSP reached moderate values and MJaT was the highest of the entire Holocene. MSP increased gradually and MJaT decreased dramatically during the period between 8000 and 4400 cal. BP. The interval from 4400 to 3350 cal. BP was marked by drastic declines in MSP and by slight decreases in MJaT. MSP recovered from 3350 to 1000 cal. BP and MJaT rose from 3350 to 2050 cal. BP. MSP dropped from 1000 to 500 cal. BP and MJaT dropped from 2050 to 500 cal. BP. both. During the last 500 years, both MSP and MJaT exhibited gradual increases. However, there exist obvious differences between the reconstructions on the final and the original datasets in both MSP and MJaT reconstructions. The sample-specific errors of prediction for MSP and MJaT based on

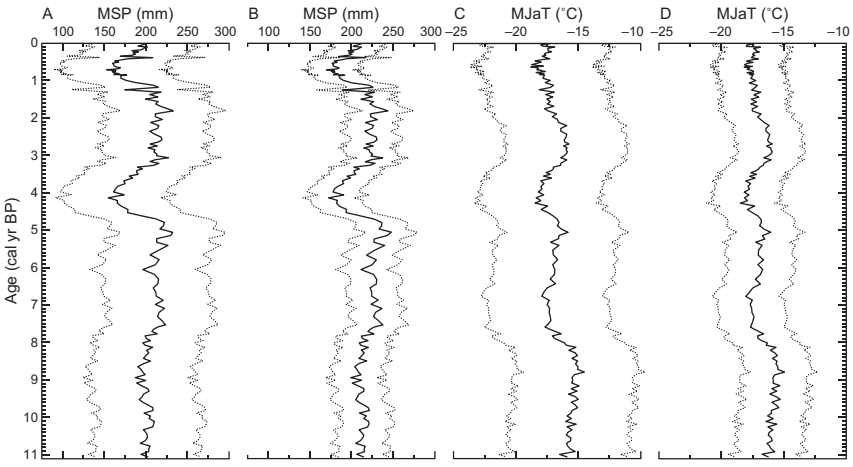


**Fig. 6.** Paleoclimates reconstructed with the weighted averaging partial least squares method on the pollen profile of the HL06 sediment core at Hulun Lake spanning the last 11,000 cal yr (Wen et al., 2010). (A) Mean summer precipitation (MSP, mm) based on the original dataset. (B) Mean summer precipitation (MSP, mm) based on the final dataset. (C) Mean January temperature (MJaT, °C) based on the original dataset. (D) Mean January temperature (MJaT, °C) based on the final dataset. Solid lines represent the predicted values and dashed lines show the corresponding error estimates. The chronology was derived from calibrated ages of reservoir-effect-free radiocarbon dates (Wen et al., 2010).

the final datasets are smaller than those on the original datasets. The MSP values based on the final dataset are 5–10 mm higher than those on the original dataset for the entire Holocene. The MJaT values based on the final dataset are 0.5 °C lower than those based on the original dataset for the early to mid-Holocene.

Both reconstructions, one based on the original dataset for the samples from the core top and another based on the final datasets for the same samples, were compared with the observed data from a meteorological observatory near Hulun Lake in order to evaluate the performance of the transfer functions. In the Hulun Lake region, mean annual precipitation is 285 mm with a summer average of 203 mm and a winter average of 6 mm, and mean annual temperature is 0.3 °C with a July average of 20.3 °C and a January average of −21.2 °C (Xu et al., 1989). The reconstructed values of all the climatic variables based on the final datasets (MSP, 201 mm; MWP, 7 mm; MAP, 298 mm; MJaT, −17.4 °C; MJuT, 19.7 °C; MAT, 1.6 °C) are closer to the observed data than those based on the original datasets (MSP, 200 mm; MWP, 8 mm; MAP, 307 mm; MJaT, −17.3 °C; MJuT, 19.0 °C; MAT, 2.1 °C), suggesting that the transfer functions derived from the final datasets have better performances in the quantitative reconstruction of paleoclimates.

## 5. Conclusions

This study suggests that precipitation is the dominant factor controlling the spatial distribution of the surface pollen in northern China and Mongolia. MSP, MJaT, MJuT and MWP representing the seasonality of precipitation and temperature are important climatic parameters with respect to the variance in the surface pollen data, but MSP and MJaT are the most important parameters. The surface samples that might have suffered from human disturbances or/and from topographic distortion can be detected and then excluded through two-step screenings of the original dataset by the methods of CCA and WAPLS to improve the pollen–climate transfer functions. After exclusion of the human-disturbed and topography-distorted components, screened datasets were formed and more reliable transfer functions were developed for MSP, MJaT, MJuT, MWP, MAP and MAT.

## Acknowledgements

We thank an anonymous reviewer for valuable comments and suggestions that helped improve the early version of the manuscript. This study was financially supported by the National Natural Science Foundation of China (Grants 41130101 and 41002054) and the Ministry of Science and Technology of China (Grant 2010CB833402).

## References

Andersen, S.T., 1970. The relative pollen productivity and pollen representation of North European trees, and correction factors for tree pollen spectra: determined by surface pollen analyses from forests. C.A. Reitzel, København, p. 99.

Bartlein, P.J., Webb III, T., Fleri, E., 1984. Holocene climatic change in the northern Midwest: pollen-derived estimates. Quaternary Research 22, 361–374.

Birks, H.J.B., 1995. Quantitative palaeoenvironmental reconstructions. In: Maddy, D., Brews, J.S. (Eds.), Statistical Modeling of Quaternary Science Data. Quaternary Research Association, Technical Guide, vol. 5, pp. 161–254.

Birks, H.J.B., 1998. Numerical tools in palaeolimnology – progress, potentialities, and problems. Journal of Paleolimnology 20, 307–332.

Chinese Academy of Sciences (Compilatory Commission of Physical Geography of China), 1984. Physical Geography of China: Climate. Science Press, Beijing, p. 161 (in Chinese).

Chinese Academy of Sciences (Compilatory Commission of the Vegetation Map of China), 2007. The Vegetation Map of the People's Republic of China (1:1 000 000). Geological Publishing House, Beijing, p. 1270 (in Chinese).

Fægri, K., Kaland, P.E., Krzywinski, K., 1989. Textbook of Pollen Analysis, fourth ed. John Wiley & Sons, Chichester, p. 328.

Finsinger, W., Heiri, O., Valsecchi, V., Tinner, W., Lotter, A.F., 2007. Modern pollen assemblages as climate indicators in south Europe. Global Ecology and Biogeography 16, 567–582.

Guiot, J., 1987. Late Quaternary climatic change in France estimated from multivariate pollen time series. Quaternary Research 28, 100–118.

Hilbig, W., 1995. The Vegetation of Mongolia. SPB Academic Publishing, Amsterdam, p. 258.

Imbrie, J., Kipp, N.G., 1971. A new micropaleontological method for quantitative paleoclimatology: application to a late Pleistocene Caribbean core. In: Turekian, K.K. (Ed.), The Late Cenozoic Glacial Ages. Yale University Press, New Haven, CT, pp. 71–181.

Jongman, R.H.G., ter Braak, C.J.F., van Tongeren, O.F.R., 1995. Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge, p. 299.

Li, W.Y., Yao, Z.J., 1990. A study on the quantitative relationship between *Pinus* pollen in surface sample and *Pinus* vegetation. Acta Botanica Sinica 32, 943–950 (in Chinese).

Li, Y.C., Xu, Q.H., Xiao, J.L., Yang, X.L., 2005. Indication of some major pollen taxa in surface samples to their parent plants of forest in northern China. Quaternary Sciences 25, 598–608 (in Chinese).

Lotter, A.F., Birks, H.J.B., Hofmann, W., Marchetto, A., 1997. Modern diatom, cladocera, chironomid, and chrysophyte cyst assemblages as quantitative indicators for the reconstruction of past environmental conditions in the Alps I. Climate. Journal of Paleolimnology 18, 395–420.

Lu, H.Y., Wu, N.Q., Yang, X.D., Jiang, H., Liu, K.B., Liu, T.S., 2006. Phytoliths as quantitative indicators for the reconstruction of past environmental conditions in China I: phytolith-based transfer functions. Quaternary Science Reviews 25, 945–959.

Luo, C.X., Zheng, Z., Tarasov, P., Nakagawa, T., Pan, A.D., Xu, Q.H., Lu, H.Y., Huang, K.Y., 2010. A potential of pollen-based climate reconstruction using a modern pollen–climate dataset from arid northern and western China. Review of Palaeobotany and Palynology 160, 111–125.

Ma, Y.Z., Liu, K.B., Feng, Z.D., Sang, Y.L., Wang, W., Sun, A.Z., 2008. A survey of modern pollen and vegetation along a south–north transect in Mongolia. Journal of Biogeography 35, 1512–1532.

Nakagawa, T., Tarasov, P.E., Nishida, K., Gotanda, K., Yasuda, Y., 2002. Quantitative pollen-based climate reconstruction in central Japan: application to surface and Late Quaternary spectrum. Quaternary Science Reviews 21, 2099–2113.

Shen, C.M., Liu, K.B., Tang, L.Y., Overpeck, J.T., 2006. Quantitative relationships between modern pollen rain and climate in the Tibetan Plateau. Review of Palaeobotany and Palynology 140, 61–77.

ter Braak, C.J.F., 1988. CANOCO – A FORTRAN Program for Canonical Community Ordination by [Partial] [Detrended] [Canonical] Correspondence Analysis, Principal Components Analysis and Redundancy Analysis (Version 2.1). Report LWA-88-02. Agricultural Mathematics Group, Wageningen, p. 95.

ter Braak, C.J.F., Juggins, S., 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. Hydrobiologia 269, 485–592.

ter Braak, C.J.F., Prentice, I.C., 1988. A theory of gradient analysis. Advances in Ecological Research 18, 271–317.

ter Braak, C.J.F., Smilauer, P., 2002. CANOCO Reference manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (Version 4.5). Microcomputer Power, Ithaca, NY, p. 500.

Wang, F.Y., Song, C.Q., Sun, X.J., 1996. Study on surface pollen in middle Inner Mongolia, China. Acta Botanica Sinica 38, 902–909 (in Chinese).

Wen, R.L., Xiao, J.L., Chang, Z.G., Zhai, D.Y., Xu, Q.H., Li, Y.C., Itoh, S., Lomtatidze, Z., 2010. Holocene climate changes in the mid-high latitude monsoon margin reflected by the pollen record from Hulun Lake, northeastern Inner Mongolia. Quaternary Research 73, 293–303.

Xu, Q.H., Li, Y.C., Bunting, M.J., Tian, F., Liu, J.S., 2010. The effects of training set selection on the relationship between pollen assemblages and climate parameters: implications for reconstructing past climate. Palaeogeography, Palaeoclimatology, Palaeoecology 289, 123–133.

Xu, Z.J., Jiang, F.Y., Zhao, H.W., Zhang, Z.B., Sun, L., 1989. Annals of Hulun Lake. Jilin Literature and History Publishing House, Changchun, p. 691 (in Chinese).

Zhang, J.C., Lin, Z.G., 1985. Climate of China. Shanghai Scientific and Technical Publishers, Shanghai, p. 603 (in Chinese).