A chironomid-based mean July temperature inference model from the south-east margin of the Tibetan Plateau, China

Enlou Zhang[a*], Jie Chang[a],Yanmin Cao[b], Hongqu Tang[c], Pete Langdon[d], James Shulmeister[e], Rong Wang[a], Xiangdong Yang[a], Ji Shen[a]

[a] State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Science, Nanjing 210008, China

[b] College of Resources and Environmental Science, South-Central University for Nationalities, Wuhan 430074, P. R. China

[c] Research Centre of Hydrobiology, Jinan University, Guangzhou 510632, P. R. China

[d] Geography and Environment, University of Southampton, Southampton SO17 1BJ, UK

[e] School of Geography, Planning and Environmental Management, The University of Queensland, St Lucia, Brisbane, Qld 4072, Australia

*Corresponding author Email: elzhang@niglas.ac.cn

Abstract: Chironomid-based calibration training set comprised of 100 lakes from south-western China was established. Multivariate ordination analyses were used to investigate the relationship between the distribution and abundance of chironomid species and 18 environmental variables from these lakes. Canonical correspondence analyses (CCAs) and partial CCAs showed that mean July temperature is one of the independent and significant variables explaining the second largest amount of variance after potassium ions (K^+) in the 100 south-western Chinese lakes. Quantitative transfer functions were created using the chironomid assemblages for this calibration data set. The second component of the weighted average partial least square (WA-PLS) model produced a coefficient of determination ($r^2_{bootstrap}$) of 0.63, maximum bias (bootstrap) of 5.16 and root mean squared error of prediction (RMSEP) of 2.31 °C. We applied the transfer functions to a 150-year chironomid record from Tiancai Lake (26°38′3.8 N, 99°43′E, 3898 m a.s.l), Yunnan, China to obtain mean July temperature inferences. We validated these results by applying several reconstruction diagnostics and comparing them to a 50-year instrumental record from the nearest weather station (26°51'29.22"N, 100°14'2.34"E, 2390 m a.s.l). The transfer function performs well in this comparison. We argue that this 100-lake large training set is suitable for reconstruction work despite the low explanatory power of mean July temperature because it contains a complete range of modern temperature and environmental data for the chironomid taxa observed and is therefore robust.

Keywords: Chironomids; Temperature reconstruction; the south-east margin of the Tibetan Plateau; Transfer function; Quantitative paleoclimate record

1 Introduction

South-western China is an important region for examining changes in low and mid-latitudes atmospheric circulation in the Northern Hemisphere. It lies at the intersection of the influence of the Northern Hemisphere westerlies and two tropical monsoon systems, namely the Indian Ocean South-west Monsoon (IOSM) and the East Asian Monsoon (EAM) and should be able to inform us about changes in both the latitude and longitude of the influence of these respective systems through time. Reconstructing changes in circulation requires information about several climatic parameters, including past precipitation and temperature. While there are reasonable records of precipitation from this region (e.g. Wang et al., 2001, 2008; Dykoskia et al., 2005; Xiao et al., 2014), there is a paucity of information about temperature changes. In order to understand the extent and intensity of penetration of monsoonal air masses, robust summer temperature estimates are vital as this is the season that the monsoon penetrates south-western China.

Chironomid larvae are frequently the most abundant insects in freshwater ecosystems (Cranston, 1995) and subfossil chironomids are widely employed for palaeoenvironmental studies due to their sensitivity to environmental changes and ability of the head capsules to preserve well in lake sediments (Walker, 2001). A strong relationship between chironomid species assemblages and mean summer air temperature have been reported from many regions around the world and transfer functions were subsequently developed (e.g. Brooks and Birks, 2001; Larocque et al., 2001; Heiri et al., 2003; Gajewski et al., 2005; Barley et al., 2006; Woodward and Shulmeister, 2006; Langdon et al., 2008; Rees et al., 2008; Eggermont et al., 2010; Luoto, 2009; Holmes et al., 2011; Heiri et al., 2011; Chang et al., 2015a). The application of these transfer functions has provided quantitative temperature data since the last glacial period in many regions of the world (e.g. Woodward and Shulmeister, 2007; Rees and Cwynar, 2010; Samartin et al., 2012; Chang et al., 2015b; Muschitiello et al., 2015; Brooks et al., 2016). Consequently, subfossil chironomids have been the most widely applied proxy for past summer temperature reconstructions.

Merged regional chironomid training sets and combined inference models have been developed in Europe (Lotter et al., 1999; Holmes et al., 2011; Heiri et al., 2011; Luoto et al., 2014). These large datasets and models provide much more robust reconstructions than smaller local temperature inference models (Heiri et al., 2011; Luoto et al., 2014). However, the distribution of large regional inference models is limited to Europe and northern North America (e.g. Fortin et al., 2015). There is a need to build large training sets for other parts of the world where chironomids will likely be sensitive to temperature changes. Subfossil chironomids have been successfully used as paleoenvironmental indicators in China for over a decade. These included salinity studies on the Tibetan Plateau (Zhang et al, 2007) and the development of a nutrient based inference model for eastern China and parts of

Yunnan (Zhang et al., 2006, 2010, 2011, 2012). A large database of relatively undisturbed lakes, in which nutrient changes are minimal while temperature gradients are suitably large, is now available from south-western China and this provides the opportunity to develop a summer temperature inference model for this broad region.

In this study, a chironomid species assemblage training set and chironomid-based mean July air temperature inference models from 100 lakes on the south-east margin of the Tibetan Plateau are developed. We test and validate the selected transfer function models by applying it to a sediment core collected from Tiancai Lake (26°38′3.8 N, 99°43′E, 3898 m a.s.l) (Fig. 1) in Yunnan Province, south-western China for the last 120 years against a 50-year long instrumental record from Lijiang weather station (26°51'29.22"N, 100°14'2.34"E, 2390 m a.s.l) (Fig. 1), which is the closest meteorological station with the longest record.

2 Regional setting

The study area lies in the south-east margin of the Tibetan Plateau including the south-west part of Qinghai Province, the western part of Sichuan Province and the north-west part of Yunnan Province (Fig. 1). It is situated between 26 – 34° N, 99 – 104° E with elevations ranging from about 1000 m to above 5000 m a.s.l.. The study area is characterized by many north-south aligned high mountain ranges (e.g. Hengduan Mountains, Daxue Mountains, Gongga Mountains etc.) that are fault controlled and fall away rapidly into adjacent tectonic basins. The mountain ranges have been deeply dissected by major rivers including the Nujiang, Lancangjiang, Jinshajiang, Yalongjiang and Dadu rivers. Local relief in many places exceeds 3000 m a.s.l..

The climate of the study area is dominated by the westerlies in winter and by the IOSM in Yunnan and Tibet, but some of the easternmost lakes are affected by the EAM. There is a wet season that extends from May (June) to October accounting for 85-90% of total rainfall and a dry season from November to April. Annual precipitation varies greatly according to altitude and latitude. Most of the precipitation is derived from a strong south-west summer monsoonal flow that emanates from the Bay of Bengal (Fig. 1). Precipitation declines from south-east to north-west. Mean summer temperatures vary between about 6 to 22 °C from the north-west to the south-east (Institute of Geography, Chinese Academy of Sciences, 1990). Vegetation across the study area changes from warm temperate to subtropical rainforest at lower elevations in the south-west to alpine grasslands and herb meadows at high altitude.

2.1 Description of model validation site

Tiancai Lake (26°38′3.8 N, 99°43′E, 3898 m a.s.l) (Fig. 1) is in Yunnan Province, on the south-east margin of the Tibetan Plateau. It is a small alpine lake and has a maximum depth of 7 m, with a water surface area of ~ 2.1 ha and a drainage area of

~3 km^2. Tiancai Lake is dominated in summer by the IOSM, and most likely retains a tropical airflow in winter as the climate is remarkably temperate for this altitude. The mean annual and July air temperatures are approximately 2.5 °C and 8.4 °C respectively, and the annual precipitation is modelled as > 910 mm (Xiao et al. 2014). The lake is charged by 3 streams and directly from precipitation and drains into a lower alpine lake via a stream. The most common rock type in the catchment is a quartz poor granitoid (syenite). Terrestrial vegetation in the catchment consists mainly of conifer forest comprising *Abies* sp. and *Picea* sp. with an understory of *Rhododendron* spp. Above the tree-line, at about 4100 m a.s.l, Ericaceae shrubland (rhododendrons) gives way to alpine herb meadow and rock screes.

3 Methodology

3.1 Field and laboratory analysis

Surface sediment samples were collected from 100 lakes in the south-east margin of the Tibetan Plateau via six field campaigns during the autumn of each year between 2006 and 2012. The lakes in this area are mainly distributed at the top or upper slopes of the mountains and are primarily glacial in origin. Most lakes were reached by hiking or with horses and the lake investigation spanned several seasons. Small lakes (surface area c. ~1 km^2) were the primary target for sampling but some larger lakes were also included.

Surface sediments (0-1 cm) were collected from the deepest point in each lake after a survey of the bathymetry using a portable echo-sounder. Surface sediment samples were taken using a Kajak gravity corer (Renberg, 1991). The samples were stored in plastic bags and kept in the refrigerators at 4 °C before analysis. A 30 cm short core was extracted from the centre of Tiancai Lake at a water depth of 6.8 m using UWITEC gravity corer in 2008. The sediment core was sub-sampled at 0.5 cm contiguous intervals and refrigerated at 4°C prior to analysis.

Water samples were collected for chemical analysis from 0.5 m below the lake surface immediately before the sediment samples were obtained. Water samples for chemical analysis were stored in acid-washed polythene bottles and kept at 4 °C until analyses. Secchi depth was measured using a standard transparency disc. Conductivity, pH and dissolved oxygen (DO) were measured in the field using a HI-214 conductivity meter, Hanna EC-214 pH meter and JPB-607 portable DO meter. Chemical variables for the water samples including total phosphorus (TP), total nitrogen (TN), chlorophyll-*a* (chl *a*), K^+, Na^+, Mg^{2+}, Ca^{2+}, Cl^-, SO_4^{2-}, NO_3^- were determined at the Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences. The surface sediments were also analysed for percentage loss-on-ignition (% LOI) following standard methods (Dean 1974). Site-specific values for the mean July air temperature (MJT) and mean annual precipitation (MAP) were estimated using climate layers that were created using statistical downscaling of General Circulation

Model (GCM) outputs and terrain parameterization methods in a regular grid network with a grid-cell spacing of 1 km^2 (Böhner 1994, 2006; Böhner and Lehmkuhl, 2005) using reanalysis data. MJT is used to represent summer temperatures because July is the warmest month in south-western China.

3.2 Chironomid analyses

100 surface sediment samples from lakes of south-western China and 55 sub-samples from the Tiancai Lake short core were analysed for chironomids following standard methods (Brooks et al, 2007). The sediment was deflocculated in 10% potassium hydroxide (KOH) in a water bath at 75 °C for 15 minutes. The samples were then sieved at 212 µm and 90 µm and the residue was examined under a stereo-zoom microscope at x 25. Chironomid head capsules were hand-picked using fine forceps. All the head capsules found were mounted on microscope slides in a solution of Hydromatrix®. Samples produced less than 50 head capsules were not included in the subsequent analyses (Quinlan and Smol, 2001). The chironomid head capsules were identified mainly following Wiederholm (1983), Oliver and Roussel (1982), Rieradevall and Brooks (2001), Brooks et al. (2007) and a photographic guide provided in Tang (2006).

3.3 Numerical analysis

A range of numerical methods were used to determine the relative influence of the measured environmental parameters on the distribution and abundance of chironomids in the surface sediments within the training set. A total of eighteen environmental variables were considered in the initial statistical analyses (Table 1). These measurements were normalized using a $\log_{10}$ transformation prior to ordinations following a normality assessment of each data set. Chironomid species were used in the form of square root transformed percentage data in all statistical analyses. The ordinations were performed using CANOCO version 4.5 (ter Braak and Šmilauer, 2002). A detrended correspondence analysis (DCA; Hill and Gauch, 1980) with detrending by segments and nonlinear rescaling was used to explore the chironomid distribution pattern. The DCA was also used to identify the gradient length within the chironomid data and hence whether unimodal analyses were appropriate (ter Braak, 1987). Canonical correspondence analysis (CCA) down-weighted for rare taxa (with a maximum abundance of less than 2% and/or occurred in fewer than two lakes, i.e. Hill's $N_2 < 2$), with forward selection and Monte Carlo permutation tests (999 unrestricted permutations) was then used to identify the statistically significant ($p < 0.05$) variables influencing the chironomid distribution and abundance (ter Braak and Šmilauer, 2002). A preliminary CCA with all eighteen variables was used to identify redundant variables, reducing excessive co-linearity among variables (Hall and Smol, 1992), i.e. the environmental variable with highest variance inflation factor (VIF) was removed after each CCA and the CCA was repeated until all VIFs were less than 20 (ter Braak and Šmilauer, 2002). In addition, we used stepwise selection based on pseudo-F to aid the variable selection process. Only the remaining significant ($p < 0.05$) variables were included in the final CCA ordination. The relationship between the significant environmental variables and ordination axes was assessed with canonical coefficients and the associated t-values of the environmental variables with the respective axes. CCA bi-plots of sample and species scores were generated using CanoDraw (ter Braak and Šmilauer, 2002). Partial canonical correspondence analyses (pCCAs) were applied to test the direct and indirect effects of each of the significant variables in relation to the chironomid species data. These were performed for each of the significant variable with and without the remaining significant variables included as co-variables. Environmental variables that retained their significance after all pCCAs were selected for use in the analyses as they are the independent variables.

Chironomid-based transfer functions were developed for mean July temperatures using C2, version 1.5. (Juggins, 2005) for the calibration data set comprised of 100 lakes. The models were constructed using algorithms based on weighted-averaging (WA) and weighted-averaging partial-least-squares (WA-PLS) (Birks, 1995). Bootstrap cross-validation technique was tested for the dataset as previously demonstrated that it is more suitable for large datasets (Heiri et al., 2011) comparing to the jackknife technique. Transfer function models were evaluated based on the performance of the coefficient of determination (r^2_{boot}), average bias of predictions, maximum bias of predictions and root mean square error of prediction ($RMSEP_{boot}$). The number of components included in the final model was selected based on reducing the RMSEP by at least 5% (Birks, 1998). In addition, instead of using 5% as a simple threshold we also performed a t-test to further check if the additional component of the WA-PLS model is outperformed.

The transfer function models were then applied to the fossil chironomid data from Tiancai Lake. Mean July temperatures (MJT) were reconstructed from the site and three types of reconstruction diagnostics suggested in Birks (1995) were applied to assess the reliability of the results. These include goodness-of-fit, modern analogue technique (MAT) and the percentage (%) analysis of modern rare taxa in the fossil samples. For the goodness-of-fit analysis, the squared residual length (SqRL) was calculated by passively fitting fossil samples to the CCA ordination axis of the modern training set data constrained to MJT in CANOCO version 4.5 (ter Braak and Smilauer, 2002). Fossil samples with a SqRL to axis 1 higher than the extreme 10 and 5% of all residual distances in the modern calibration dataset were considered to have a ‘poor’ and ‘very poor’ fit with MJT respectively. The chi-square distance to the closest modern assemblage data for each fossil sample was calculated in C2 (Juggins, 2005) using the MAT. Fossil samples with a chi-square distance to the closest modern sample larger than the 5^{th} percentile of all chi-square distances in the modern assemblage data were identified as samples with ‘no good’ analogue. The percentage of rare taxa in the fossil samples was also calculated in C2 (Juggins, 2005), where a rare taxon has a Hill's $N_2 < 2$ in the modern data set (Hill, 1973). Fossil samples that contain > 10% of these rare taxa were likely to be poorly estimated (Brooks and Birks, 2001). Finally, the chironomid-based transfer functions inferred MJT patterns were compared to the instrumental recorded data from Lijiang weather station between the years of 1951 and 2014.

3.4. Chronology for Tiancai Lake core

The top 28 cm of the sediment core recovered from Tiancai Lake were used for ^{210}Pb dating. Sediment samples were dated using ^{210}Pb and ^{137}Cs by non-destructive gamma spectrometry (Appleby and Oldfield, 1992). Samples were counted on an Ortec HPGe GWL series well-type coaxial low background intrinsic germanium detector to determine the activities of ^{210}Pb, ^{226}Ra and ^{137}Cs. A total of 58 samples at an interval of every 0.5 cm were prepared and analysed at the Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences. Sediment chronologies were calculated using a composite model (Appleby, 2001). ^{137}Cs was used to identify the 1963 nuclear weapons peak, which was then used as part of a constant rate of supply (CRS) model to calculate a ^{210}Pb chronology for the core.

4 Results

4.1 Distribution of chironomid taxa along the temperature gradient

A total of 85 non-rare taxa (Hill's $N_2 > 2$) (Brooks and Birks, 2001) were identified from 100 south-western Chinese lakes (Fig. 2a). Only these non-rare taxa were included in the final transfer function models. Mean July temperature is an important variable driving the distribution and abundance of the chironomid taxa in this dataset (Fig. 2a). Common cold stenotherms, here defined as taxa with a preference for < 12°C MJT include *Heterotrissocladius marcidus*-type, *Tanytarsus gracilentus*-type, *Paracladius*, *Micropsectra insignilobus*-type, *Micropsectra radialis*-type, *Tanytarsus lugens*-type, *Micropsectra* Type A, *Pseudodiamesa*, *Micropsectra atrofasciata*-type and *Corynoneura lobata*-type (Fig. 2a). Taxa characterizing warmer temperatures (> 12°C) include *Polypedilum nubeculosum*-type, *Eukiefferiella devonica*-type, *Microtendipes pedellus*-type and *Tanytarsus lactescens*-type and *Chironomus plumosus*-type (Fig. 2a). Many of the remaining taxa reflect more cosmopolitan distributions, these include *Procladius*, *Corynoneura scutellata*-type, *Tanytarsus pallidicornis*-type, *Tanytarsus mendax*-type and *Paratanytarsus austriacus*-type (Fig. 2a).

4.2 Ordination analyses and model development

The detrended canonical analyses (DCAs) performed on the 100 lakes and 85 non-rare chironomid taxa had an axis 1 gradient length of 3.033 indicating a CCA approach was appropriate for modelling the chironomid taxa response (Birks, 1998). The eighteen environmental variables were tested as in the initial CCA and the results showed that TDS had the highest VIF. It was then removed from the following CCAs.

Seven of the remaining variables had significant ($p < 0.05$) explanatory power with respect to the chironomid species data. These were K^+ (4.8%), MJT (4.4%), conductivity (4.4%), Cl^- (3.4%), LOI (3.1%), Na^+ (2.7%) and depth (2%) (Table 2). A total of 14.6% of variance was explained by the four CCA axes with the seven significant variables included and the first two axes explained 10% of the total variance. Of these variables, conductivity and K^+ were significantly correlated ($p < 0.01$) with CCA axis 1 and cond, depth, Cl^-, MJT showed a significant correlation ($p < 0.01$) with CCA axis 2 (Table 2, Fig. 3a, b, based on the t-values). Potassium ions (K^+) explained the largest variance in the chironomid species data and showed the strongest correlation with CCA axis 1. MJT and conductivity explained equally the second largest amount of variance (4.4%) where MJT was significantly correlated with CCA axis 2 and conductivity was significantly correlated with both axis 1 and 2 (Table 2). The pCCAs (Table 3) demonstrated that within the significant variables K^+, MJT, Cl^-, LOI and depth remained their significance ($p < 0.01$) when the other variables were included as co-variables. Potassium ions (K^+) is the independent variable dominates the first CCA axis. MJT and Cl^- are the independent variables dominating the second CCA axis but MJT has an overall higher explanatory power (Table 2).

A bi-plot of the CCA species scores indicated that taxa such as *Heterotrissocladius marcidus*-type and *Tanytarsus lugens*-type had a significant amount of variance explained by the first two CCA axes and were negatively correlated with CCA axis 1. Taxa including *Polypedilum nubeculosum*-type, *Chironomus plumosus*-type were positively correlated with CCA axis 1 with a significant amount of variance explained by the CCA axis 1 and 2. A bi-plot of the CCA sample scores showed that a major proportion of sites distributed concentrating around depth (Fig. 3b) whereas depth only explains 2% of the total variance in the chironomid data.

The transfer functions were developed for mean July temperature (MJT). We acknowledge that MJT is not the sole independent variable on CCA axis 2 in the dataset but transfer functions based on this large regional dataset are created and applied to reconstruct MJT because it is a more useful parameter compared to K^+ and Cl^-. Both weighted averaging (WA) and weighted averaging partial least squares (WA-PLS) models were tested in the modern calibration set. Summary statistics of inference models based on these two different numerical methods are listed in Table 4a. As expected, the bootstrapped WA with inverse deshrinking (WAinv) and WA-PLS models generated similar statistical results for the calibration training set. The WAinv model produced an r^2_{boot} of 0.61, AveBiasboot of 0.06, MaxBiasboot of 5.30 and RMSEP of 2.30 °C (Table 4a). We selected the second component of WA-PLS bootstrap model as it is more robust according to the t-test results (Table 4b). It produced an r^2_{boot} of 0.63, AveBiasboot of 0.101**,** a lower MaxBiasboot of 5.16 and RMSEP of 2.31 °C. Figures 4c and 4d show the chironomid-inferred versus observed MJT and the distribution of prediction residuals for the above transfer function models respectively.

4.3 Reconstructions from Tiancai Lake

A total of 55 sub-samples were analysed for chironomid taxa throughout the top 28 cm of the core recovered from Tiancai Lake. There were 41 non-rare (Hill's N_2 > 2) taxa present (Fig. 2b). The general assemblages of these 55 sub-samples include *Heterotrissocladius marcidus*-type, *Tvetenia tamafalva*-type, *Micropsectra insignilobus*-type, *Corynoneura lobata*-type, *Paramerina divisa*-type, *Micropsectra radialis*-type, *Paratanytarsus austriacus*-type, *Thienemanniella clavicornis*-type, *Eukiefferiella claripennis*-type, *Rheocricotopus effusus*-type, *Macropelopia*, *Pseudodiamesa* and *Procladius* (Fig. 2b). All the taxa identified from this record were well represented, and most of them were recognized as cold stenotherms, in the modern calibration training sets (Fig. 2a). We acknowledge that some of the lotic taxa may result in poor temperature estimates when applying the transfer function therefore, reconstruction diagnostics were necessary.

The ^{210}Pb dating results demonstrated that the top 28 cm of the short core recovered from Tiancai Lake represent the last c. ~150 years (Fig 5). We applied both new transfer function models (WA and WA-PLS based on100 lakes) to reconstruct the MJT changes between 1860 AD and 2008 (Fig. 6a). The WA and WA-PLS models constructed showed identical trends in the MJT reconstructions over the last c. ~150 years (Fig. 6a). There were small deviations in terms of absolute values but the variations in the reconstructed MJT between the two models were within 0.1 °C for each sample (Fig. 6a). Goodness-of-fit analysis on the reconstruction results showed that out of the 55 fossil samples, eight samples from the years between 2000 and 2007 AD have 'poor' and 'very poor' fit to MJT (Fig. 6b). The modern analogue analysis showed that only four fossil samples have 'no good' analogues in the 100 lake dataset (Fig. 6c). All 55 fossil samples contain less than 10% of the taxa that were rare in the modern 100 lake training set (Fig. 6d). Finally, the reconstructed results also showed a comparable MJT trend and a statistical significant correlation ($p < 0.05$, $r = 0.45$, $n = 31$) with the instrumental measured data between 1951 and 2007 AD from Lijiang weather station (Fig. 6e).

5 Discussion

5.1 Reliability of the environmental and chironomid data

Obtaining reliable estimates of the modern climate data has been challenging in south-western China. There are very few meteorological stations and climate monitoring in the high mountains of our study area is virtually non-existent. Climate parameters including mean July temperatures and mean annual precipitation used in this study are interpolated from climate surfaces derived from a mathematical climate surface model based on the limited meteorological data and a digital terrain model (DTM) applied to the whole of the wider Tibetan region (4000 x 3000 km) (Böhner, 2006). We acknowledge that there are limitations in these data due to the sparse distribution of observations from meteorological stations. Modelling precipitation in topographically complex parts of this region such as Yunnan is problematic due to the orographic interception (or non-interception) of monsoonal air masses upwind of the sites, but the scale of the DTM means that mean temperature data should be reasonably robust, except in the most topographically complex areas. Further meteorological observations are required to refine this and other studies. We suspect that this is potentially an issue resulting the relatively low transfer function model coefficient (r^2_{boot}).

We examined the chironomid taxa that appeared as temperature indicators in the calibration set. A number of taxa, namely *Pseudodiamesa*, *Pseudosmittia* and *Corynoneura lobata*-type emerge as cold stenotherms. Further examination of these taxa show that these three taxa are all likely lotic (Cranston, 2010). These taxa would possibly have washed in to the lakes from streams and therefore it is not appropriate to make temperature inferences based on them. We also observed that another cold stenotherm *Tanytarsus gracilentus*-type is closely related to lake depth, while both *Tvetenia tamafalva*-type and *Micropsectra* show closer correlation with LOI and Cl^- in the CCA biplot (Fig. 3a). The observations match with the ecological recognition and interpretation of these taxa in literature where *Tanytarsus gracilentus*-type was identified as a benthic species in the arctic and is sometimes found in temperate shallow eutrophic ponds (Einarsson et al., 2004; Ives et al., 2008); *Tvetenia tamafalva*-type was often found in streams and this is likely related to the organic content (LOI) of the substrates as they are detritus feeders (Brennan and McLachlan, 1979); while *Micropsectra* was found in thermal springs and pools (Hayford et al., 1995; Batzer and Boix, 2016) and this is reflected in this dataset with having a close relationship with Cl^-. It presents in lakes such as Lake Tengchongqinghai, Qicai Lake and Lake Haizibian that have high levels of Cl^- ions. These sites are located in geothermal spring region of Sichuan and Yunnan Provinces.

Well-known warm stenotherms that are distributed along the MJT gradient of the CCA species bi-plot (Fig. 3a) include *Dicrotendipes*, *Microchironomus*, *Polypedilum* and *Microtendipes*. Many studies (e.g. Walker et al. 1991; Larocque et al. 2001; Rosenberg et al., 2004; Brodersen and Quinlan, 2006; Woodward and Shulmeister, 2006) show that these taxa are warm temperature indicators worldwide. We therefore argue that this large calibration training set contains a relatively complete range of temperatures and environments expected to have been experienced by lakes and their chironomid fauna in the past (Brooks and Birks, 2001). This will be particularly useful when applying the models to reconstruct changes in the late Pleistocene and Holocene when climates were different (Heiri et al., 2011).

This 100-lake training set covers a temperature gradient ranging from 4.2 °C to 20.8 °C (MJT gradient of 16.6 °C). Based on the CCAs, we observed that the MJT signal in this larger training set is partially masked by a salinity gradient. This is represented by potassium ions (K^+) and conductivity (Fig. 3a, b). CCA axis 1 is dominated by K^+ and this may be related to weak weathering. This is because (1) the first CCA axis is driven by lakes that have low precipitation but intermediate level of evaporation, examples of these sites include Lake Xiniuhaijiuzhai, Lake Muchenghai and Lake Kashacuo, from the north margin of Sichuan Province. These lakes indicate cool, dry and low windiness conditions that lead to a weak weathering environment. We highlight that this area is different from the high Tibetan Plateau where aridity and salinity dominates. (2) In chemical weathering sequences, K^+ is an early stage weathering product (Meunier and Velde, 2013) and K^+ is often associated with primary minerals, such as feldspars and micas in the bedrock (Hinkley, 1996). Salinity is responding to both temperature and aridity but further pCCAs (Table 3) indicate that both K^+ and MJT are independent variables in this training set.

The second CCA axis is co-dominated by MJT and Cl^- with very similar gradient lengths. Lakes distributed along the warmer end of the MJT gradient include Lake Longtan, Lake Lutu, Lake Luoguopingdahaizi and Lake Jianhu. Most of these sites are lower to intermediate altitude sites in the dataset (below 2700 m a.s.l) because elevation is correlated with temperature. Sodium ions (Na^+) largely follow the same axis as MJT as evaporation is related in part to temperature. In summary, MJT and Cl^- are both independent variables that drive the second CCA axis and Cl^-, and Na^+ partially reflect evaporation effects because, on average, lakes in warmer climates evaporate more than those in colder ones. In addition, Cl^- concentration may also relate to the characteristics of the bedrock geology of the region. We highlight that there are very few lakes on the Cl^- gradient and these lakes are from the border of Sichuan and Yunnan Provinces, where geothermal springs are widespread. We argue that developing a MJT transfer function is appropriate for this large lake training set because MJT is independent of other variables (e.g. Rees et al., 2008; Chang et al., 2015a). Although Cl^- is also independent and co-dominates CCA axis 2, the overall explanatory power is lower (Table 2) and also the lambda ratio ($\lambda1/\lambda2$) is smaller than MJT (Table 3). We retained all 100 lakes from the region without removing sites to artificially enhance the MJT gradient in the ordination analyses and model development because this large dataset is an accurate reflection of the natural environment of south-western China.

We selected the WA-PLS based transfer function models over the WAinv based approach for both training sets because the addition of PLS components can reduce the prediction error in datasets with moderate to large noise (ter Braak and Juggins, 1993). The training set has a MJT gradient of 16.6 °C and the RMSEP represents 13.8% of the scalar length of the MJT gradient. This is comparable with most chironomid-based transfer function models including those developed from Northern Sweden with 100 lakes ($r^2 = 0.65$, Larocque et al., 2001), western Ireland with 50 lakes ($r^2 = 0.60$, Potito et al., 2014) and Finland with 77 lakes ($r^2 = 0.78$, Luoto, 2009) representing 14.7%, 15% and 12.5% of the scalar length of the temperature gradient respectively but less robust than the combined 274-lakes transfer function developed from Europe ($r^2 = 0.84$, RMSEP representing 10.4% of the scalar length of the MJT gradient) (Heiri et al., 2011). Despite of the relatively lower model coefficient (r_{boot} = 0.63), we observe that by having a large number of lakes in the calibration set, the distribution of the sites along the MJT gradient is relatively even (Fig. 4d). The distribution of the error residuals generates a smooth curve (Fig. 4d). The model leads to overestimation of low and underestimation of high temperature values which is typical of the WA models (ter Braak and Juggins, 1993). We acknowledge that the lower model coefficient (r_{boot}) may also relate to the low explanatory power of MJT in the chironomid species data and large number of independent and significant variables in the training set when a wide range of lakes were included. However, the extensive temperature gradient length allowed the incorporation of full potential abundance and distributional ranges for each of the chironomid taxa.

5.2 Tiancai Lake reconstructions

All three types of diagnostic techniques applied (Fig 6 b-d) suggest that a reliable MJT reconstruction was provided by the two-component WA-PLS model based on this 100-lake dataset overall. We highlight that the eight samples from the years between 2000 and 2007 AD have ‘poor’ and ‘very poor’ fit to MJT may suggest that it is possible a second gradient other than MJT influencing the chironomid species distribution and abundance in the most recent fossil samples of Tiancai Lake. In the comparison for the MJT reconstruction results with the instrumental record from Lijiang weather station (Fig. 6a), we do not expect the absolute MJT values to be identical because Lijiang is located ~55 km east-northeast (ENE) and ~1600 m lower in altitude than Tiancai Lake. We applied a typical environmental lapse rate of temperature (change with altitude) for Alpine regions, which is 0.58 °C per 100 metres (Rolland, 2003) to estimate the equivalent MJT values from Lijiang station. If the chironomid-based transfer functions are able to provide reliable estimates for MJT, we expect the records demonstrate a similar trend with the instrumental data (Fig. 6e).

The reconstruction results are well matched with the expected outcomes as the transfer function models based on 100 lakes for the broad area of south-western China reconstructs MJT broadly match the trend recorded by the instrument. By applying the environmental lapse rate, we observe a temperature depression from Lijiang to Tiancai Lake of about 9.3 °C (giving an inferred MJT at Tiancai Lake of 8.1 °C in the year of 2004). This magnitude of change is consistent with the chironomid-based reconstructions from Tiancai Lake (at an average of 7.8 °C for the samples representing the years of 2004-2005), where the difference in mean is 0.3 °C when compared. The implication is that the transfer function model is able to reconstruct the MJT that closely reflects the actual climate record. We observe there are minor out of phase patterns (Fig. 6e) and this may reflect the uncertainties of applying the ^{210}Pb chronology to very recent lake sediments (Binford, 1990). Furthermore, we note that sediment samples reflect more than one season and consequently the total range of the temperature reconstructions from the chironomid samples is likely to be slightly less than the meteorological data because of the smearing out of extreme years. While we expect overall trends between Lijiang and Tiancai Lake to be similar, the sites are not closely co-located and some natural variability between the sites is expected. Nevertheless, a significant correlation ($p < 0.05$) was obtained between the instrumental data and the WA-PLS model inferred MJT data for the last ~ 50 years. We highlight that in addition to the record validation produced by the reconstruction diagnostic techniques, the well-compared trend with the instrumental record is reassuring that the model is capable to provide realistic pattern of the long-term mean July temperature changes. In summary, the chironomid-based transfer function developed using the 100 lakes calibration training set has generated reliable quantitative temperature records and can be applied to reconstructing past climate in south-western China.

6 Conclusions

Chironomid-based summer temperature transfer functions using 100 lakes from south-western China have been constructed and applied to Yunnan region in the south-eastern margin of the Tibetan Plateau. Both the ordination and transfer function statistics show that the chironomid-based transfer function is reliable. This large regional training set allowed insight into the regional chironomid distribution and species abundance despite having many more independent environmental gradients. The test of the transfer function models against the modern data suggest that the two-component WA-PLS model provided reconstructions that match the trend of the local instrumental record for the last 50 years. As also demonstrated from pan-European chironomid based transfer functions (e.g. Brooks and Birks, 2001; Heiri et al., 2011), this broadly based 100 Chinese lakes is likely robust and is appropriate for use reconstructing long-term summer temperature changes of south-western China.

Acknowledgement: We thank X Chen, E.F. Liu, M. Ji, R. Chen, Y.L. Li, X.Y. Xiao, J.J. Wang, Q. Lin and B.Y. Zheng (Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences) for field assistance, Jürgen Böhner (Georg-August-University Göttingen, Germany) for help with climate data. This research was supported by the Program of Global Change and Mitigation (2016YFA0600502), the National Natural Science Foundation of China (No. 41272380, 41572337).

References

Appleby, P.G., 2001.Chronostratigraphic techniques in recent sediment, In: Last WM, Smol J P. Tracking environmental change using lake sediments, Volume 1: basin analysis, coring, and chronological techniques. Klauwer Academic Publishers, pp. 171-196.

Appleby, P.G., Oldfield, F., 1992. Application of 210Pb to sedimentation studies. In: Ivanovich M, Harmon RS (Eds.), Uranium series disequilibrium. OUP, pp. 731-778.

Batzer, D., Boix, D., 2016. Invertebrates in Freshwater Wetlands: An International Perspective on their Ecology. Springer. pp. 361 – 370.

Barley, E.M., Walker, I.R., Kurek, J., Cwynar, L.C., Mathewes, R.W., Gajewski, K., Finney, B.P., 2006. A northwest North American training set: Distribution of freshwater midges in relation to air temperature and lake depth. Journal of Paleolimnology 36, 295-314

Binford, M.W., 1990. Calculation and uncertainty analysis of 210Pb dates for PIRLA project lake sediment cores. Journal of Paleolimnology 3, 253-267.

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

Birks, H.J.B., 1995. Quantitative palaeoenvironmental reconstructions. In: Maddy, D., Brew, J.S. (Eds.), Statistical Modelling of Quaternary Science Data. Technical Guide 5. Quaternary Research Association, Cambridge, pp. 116–254

Böhner J., 1994. Circulation and representativeness of precipitation and air temperature in the southeast of the Qinghai-Xizang Plateau. GeoJournal 34, 55-66.

Böhner, J., 2006. General climatic controls and topoclimatic variations in Central and High Asia. Boreas 35, 279-295.

Böhner, J., Lehmkuhl, F., 2005. Environmental change modelling for Central and High Asia: Pleistocene, present and future scenarios. Boreas 34, 220-231.

Brennan, A., McLachlan, A.J., 1979. Tubes and tube-building in a lotic Chironomid (Diptera) community. Hydrobiologia 67, 173-178.

Brodersen, K.P., Quinlan, R., 2006. Midges as palaeoindicators of lake productivity, eutrophication and hypolimnetic oxygen. Quaternary Science Reviews 25, 1995-2012.

Brooks, S.J., Birks, H.J.B., 2001. Chironomid-inferred air temperatures from Lateglacial and Holocene sites in north-west Europe: progress and problems. Quaternary Science Reviews 20, 1723-1741.

Brooks, S.J., Davies, K.L., Mather, K.A., Matthews, I.P., Lowe, J.J., 2016. Chironomid-inferred summer temperatures for the Last Glacial–Interglacial Transition from a lake sediment sequence in Muir Park Reservoir, west-central Scotland. Journal of Quaternary Science 31, 214-224.

Brooks, S.J., Langdon, P.G., Heiri, O., 2007. The Identification and Use of Palaearctic Chironomidae Larvae in Palaeoecology. Quaternary Research Association.

Chang, J.C., Shulmeister, J., Woodward, C., 2015. A chironomid based transfer function for reconstructing summer temperatures in southeastern Australia. Palaeogeography, Palaeoclimatology, Palaeoecology 423, 109-121.

Chang, J.C., Shulmeister, J., Woodward, C., Steinberger, L., Tibby, J., Barr, C., 2015. A chironomid-inferred summer temperature reconstruction from subtropical Australia during the last glacial maximum (LGM) and the last deglaciation. Quaternary Science Reviews 122, 282-292.

Cranston. P.S., 1995. Chironomids: from Genes to Ecosystems, CSIRO, Melbourne, pp. 482.

Cranston, P.S., 2000. Electronic guide to the chironomidae of Australia. http://www.science.uts.edu.au/sasb/chiropage/

Dean Jr, W.E., 1974. Determination of carbonate and organic matter in calcareous sediments and sedimentary rocks by loss on ignition: comparison with other methods. Journal of Sedimentary Research, 44, 242-248.

Dykoski, C.A., Edwards, R.L., Cheng, H., Yuan, D., Cai, Y., Zhang, M., Lin, Y., Qing, J., An, Z., Revenaugh, J., 2005. A high-resolution, absolute-dated Holocene and deglacial Asian monsoon record from Dongge Cave, China. Earth and Planetary Science Letters 233, 71-86.

Eggermont, H., Heiri, O., Russell, J., Vuille, M., Audenaert, L., Verschuren, D., 2010. Paleotemperature reconstruction in tropical Africa using fossil Chironomidae (Insecta: Diptera). Journal of Paleolimnology 43, 413-435.

Einarsson, Á., Stefánsdóttir, G., Jóhannesson, H., Ólafsson, J.S., Már Gíslason, G., Wakana, I., Gudbergsson, G., Gardarsson, A., 2004. The ecology of Lake Myvatn and the River Laxá: Variation in space and time. Aquatic Ecology 38, 317-348.

Fortin, M.C., Medeiros, A.S., Gajewski, K., Barley, E.M., Larocque-Tobler, I., Porinchu, D.F., Wilson, S.E., 2015. Chironomid-environment relations in northern North America. Journal of Paleolimnology 54, 223-237.

Gajewski, K., Bouchard, G., Wilson, S.E., Kurek, J., Cwynar, L.C., 2005. Distribution of Chironomidae (Insecta: Diptera) Head Capsulesin Recent Sediments of Canadian Arctic Lakes. Hydrobiologia 549, 131-143.

Hall, R.I., Smol, J.P., 1992. A weighted—averaging regression and calibration model for inferring total phosphorus concentration from diatoms in British Columbia (Canada) lakes. Freshwater Biology 27, 417-434.

Hayford, B.L., Sublette, J.E., Herrmann, S.J., 1995. Distribution of Chironomids (Diptera: Chironomidae) and Ceratopogonids (Diptera: Ceratopogonidae) along a Colorado Thermal Spring Effluent. Journal of the Kansas Entomological Society 68, 77-92.

Heiri, O., Brooks, S.J., Birks, H.J.B., Lotter, A.F., 2011. A 274-lake calibration data-set and inference model for chironomid-based summer air temperature reconstruction in Europe. Quaternary Science Reviews 30, 3445-3456.

Heiri, O., Lotter, A.F., Hausmann, S., Kienast, F., 2003. A chironomid-based Holocene summer air temperature reconstruction from the Swiss Alps. The Holocene 13, 477-484.

Hill, M.O., Gauch, H.G., 1980. Detrended Correspondence Analysis: An Improved Ordination Technique. Vegetatio 42, 47–58.

Hinkley, T.K., 1996. Preferential Weathering of Potassium Feldspar in Mature Soils, Earth Processes: Reading the Isotopic Code. American Geophysical Union, pp.

377-389.
Holmes, N., Langdon, P.G., Caseldine, C., Brooks, S.J., Birks, H.J.B., 2011. Merging chironomid training sets: implications for palaeoclimate reconstructions. Quaternary Science Reviews 30, 2793-2804.
Ives, A.R., Einarsson, A., Jansen, V.A.A., Gardarsson, A., 2008. High-amplitude fluctuations and alternative dynamical states of midges in Lake Myvatn. Nature 452, 84-87.
Kim, N., Han, J.K., 1997. Assessing the integrity of cross-validation: a case for small sample-based research. Hong Kong University of Science and Technology, Department of Marketing, Working Paper Series no. MKTG 97.096
Langdon, P.G., Holmes, N., Caseldine, C.J., 2008. Environmental controls on modern chironomid faunas from NW Iceland and implications for reconstructing climate change. Journal of Paleolimnology 40, 273-293.
Larocque, I., Hall, R.I., Grahn, E., 2001. Chironomids as indicators of climate change: a 100 ‐ lake training set from a subarctic region of northern Sweden (Lapland). Journal of Paleolimnology 26, 307-322.
Lotter, A.F., Walker, I.R., Brooks, S.J., Hofmann, W., 1999. An intercontinental comparison of chironomid palaeotemperature inference models: Europe vs North America. Quaternary Science Reviews 18, 717-735.
Luoto, T.P., 2009. A Finnish chironomid- and chaoborid-based inference model for reconstructing past lake levels. Quaternary Science Reviews 28, 1481-1489.
Luoto, T.P., Kaukolehto, M., Weckström, J., Korhola, A., Väliranta, M., 2014. New evidence of warm early-Holocene summers in subarctic Finland based on an enhanced regional chironomid-based temperature calibration model. Quaternary Research 81, 50-62.
Institute of Geography, Chinese Academy of Sciences, 1990. Atlas of Tibet Plateau. Science Press. (in Chinese).
Juggins, S., 2005. C2 Version 1.5: software for ecological and palaeoecological data analysis and visualisation. University of Newcastle, Newcastle-upon-Tyne.
Juggins, S., 2013. Quantitative reconstructions in palaeolimnology: new paradigm or sick science? Quaternary Science Reviews 64, 20-32.
Meunier, A. and Velde, B.D., 2013. Illite: Origins, evolution and metamorphism. Springer Science & Business Media, pp. 65 – 76.
Muschitiello, F., Pausata, F.S.R., Watson, J.E., Smittenberg, R.H., Salih, A.A.M., Brooks, S.J., Whitehouse, N.J., Karlatou-Charalampopoulou, A., Wohlfarth, B., 2015. Fennoscandian freshwater control on Greenland hydroclimate shifts at the onset of the Younger Dryas. Nature Communications 6.
Oliver, D.R., Roussel, M.E. 1982. The larvae of Pagastia Oliver (Diptera: Chironomidae) with descriptions of three new species. The Canadian Entomologist, 114: 849–854.
Potito, A.P., Woodward, C.A., McKeown, M., Beilman, D.W., 2014. Modern influences on chironomid distribution in western Ireland: potential for palaeoenvironmental reconstruction. Journal of Paleolimnology 52, 385-404.
Quinlan, R., Smol, J.P., 2001. Setting minimum head capsule abundance and taxa

deletion criteria in chironomid-based inference models. Journal of Paleolimnology 26, 327-342.

Rees, A.B.H., Cwynar, L.C., 2010. Evidence for early postglacial warming in Mount Field National Park, Tasmania. Quaternary Science Reviews 29, 443-454.

Rees, A.B.H., Cwynar, L.C., Cranston, P.S., 2008. Midges (Chironomidae, Ceratopogonidae, Chaoboridae) as a temperature proxy: a training set from Tasmania, Australia. Journal of Paleolimnology 40, 1159-1178.

Renberg, I., 1991. The HON-Kajak sediment corer. Journal of Paleolimnology 6, 167-170.

Rieradevall, M., Brooks, S.J., 2001. An identification guide to subfossil Tanypodinae larvae (Insecta: Diptera: Chrironomidae) based on cephalic setation. Journal of Paleolimnology 25, 81-99.

Rolland, C., 2003. Spatial and Seasonal Variations of Air Temperature Lapse Rates in Alpine Regions. Journal of Climate 16, 1032-1046.

Rosenberg, S.M., Walker, I.R., Mathewes, R.W., Hallett, D.J., 2004. Midge-inferred Holocene climate history of two subalpine lakes in southern British Columbia, Canada. The Holocene 14, 258-271.

Samartin, S., Heiri, O., Vescovi, E., Brooks, S.J., Tinner, W., 2012. Lateglacial and early Holocene summer temperatures in the southern Swiss Alps reconstructed using fossil chironomids. Journal of Quaternary Science 27, 279-289.

Tang, H. Q. 2006. Biosystematic study on the chironomid larvae in China (Diptera: Chironomidae). Nankai University, Tianjin, China, 945pp.

ter Braak, C. J. F., 1987. Unimodal models to relate species to environment. Agricultural Mathematics Group, Wageningen, The Netherlands. pp. 152

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, 485-502.

ter Braak, C.J.F., Šmilauer, P., 2002. CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5), Microcomputer power, Itaca, www.canoco.com.

Walker, I. R., 2001. Midges: Chironomidae and related Diptera. In: J. P. Smol, H. J. B. Birks, and W. M. Last (Eds), Tracking Environmental Change Using Lake Sediments. Volume 4. Zoological Indicators. Kluwer Academic Publishers, Dordrecht. pp. 43-66.

Walker, I.R., Smol, J.P., Engstrom, D.R., Birks, H.J.B., 1991. An Assessment of Chironomidae as Quantitative Indicators of Past Climatic Change. Canadian Journal of Fisheries and Aquatic Sciences 48, 975-987.

Wang, Y.J., Cheng, H., Edwards, R.L., An, Z.S., Wu, J.Y., Shen, C.C., Dorale, J.A., 2001. A High-Resolution Absolute-Dated Late Pleistocene Monsoon Record from Hulu Cave, China. Science 294, 2345-2348.

Wang, Y., Cheng, H., Edwards, R.L., Kong, X., Shao, X., Chen, S., Wu, J., Jiang, X., Wang, X., An, Z., 2008. Millennial- and orbital-scale changes in the East Asian monsoon over the past 224,000 years. Nature 451, 1090-1093.

Wiederholm T., 1983. Chironomidae of the Holarctic region. Keys and diagnoses. Part 1 Larvae. In: Wiederholm T (ed). Scandinavian Entomology Ltd, 1-482.

Woodward, C.A., Shulmeister, J., 2006. New Zealand chironomids as proxies for human-induced and natural environmental change: Transfer functions for temperature and lake production (chlorophyll a). Journal of Paleolimnology 36, 407-429.

Woodward, C.A., Shulmeister, J., 2007. Chironomid-based reconstructions of summer air temperature from lake deposits in Lyndon Stream, New Zealand spanning the MIS 3/2 transition. Quaternary Science Reviews 26, 142-154.

Xiao, X., Haberle, S.G., Shen, J., Yang, X., Han, Y., Zhang, E., Wang, S., 2014. Latest Pleistocene and Holocene vegetation and climate history inferred from an alpine lacustrine record, northwestern Yunnan Province, southwestern China. Quaternary Science Reviews 86, 35-48.

Zhang, E., Bedford, A. Jones,R., Shen, J., Wang, S., Tang, H., 2006. A subfossil chironomid-total phosphorus inference model from the middle and lower reaches of Yangtze River lakes. Chinese Science Bulletin. 51: 2125-2132

Zhang, E., Cao, Y., Langdon, P., Jones, R., Yang, X., Shen, J., 2012. Alternate trajectories in historic trophic change from two lakes in the same catchment, Huayang Basin, middle reach of Yangtze River, China. Journal of Paleolimnology 48, 367-381.

Zhang, E., Jones, R., Bedford, A., Langdon, P., Tang, H., 2007. A chironomid-based salinity inference model from lakes on the Tibetan Plateau. Journal of Paleolimnology 38, 477-491.

Zhang, E., Langdon, P., Tang, H., Jones, R., Yang, X., Shen, J., 2011. Ecological influences affecting the distribution of larval chironomid communities in the lakes on Yunnan Plateau, SW China. Fundamental and Applied Limnology 179, 103-113.

Zhang, E., Liu, E., Jones, R., Langdon, P., Yang, X., Shen, J., 2010. A 150-year record of recent changes in human activity and eutrophication of Lake Wushan from the middle reach of the Yangze River, China. Journal of Limnology 69, 235-241.

FIGURE LEGENDS

Fig. 1 Map of south-west China (a) showing the location of 100 lakes included in the calibration training set (square box). (b) Lakes from Yunnan Province are shown in the square box and (c) the location of Tiancai Lake is marked with yellow triangle.

Fig 2a. Chironomid species stratigraphy diagram of the 85 non-rare taxa with $N_2 > 2$. Mean July temperature is on the y-axis and taxon abundance is in percentage. The taxon code is correspondent to the code used in Figure 3a. Warm and cold stenotherms were identified and grouped based on optical observation and the Beta coefficient (from low to high) calculated based on the bootstrap weighted average partial least square (WA-PLS) model for each species in C2 software (Juggins, 2005).

Fig 2b. Forty-one (41) non-rare chironomid species present in the short core (28 cm) from Tiancai Lake where the calibrated ^{210}Pb based age is on the y-axis and taxon abundance is in percentage.

Fig 3 CCA bip-lots of sample and species scores constrained to environmental variables that individually explain a significant ($p < 0.05$) proportion of the chironomid species data. (a) species and (b) sample scores constrained to seven significant environmental variables in the 100 lakes of southwestern China. The species codes are correspondent to the taxa names shown in Fig. 2a.

Fig 4 Performance of the weighted average models with inverse deshrinking (WAinv) and partial least square (WA-PLS) models using the 100 lakes calibration data sets: (a) WAinv bootstrap model (b) the second component of the WA-PLS bootstrap model. Diagrams on the left show the predicted versus observed mean July temperature (MJT) and diagrams on the right display residuals of the predicted versus observed mean July temperature. Note that both models have a tendency to over-predict temperatures from the cold end of the gradient and underestimate temperatures at the warm end. This is typical for the WA based models.

Fig 5 The age and depth model for ^{210}Pb dating results of the short core (28 cm) from Tiancai Lake. The concentration of ^{137}Cs (circle), excess ^{210}Pb (triangle) and the calibrated age (AD years) (square) were plotted against core sample depth, respectively.

Fig 6 (a) Chironomid-based mean July temperature (MJT) reconstruction results from Tiancai Lake based on two transfer function models: solid black line is the reconstruction based on the weighted average partial least square (WA-PLS) bootstrap model with two components, dashed black line is the reconstruction based on the weighted average with inverse deshrinking (WAinv) bootstrap model. Red solid line is the instrumental data from Lijiang weather station, corrected applying the lapse rate and solid grey line is the three-sample moving average of the dataset.

Reconstruction of diagnostic statistics for the 100 lake dataset where (b) displays the goodness-of-fit statistics of the fossil samples with MJT. Dashed lines are used to identify samples with ‘poor fit’ (> 95th percentile) and ‘very poor fit’ (> 90th percentile) with temperature (c) Nearest modern analogues for the fossil samples in the calibration data set, where dashed line is used to show fossil samples with ‘no good’ (5%) modern analogues. (d) Percentage of chironomid taxa in fossil samples that are rare in the modern calibration data set (Hill’s $N_2 < 2$). (e) Comparison between the chironomid-based transfer function reconstructed trends (represented by MJT anomalies) with the instrumental data from Lijiang weather station (in red solid line, with three-sample moving average). Black solid line represents the reconstruction based on the WA-PLS bootstrapped model with two components using 100 lakes calibration set.

TABLE LEGENDS

Table 1. List of all the 18 environmental and climate variables measured from 100 south-western Chinese lakes, with mean, minimum and maximum values.

Table 2. CCA summary of the seven significant variables ($p < 0.05$) including canonical co-efficients and t-values of the environmental variables with the ordination axes including 100 lakes and 85 non-rare species

Table 3. Partial Canonical Correspondence Analysis (pCCA) result with environmental variables that showed a significant correlation ($p < 0.05$) in CCAs with chironomid species data included based on the 100 lakes calibration training set. Depth, K^+, Cl^-, LOI and MJT (bold) maintained their significance ($p < 0.01$) after each step of the pCCAs.

Table 4. (a) Results of the transfer function output shows the performance of the weighted average model with inverse and classical deshrinking (WAinv, WAcla), weighted average partial least squares (WA-PLS) models for reconstructing mean July temperature using 100 lakes from south-western China and 85 non-rare chironomid species. The bold indicates the models that are tested for reconstructing the mean July temperatures from Tiancai Lake. (b) The t-Test (Two-Sample assuming unequal variances) performed on the RMSEP output values of the WA-PLS component 1 and component 2 shows that the result is significant at $p < 0.05$. This suggests there is a difference between the RMSEP of the two models. We therefore selected the second component of the WA-PLS because it produced a lower RMSEP value.

Figure 1.

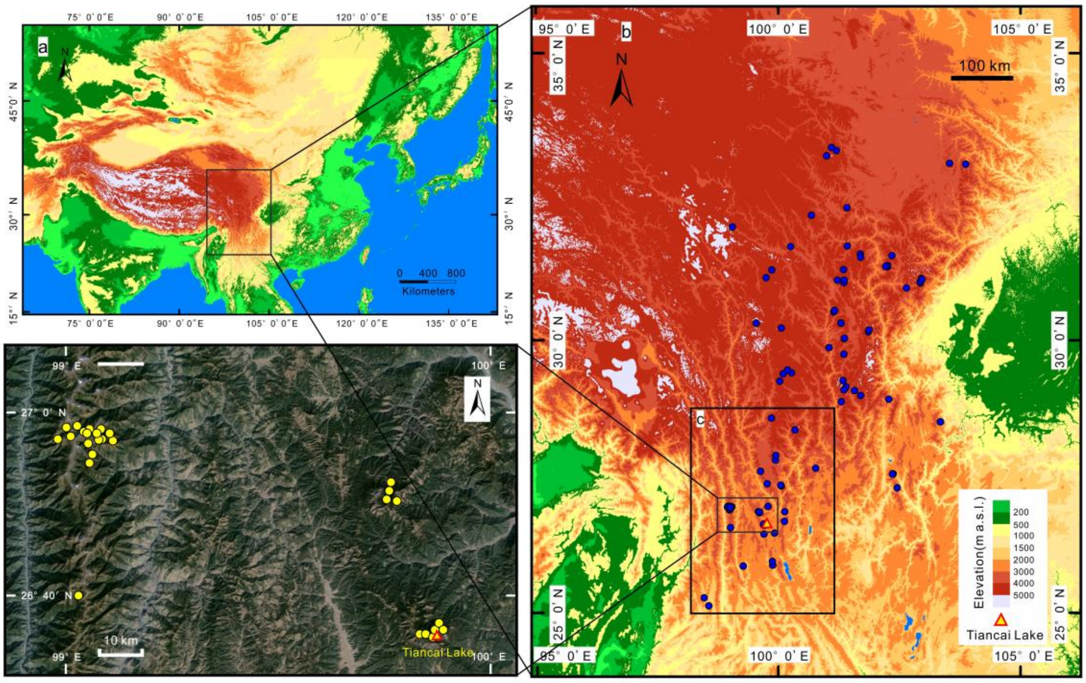

a
b
c
N
0 400 800
Kilometers
100 km
10 km
Elevation(m a.s.l.)
200
500
1000
1500
2000
3000
4000
5000
Tiancai Lake

Figure 2a

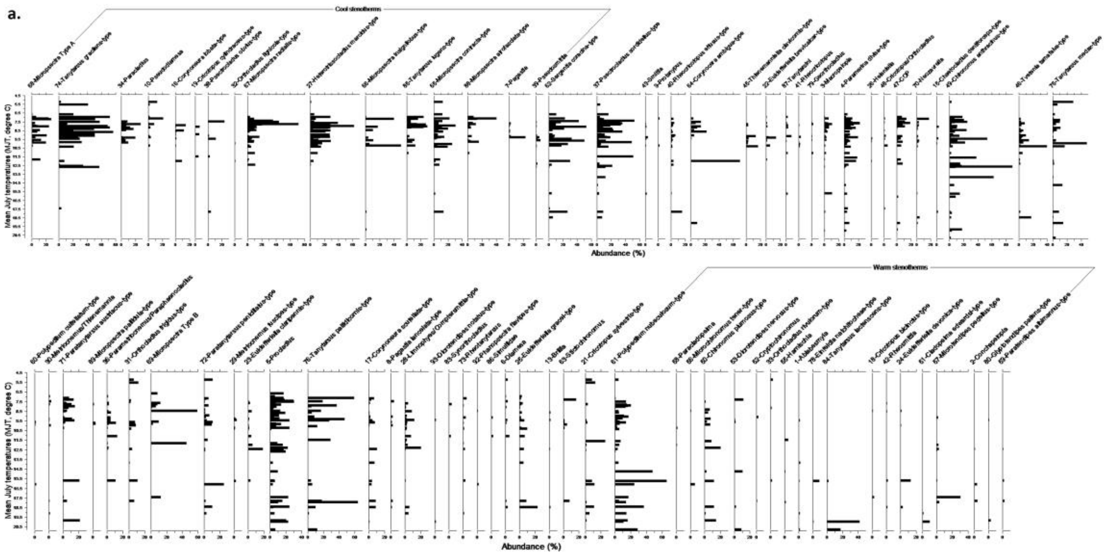


Figure 2b

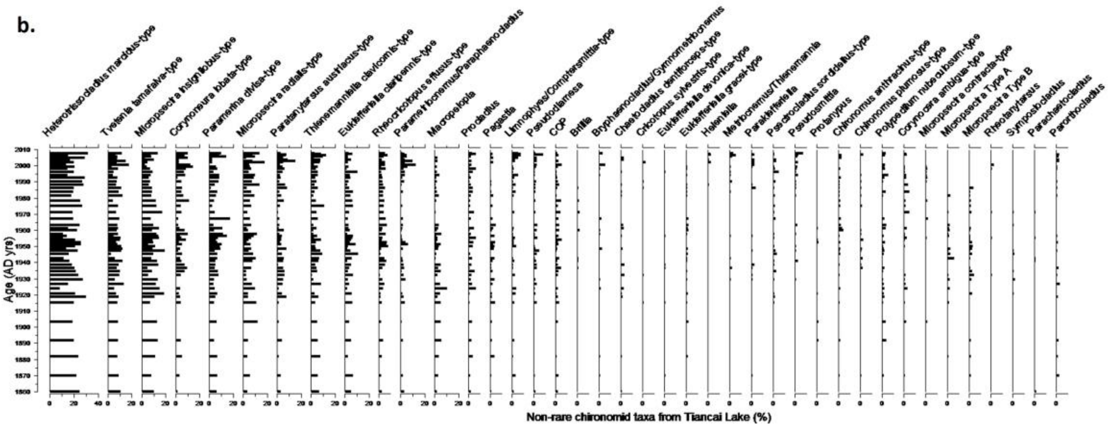


Figure 3

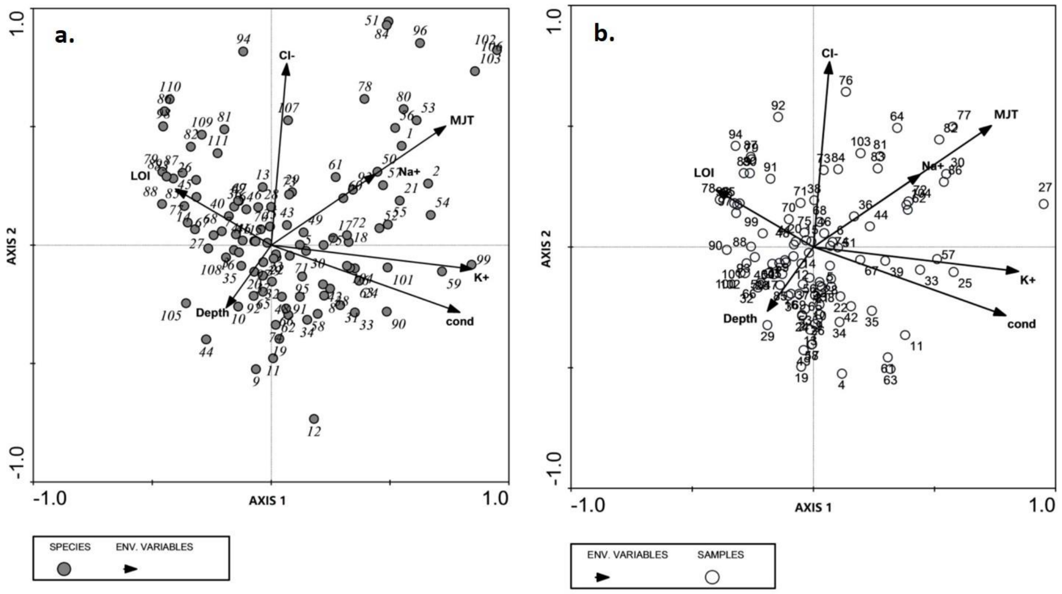

a.
b.
Cl-
MJT
Na+
K+
cond
LOI
Depth
AXIS 1
AXIS 2
1.0
-1.0
SPECIES
ENV. VARIABLES
SAMPLES

Figure 4

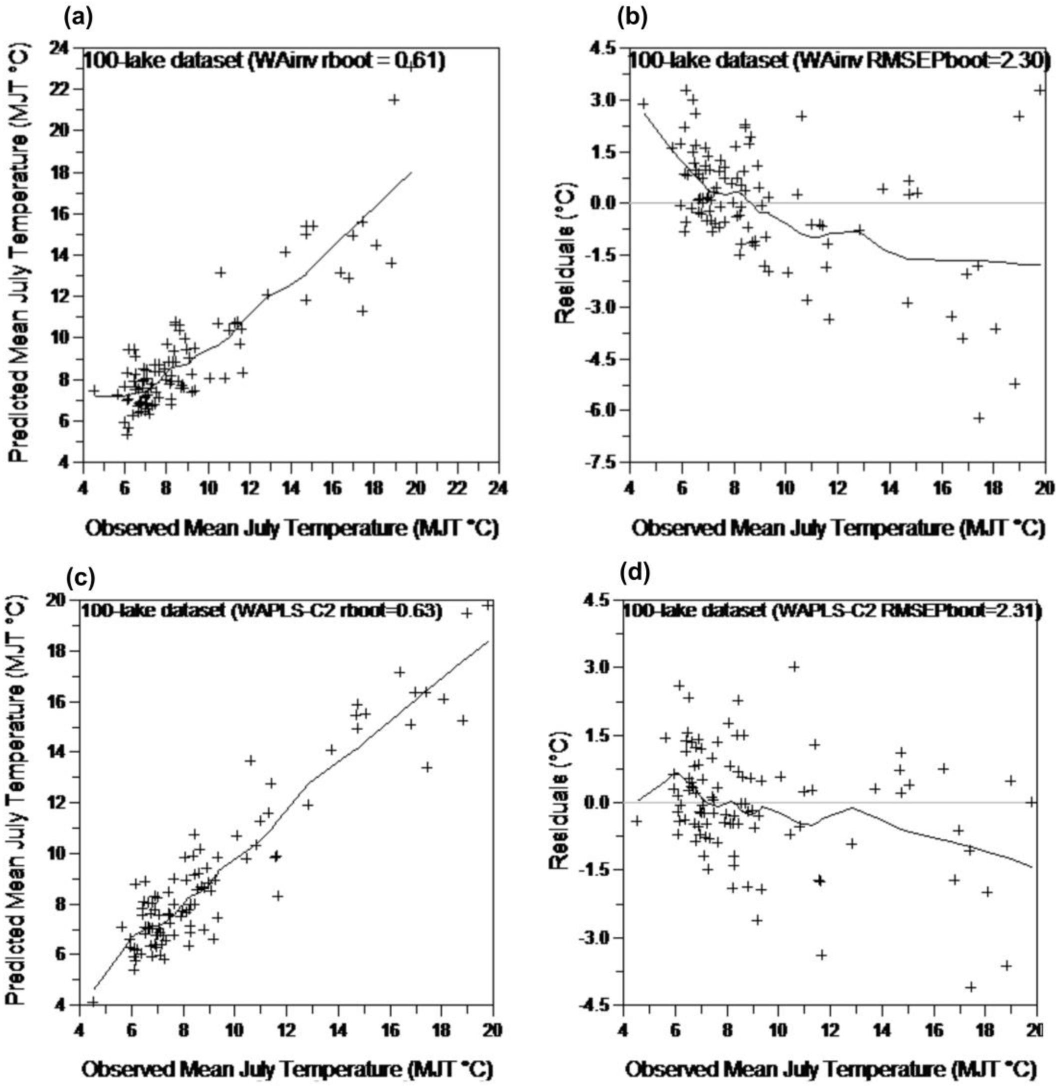
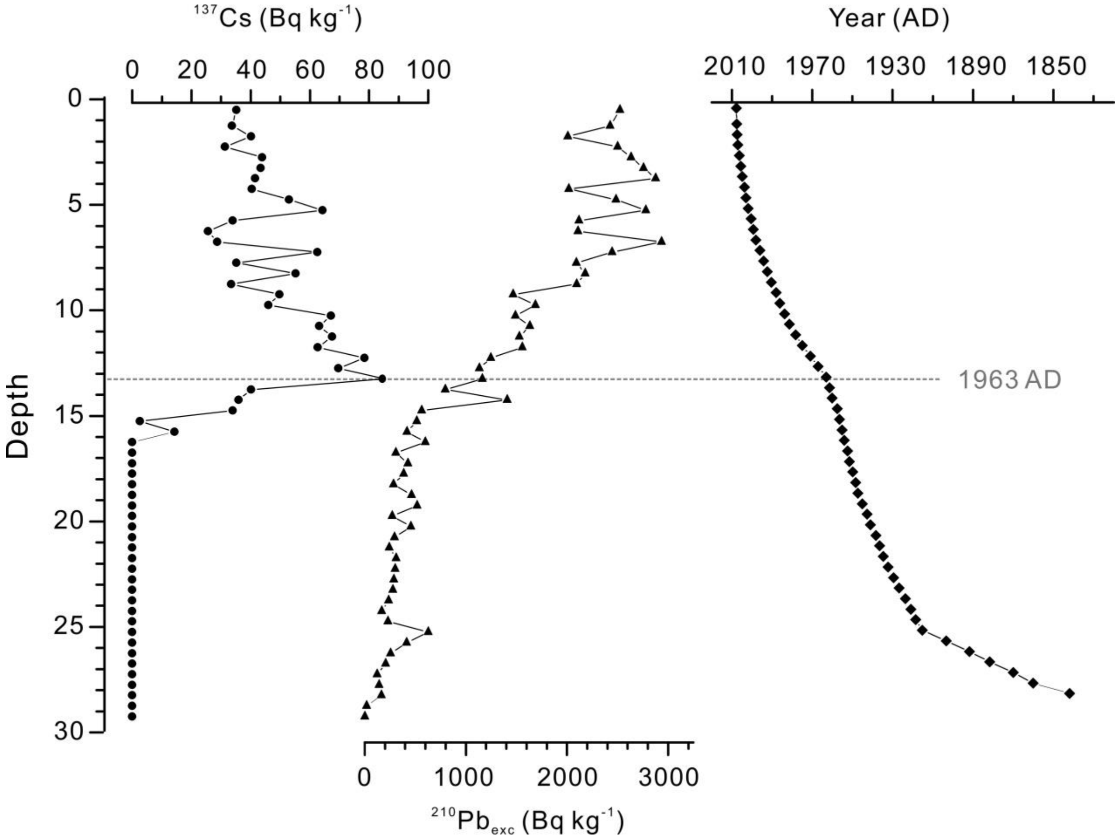

(a)
100-lake dataset (WAinv rboot = 0.61)
Predicted Mean July Temperature (MJT °C)
Observed Mean July Temperature (MJT °C)
(b)
100-lake dataset (WAinv RMSEPboot=2.30)
Residuals (°C)
Observed Mean July Temperature (MJT °C)
(c)
100-lake dataset (WAPLS-C2 rboot=0.63)
Predicted Mean July Temperature (MJT °C)
Observed Mean July Temperature (MJT °C)
(d)
100-lake dataset (WAPLS-C2 RMSEPboot=2.31)
Residuals (°C)
Observed Mean July Temperature (MJT °C)

Figure 5



^{137}Cs (Bq kg^{-1})
0
20
40
60
80
100
Year (AD)
2010
1970
1930
1890
1850
Depth
0
5
10
15
20
25
30
1963 AD
0
1000
2000
3000
^{210}Pb$_{exc}$ (Bq kg^{-1})

Figure 6

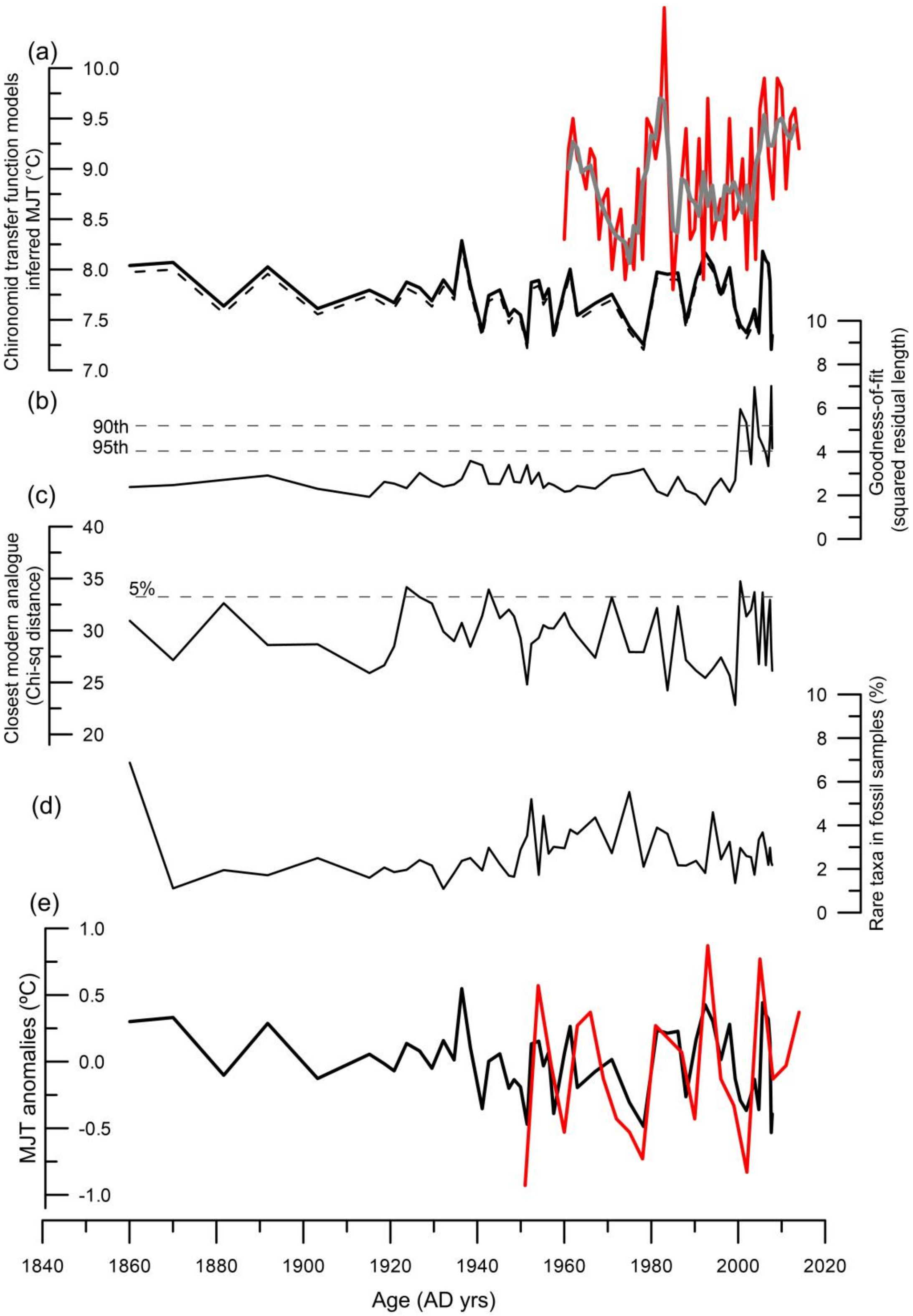

(a)
Chironomid transfer function models inferred MJT (°C)
10.0
9.5
9.0
8.5
8.0
7.5
7.0
(b)
90th
95th
Goodness-of-fit (squared residual length)
10
8
6
4
2
0
(c)
Closest modern analogue (Chi-sq distance)
40
35
5%
30
25
20
(d)
Rare taxa in fossil samples (%)
(e)
MJT anomalies (ºC)
1.0
0.5
0.0
-0.5
-1.0
1840
1860
1880
1900
1920
1940
1960
1980
2000
2020
Age (AD yrs)

Table 1

Variable	Unit	Symbol	Mean	Min	Max
Altitude	m	alt	3785	1769	4608
Mean July precipitation	mm	MJP	392	104	1055
Mean annual precipitation	mm	MAP	1820	505	5228
Mean July temperature	°C	MJT	9.1	4.2	19.8
Secchi depth	m	SD	3.5	0.2	12.5
Conductivity	µm cm^{-1}	Cond	55.8	5	336
Total dissolved solids	mg L^{-1}	TDS	18.4	1.9	79.7
pH	-	pH	8.5	7.23	10
Depth	m	Depth	10.7	0.25	52
Total Nitrogen	mg L^{-1}	TN	0.3	0.01	3.4
Total Phosphorus	mg L^{-1}	TP	0.05	0	1.6
Sodium	mg L^{-1}	Na+	2.7	0.22	37.2
Potassium	mg L^{-1}	K+	0.5	0	4.5
Magnesium	mg L^{-1}	Mg^{2+}	2.2	0	20
Calcium	mg L^{-1}	Ca^{2+}	7.3	0.8	34.6
Chlorine	mg L^{-1}	Cl-	1.7	0	9
Sulfate	mg L^{-1}	SO42-	3.9	0.1	31.6
Loss-on-ignition	%	LOI	24.3	1.92	77.1

Table 2

		Axis 1	Axis 2	Axis 3	Axis 4
Eigenvalues		0.24	0.17	0.10	0.08
Cum % var. spp.		5.90	10.0	12.5	14.6
Cum% var. spp. - env. relation		33.5	57.0	71.2	82.7
Variable	**Total variance explained**	**Regression/canonical coefficeints**		**t-values of regression coefficients**	
	Axis 1	Axis 2		Axis 1	Axis 2
cond	4.4%	0.44	-0.27	3.99	-2.65
depth	2.0%	-0.15	-0.21	-1.90	-2.82
Na+	2.7%	0.10	0.02	0.91	-0.17
K+	4.8%	0.49	-0.07	4.67	-0.65
Cl-	3.4%	-0.21	0.65	-2.18	6.94
MJT	***4.4%***	***0.14***	***0.62***	***1.49***	***6.90***
LOI	3.1%	-0.09	0.04	-1.02	0.48

Table 3

Variable	**Covariable**	**% var. axis 1**	**% var. axis 2**	**p-value**	**λ1**	**λ2**	**λ1/λ2**
cond	none	4.40	7.90	0.001	0.179	0.317	0.560
	depth	4.60	7.90	0.001	0.181	0.315	0.570
	Na+	4.10	7.70	0.001	0.159	0.305	0.520
	K+	1.80	8.20	0.004	0.069	0.316	0.220
	Cl-	4.60	7.50	0.001	0.179	0.293	0.610
	MJT	3.60	8.10	0.001	0.140	0.313	0.450
	LOI	3.60	7.90	0.001	0.140	0.310	0.450
	ALL	1.70	7.60	0.016	0.057	0.259	0.220
depth	**none**	**2.00**	**9.80**	**0.001**	**0.082**	**0.397**	**0.210**
	cond	**2.20**	**8.10**	**0.002**	**0.083**	**0.315**	**0.260**
	Na+	**2.10**	**9.90**	**0.001**	**0.083**	**0.387**	**0.210**
	K+	**2.20**	**8.30**	**0.001**	**0.083**	**0.321**	**0.260**
	Cl-	**2.00**	**10.0**	**0.002**	**0.079**	**0.390**	**0.200**
	MJT	**2.00**	**9.60**	**0.001**	**0.077**	**0.371**	**0.210**
	LOI	**2.10**	**9.50**	**0.001**	**0.082**	**0.372**	**0.220**
	ALL	**2.20**	**7.60**	**0.001**	**0.074**	**0.259**	**0.290**
Na+	none	2.70	9.60	0.001	0.111	0.388	0.290
	Cond	2.40	7.80	0.001	0.091	0.305	0.300
	depth	2.80	9.80	0.001	0.112	0.387	0.290
	K+	2.30	7.70	0.001	0.089	0.296	0.300
	Cl-	2.70	8.90	0.001	0.106	0.347	0.310
	MJT	1.90	9.60	0.008	0.072	0.371	0.190
	LOI	2.40	9.60	0.001	0.093	0.375	0.250
	ALL	1.70	7.70	0.011	0.058	0.259	0.220
K+	**none**	**4.80**	**7.90**	**0.001**	**0.192**	**0.322**	**0.600**
	cond	**2.10**	**8.20**	**0.002**	**0.082**	**0.316**	**0.260**
	Na+	**4.30**	**7.60**	**0.001**	**0.171**	**0.296**	**0.580**
	Cl-	**5.00**	**7.40**	**0.001**	**0.195**	**0.290**	**0.670**
	LOI	**4.10**	**8.20**	**0.001**	**0.160**	**0.320**	**0.500**
	Depth	**4.90**	**8.10**	**0.001**	**0.193**	**0.321**	**0.600**
	MJT	**3.30**	**8.20**	**0.001**	**0.129**	**0.314**	**0.410**
	ALL	**2.00**	**7.70**	**0.003**	**0.069**	**0.259**	**0.270**
Cl-	**none**	**3.40**	**9.70**	**0.001**	**0.137**	**0.393**	**0.350**
	cond	**3.50**	**7.60**	**0.001**	**0.137**	**0.293**	**0.470**
	K+	**3.60**	**7.60**	**0.001**	**0.140**	**0.290**	**0.480**

	MJT	**3.20**	**8.60**	**0.001**	**0.125**	**0.332**	**0.380**
	LOI	**3.50**	**9.40**	**0.001**	**0.137**	**0.366**	**0.370**
	Depth	**3.40**	**9.90**	**0.001**	**0.134**	**0.390**	**0.340**
	Na+	**3.40**	**8.80**	**0.001**	**0.132**	**0.347**	**0.380**
	ALL	**2.80**	**7.60**	**0.001**	**0.098**	**0.259**	**0.380**
LOI	**none**	**3.10**	**9.30**	**0.001**	**0.124**	**0.377**	**0.330**
	Na+	**2.70**	**9.60**	**0.001**	**0.107**	**0.375**	**0.290**
	cond	**2.20**	**8.00**	**0.001**	**0.086**	**0.310**	**0.280**
	K+	**2.40**	**8.30**	**0.001**	**0.092**	**0.320**	**0.290**
	MJT	**3.00**	**9.30**	**0.001**	**0.116**	**0.361**	**0.320**
	Cl-	**3.20**	**9.40**	**0.001**	**0.124**	**0.366**	**0.340**
	Depth	**3.10**	**9.40**	**0.001**	**0.124**	**0.372**	**0.330**
	ALL	**2.20**	**7.60**	**0.001**	**0.074**	**0.259**	**0.290**
MJT	**none**	**4.40**	**9.10**	**0.001**	**0.176**	**0.371**	**0.470**
	Na+	**3.50**	**9.40**	**0.001**	**0.137**	**0.371**	**0.370**
	cond	**3.50**	**8.10**	**0.001**	**0.137**	**0.313**	**0.440**
	K+	**2.90**	**8.20**	**0.001**	**0.113**	**0.314**	**0.360**
	LOI	**4.30**	**9.20**	**0.001**	**0.168**	**0.361**	**0.470**
	Cl-	**4.20**	**8.50**	**0.001**	**0.164**	**0.332**	**0.490**
	Depth	**4.30**	**9.40**	**0.001**	**0.171**	**0.371**	**0.460**
	ALL	**2.70**	**7.50**	**0.001**	**0.091**	**0.259**	**0.350**

Table 4

a.

#	**Model type**	**Bootstrap R2**	**Bootstrap Average Bias**	**Bootstrap Maximum Bias**	**RMSE_ s1**	**RMSE_ s2**	**RMSEP**
1	**WA_Inv**	**0.61**	**0.06**	**5.30**	**0.69**	**2.19**	**2.30**
2	WA_Cla	0.61	0.07	4.78	0.86	2.20	2.36
Component 1	WA-PLS	0.60	0.02	5.28	0.71	2.22	2.33
Component 2	**WA-PLS**	**0.63**	**0.10**	**5.16**	**0.89**	**2.14**	**2.31**
Component 3	WA-PLS	0.60	0.07	5.08	1.03	2.19	2.41

b.

t-Test: Two-Sample Assuming Unequal Variances	***RMSEP of WA-PLS_C1***	***RMSEP of WA-PLS_C2***
Mean	0.0645	-0.0524
Variance	2.8822	1.5186
Observations	100	100
Hypothesized Mean Difference	0	
df	181	
t Stat	0.5570	
P(T<=t) one-tail	0.2891	
t Critical one-tail	2.3471	
P(T<=t) two-tail	0.5782	
t Critical two-tail	2.6033	
	The P-Value is 0.01 < 0.05	**Reject null hypothesis**
The RMSEPs of WA-PLS C2 and WA-PLS C1 are different		