Review and test of reproducibility of sub-decadal resolution palaeoenvironmental reconstructions from microfossil assemblages

Richard J. Telford*,[a]

[a] *Department of Biological Sciences, University of Bergen and Bjerknes Centre for Climate Research, Postboks 7803, N-5020 Bergen, Norway*

Abstract

Very high, even annual, resolution quantitative palaeoenvironmental reconstructions have been generated from microfossil assemblage data using transfer functions. Evidence of the utility of some of these reconstructions is given by the high correlations between the reconstructions and instrumental records. Such performance is despite several ecological and taphonomic issues that can become more problematic at sub-decadal resolution than at multidecadal resolution. This paper explores these timescale dependent issues and tests the reproducibility of the reconstructions. It demonstrates that the exceptionally good performance of several published reconstructions is not reproducible.

Highlights

- Microfossil assemblage-based palaeoenvironmental reconstructions with sub-decadal resolution that are validated against instrumental records are reviewed.
- Ecological and taphonomic processes and chronological uncertainty mean that only in favourable circumstances will precise reconstructions be possible.
- Conditions are favourable for reconstructing recent changes in lake nutrient status with diatoms.
- Conditions are not favourable for sub-decadal chironomid-temperature reconstructions.
- Serious problems with several of the sub-decadal chironomid-temperature reconstructions are discussed.
- Chrysophyte-based reconstructions are difficult to evaluate due to a dearth of archived data.

Keywords: Microfossils; Quantitative palaeoenvironmental reconstructions; Transfer functions; Reproducibility; Anthropocene; Palaeoclimatology; Palaeolimnology; Global; Data treatment, data analysis

*Corresponding Author
Email address: richard.telford@uib.no (Richard J. Telford)

1. Introduction

Palaeoecologists, amongst others, value high-resolution palaeoecological data (PAGES2k Consortium, 2013; Perga et al., 2015; von Gunten et al., 2012). Of the 17 very high resolution palaeolimnological time series in the PAGES2k Consortium's (2017) compilation of temperature-sensitive proxy records, most are the thickness or other physical property of varves, which can be difficult to interpret (Blass et al., 2007), but two series are derived from microfossil assemblages using transfer functions. Evidence for the utility of these high, even annual, resolution microfossil-based palaeoenvironmental reconstructions comes from comparisons of these reconstructions, typically from varved sediments, with instrumental records. As such records require large investments of time (Green, 1983), it is important to understand the circumstances in which they are likely to yield good results. All validated sub-decadal microfossil-assemblage based reconstructions identified by a literature search are listed in Table 1. Lotter (1998) and Alefs and Müller (1999) reconstruct the slowly varying total phosphorus concentrations from annually-resolved, mainly planktic, lacustrine diatom assemblages. High-resolution chrysophyte stomatocyst stratigraphies have been used to reconstruct winter conditions as reflected by the date of spring mixing (de Jong and Kamenik, 2011) and the duration of cold water (Hernández-Almeida et al., 2015a). Since there are no long instrumental time series of these limnological variables, the reconstructions are correlated with winter air temperatures. Chrysophyte stomatocysts have also been used to reconstruct summer calcium concentrations (Hernández-Almeida et al., 2015b), which might reflect summer mixing and so were correlated with summer wind. High-resolution chironomid head capsule stratigraphies have been used to reconstruct summer air (e.g., Larocque et al., 2009a; Luoto and Ojala, 2016; Zhang et al., 2017a) and water (Medeiros et al., 2012) temperature. High-resolution pollen records have been used to reconstruct summer air temperature (Kamenik et al., 2009).

Table 1: Location, proxy, resolution, and correlation with the instrumental record of the validated reconstructions derived from microfossil stratigraphies. Where two numbers are given for the correlation, the first is for a calibration-in-space model, the second a calibration-in-time model.

Study	Proxy	Variable	Site	Resolution	Instrumental correlation
Lotter (1998)	Diatoms	Total phosphorus	Baldegersee	Annual	“compares well”*
Alefs and Müller (1999)	Diatoms	Total phosphorus	Ammersee	Annual	0.91*
			Starnberger See	Annual	0.77*
De Jong and Kamenik (2011)	Chrysophytes	Date of spring mixing	Lake Silvaplana	Annual (3 year smooth)	-0.58 vs Oct.–May air T*
Hernández-Almeida et al. (2015b)	Chrysophytes	Calcium concentration	Lake Żabińskie	Annual (3 year smooth)	0.5 vs May–Oct. zonal wind velocity*
Hernández-Almeida et al. (2015a)	Chrysophytes	No. consecutive days with water < 4°C	Lake Żabińskie	Annual (3 year smooth)	0.35 vs Jan.–March air T*
Larocque and Hall (2003)	Chironomids	July air-T	Lake Njulla	2–7 years	0.39
			Lake 850	2–6 years	0.365
			Alanen Laanijavri	1–5 years	0.35
			Vuoskkujavri	1–6 years	0.37

Study	Proxy	Variable	Site	Resolution	Instrumental correlation
Larocque et al. (2009a), Larocque-Tobler et al. (2011a)	Chironomids	July air-T	Lake Silvaplana	Near annual	0.65/0.53*
Larocque-Tobler et al. (2011b)	Chironomids	July air-T	Seebergsee	3–8 years	0.64/0.73
Larocque-Tobler et al. (2016a, 2015)	Chironomids	August air-T	Lake Żabińskie	1–5 years	0.76*
Medeiros et al. (2012)	Chironomids	Mid-summer water T	Baker Lake	1–8 years	“corresponded well” vs annual air T
Luoto and Ojala (2016)	Chironomids	July air-T	Nurmijärvi	1–11 years	0.38/0.51*
Zhang et al. (2017a)	Chironomids	July air-T	Tiancai Lake	1–4 years	0.45
Lang et al. (2017)	Chironomids	July air-T	Speke Hall Lake	~8 years	0.62
Kamenik et al. (2009)	Pollen	April-November air T	Mauntschas Mire	1–5 years	–/0.66

* - lakes with varve chronologies. T - temperature

Several of the correlations shown in Table 1 are surprisingly good given the ecological sensitivity of the proxy, taphonomic issues, and the inherent chronological uncertainty that occurs even in the best-dated palaeoecological records. Some are comparable with the best tree ring-width–climate correlations where the 99th percentile of statistically significant ($p < 0.05$) correlations is 0.7 for winter precipitation and 0.63 for summer temperature (St. George, 2014). Although the tree ring-width–climate relationship has its complexities (D'Arrigo et al., 2008), it is much simpler than the multi-species relationship between microfossil assemblages and climate, usually has no chronological error, and no taphonomic complications.

Given that the exceptional performances shown in Table 1 are used to justify the development of millennium-long temperature reconstructions (Hernández-Almeida et al., 2017; Larocque-Tobler et al., 2010, 2012), which have been used in regional (Trachsel et al., 2012) climate compilations, and could be used in hemispheric or global compilations with policy relevance (Masson-Delmotte et al., 2013), it is critical to ascertain that the results are reproducible. Pre-publication peer-review is not a sufficient guarantee of reproducibility, not least because the raw data are typically not made available to reviewers. Recent replication studies in other disciplines have shown low levels of reproducibility (Fidler et al., 2017). Full direct replication of palaeoecological studies would be costly and time consuming, but re-analysis of archived data is an important alternative for testing for reproducibility (Fidler et al., 2017). Fidler et al. (2017) argue for a systematic evaluation of the reproducibility of the evidence base in ecology. Papers with surprisingly good performance are an obvious target for evaluation.

Chironomid studies are the most numerous of the studies in table 1, are from diverse settings, and several have exceptional performance, so this paper includes an attempt to reproduce the chironomid-inferred reconstructions. Because the best correlation between chironomid-inferred temperature and the instrumental record is from Lake Żabińskie (Larocque-Tobler et al., 2015), this study is used to exemplify the issues with high-resolution palaeoenvironmental reconstructions. Interpretation of this reconstruction of August air temperature since 1896 at ~annual resolution, calculated using a Polish-Canadian chironomid calibration set, is helped by information from several other studies on the same varved lake (Amann et al., 2014; Bonk et al., 2016, 2015a, 2015b; Tylmann et al., 2016; Witak et al., 2017).

2. Challenges for high-resolution reconstructions

This section explores the ecological, chronological, and taphonomic challenges faced by sub-decadal resolution reconstructions, with a focus on chironomid inferred-air temperature reconstructions, and considering other reconstructions when they differ in important ways. Many of these challenges discussed below will also affect sedimentary geochemical proxies to some extent.

2.1. Ecological sensitivity

With the exception of the diatom-total phosphorus reconstructions, all the work cited in Table 1 attempts to reconstruct climate for part of a year, most often air temperature from a single month in summer. It is possible to conceive of a biological proxy that is sensitive to, and only to, a single month's temperature. Such a proxy would need to have a short generation time, so it can rapidly respond to temperatures in that month, and to live mainly within that month and be dormant throughout the remainder of the year. It is not clear that any proxy meets these ideal standards, but bloom-forming algae such as chrysophytes might, especially in lakes with a short ice-free period. Chironomids would seem to fall far short of these ideals, with the larvae taking months, or even years, to mature (Tokeshi, 1995).

Larocque-Tobler et al. (2015) reconstruct August air temperature because this variable explains the most variance in a Canadian calibration set (Larocque et al., 2006); several other papers reconstruct July air temperature (Lang et al., 2017; Larocque and Hall, 2003; Luoto and Ojala, 2016). While it is widely believed (Eggermont and Heiri, 2011) that summer air temperature is a key variable affecting chironomid assemblages, there is no particular reason to suspect July or August temperatures to be uniquely important. August is not, for example, the warmest month in Lake Żabińskie (Bonk et al., 2015a). The phenology of chironomid emergence (Tokeshi, 1995) suggests that many chironomids will have emerged before August, precluding August temperatures of their assigned year from having any influence, a problem recognised by Larocque-Tobler et al. (2012). At Polish latitudes, many of the chironomids in shallow water are likely to be bivoltine (Tokeshi, 1995) with two generations a year. The first generation of adults, along with some univoltine species from deeper water, will emerge in spring (Tokeshi, 1995) and so might be influenced by the previous year's August temperatures, introducing a lagged response. The second generation, emerging in late summer/autumn (Tokeshi, 1995), will be affected by August temperatures but also temperatures throughout the summer and by the original species composition of the eggs. Lags can also be expected for sub-decadal pollen-climate reconstructions: van der Knaap et al. (2010) used annual pollen trap data to show that the pollen accumulation rate for many taxa tend to depend on the conditions of the previous summer.

With a few exceptions (Kamenik et al., 2009; Larocque-Tobler et al., 2011a, 2011b; Luoto and Ojala, 2016), transfer-function models for quantitative palaeoclimate reconstructions are calibrated on a spatial calibration set of sites with paired microfossil and climatological normals that span a wide climatic range (Birks et al., 2010). Over a spatial calibration set, the correlation between air temperature in different months can be expected to be high. For example, in the Polish-Canadian chironomid calibration set, the Pearson correlation between June and August mean air temperatures (Fick and Hijmans, 2017) is 0.97 (Fig. 1). Hence, it makes little difference if we calibrate the model against June, August or mean summer temperatures, as the model performance statistics will be very similar (Fig. 2a-b) and the reconstructions will all have a similar shape (Fig. 3). Any of these months can be reconstructed, provided the assumption that the relationship between monthly temperatures is invariant across time (Juggins, 2013). It makes sense to try to reconstruct temperature for the most ecologically important time of year, but the calibration set may give little information on when this is.

While the calibration set is calibrated against climate normals, the target for validating the reconstruction is a time series of weather data. At an annual resolution, the correlations between months in the time series are low (Fig. 1): a warm August does not necessarily follow a warm July. Consequently, the choice of calibration target has a large impact on the correlation between the reconstruction and the instrumental data. If a perfect reconstruction is validated against the wrong month, the performance will seem poor; a good correlation can only be expected for the portion of the year that the organisms are sensitive to. Surprisingly, despite the presumed sensitivity of chironomids to summer temperatures, the air-temperature reconstruction from the Lake Żabińskie chironomid stratigraphy is high only in August (Fig. 2c). The correlation between June and August air temperature in a composite of climate stations in the Lake Żabińskie region (similar to the composite used in Larocque-Tobler et al. (2015)) in the period 1896-2010, is only 0.17. The correlation between mean summer air temperature and August air temperature in the same composite is 0.74. This is as good a correlation as can be expected for predicting August temperature from mean summer data collected with a thermometer. This problem is time-scale dependent: if temperature is averaged over several decades, as would be typical for a low- to medium-resolution palaeoecological analysis, the correlation between June and August temperatures, or August and the whole summer, increases because radiative forcing or long-term internal variability such as the Atlantic Multidecadal Oscillation (Schlesinger and Ramankutty, 1994) tends to warm the whole season and weather-related variability is suppressed (Fig. 1).

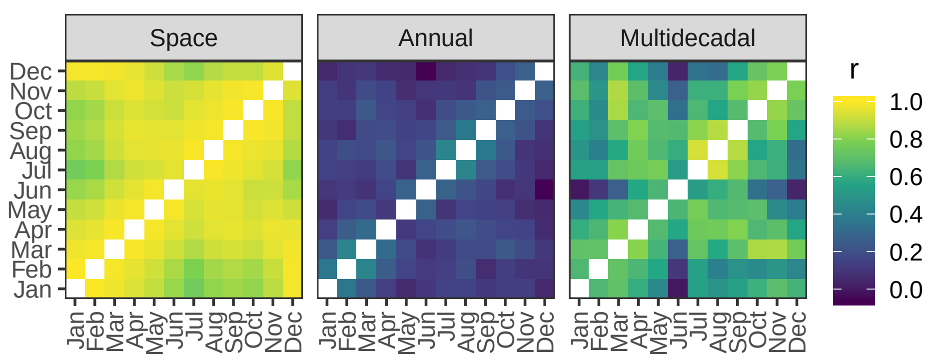

Figure 1: Correlation between mean monthly temperatures in the Canadian-Polish calibration set (left), in single years (centre) and 30-year means from the central England temperature series (Parker et al., 1992).

Chironomids are probably more sensitive to water temperature than air temperature directly (Eggermont and Heiri, 2011); transfer-function models

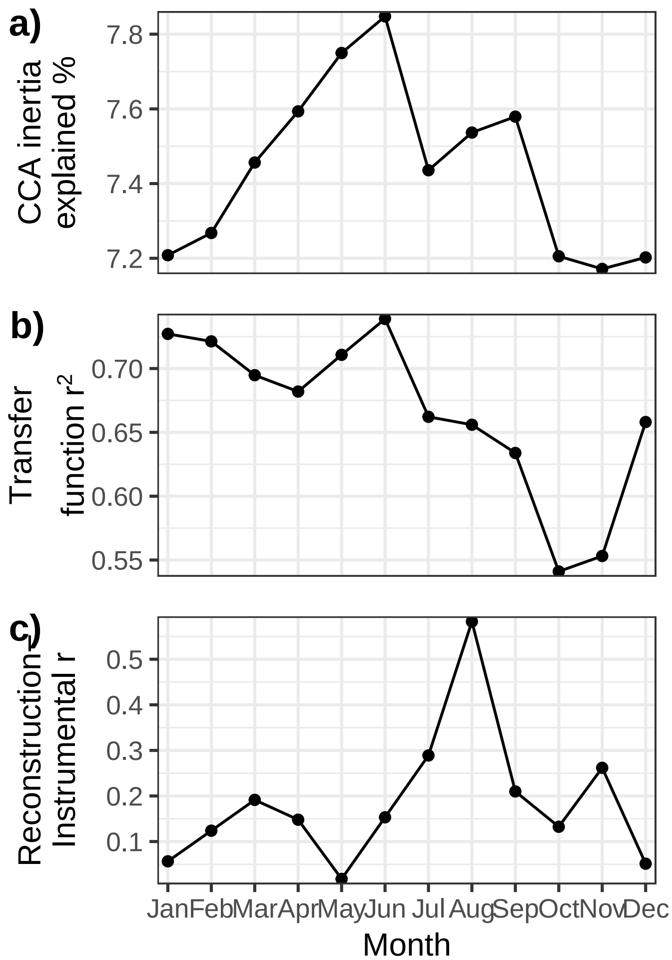

Figure 2: By month, a) the percent of intertia in the Canadian-Polish calibration set explained in a cannonical correspondence analysis (CCA) constrained by temperature, b) transfer function leave-one-out cross-validation r^2 calibrated against temperature, c) correlation between the reconstructed temperature from Lake Żabińskie and the instrumental temperature

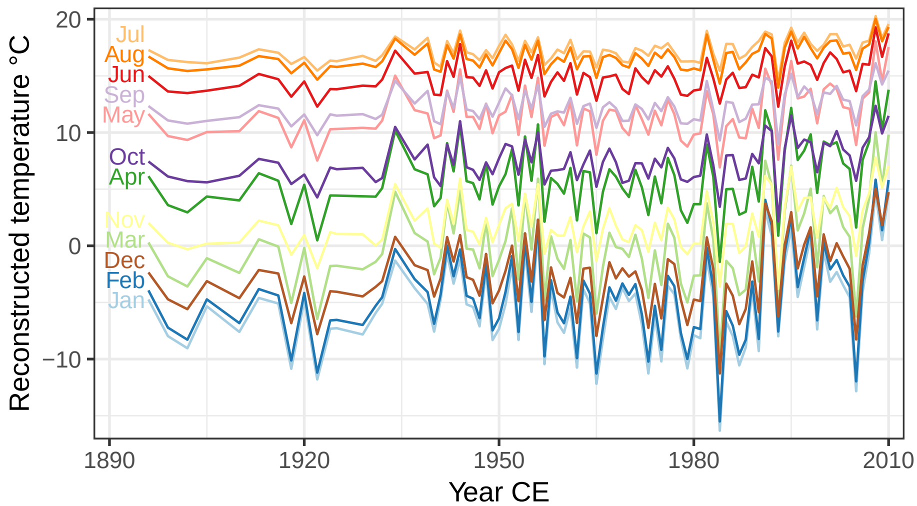

Figure 3: Temperature reconstructions for different months at Lake Żabińskie

reconstructing air temperature depend on the strong relationship between air and water temperature. Air temperature is the most important determinant of water temperature (O'Reilly et al., 2015) but other factors such as wind speed and cloudiness also have an influence, so the inter-year correlation between air and water temperature is less than one. In Lake Zurich, this correlation is 0.67 in July (Livingstone and Lotter, 1998), varies between 0.59 and 0.76 for mean summer temperatures in eight Austrian lakes (Livingstone and Dokulil, 2001), and is about 0.75 throughout the summer in the small Plußsee in northern Germany (Rösner et al., 2012). An analysis of summer lake and air temperatures compiled by Sharma et al. (2015) shows that for small (area $< 2\text{km}^2$), deep (mean depth > 5m), mid- to high-latitude lakes, the median correlation is 0.8 ($n = 41$). This correlation is as good as should be expected when comparing a thermometer in the lake with the instrumental air temperature.

While temperature undoubtedly has a strong effect on chironomid physiology, growth, and community structure, at least some of the apparent effect of temperature on chironomid assemblages in the calibration set is due to indirect effects of temperature (Eggermont and Heiri, 2011) acting through, for example, the relationship between temperature and productivity (Velle et al., 2010) or dissolved organic carbon (Larocque et al., 2001). These indirect effects will tend to strengthen the relationship between chironomid assemblages and temperature in the calibration set, but cannot be expected to fully operate on a monthly basis. As such, the cross-validation performance of the transfer function will tend to overestimate the model's predictive power on short time-scales.

Conversely, ecologically important variables other than the variable being reconstructed may have changed over time, which will bias the reconstructions (Juggins, 2013). For example, nutrient status (Amann et al., 2014; Witak et al.,

2017) and anoxia (Bonk et al., 2015a), both important predictors of chironomid assemblages (Eggermont and Heiri, 2011), changed independently of August temperatures at Lake Żabińskie, and the naturally oligotrophic Lake Silvaplana became eutrophic due to human impact before recovering (Bigler et al., 2007).

2.2. Taphonomic processes

A strong correlation between an annual-resolution reconstruction and instrumental data can only be achieved if most of the microfossils in the sediment attributed to a given year, lived in that year. This could be achieved if the organisms lived at the core site (which might prevent the preservation of varves), or if direct deposition of planktic organisms dominated the record.

Varved sediments can capture the seasonal succession of diatom communities (Bonk et al., 2015b; Simola, 1977), although seasonal mixing can transport more diatoms to the centre of the lake (Raubitschek et al., 1999). This suggests that the largely planktic diatom record from the varved Baldeggersee (Lotter, 1998), marked by abrupt changes in the assemblages, might be an example of the more promising case of direct deposition of much of the assemblage.

By analogy with Lake De Waay (van Hardenbroek et al., 2011), the seasonally anoxic hypolimnion (Bonk et al., 2015a) of Lake Żabińskie will be inhabited by few chironomids (perhaps *Chironomus plumosus* which Nagell (1978) showed can withstand long periods of anoxia, but without feeding), and the chironomid assemblage in the centre of the lake will have been transported and redeposited from the oxygenated shallows by wave-induced turbulence or currents during seasonal mixing (Eggermont and Heiri, 2011). Lake Żabińskie experiences mixing in spring and late autumn (Bonk et al., 2015a): complete mixing results in a large sediment flux to the centre of the lake (Bonk et al., 2015a), but this does not happen every year.

The importance of intense sediment focusing in Lake Żabińskie is evident in the high ^{137}Cs and ^{210}Pb inventories (Tylmann et al., 2016). The diffused ^{137}Cs peaks (Tylmann et al., 2016) indicate that not all the sediment in one varve was generated in that year. The five year spread in the Chernobyl ^{137}Cs peak might give an indication of the degree of smoothing that can be expected in redeposited proxies. This is only a problem for very high resolution analyses.

The importance of sediment reworking in transporting chironomids to the centre of Lake Żabińskie is demonstrated by the high proportion (73% on average) of littoral taxa in the samples analysed (Larocque-Tobler et al., 2015). In Lake Silvaplana, an average of 50% of the assemblage is littoral taxa (Larocque et al., 2009a).

In lakes with oxic hypolimnia, there are other taphonomic and ecological processes that will degrade the proxy record. Bioturbation and other processes will mix the top few centimetres of sediment, acting as a low-pass filter. The smooth changes in chironomid assemblages at Baker Lake (Medeiros et al., 2012) may indicate the importance of such processes in this large deep lake. Stratigraphy can potentially be further compromised as chironomids can burrow several centimetres into the sediment (Charbonneau and Hare, 1998). Chironomids living in the hypolimion are isolated from summer temperature by the thermocline (Livingstone and Lotter, 1998), and cannot be expected to respond directly to it.

2.3. Chronological precision

Reliable chronologies are essential for any proxy-proxy or proxy-instrumental data comparison (Trachsel and Telford, 2017). The impact of chronological error will depend on the magnitude of the chronological error relative to the persistence in the environmental variable being reconstructed. In cases such as the diatom reconstructions of total phosphorus (Alefs and Müller, 1999; Lotter, 1998), where the target environmental variable rises and declines over decades, a small chronological error will have little effect on the agreement between the reconstruction and the instrumental records. If, however, the target variable has little or no autocorrelation, as for example with the Lake Żabińskie regional August air-temperature composite which has a lag-1 autocorrelation coefficient of 0.16, then even an error of a single year would seriously degrade the agreement.

The chronologies supporting the sub-decadal resolution reconstructions listed in Table 1 are either ^{210}Pb or varve based, with the exception of Kamenik et al. (2009) who use bomb-^{14}C to date the peat sequence from Mauntschas Mire and report a dating uncertainty of 1–2 years. ^{210}Pb chronologies typically have an uncertainty of 5–10 years on sediment 100 years old (Appleby and Piliposian, 2006). Varved chronologies are usually more precise, with a typical uncertainty of 1–3% (Ojala et al., 2012) and allow the sediment from a single year (or known number of years) to be analysed. The error in varved chronologies tends to accumulate, becoming higher further back in time.

The impact of chronological uncertainty on the validation of an ~annual resolution reconstruction can be illustrated with Lake Żabińskie. The varve-based chronology for Lake Żabińskie is supported by ^{14}C dates, a ^{137}Cs profile, and a microtephra (Bonk et al., 2015b). The counting uncertainty on the chronology is low: there are two possible extra varves in the mid-1960s and five possible missing varves between 1926 and 1899 CE. The impact of this uncertainty can be explored with a Monte Carlo procedure. Assuming that the probability of making a wrong decision at each uncertain varve is 1/3 and zero otherwise, the probability that the chronology is entirely correct is 0.06. If the reconstruction was perfect, chronological error is expected to reduce the correlation from 1 to 0.8. This reduction is relatively small because most of the uncertainty occurs in the lower half of the record, so any error affects fewer years. Furthermore, because chironomid concentrations were low, samples before 1939 span two to five years which reduces both the number of samples and their sensitivity to a single year's error. The expected correlation in the pre-1939 section of the record for a perfect reconstruction is 0.69. This is lower than the reported correlation between the reconstruction and the instrumental record (0.81) for this period. Where the chronological uncertainty is higher than at Lake Żabińskie, for example at Lake Silvaplana where the uncertainty of the varve chronology is estimated to be 15% (Blass et al., 2007), the effect of chronological uncertainty will be greater.

There are some possible solutions to chronological uncertainty. Von Gunten et al. (2012) recommend smoothing the proxy data to allow for chronological errors. The three-year running-mean used by de Jong and Kamenik (2011) to smooth their annually resolved chrysophyte data will reduce the impact of a one year dating error, and slightly reduce the impact of a two year error, however the loss of independence between consecutive years complicates the calculation of the significance of the correlation between the reconstruction and the instrumental data (von Gunten et al., 2012). Since the smoothed data have fewer degrees of freedom, they will have less statistical power if the chronology is correct. Smoothing annual resolution will generally have more statistical power than sampling the microfossil data at a lower resolution, but at the cost of having more samples to process and count. Either smoothing or sampling at a lower than annual resolution are valid strategies when the sediments are not laminated and single years cannot be precisely sampled.

An alternative strategy for dealing with chronological errors in varved sediment would be to identify which of an ensemble of possible chronologies are most plausible given the relationship between the proxy and the instrumental record (Werner and Tingley, 2015).

2.4. Calibration-in-time

Calibration-in-time models are routinely used in dendroclimatology (Sheppard, 2010). They have been used in palaeolimnology where calibration-in-space models are not appropriate because the models need tailoring to a specific site (von Gunten et al., 2012). For example, they have been used to calibrate varve thickness against climate, especially for reconstructing summer temperature from proglacial lakes (Leemann and Niessen, 1994; Thomas and Briner, 2009); mass accumulation rate against winter precipitation (Elbert et al., 2012); and reflectance spectroscopy measurements against temperature (Saunders et al., 2013; Trachsel et al., 2010; von Gunten et al., 2009).

Kamenik et al. (2009) introduced calibration-in-time to microfossil data, using it to reconstruct April-November temperatures from the pollen stratigraphy from Mauntschas Mire, Switzerland. Subsequently, calibration-in-time models have been used to reconstruct summer temperatures from chironomid stratigraphies (Larocque-Tobler et al., 2011a, 2011b; Luoto and Ojala, 2016).

Calibration-in-time models are particularly sensitive to chronological errors and ecological or taphonimic lags as these affect both the calibration set and the fossil data to which the transfer function is applied.

Calibration-in-time models need to be checked against all the usual transfer function and reconstruction diagnostics. So far, this aspect of model development has been neglected. Because of the limited range of environmental conditions in the calibration set, there is a risk of non-analogue conditions downcore. Temporal autocorrelation needs to be checked for as it may seriously bias transfer function performance statistics (Telford and Birks, 2009, 2005). The pollen stratigraphy in Kamenik et al. (2009) is strongly autocorrelated, partly because of the persistence of vegetation in the landscape, partly due to mixing of pollen in the acrotelm (Joosten and de Klerk, 2007), and partly due to linear interpolation and the three-year triangular filter used to smooth the pollen and climate data. Leave-one-out cross-validation cannot be expected to give credible performance statistics when adjacent samples have so little independence. With *h*-block cross-validation (Trachsel and Telford, 2017), a scheme that is robust to autocorrelation, or split sampling, the performance statistics would be worse, but more realistic.

3. Discussion

Some proxy-environment systems appear more promising for sub-decadal to annual resolution reconstructions than others. Some attributes of lakes, proxies and environmental variables associated with promising or unpromising systems are given in Table 2.

Table 2: Aspects of proxies, sites and environmental variables leading to favourable or unfavourable reconstructions, with a brief explanation

Aspect	Favourable	Unfavourable	Explanation
Proxy sensitivity to target variable	High	Low	Better reconstruction
Habitat	Planktic	Benthic	Simple taphonomy
Generation time	Hours-Days	Months-Years	Rapid response
Ecological lags	None	Proxy responds to previous year	Lagged proxy response
Target variable resolution	Seasonal	Monthly	More ecologically relevant
Other environmental change	Minimal	Substantial	Minimise secondary gradients
Dominant frequency in target	Trends and low frequencies	High-frequencies	Less sensitive to lags and chronological error
Sediment	Varved	Not-varved	No bioturbation, demarcate one year's sediment
Chronology	Varves	Other	More precise

The diatom-total phosphorus reconstructions (Alefs and Müller, 1999; Lotter, 1998) derived from planktic diatoms hold the most promise. The planktic diatoms have a short generation time; nutrient availability has a strong control on diatom communities (Tilman, 1981); the taphonomy of sinking diatoms is relatively simple; and the slow changes in the target environmental variable mean that small chronological errors are inconsequential (conversely, this might make it difficult to demonstrate skill on sub-decadal rather than on decadal scales).

The chrysophyte-cyst climate system shares several of the advantages of the diatoms, but is less promising because the target climate variable has strong high-frequency variability so any chronological error or taphonomic mixing will degrade the reconstruction. The three-year running mean used in all the chrysophyte studies will partially mitigate the high-frequency variability. The apparently successful reconstruction of summer wind speed (Hernández-Almeida et al., 2015b) and winter severity (Hernández-Almeida et al., 2015a) from the same stratigraphy is surprising given that these two climatic variables cannot be expected to be highly correlated. If both variables are ecologically important then the assumption (Birks et al., 2010) that secondary environmental variables are unchanging is violated and following Juggins (2013), good reconstructions cannot be expected. The mean summer wind reconstruction, relying on wind-induced upwelling of calcium to change species composition, seems particularly fragile, and is physically dubious as it seeks to find a relationship with mean wind velocity rather than mean wind speed. No data have been archived for either of these studies so their reproducibility cannot be verified.

Chironomids are even less promising for sub-decadal reconstructions. In addition to the high-frequency variability in the target climatic variable, a chironomid's life-span is long relative to the month-long target variable, the taphonomy is complex as most head capsules need to be transported from the littoral zone, and chironomid phenology means that many chironomids will have emerged before the summer so cannot be affected by summer temperatures in the year they emerge.

Doubts about the plausibility of sub-decadal resolution chironomid reconstructions do not imply that reconstructions on longer timescales (Brooks et al., 2012) are flawed (but see Velle et al. (2010)). At multi-decadal to centennial time scales, most of the issues discussed above are mitigated: single month and whole summer temperatures correlate well with each other as climate forcings rather than weather dominate; secondary variables such as dissolved organic carbon might correlate with climate change in time as they do in space and exert some influence on the assemblages; and lags due to taphonomic processes and phenology become largely irrelevant. It is not yet clear at what time scale the signal in the chironomid-inferred reconstruction can be expected to emerge from the noise.

While sub-decadal resolution reconstructions from microfossil assemblages can only be expected to correlate well with instrumental data only in favourable circumstances, very high-resolution palaeoecological analyses from varved sediments can be extremely valuable (Lotter and Anderson, 2012). For example, Allison et al. (1986) and Peglar (1993) use sub-decadally (even annually) resolved pollen data to show the mid-Holocene *Tsuga* (Hemlock) and *Ulmus* (Elm) declines in North America and Europe, respectively, took just a few years.

3.1. Reproducibility issues with high-resolution chironomid reconstructions

Given the high correlations between chironomid-inferred temperatures and air temperature shown in Table 1 despite the severe problems such high-resolution reconstructions face, it is necessary to examine each reconstruction to try to determine the reasons for these apparently good performances. Below, I show there are multiple problems with several of these reconstructions or their validation.

3.1.1. Lake Żabińskie

One might expect that a reconstruction as excellent as the Lake Żabińskie chironomid-temperature reconstruction would have excellent reconstruction diagnostics (Juggins and Birks, 2012). Indeed, the Żabińskie paper reports that all the fossil samples had good or fair analogues in the modern calibration set. However, using the methods described in the Żabińskie paper with the archived data (Larocque-Tobler et al., 2016b), 39 of the 89 fossil samples have no analogues (distance > 95% of calibration-set distances to nearest neighbour; Fig. S1). Another reconstruction diagnostic, the squared residual distance (Juggins and Birks, 2012), shows that 19 samples have a very poor fit (> 95% of calibration-set squared residual distances; Fig. S2), and a test of reconstruction significance (Telford and Birks, 2011) is not statistically significant ($p = 0.1$). This lack of significance is reflected in the relatively weak correlation ($r = 0.36$) between principal component axis 1 of the fossil data and August air temperature reported in the Żabińskie paper.

Since both the assemblage and instrumental temperature data are available, more direct tests can be applied than the reconstruction diagnostics used above. A canonical correspondence analysis fitted to the fossil assemblage data with August temperature as the sole predictor explains 2.47% of the variance in the assemblage data. This is statistically significant ($p = 0.001$) but the ratio of the constrained eigenvalue to the first unconstrained eigenvalue is far below one ($\lambda_1/\lambda_2 = 0.38$) suggesting that August temperature is not the most important axis of variability in the assemblage data and hence precise reconstructions cannot be expected (Juggins, 2013).

The surprisingly poor reconstruction diagnostics are not the only problematic aspect of the Żabińskie paper. The instrumental temperature series used by the paper was miscalculated for the period before 1939, the period in which chironomid samples were merged because of low abundances (Larocque-Tobler, pers comm). Rather than using the mean temperature for the interval corresponding to each sample, the temperature of the first year of the interval was used. Despite this, the pre-1939 section of the reconstruction has a better correlation ($r = 0.81$) with the incorrectly calculated temperature data than with the correct temperature data ($r = 0.53$).

The two ordinations in the Żabińskie paper cannot be reproduced from the archived data (seven substantially different versions of the fossil data were archived without explanation before the current version). Figure 2 in the Żabińskie paper is a principal component analysis of the calibration set. It shows only 103 of the 121 lakes in the calibration set (the paper reports 122): 18 lakes have been omitted, including all four lakes warmer than 23°C (Fig. 4a). The omission of these lakes makes the relationship between the chironomid assemblages and temperature appear to be stronger than it is. In the pre-publication versions of this figure, several lakes had been moved rather than omitted.

Conversely, Supplementary Data Figure 1 in the Żabińskie paper, a redundancy analysis of the calibration set constrained to temperature with the fossil

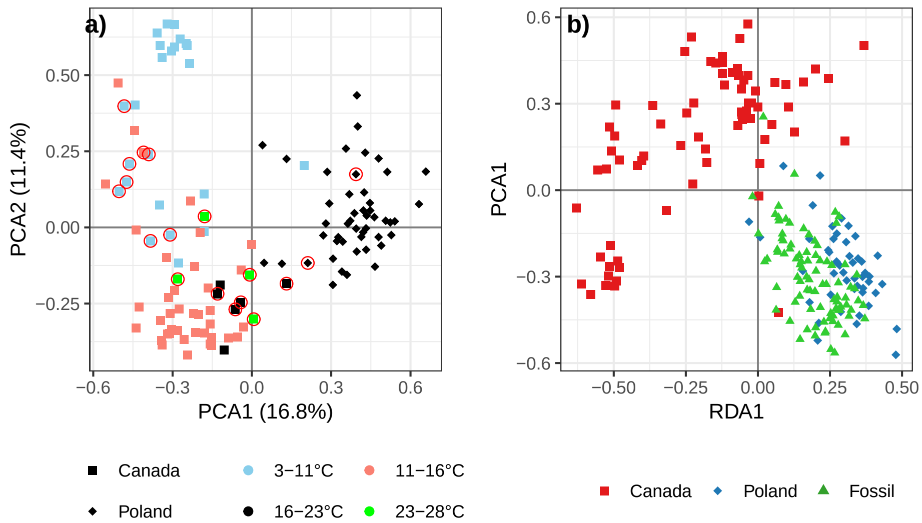

Figure 4: a) Attempt to reproduce the principal component analysis shown in Figure 2 from the the Żabińskie paper. Lakes enclosed by a red circle are missing from the published figure. Axis labels indicate the percent of variance in the calibration set explained (the Żabińskie paper reports 22.3% and 17.5% for the first two axes). b) Attempt to reproduce the redundancy analysis shown in Supplementary Data Figure 1 from the Żabińskie paper. The original figure has six more Canadian lakes in the lower right quadrant.

data added passively, has more lakes than expected (Fig. 4b). There are 73 Canadian lakes in the calibration set (the paper reports 72) but the figure shows 78. The extra Canadian lakes are mostly in the same quadrant as the Polish lakes and falsely suggest "a similarity between the assemblages in warmer lakes in Canada and those which cover the same temperature gradient in Poland" (Larocque-Tobler et al., 2015).

The Żabińskie paper reports that the count sums in both the fossil and modern data sets were at least 50 chironomid head capsules. In a corrigendum, Larocque-Tobler et al. (2016a) subsequently acknowledged that many samples had lower count sums and removed nine of the 48 Polish lakes from the calibration set and reported that the count sums for the fossil data are as low as 19. Eighty eight percent of fossil samples have a count sum lower than that initially reported. It is well known (Heiri and Lotter, 2001; Larocque, 2001; Quinlan and Smol, 2001) that low count sums are associated with increased uncertainty in reconstructions. The magnitude of the increase in uncertainty depends on the interaction between species diversity and the differences in optima between taxa. Here, I estimate the uncertainty due to low counts by finding the standard deviation of the transfer function predictions for the 59 lakes in the calibration set which appear to have large count sums (> 100 chironomids) when resampled to count sums spanning the range observed in the fossil data. For each fossil sample from Lake Żabińskie, I estimate the reconstruction uncertainty from the mean of these standard deviations. The expected uncertainty in the reconstruction given the count sums is 1.15°C. This is slightly larger than the observed standard deviation of the residuals, 0.95°C, which is due to all sources of error.

The chironomid counts from Lake Żabińskie are unusual. Typically the rarest species in any assemblage is represented by a single individual. In the Lake Żabińskie counts, 38% of samples have a minimum abundance greater than one. In one sample, all taxa occur at least five times, in another, all counts are multiples of four.

The calibration set also has problems. Lac A and Lac H have identical assemblages, except that the percentages in Lac H have been multiplied by 1.79 such that they sum to 175%. Errors of up to 6.5°C in the August temperatures of some of the Canadian lakes were corrected by Bajolle et al. (2018).

3.1.2. Abisko lakes

Larocque and Hall (2003) report reconstructions of July air temperature for four lakes in northern Sweden. The correlations of the reconstructions with instrumental data, all between 0.35 and 0.39, are surprisingly similar. Using a simulation analogous to that used by Simonsohn (2013), the probability of obtaining a standard deviation of the correlations as low as that observed is $<$ 0.001.

Of the four correlations, only one matches the correlation of reconstruction and instrumental data digitised from the authors' figure 4 – Alanen Laanijavri. For the other lakes, the correlations are 0.04, 0.45, and 0.69 for Lake Njulla, Lake 850, and Vuoskkujavri, respectively (Fig. S 3). No assemblage data have been archived, so the reconstructions cannot be verified, but there are some discrepancies between the published stratigraphies and the reconstructions. Lake 850 lacks an assemblage for 1999 CE but has a reconstruction. Similarly, Lake Njulla has a reconstruction but no assemblage for 1923 CE and the converse for 1949 CE.

3.1.3. Lake Silvaplana

The chironomid-based July temperature reconstruction from the proglacial Lake Silvaplana for the period 1850–2001 CE is first presented in Larocque et al. (2009a). Subsequently, Larocque-Tobler et al. (2009) and Larocque-Tobler et al. (2010) published 420 and 1000 year long reconstructions, respectively, which both include the 1850–2001 CE period. The three reconstructions for the last 150 years should be identical, or at least nearly so, however they are very different. Furthermore, archived reconstructions for the 150-year reconstruction (Larocque et al., 2008) and the 420-year reconstruction (Larocque et al., 2009b) are different from their associated papers, giving five different reconstructions for this time interval. None of the papers include any explanation of these differences. The overlapping portions of the archived 150- and 420-year reconstructions can largely be reconciled by aligning by sample number rather than age, which stretches the time series by up to 42 years (Fig. 5) – far more than the chronological uncertainty of the varve record (Blass et al., 2007). Remaining differences between the aligned reconstructions are nearly all integer °C differences.

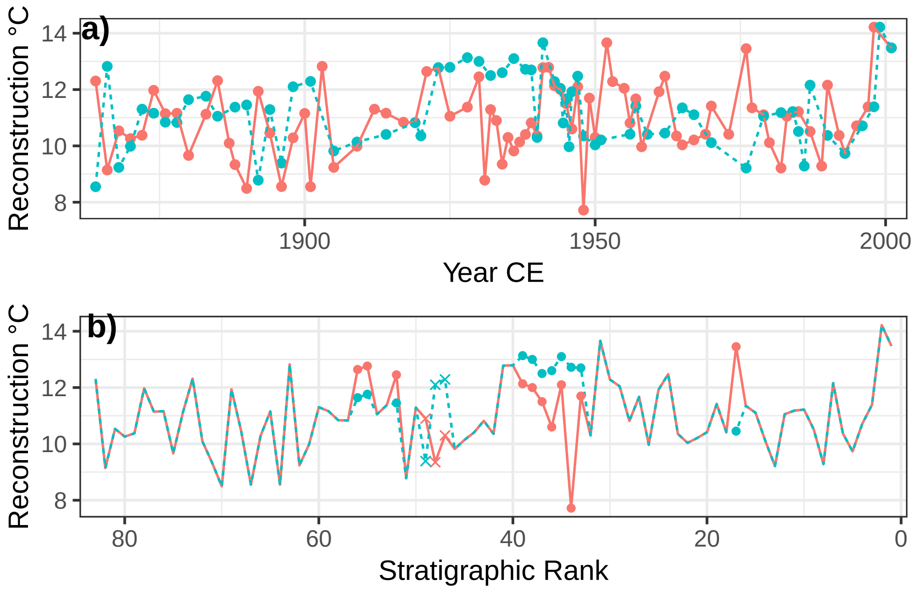

Figure 5: The 150-yr reconstruction (solid, red) and the overlapping portion of the 420-yr reconstruction (dashed, blue) from Silvaplana plotted against a) date, and b) stratigraphic rank. In b) circles indicate integer °C differences, crosses show non-integer differences between reconstructions.

Fossil assemblage percent data have been archived for the 420-year reconstruction. Unfortunately the count sums have not been archived but were reported to be at least 30. However, the data suggest that many of the samples have much smaller counts. For example, the sample for 1580 CE has two taxa with relative abundances of 16.67% and 83.33%. Given that the rarest taxa is probably represented by a single individual, it seems that this represents a count of 6 individuals. Inspection of Figure 3a from Larocque-Tobler et al. (2009), which shows the raw counts, supports this interpretation. Forty one of the 134 samples appear to have a count of fewer than 20 head capsules. In older material from Lake Silvplana (Stewart et al., 2011), about a quarter of the fossil samples are represented by a single taxon with an abundance of 100%: these counts are likely to be of a single individual. Such extremely low counts cannot be expected to yield precise reconstructions.

The instrumental temperature series used for the calibration-in-time reconstruction (Larocque-Tobler et al., 2011a) differs substantially from the series used in the 150-year calibration-in-space paper, and both are different from the instrumental data in Laroque-Tobler et al. (2010). No version of the instrumental data can be reconciled with the July temperature data from Segl-Maria (Fig S4).

3.1.4. Seebergsee

No fossil data have been archived from the anoxic Seebergsee, but the raw count data from Figure 4a in Larocque-Tobler et al. (2011b) can be digitised and samples merged to match Figure 4b. The paper reports that adjacent fossil samples were merged until there were at least 30 chironomids. However, 69% of the merged samples have count sums lower than 30 with a minimum count sum of twelve. Curiously, all chironomid counts in Figure 4a of the Seebergsee paper are multiples of four. Since the rarest taxa in the fossil data is almost certainly present as a single individual, the probability that all counts are multiples of four is infinitesimally small (under very optimistic assumptions, the probability that a given count is divisible by four is 0.25. The probability that all 186 counts are divisible by four is 10^{-112}). This suggests that the counts have been multiplied by four. If so, all the count sums are below 30 and the minimum count sum is only three.

Digitising the July instrumental temperature data from figure 6 in the Seebergsee paper allows the reproducibility of the calibration-in-time result to be tested. The paper does not report which transfer function method was used, but a two component WAPLS model on square-root transformed assemblage data yields a reconstruction that has a correlation of 0.73 with the instrumental data, identical to that reported, and the reconstruction appears to be identical. However, this performance is only achieved if the model is not cross-validated. With leave-one-out cross-validation, which is essential to prevent overfitting and give an unbiased estimate of performance, the correlation drops to 0.13, and the reconstruction has no skill. The importance of cross-validation can be demonstrated by trying to reconstruct random noise: the observed correlation is exceeded by 16.7% of 1000 trials reconstructing Gaussian noise, and so cannot be considered remarkable.

There are some other issues with the Seebergsee paper. These include substantial discrepancies between the age-depth model shown in its Figure 3 and Figure 4b; difficulties reconciling the reported instrumental data with the series from Château d'Oex (Fig. S 6); and the report that the first and second axes of a correspondence analysis (CA) of the fossil data explain 45% and 27% of the inertia, respectively. My attempt to reproduce the CA finds that the first and second axes explain only 13% and 12%, respectively, of the inertia.

Larocque-Tobler et al. (2012) use the calibration-in-time model from Larocque-Tobler et al. (2011b) to reconstruct July air temperature from a millennium-long chironomid stratigraphy from Seebergsee. In view of the very poor cross-validated performance of the calibration-in-time model ($r^2 = 0.02$), the millennium-long reconstruction cannot be expected to be skilful.

3.1.5. Baker Lake

The main features of the reconstruction of mid-summer water temperature from Baker Lake (Medeiros et al., 2012) can be reproduced from the raw data (Andrew S. Medeiros, pers comm), and the correlation with the smoothed annual temperature anomaly is high. Baker Lake is a relatively favourable study system: changes in the short ice-free period of this arctic lake due to climate change will have ecological consequences; and the lake experienced a large, year-round warming trend during the studied interval. However, chronological uncertainties, the relatively low resolution, and autocorrelation due to both the relatively smooth chironomid stratigraphy, which may be due to taphonomic processes, and the smoothing of the instrumental data, act together to preclude the detection of a sub-decadal signal so the authors never attempted to interpret one.

3.1.6. Nurmijärvi

No data have been archived for Luoto and Ojala (2016), but it is possible to digitise the fossil stratigraphy and test the reproducibility of the calibration-in-time reconstruction of July air temperature. The calibration-in-time model is reported as having a leave-one-out r^2 of 0.64, whereas the reconstruction has a reported correlation with instrumental temperature of only 0.51 (i.e. $r^2 = 0.26$): these two performance statistics should be identical. With the digitised data, I find the leave-one-out r^2 to be only 0.13 (Fig. S7). It is possible that the apparent performance was reported by mistake.

The Nurmijärvi paper reports that there is a 4–8 year lag (perhaps caused by the ecological and taphonomic processes described above) in the chironomid response to temperature, which would seem to preclude the development of a good calibration-in-time model.

3.1.7. Tiancai Lake

Zhang et al. (2017a) report that they correlate their July air temperature reconstruction with a three-year moving average of the instrumental data (Zhang et al., 2017b). However, the instrumental time series has been mis-processed. Rather than using a three-year moving average, the authors take the July air temperature of every third year and interpolate. The significant correlation between the reconstruction and the interpolated triennial data can only be a Type 1 error, unless it is believed that chironomids are only sensitive to July temperature every third year and that they can predict the temperature up to two years ahead. With a three-year moving average, the correlation is weak ($r = 0.12$) and not statistically significant ($p = 0.54$).

3.1.8. Speke Hall Lake

The polluted, eutrophic Speke Hall Lake (Lang et al., 2017) is a curious choice of site on which to apply the Norwegian chironomid calibration set (Brooks and Birks, 2001), conflicting with the usual guidance that sites should be within the environmental space covered by the calibration set to minimise non-analogue problems. Perhaps not surprisingly, none of the fossil samples have good analogues in the calibration set (Fig. S8), and other diagnostics are generally very poor (Fig. S9). The reported lack of correlation between the detrended correspondence analysis scores and the reconstruction (which can be reproduced from the raw data (Andrew S. Medeiros, pers comm)) suggests that temperature is not the most important variable driving the variability in assemblage composition. The reported correlation between the reconstruction and instrumental data, based on sixteen samples, is statistically significant at the $p < 0.05$ level, but given the lack of correction for temporal autocorrelation, the inverse correlation between July temperature and most of the heavy metals measured, and the very poor reconstruction diagnostics, it is more likely that the correlation is due to chance than that the reconstruction has any real skill.

3.2. Recommendations for increasing reproducibility

One of the main impediments to investigating the issues with sub-decadal resolution reconstructions was the lack of archived data. Some papers had no data archived (Hernández-Almeida et al., 2015a; Lang et al., 2017; Larocque and Hall, 2003; Larocque-Tobler et al., 2011b; Luoto and Ojala, 2016), others had partial data (Larocque-Tobler et al., 2015; Zhang et al., 2017a). Some data were provided on request, and some could be digitised, which is slow and imprecise.

Current data archiving policies in Quaternary science journals are either inadequate or unenforced. For example, *Quaternary Science Reviews* "encourages" authors to share data "where appropriate"[1]. In contrast, *Ecological Monographs* makes the archiving of "all data" a "condition for publication"[2]. Quaternary science journals should adopt and enforce a similar policy (Simonsohn, 2013). Stricter data archiving policies have been shown to increase the proportion of papers archiving data (Giofrè et al., 2017; Nuijten et al., 2017). Data should also be made available to reviewers.

A second impediment to investigating the issues was the absence of available code to reproduce the analyses. For example, Figure 5 in Larocque-Tobler et al. (2012), which shows that ordinations of the fossil data from both Seebergsee and Lake Silvaplana have a large separation between the pre- and post- 1950 samples, cannot be reproduced with the methods described in the paper (or similar methods). If computer code, or equivalent, had been archived, it would probably be trivial to understand this problem. None of the other published ordinations and zonations of the Lake Silvaplana chironomid stratigraphy show a discontinuity at 1950.

Reproducibility is a central requirement for testing scientific theories and therefore testing the reproducibility of published literature should be encouraged (Fidler et al., 2017). Using state-of-the-art methods for generating reproducible analyses (Cooper and Hsing, 2017) is strongly recommended.

4. Conclusions

Reliable sub-decadal resolution reconstructions depend on the conjunction of a strong direct ecological link to the variable being reconstructed and a simple taphonomy. Accurate and precise chronologies are also needed for a direct comparison with instrumental records. These requirements might be met by diatom-total phosphorus reconstructions, and perhaps by chrysophyte reconstructions of winter climate. It appears unlikely that they will be met by

[1] https://www.elsevier.com/journals/quaternary-science-reviews/0277-3791/guide-for-authors

[2] http://esajournals.onlinelibrary.wiley.com/hub/journal/10.1002/(ISSN)1557-7015/resources/data-availability-policy-ecm.html

the relatively long-lived, predominately littoral chironomids for reconstructions of a single month's temperature.

The apparently extensive evidence for the utility of sub-decadal resolution temperature reconstructions is shown to be illusory. The papers reporting chironomid-temperature reconstructions from Lake Żabińskie, the lakes near Abisko, Lake Silvaplana, and Seebergsee are plagued by pervasive errors and improbabilities and key results cannot be reproduced. The reported correlation between instrumental data and the reconstruction from Tiancai Lake cannot be reproduced and at least some results from Nurmijärvi cannot be reproduced. The Speke Hall Lake reconstruction can be reproduced, but this study is unlikely to replicate in a similar site. The Baker Lake reconstruction is reproducible and it is plausible that the result could be replicated at a similar site, but there is no sub-decadal signal.

The number and severity of errors in some of the papers reviewed highlights the importance of adopting reproducible practices and archiving all the raw data and computer code needed to reproduce the results.

5. Methods

All analyses were done in R version 3.4.4 (R Core Team, 2018). Ordinations were fitted with vegan version 2.5.3 (Oksanen et al., 2018) with square-root transformed assemblage data. Transfer functions were fitted with rioja version 0.9.15.1 (Juggins, 2017) using square-root transformed species data. Some diagnostics were performed using analogue version 0.17.1 (Simpson and Oksanen, 2018). The reconstruction significance test was run with randomTF from palaeoSig version 1.1.3 (Telford, 2015) with 999 trials for the null distribution.

Spatial climate data were extracted from WorldClim2 (Fick and Hijmans, 2017); some of the lake positions are approximate as their locations are not all archived. The regional Polish temperature composite included data from Kaliningrad, Kaunas, Vilnius, and Warsaw GHCN series (Lawrimore et al., 2011) extracted using the KNMI Climate Explorer. The 1951-1980 mean was subtracted from each series. Code to reproduce this paper is archived at https://doi.org/xxxx.xxxx.

Acknowledgements

I thank all my colleagues in Bergen and elsewhere who have given the help and encouragement needed to write this paper. I thank Dr Gavin Simpson and three anonymous reviewers for their constructive comments. This work was partially funded by the Norwegian Research Council FriMedBio project palaeoDrivers (213607).

References

Alefs, J., Müller, J., 1999. Differences in the eutrophication dynamics of Ammersee and Starnberger See (southern Germany), reflected by the diatom succession in varve-dated sediments. Journal of Paleolimnology 21, 395–407. https://doi.org/10.1023/A:1008098118867

Allison, T.D., Moeller, R.E., Davis, M.B., 1986. Pollen in laminated sediments provides evidence for a mid-Holocene forest pathogen outbreak. Ecology 67, 1101–1105. https://doi.org/10.2307/1939835

Amann, B., Lobsiger, S., Fischer, D., Tylmann, W., Bonk, A., Filipiak, J., Grosjean, M., 2014. Spring temperature variability and eutrophication history inferred from sedimentary pigments in the varved sediments of Lake Zabinskie, north-eastern Poland, AD 1907-2008. Global and Planetary Change 123, 86–96. https://doi.org/10.1016/j.gloplacha.2014.10.008

Appleby, P.G., Piliposian, G.T., 2006. Radiometric dating of sediment records from mountain lakes in the Tatra Mountains. Biologia 61, S51–S64. https://doi.org/10.2478/s11756-006-0119-4

Bajolle, L., Larocque-Tobler, I., Gandouin, E., Lavoie, M., Bergeron, Y., Ali, A.A., 2018. Major postglacial summer temperature changes in the central coniferous boreal forest of Quebec (Canada) inferred using chironomid assemblages. Journal of Quaternary Science 33, 409–420. https://doi.org/10.1002/jqs.3022

Bigler, C., von Gunten, L., Lotter, A.F., Hausmann, S., Blass, A., Ohlendorf, C., Sturm, M., 2007. Quantifying human-induced eutrophication in Swiss mountain lakes since AD 1800 using diatoms. The Holocene 17, 1141–1154. https://doi.org/10.1177/0959683607082555

Birks, H.J.B., Heiri, O., Seppä, H., Bjune, A.E., 2010. Strengths and weaknesses of quantitative climate reconstructions based on late-Quaternary biological proxies. The Open Ecology Journal 3, 68–110. https://doi.org/10.2174/1874213001003020068

Blass, A., Grosjean, M., Troxler, A., Sturm, M., 2007. How stable are twentieth-century calibration models? A high-resolution summer temperature reconstruction for the eastern Swiss Alps back to AD 1580 derived from proglacial varved sediments. The Holocene 17, 51–63. https://doi.org/10.1177/0959683607073278

Bonk, A., Kinder, M., Enters, D., Grosjean, M., Meyer-Jacob, C., Tylmann, W., 2016. Sedimentological and geochemical responses of Lake Zabinskie (north-eastern Poland) to erosion changes during the last millennium. Journal of Paleolimnology 56, 239–252. https://doi.org/10.1007/s10933-016-9910-6

Bonk, A., Tylmann, W., Amann, B., Enters, D., Grosjean, M., 2015a. Modern limnology and varve-formation processes in Lake Zabinskie, northeastern Poland: comprehensive process studies as a key to understand the sediment record. Journal of Limnology 74, 358–370. https://doi.org/10.4081/jlimnol.2014.1117

Bonk, A., Tylmann, W., Goslar, T., Wacnik, A., Grosjean, M., 2015b. Comparing varve counting and C-14-AMS chronologies in the sediments of Lake Zabinskie, northeastern Poland: implications for accurate C-14 dating of lake sediments. Geochronometria 42, 159–171. https://doi.org/10.1515/geochr-2015-0019

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. https://doi.org/10.1016/s0277-3791(01)00038-5

Brooks, S.J., Matthews, I.P., Birks, H.H., Birks, H.B., 2012. High resolution Lateglacial and early-Holocene summer air temperature records from Scotland inferred from chironomid assemblages. Quaternary Science Reviews 41, 67–82. https://doi.org/10.1016/j.quascirev.2012.03.007

Charbonneau, P., Hare, L., 1998. Burrowing behavior and biogenic structures of mud-dwelling insects. Journal of the North American Benthological Society 17, 239–249. https://doi.org/10.2307/1467965

Cooper, N., Hsing, P.-Y., 2017. A guide to reproducible code in ecology and evolution. British Ecological Society. http://www.britishecologicalsociety.org/wp-content/uploads/2017/12/guide-to-reproducible-code.pdf

de Jong, R., Kamenik, C., 2011. Validation of a chrysophyte stomatocyst-based cold-season climate reconstruction from high-alpine Lake Silvaplana, Switzerland. Journal of Quaternary Science 26, 268–275. https://doi.org/10.1002/jqs.1451

D'Arrigo, R., Wilson, R., Liepert, B., Cherubini, P., 2008. On the "divergence problem" in northern forests: a review of the tree-ring evidence and possible causes. Global and Planetary Change 60, 289–305. https://doi.org/10.1016/j.gloplacha.2007.03.004

Eggermont, H., Heiri, O., 2011. The chironomid-temperature relationship: expression in nature and palaeoenvironmental implications. Biological Reviews 87, 430–456. https://doi.org/10.1111/j.1469-185x.2011.00206.x

Elbert, J., Grosjean, M., von Gunten, L., Urrutia, R., Fischer, D., Wartenburger, R., Ariztegui, D., Fujak, M., Hamann, Y., 2012. Quantitative high-resolution winter (JJA) precipitation reconstruction from varved sediments of Lago Plomo 47°S, Patagonian Andes, AD 1530–2002. The Holocene 22, 465–474. https://doi.org/10.1177/0959683611425547

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302–4315. https://doi.org/10.1002/joc.5086

Fidler, F., Chee, Y.E., Wintle, B.C., Burgman, M.A., McCarthy, M.A., Gordon, A., 2017. Metaresearch for evaluating reproducibility in ecology and evolution. BioScience 67, 282–289. https://doi.org/10.1093/biosci/biw159

Giofrè, D., Cumming, G., Fresc, L., Boedker, I., Tressoldi, P., 2017. The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. PLoS ONE 12, e0175583. https://doi.org/10.1371/journal.pone.0175583

Green, D.G., 1983. The ecological interpretation of fine resolution pollen records. New Phytologist 94, 459–477. https://doi.org/10.1111/j.1469-8137.1983.tb03459.x

Heiri, O., Lotter, A.F., 2001. Effect of low count sums on quantitative environmental reconstructions: an example using subfossil chironomids. Journal of Paleolimnology 26, 343–350. https://doi.org/10.1023/a:1017568913302

Hernández-Almeida, I., Grosjean, M., Gómez-Navarro, J.J., Larocque-Tobler, I., Bonk, A., Enters, D., Ustrzycka, A., Piotrowska, N., Przybylak, R., Wacnik, A., Witak, M., Tylmann, W., 2017. Resilience, rapid transitions and regime shifts: fingerprinting the responses of Lake Żabińskie (NE Poland) to climate variability and human disturbance since AD 1000. The Holocene 27, 258–270. https://doi.org/10.1177/0959683616658529

Hernández-Almeida, I., Grosjean, M., Przybylak, R., Tylmann, W., 2015a. A chrysophyte-based quantitative reconstruction of winter severity from varved lake sediments in NE Poland during the past millennium and its relationship to natural climate variability. Quaternary Science Reviews 122, 74–88. https://doi.org/10.1016/j.quascirev.2015.05.029

Hernández-Almeida, I., Grosjean, M., Tylmann, W., Bonk, A., 2015b. Chrysophyte cyst-inferred variability of warm season lake water chemistry and climate in northern Poland: training set and downcore reconstruction. Journal of Paleolimnology 53, 123–138. https://doi.org/10.1007/s10933-014-9812-4

Joosten, H., de Klerk, P., 2007. In search of finiteness: The limits of fine-resolution palynology of sphagnum peat. The Holocene 17, 1023–1031. https://doi.org/10.1177/0959683607082416

Juggins, S., 2017. Rioja: Analysis of Quaternary science data. R package version 0.9-15.1. http://www.staff.ncl.ac.uk/stephen.juggins/

Juggins, S., 2013. Quantitative reconstructions in palaeolimnology: New paradigm or sick science? Quaternary Science Reviews 64, 20–32. https://doi.org/10.1016/j.quascirev.2012.12.014

Juggins, S., Birks, H.J.B., 2012. Quantitative environmental reconstructions from biological data, in: Birks, H.J.B., Lotter, A.F., Juggins, S., Smol, J.P. (Eds.), Tracking Environmental Change Using Lake Sediments, Vol. 5: Data Handling and Numerical Techniques. Springer, Dordrecht, pp. 431–494. https://doi.org/10.1007/978-94-007-2745-8_14

Kamenik, C., van der Knaap, W.O., van Leeuwen, J.F.N., Goslar, T., 2009. Pollen/climate calibration based on a near-annual peat sequence from the Swiss Alps. Journal of Quaternary Science 24, 529–546. https://doi.org/10.1002/jqs.1266

Lang, B., Medeiros, A.S., Worsley, A., Bedford, A., Brooks, S.J., 2017. Influence of industrial activity and pollution on the paleoclimate reconstruction from a eutrophic lake in lowland England, UK. Journal of Paleolimnology. https://doi.org/10.1007/s10933-017-9995-6

Larocque, I., 2001. How many chironomid head capsules are enough? A statistical approach to determine sample size for palaeoclimatic reconstructions. Palaeogeography, Palaeoclimatology, Palaeoecology 172, 133–142. https://doi.org/10.1016/s0031-0182(01)00278-4

Larocque, I., Grosjean, M., Heiri, O., Bigler, C., Blass, A., 2009a. Comparison between chironomid-inferred July temperatures and meteorological data AD 1850–2001 from varved Lake Silvaplana, Switzerland. Journal of Paleolimnology 41, 329–342. https://doi.org/10.1007/s10933-008-9228-0

Larocque, I., Grosjean, M., Heiri, O., Bigler, C., Blass, A., 2008. Larocque et al. 2008 Lake Silvaplana, Switzerland chironomid temperature [WWW Document]. https://www.ncdc.noaa.gov/paleo/study/6394

Larocque, I., Grosjean, M., Heiri, O., Trachsel, M., 2009b. Larocque et al.

2009 Lake Silvaplana 420yr chironomid inferred temperature [WWW Document]. https://www.ncdc.noaa.gov/paleo/study/8737

Larocque, I., Hall, R.I., 2003. Chironomids as quantitative indicators of mean July air temperature: Validation by comparison with century-long meteorological records from northern Sweden. Journal of Paleolimnology 29, 475–493. https://doi.org/10.1023/A:1024423813384

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. https://doi.org/10.1023/a:1017524101783

Larocque, I., Pienitz, R., Rolland, N., 2006. Factors influencing the distribution of chironomids in lakes distributed along a latitudinal gradient in northwestern Quebec, Canada. Canadian Journal of Fisheries and Aquatic Sciences 63, 1286–1297. https://doi.org/10.1139/f06-020

Larocque-Tobler, I., 2010. Reconstructing temperature at Egelsee, Switzerland, using North American and Swedish chironomid transfer functions: Potential and pitfalls. Journal of Paleolimnology 44, 243–251. https://doi.org/10.1007/s10933-009-9400-1

Larocque-Tobler, I., Filipiak, J., Tylmann, W., Bonk, A., Grosjean, M., 2016a. Corrigendum to "Comparison between chironomid-inferred mean-August temperature from varved Lake Żabińskie (Poland) and instrumental data since 1896 AD" [Quat. Sci. Rev. 111 (2015) 35–50]. Quaternary Science Reviews 140, 163–167. https://doi.org/10.1016/j.quascirev.2016.01.020

Larocque-Tobler, I., Filipiak, J., Tylmann, W., Bonk, A., Grosjean, M., 2016b. Lake Żabińskie, Poland 115 year August chironomid-inferred temperature [WWW Document]. https://www.ncdc.noaa.gov/paleo-search/study/19501

Larocque-Tobler, I., Filipiak, J., Tylmann, W., Bonk, A., Grosjean, M., 2015. Comparison between chironomid-inferred mean-August temperature from varved Lake Żabińskie (Poland) and instrumental data since 1896 AD. Quaternary Science Reviews 111, 35–50. https://doi.org/10.1016/j.quascirev.2015.01.001

Larocque-Tobler, I., Grosjean, M., Heiri, O., Trachsel, M., 2009. High-resolution chironomid-inferred temperature history since AD 1580 from varved Lake Silvaplana, Switzerland: comparison with local and regional reconstructions. The Holocene 19, 1201–1212. https://doi.org/10.1177/0959683609348253

Larocque-Tobler, I., Grosjean, M., Heiri, O., Trachsel, M., Kamenik, C., 2010. Thousand years of climate change reconstructed from chironomid subfossils preserved in varved Lake Silvaplana, Engadine, Switzerland. Quaternary Science Reviews 29, 1940–1949. https://doi.org/10.1016/j.quascirev.2010.04.018

Larocque-Tobler, I., Grosjean, M., Kamenik, C., 2011a. Calibration-in-time versus calibration-in-space (transfer function) to quantitatively infer July air temperature using biological indicators (chironomids) preserved in lake sediments. Palaeogeography, Palaeoclimatology, Palaeoecology 299, 281–288. https://doi.org/10.1016/j.palaeo.2010.11.008

Larocque-Tobler, I., Quinlan, R., Stewart, M.M., Grosjean, M., 2011b. Chironomid-inferred temperature changes of the last century in anoxic Seebergsee, Switzerland: assessment of two calibration methods. Quaternary Science

Reviews 30, 1770–1779. https://doi.org/10.1016/j.quascirev.2011.04.008

Larocque-Tobler, I., Stewart, M.M., Quinlan, R., Trachsel, M., Kamenik, C., Grosjean, M., 2012. A last millennium temperature reconstruction using chironomids preserved in sediments of anoxic Seebergsee (Switzerland): consensus at local, regional and Central European scales. Quaternary Science Reviews 41, 49–56. https://doi.org/10.1016/j.quascirev.2012.03.010

Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S., Rennie, J., 2011. An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. Journal of Geophysical Research: Atmospheres 116, D19121. https://doi.org/10.1029/2011JD016187

Leemann, A., Niessen, F., 1994. Varve formation and the climatic record in an Alpine proglacial lake: calibrating annually-laminated sediments against hydrological and meteorological data. The Holocene 4, 1–8. https://doi.org/10.1177/095968369400400101

Livingstone, D.M., Dokulil, M.T., 2001. Eighty years of spatially coherent Austrian lake surface temperatures and their relationship to regional air temperature and the North Atlantic Oscillation. Limnology and Oceanography 46, 1220–1227. https://doi.org/10.4319/lo.2001.46.5.1220

Livingstone, D.M., Lotter, A.F., 1998. The relationship between air and water temperatures in lakes of the Swiss Plateau: a case study with palaeolimnological implications. Journal of Paleolimnology 19, 181–198. https://doi.org/10.1023/A:1007904817619

Lotter, A.F., 1998. The recent eutrophication of Baldeggersee (Switzerland) as assessed by fossil diatom assemblages. The Holocene 8, 395–405. https://doi.org/10.1191/095968398674589725

Lotter, A.F., Anderson, N.J., 2012. Limnological responses to environmental changes at inter-annual to decadal time-scales, in: Birks, H.J.B., Lotter, A.F., Juggins, S., Smol, J.P. (Eds.), Tracking Environmental Change Using Lake Sediments Vol 5: Data Handling and Numerical Techniques. Springer, Dordrecht, pp. 557–578. https://doi.org/10.1007/978-94-007-2745-8_18

Luoto, T.P., Ojala, A.E., 2016. Meteorological validation of chironomids as a paleotemperature proxy using varved lake sediments. The Holocene 27, 870–878. https://doi.org/10.1177/0959683616675940

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Rouco, J.G., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., Timmermann, A., 2013. Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, in: Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. (Eds.),. Cambridge University Press, Cambridge, United Kingdom; New York, NY, USA.

Medeiros, A.S., Friel, C.E., Finkelstein, S.A., Quinlan, R., 2012. A high resolution multi-proxy record of pronounced recent environmental change at Baker Lake, Nunavut. Journal of Paleolimnology 47, 661–676. https://doi.org/10.1007/s10933-012-9589-2

Nagell, B., Landahl, C.-C., 1978. Resistance to anoxia of *Chironomus* *plumosus* and *Chironomus anthracinus* (Diptera) larvae. Holarctic Ecology 1, 333–336. http://www.jstor.org/stable/3682403

Nuijten, M.B., Borghuis, J., Veldkamp, C.L.S., Dominguez-Alvarez, L., van Assen, M.A.L.M., Wicherts, J.M., 2017. Journal data sharing policies and statistical reporting inconsistencies in psychology. Collabra: Psychology 3, 31. https://doi.org/10.1525/collabra.102

Ojala, A., Francus, P., Zolitschka, B., Besonen, M., Lamoureux, S., 2012. Characteristics of sedimentary varve chronologies – a review. Quaternary Science Reviews 43, 45–60. https://doi.org/10.1016/j.quascirev.2012.04.006

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2018. Vegan: Community ecology package. R package version 2.5-3. https://CRAN.R-project.org/package=vegan

O'Reilly, C.M., Sharma, S., Gray, D.K., Hampton, S.E., Read, J.S., Rowley, R.J., Schneider, P., Lenters, J.D., McIntyre, P.B., Kraemer, B.M., Weyhenmeyer, G.A., Straile, D., Dong, B., Adrian, R., Allan, M.G., Anneville, O., Arvola, L., Austin, J., Bailey, J.L., Baron, J.S., Brookes, J.D., Eyto, E. de, Dokulil, M.T., Hamilton, D.P., Havens, K., Hetherington, A.L., Higgins, S.N., Hook, S., Izmest'eva, L.R., Joehnk, K.D., Kangur, K., Kasprzak, P., Kumagai, M., Kuusisto, E., Leshkevich, G., Livingstone, D.M., MacIntyre, S., May, L., Melack, J.M., Mueller-Navarra, D.C., Naumenko, M., Noges, P., Noges, T., North, R.P., Plisnier, P.-D., Rigosi, A., Rimmer, A., Rogora, M., Rudstam, L.G., Rusak, J.A., Salmaso, N., Samal, N.R., Schindler, D.E., Schladow, S.G., Schmid, M., Schmidt, S.R., Silow, E., Soylu, M.E., Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A., Williamson, C.E., Zhang, G., 2015. Rapid and highly variable warming of lake surface waters around the globe. Geophysical Research Letters 42, 10, 773–10, 781. https://doi.org/10.1002/2015GL066235

PAGES2k Consortium, 2017. A global multiproxy database for temperature reconstructions of the common era. Scientific Data 4, 170088. https://doi.org/10.1038/sdata.2017.88

PAGES2k Consortium, 2013. Continental-scale temperature variability during the past two millennia. Nature Geoscience 6, 339–346. https://doi.org/10.1038/ngeo1797

Parker, D.E., Legg, T.P., Folland, C.K., 1992. A new daily central England temperature series, 1772–1991. International Journal of Climatology 12, 317–342. https://doi.org/10.1002/joc.3370120402

Peglar, S.M., 1993. The mid-Holocene *Ulmus* decline at Diss Mere, Norfolk, UK: a year-by-year pollen stratigraphy from annual laminations. The Holocene 3, 1–13. https://doi.org/10.1177/095968369300300101

Perga, M.-E., Frossard, V., Jenny, J.-P., Alric, B., Arnaud, F., Berthon, V., Black, J., Domaizon, I., Giguet-Covex, C., Kirkham, A., Magny, M., Manca, M., Marchetto, A., Millet, L., Paillès, C., Pignol, C., Poulenard, J., Reyss, J.-L., Rimet, F., Savichtcheva, O., Sabatier, P., Sylvestre, F., Verneaux, V., 2015. High-resolution paleolimnology opens new management perspectives for lakes adaptation to climate warming. Frontiers in Ecology and Evolution 3, 72.

https://doi.org/10.3389/fevo.2015.00072

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. https://doi.org/10.1023/a:1017546821591

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raubitschek, S., Lücke, A., Schleser, G.H., 1999. Sedimentation patterns of diatoms in Lake Holzmaar, Germany - (on the transfer of climate signals to biogenic silica oxygen isotope proxies). Journal of Paleolimnology 21, 437–448. https://doi.org/10.1023/A:1008022532458

Rösner, R.R., Müller-Navarra, D.C., Zorita, E., 2012. Trend analysis of weekly temperatures and oxygen concentrations during summer stratification in Lake Plußsee: a long-term study. Limnology and Oceanography 57, 1479–1491. https://doi.org/10.4319/lo.2012.57.5.1479

Saunders, K.M., Grosjean, M., Hodgson, D.A., 2013. A 950 yr temperature reconstruction from Duckhole Lake, southern Tasmania, Australia. The Holocene 23, 771–783. https://doi.org/10.1177/0959683612470176

Schlesinger, M.E., Ramankutty, N., 1994. An oscillation in the global climate system of period 65-70 years. Nature 367, 723–726. https://doi.org/10.1038/367723a0

Sharma, S., Gray, D.K., Read, J.S., O'Reilly, C.M., Schneider, P., Qudrat, A., Gries, C., Stefanoff, S., Hampton, S.E., Hook, S., Lenters, J.D., Livingstone, D.M., McIntyre, P.B., Adrian, R., Allan, M.G., Anneville, O., Arvola, L., Austin, J., Bailey, J., Baron, J.S., Brookes, J., Chen, Y., Daly, R., Dokulil, M., Dong, B., Ewing, K., de Eyto, E., Hamilton, D., Havens, K., Haydon, S., Hetzenauer, H., Heneberry, J., Hetherington, A.L., Higgins, S.N., Hixson, E., Izmest'eva, L.R., Jones, B.M., Kangur, K., Kasprzak, P., Köster, O., Kraemer, B.M., Kumagai, M., Kuusisto, E., Leshkevich, G., May, L., MacIntyre, S., Müller-Navarra, D., Naumenko, M., Noges, P., Noges, T., Niederhauser, P., North, R.P., Paterson, A.M., Plisnier, P.-D., Rigosi, A., Rimmer, A., Rogora, M., Rudstam, L., Rusak, J.A., Salmaso, N., Samal, N.R., Schindler, D.E., Schladow, G., Schmidt, S.R., Schultz, T., Silow, E.A., Straile, D., Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A., Weyhenmeyer, G.A., Williamson, C.E., Woo, K.H., 2015. A global database of lake surface temperatures collected by in situ and satellite methods from 1985-2009. Scientific Data 2, 150008. https://doi.org/10.1038/sdata.2015.8

Sheppard, P.R., 2010. Dendroclimatology: Extracting climate from trees. Wiley Interdisciplinary Reviews: Climate Change 1, 343–352. https://doi.org/10.1002/wcc.42

Simola, H., 1977. Diatom succession in the formation of annually laminated sediment in Lovojärvi, a small eutrophicated lake. Annales Botanici Fennici 14, 143–148. http://www.jstor.org/stable/43922137

Simonsohn, U., 2013. Just post it: The lesson from two cases of fabricated data detected by statistics alone. Psychological Science 24, 1875–1888. https:

//doi.org/10.1177/0956797613480366

Simpson, G.L., Oksanen, J., 2018. Analogue: Analogue and weighted averaging methods for palaeoecology. R package version 0.17-1. https://cran.r-project.org/package=analogue

St. George, S., 2014. An overview of tree-ring width records across the Northern Hemisphere. Quaternary Science Reviews 95, 132–150. https://doi.org/10.1016/j.quascirev.2014.04.029

Stewart, M.M., Larocque-Tobler, I., Grosjean, M., 2011. Quantitative interannual and decadal June-July-August temperature variability ca. 570 BC to AD 120 (Iron Age-Roman Period) reconstructed from the varved sediments of Lake Silvaplana, Switzerland. Journal of Quaternary Science 26, 491–501. https://doi.org/10.1002/jqs.1480

Telford, R., 2015. PalaeoSig: Significance tests of quantitative palaeoenvironmental reconstructions. R package version 1.1-3. http://cran.r-project.org/package=palaeoSig

Telford, R.J., Birks, H.J.B., 2011. A novel method for assessing the statistical significance of quantitative reconstructions inferred from biotic assemblages. Quaternary Science Reviews 30, 1272–1278. https://doi.org/10.1016/j.quascirev.2011.03.002

Telford, R.J., Birks, H.J.B., 2009. Evaluation of transfer functions in spatially structured environments. Quaternary Science Reviews 28, 1309–1316. https://doi.org/10.1016/j.quascirev.2008.12.020

Telford, R.J., Birks, H.J.B., 2005. The secret assumption of transfer functions: Problems with spatial autocorrelation in evaluating model performance. Quaternary Science Reviews 24, 2173–2179. https://doi.org/10.1016/j.quascirev.2005.05.001

Thomas, E.K., Briner, J.P., 2009. Climate of the past millennium inferred from varved proglacial lake sediments on northeast Baffin Island, Arctic Canada. Journal of Paleolimnology 41, 209–224. https://doi.org/10.1007/s10933-008-9258-7

Tilman, D., 1981. Tests of resource competition theory using four species of Lake Michigan algae. Ecology 62, 802–815. https://doi.org/10.2307/1937747

Tokeshi, M., 1995. Life cycles and population dynamics, in: Armitage, P.D., Cranston, P.S., Pinder, L.C.V. (Eds.), The Chironomidae: Biology and Ecology of Non-Biting Midges. Springer, Dordrecht, pp. 225–268. https://doi.org/10.1007/978-94-011-0715-0_10

Trachsel, M., Grosjean, M., Schnyder, D., Kamenik, C., Rein, B., 2010. Scanning reflectance spectroscopy (380–730 nm): A novel method for quantitative high-resolution climate reconstructions from minerogenic lake sediments. Journal of Paleolimnology 44, 979–994. https://doi.org/10.1007/s10933-010-9468-7

Trachsel, M., Kamenik, C., Grosjean, M., McCarroll, D., Moberg, A., Brázdil, R., Büntgen, U., Dobrovolný, P., Esper, J., Frank, D.C., Friedrich, M., Glaser, R., Larocque-Tobler, I., Nicolussi, K., Riemann, D., 2012. Multi-archive summer temperature reconstruction for the European Alps, AD 1053–1996. Quaternary Science Reviews 46, 66–79. https://doi.org/10.1016/j.quascirev.2012.04.021

Trachsel, M., Telford, R.J., 2017. All age–depth models are wrong, but are getting better. The Holocene 27, 860–869. https://doi.org/10.1177/0959683616675939

Tylmann, W., Bonk, A., Goslar, T., Wulf, S., Grosjean, M., 2016. Calibrating Pb-210 dating results with varve chronology and independent chronostratigraphic markers: problems and implications. Quaternary Geochronology 32, 1–10. https://doi.org/10.1016/j.quageo.2015.11.004

van der Knaap, W.O., van Leeuwen, J.F.N., Svitavská-Svobodová, H., Pidek, I.A., Kvavadze, E., Chichinadze, M., Giesecke, T., Kaszewski, B.M., Oberli, F., Kalnina, L., Pardoe, H.S., Tinner, W., Ammann, B., 2010. Annual pollen traps reveal the complexity of climatic control on pollen productivity in Europe and the Caucasus. Vegetation History and Archaeobotany 19, 285–307. https://doi.org/10.1007/s00334-010-0250-6

van Hardenbroek, M., Heiri, O., Wilhelm, M.F., Lotter, A.F., 2011. How representative are subfossil assemblages of Chironomidae and common benthic invertebrates for the living fauna of Lake De Waay, the Netherlands? Aquatic Sciences 73, 247–259. https://doi.org/10.1007/s00027-010-0173-4

Velle, G., Brodersen, K.P., Birks, H.J.B., Willassen, E., 2010. Midges as quantitative temperature indicator species: lessons for palaeoecology. The Holocene 20, 989–1002. https://doi.org/10.1177/0959683610365933

von Gunten, L., Grosjean, M., Kamenik, C., Fujak, M., Urrutia, R., 2012. Calibrating biogeochemical and physical climate proxies from non-varved lake sediments with meteorological data: methods and case studies. Journal of Paleolimnology 47, 583–600. https://doi.org/10.1007/s10933-012-9582-9

von Gunten, L., Grosjean, M., Rein, B., Urrutia, R., Appleby, P., 2009. A quantitative high-resolution summer temperature reconstruction based on sedimentary pigments from Laguna Aculeo, central Chile, back to AD 850. The Holocene 19, 873–881. https://doi.org/10.1177/0959683609336573

Werner, J.P., Tingley, M.P., 2015. Technical note: Probabilistically constraining proxy age–depth models within a Bayesian hierarchical reconstruction model. Climate of the Past 11, 533–545. https://doi.org/10.5194/cp-11-533-2015

Witak, M., Hernández-Almeida, I., Grosjean, M., Tylmann, W., 2017. Diatom-based reconstruction of trophic status changes recorded in varved sediments of Lake Żabińskie (northeastern Poland), AD 1888-2010. Oceanological and Hydrobiological Studies 46, 1–17. https://doi.org/10.1515/ohs-2017-0001

Zhang, E., Chang, J., Cao, Y., Tang, H., Langdon, P., Shulmeister, J., Wang, R., Yang, X., Shen, J., 2017a. A chironomid-based mean July temperature inference model from the south-east margin of the Tibetan Plateau, China. Climate of the Past 13, 185–199. https://doi.org/10.5194/cp-13-185-2017

Zhang, E., Chang, J., Cao, Y., Tang, H., Langdon, P., Shulmeister, J., Wang, R., Yang, X., Shen, J., 2017b. Supplementary information for: A chironomid-based mean July temperature inference model from the south-east margin of the Tibetan Plateau, China. [WWW Document]. https://www.dropbox.com/s/0z2sl6zwzj69f7h/Zhang%20et%20al%202017_Climate%20of%20the%20Past_dataset.xlsx?dl=0

Supplementary Data for 'Review and test of reproducibility of sub-decadal resolution palaeoenvironmental reconstructions from microfossil assemblages'

Richard J. Telford

07 January 2019

Lake Żabińskie

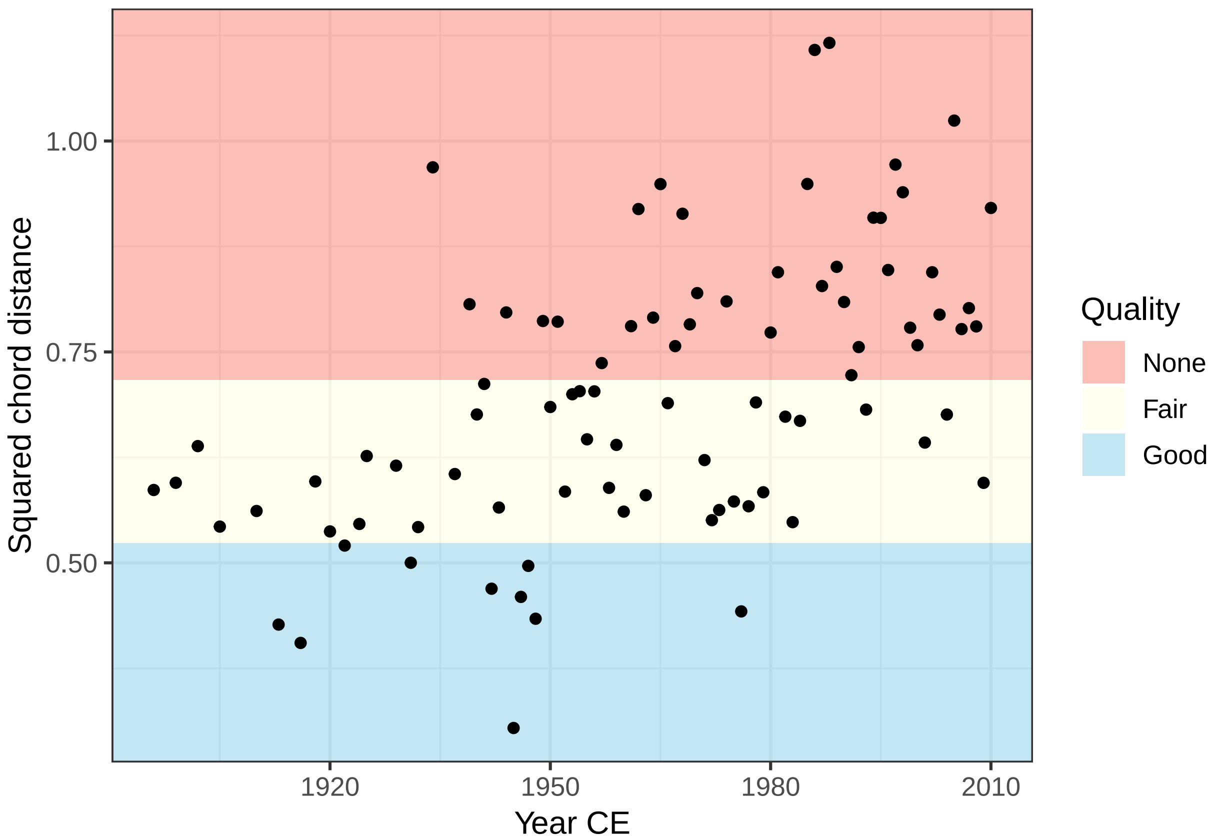

Figure 1: Squared chord distance from Lake Żabińskie fossil samples to nearest analogue in the Canadian-Polish calibration set. Analogue quality indicated by background colour estimated with the methods and thresholds used in the original paper. Calibration set samples with low count sums are omitted, but the results are similar if they are included.

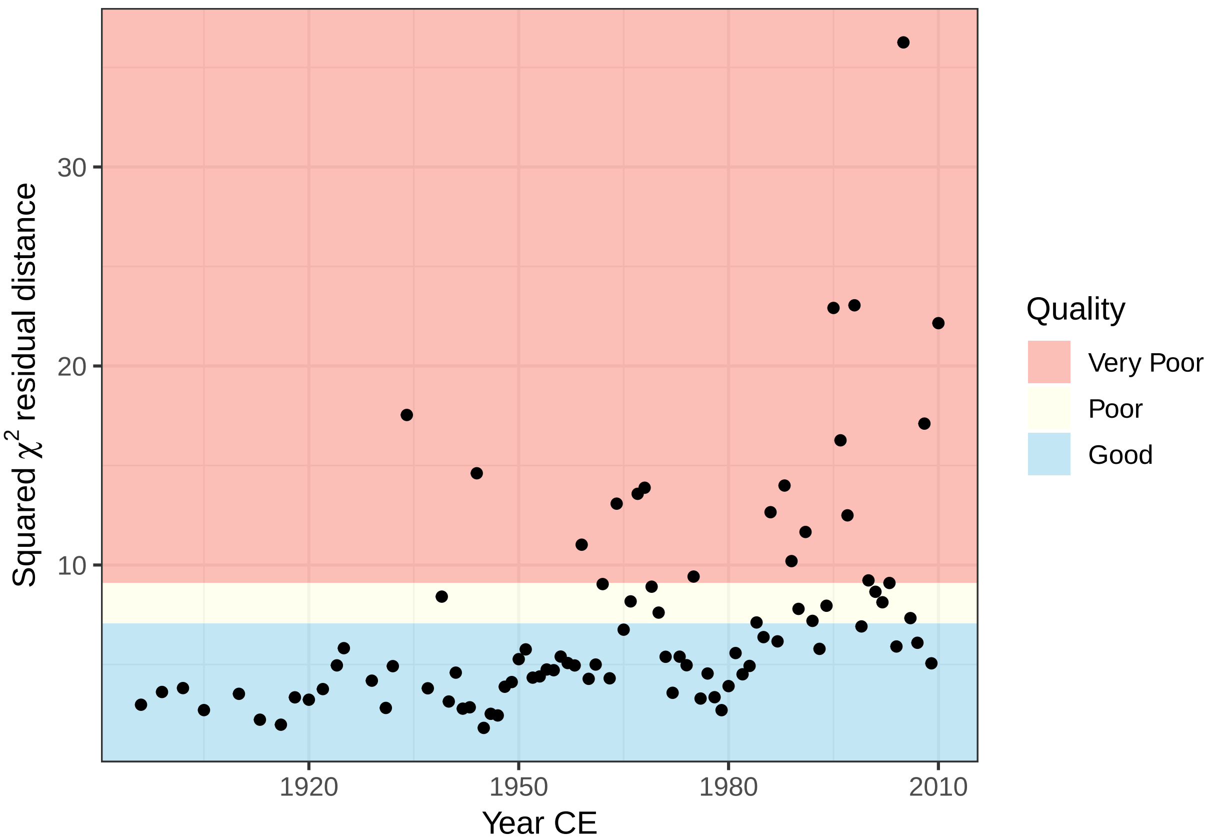

Figure 2: Goodness of fit of the Lake Żabińskie fossil samples added passively to a CCA of the Canadian-Polish calibration set constrained by temperature.

Abisko lakes

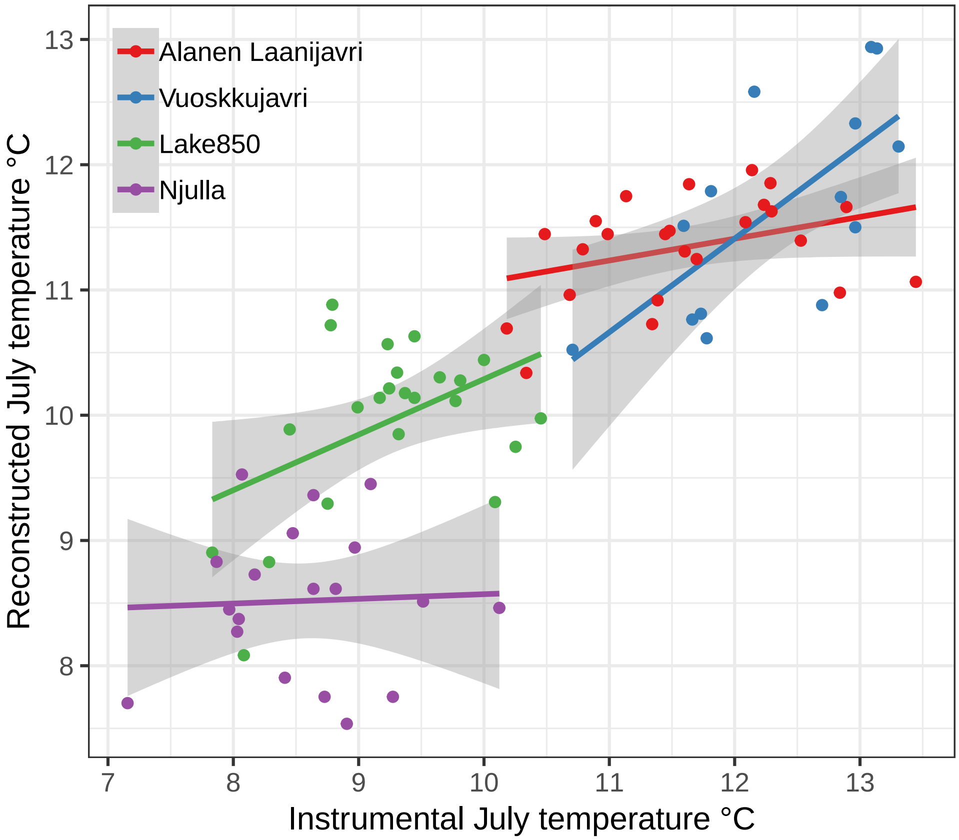

Figure 3: Twentieth century reconstructed and instrumental July air temperatures for four lakes near Abisko. Data digitised from the original paper which reported that all the lakes had a similar correlation between the reconstructed and instrumental record.

Lake Silvaplana

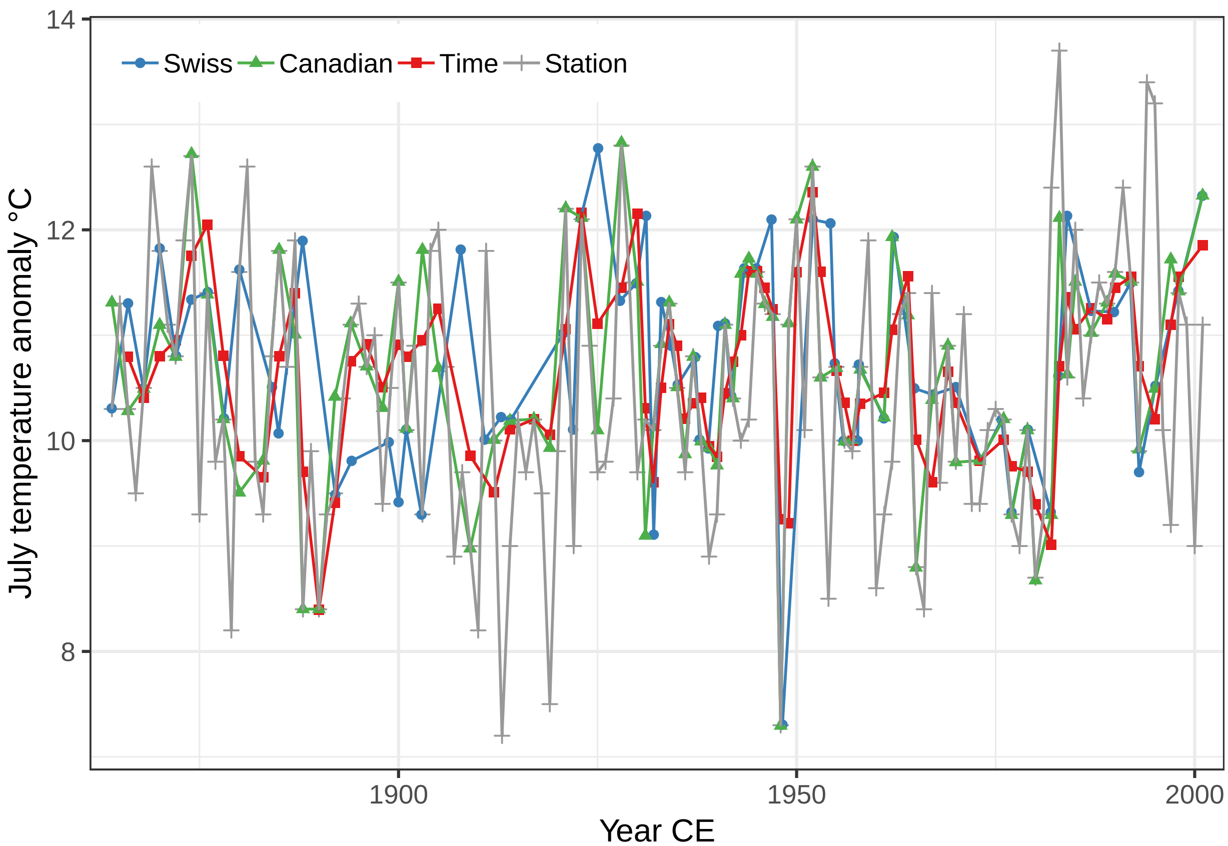

Figure 4: Instrumental July air temperatures from Selg-Maria (Begert et al., 2005) and the instrumental data as digitised from different papers. Digitised data are labelled after the calibration set used to make the reconstruction the instrumental data were compared against. All the digitised instrumental data should be identical, unless fossil samples have been merged/unmerged.

Long Swiss climate series were homogenised by Berget et al. (2005). They were reprocessed in 2009 (MeteoSwiss, 2010). The Silvaplana papers probably all used the version of the Segl Maria series by Berget et al. (2005), which is shown in Fig. S4. The differences between these data and the current version of the data are minimal (r^2 = 0.997).

Seebergsee

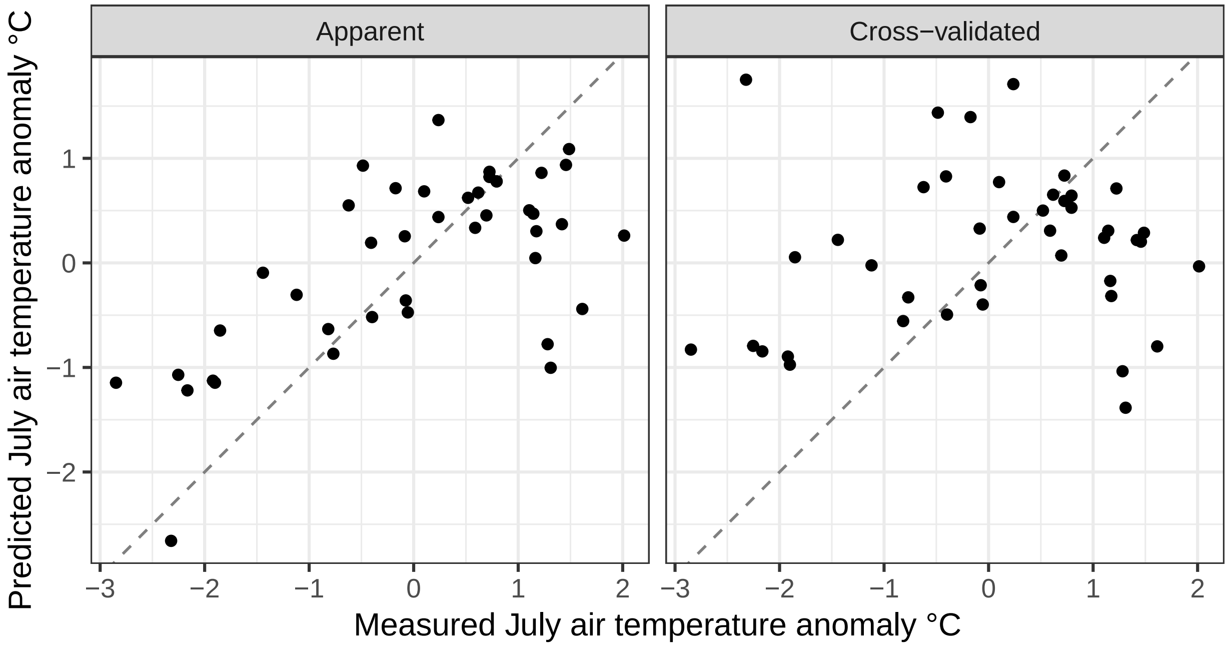

Figure 5: Apparent and cross-validated performance of the calibration-in-time WAPLS-2 model from Seebergsee.

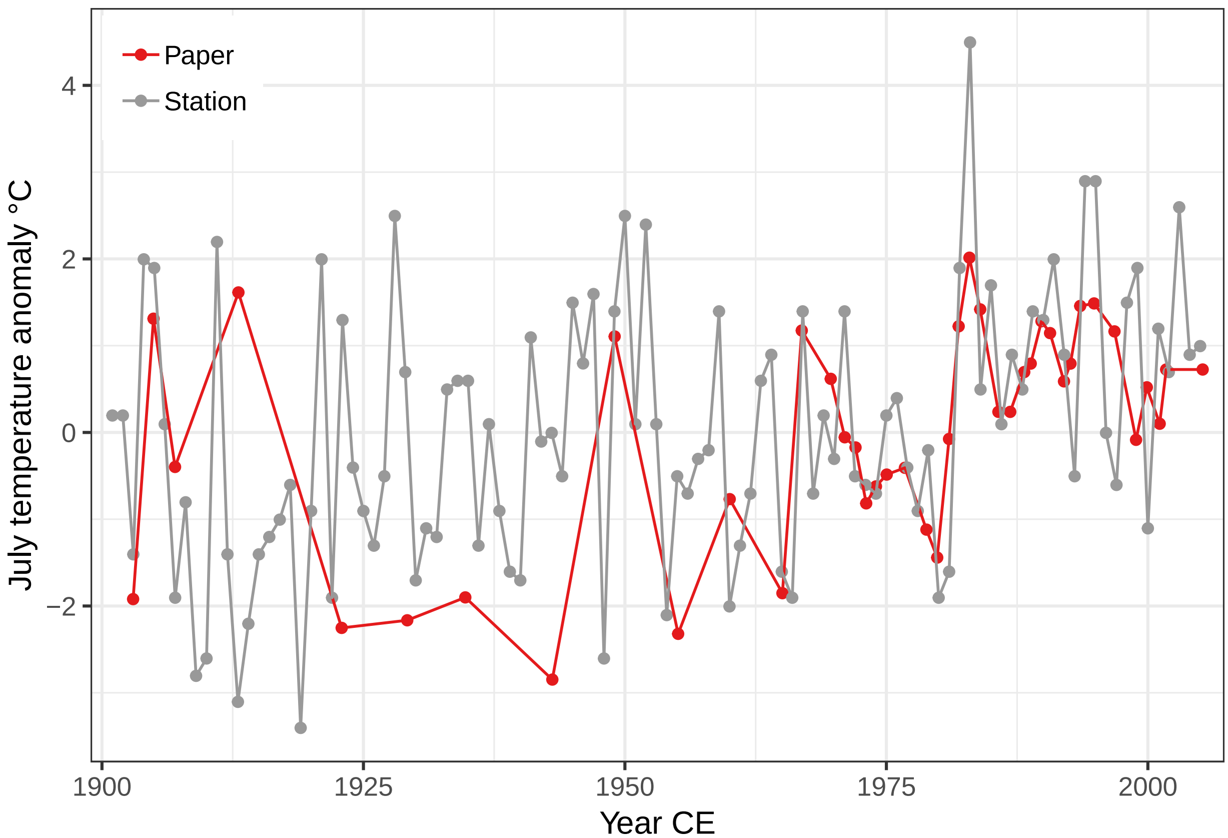

Figure 6: Instrumental July air temperature anomalies from Château-d'Oex (MeteoSwiss, 2010) and the instrumental data as digitised from the Seebergsee paper.

There may be, probably small, differences between the data shown in Fig. S6 and the version of the Château-d'Oex temperature series homogenised by Berget et al. (2005).

Nurmijärvi

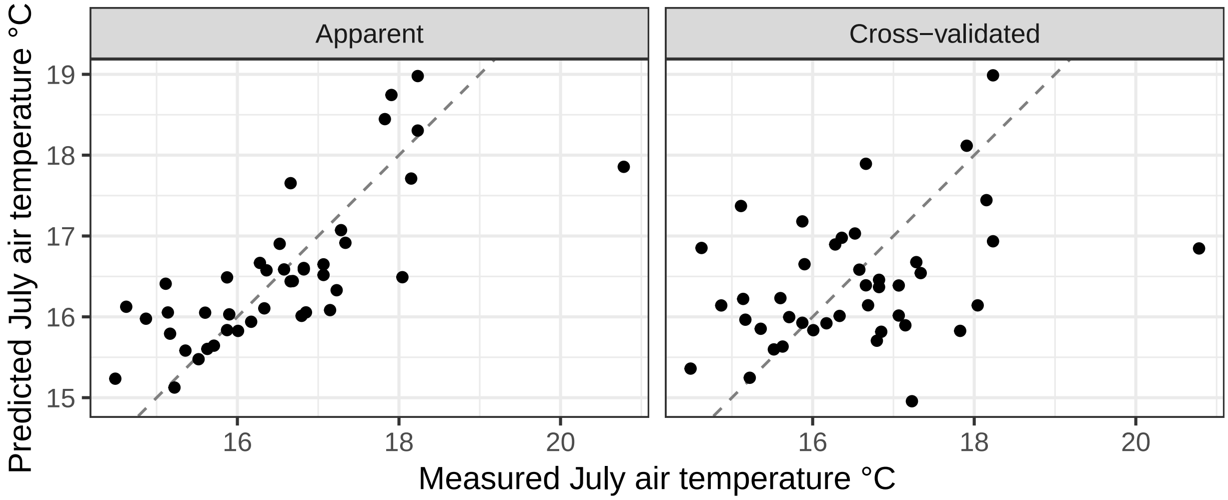

Figure 7: Apparent and cross-validated performance of the calibration-in-time tolerance-weighted weighted averaging model from Nurmijärvi.

Speke Hall Lake

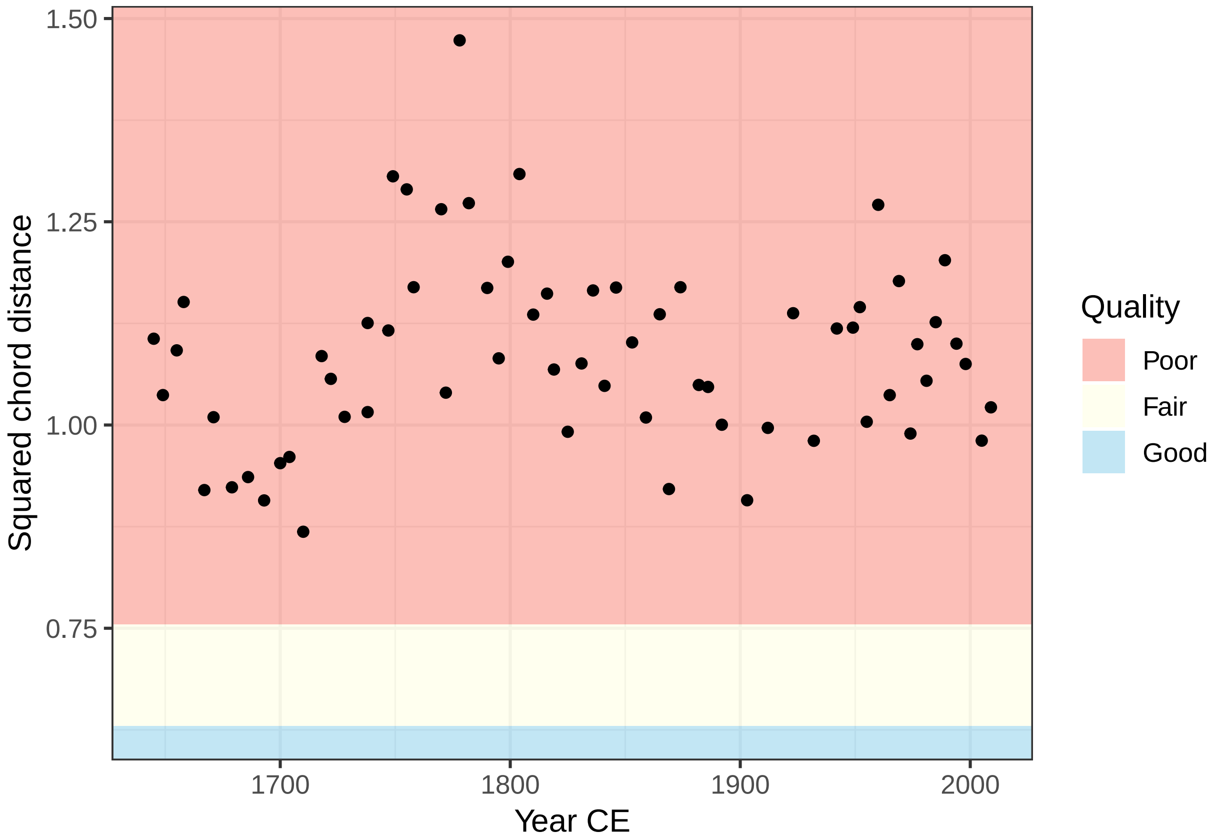

Figure 8: Squared chord distance from Speke Hall Lake fossil samples to nearest analogue in the Norwegian calibration set. Analogue quality indicated by background colour. Thresholds derived from quantiles of all calibration set distances (Good < 0.05, Fair < 0.1).

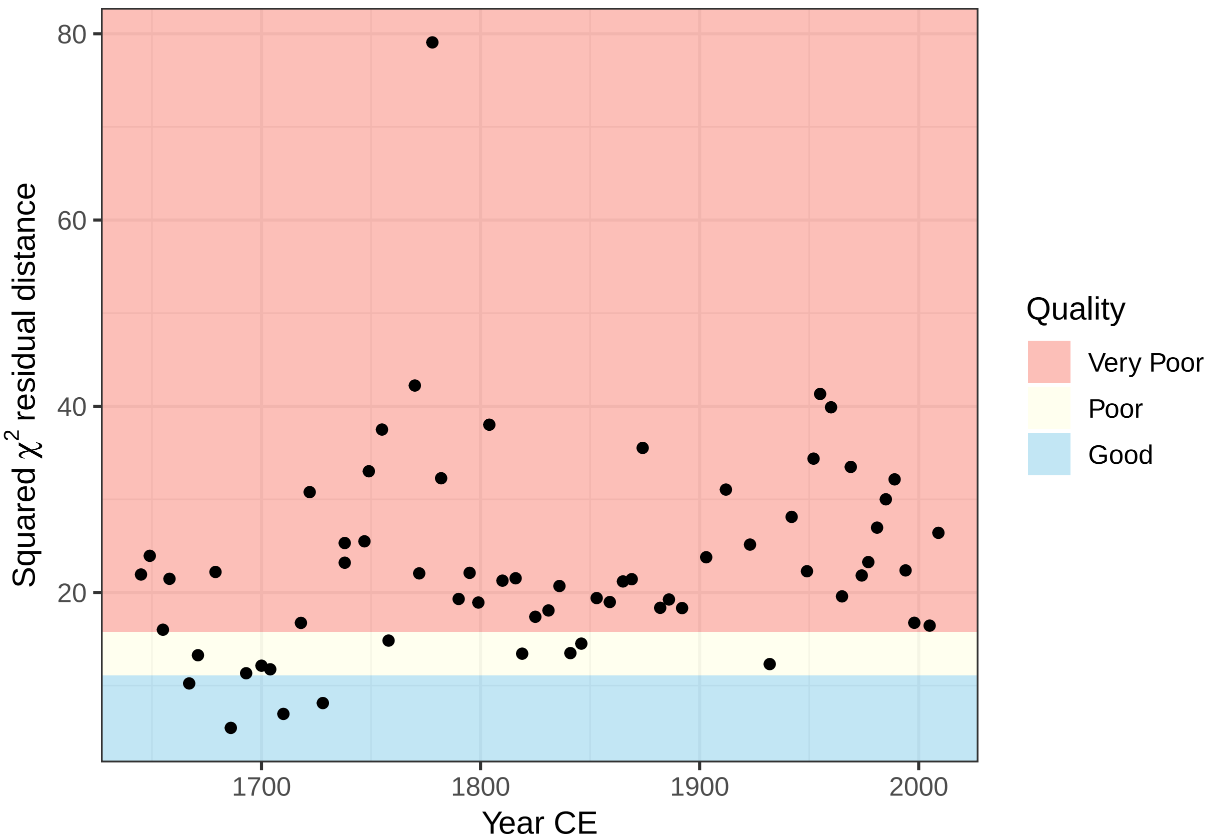

Figure 9: Goodness of fit of the Speke Hall Lake fossil samples added passively to a CCA of the Norwegian calibration set constrained by temperature. Thresholds derived from quantiles of all calibration set squared residual lengths (Good < 0.9, Fair < 0.95). An equivalent figure in the original paper used an inappropriate data tranformation prior to the analysis.

References

Begert, M., Schlegel, T., Kirchhofer, W., 2005. Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. International Journal of Climatology 25, 65–80. https://doi.org/10.1002/joc.1118

MeteoSwiss, 2010. Originale und homogene reihen im vergleich. https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/klima/klima-im-detail/doc/klima-vergleich-original-homogen.pdf