

# Testing intra-site transfer functions: an example using chironomids and water depth

Gaute Velle · Richard J. Telford · Oliver Heiri ·  
Joshua Kurek · H. John B. Birks

Received: 21 August 2011 / Accepted: 28 June 2012 / Published online: 29 July 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** Most calibration data sets used to infer past environmental conditions from biological proxies are derived from many sites. An alternative strategy is to derive the calibration data set from within a single site. Transfer functions derived from such intra-site calibration data sets are usually applied to fossil assemblages from the focal lake, but a recent development has been to apply these transfer functions to other sites. Transfer functions derived from intra-site calibration data sets can have impressive cross-validation performance, but that gives little indication of their performance when applied to other sites. Here, we develop transfer functions for lake depth from intra-lake chironomid calibration data sets in Norway and Alaska and test the resulting models by cross-validation and against known depth in external lakes.

Lake depth is a statistically significant predictor of chironomid assemblages at all these lakes, and most intra-lake transfer functions perform reasonably well under cross-validation, but their performance against external data is erratic. Downcore reconstructions from transfer functions developed on different lakes are dissimilar. Ignoring the poorly performing transfer functions, only 3 of 14 downcore reconstructions are statistically significant. Few assemblages downcore had good modern analogues in the calibration data set, even when the core was from the same lake as the calibration data set. We conclude that intra-site calibration data sets can find site-specific rather than general relationships between species and the environment and thus should be applied with care and to external sites only after careful and critical validation.

---

G. Velle (✉)  
Bergen Museum, University of Bergen, Post Box 7800,  
5020 Bergen, Norway  
e-mail: gaute.velle@bio.uib.no

G. Velle  
Uni Environment, Uni Research, Thormøhlensgt. 49b,  
5006 Bergen, Norway

R. J. Telford · H. J. B. Birks  
Department of Biology, University of Bergen,  
Post Box 7803, 5020 Bergen, Norway

R. J. Telford · H. J. B. Birks  
Bjerknes Centre for Climate Research, Allégaten 55,  
5007 Bergen, Norway

O. Heiri  
Institute of Plant Sciences and Oeschger Centre  
for Climate Change Research, University of Bern,  
Altenbergrain 21, 3013 Bern, Switzerland

J. Kurek  
Department of Biology, Queen's University,  
116 Barrie St., Kingston, ON K7L 3N6, Canada

H. J. B. Birks  
School of Geography and the Environment,  
University of Oxford, Oxford OX1 3QY, UK

H. J. B. Birks  
Environmental Change Research Centre, University  
College London, London WC1E 6BT, UK

**Keywords** Calibration data set · Cross-validation · Model performance · Sub-fossils · Analogue matching

## Introduction

Transfer functions provide a means of inferring past environmental conditions quantitatively from fossil biota based on the relationship between taxa and the environment as observed in a modern calibration data-set (Birks 1995; Birks et al. 2010). Taxa responses may be statistically significant along several environmental gradients in modern calibration data sets. However, statistical significance of relationships in calibration data does not necessarily imply that these environmental variables are ecologically important or will produce useful reconstructions at particular sites (Anderson 2000; Birks et al. 2010; Brodersen and Anderson 2002; Telford and Birks 2011; Telford 2006; Velle et al. 2005a; Velle et al. 2010a, b; Velle et al. in press). Most calibration data sets are derived from observations from multiple sites (e.g. typically 40–160 sites), with one observation per site (Birks 1995). An alternative strategy is to develop intra-site calibration data sets and transfer functions and exploit intra-lake gradients, as suggested by Birks (1998). Intra-site calibration data sets have many observations (typically 15–40) from a single site and are being increasingly used as a paleoenvironmental tool (Engels and Cwynar 2011; Kurek and Cwynar 2009a, b; Laird et al. 2011; Larocque-Tobler et al. 2011; Mackay et al. 2003, 2005; Mitchell et al. 2008).

There are many motivations for using intra-site calibration data sets. They have been used when environmental gradients within sites, such as water depth, are large relative to the differences between sites. In contrast, many variables reconstructed using multi-lake calibration data sets, such as pH, are effectively constant at each site. In intra-site calibration data sets, secondary gradients, such as catchment-related variables, are minimized. Variables such as water depth are an intrinsic morphometric feature of the lake, and transfer functions from multi-lake calibration data sets can give physically impossible values (Birks 1998). Other potential motivations are the relatively small sampling effort in generating an intra-site calibration data set compared to the sampling effort required for multi-lake calibration data sets, and the very good performance statistics reported

for intra-site data sets (Kurek and Cwynar 2009b). Diatom-based multi-lake depth models have been found to not always be applicable to a particular sediment record because of complex interactions between the distribution of diatoms with depth and the distribution of aquatic macrophytes, water clarity, and other variables (Brugam et al. 1998; Moser et al. 2000). Because of this, a number of studies have used the relationship between the distribution of diatom species assemblages and water depth within a lake as a basis for reconstruction of changes in past lake levels (Brugam et al. 1998; Laird and Cumming 2009; Laird et al. 2011).

There are grounds for suspecting that the strong cross-validation performance of an intra-lake transfer function is not always reflected in the quality of the reconstruction. Multi-site transfer functions assess the average influence of the environmental variable of interest on species assemblages across a range of secondary nuisance gradients. In contrast, intra-site calibration data sets minimise nuisance gradients that vary between sites, such as pH and mean air temperature. This may result in non-analogue assemblages in downcore records and unrealistic reconstructions if these variables have changed substantially over time. Nuisance variables that vary within a site are likely to co-vary with the variable of interest (Kurek and Cwynar 2009b). If this covariance structure is not stable over time, any reconstruction may be in error (see discussion for assumptions of transfer functions). The nuisance variables are also likely to be spatially autocorrelated. A second source of spatial autocorrelation is any mixing of biotic remains from different habitats within the site (van Hardenbroek et al. 2011). Spatial autocorrelation from either of these processes can cause transfer function performance statistics to be over-optimistic (Telford and Birks 2005).

Intra-site transfer functions have mainly been developed to track past changes in lake depth (Kurek and Cwynar 2009a, b; Laird et al. 2011; Luoto 2010), but have also been used to reconstruct distance to littoral vegetation (Luoto 2010), lotic index—a 1/0 index of turbulence (Luoto 2010), snow thickness (Mackay et al. 2003), and bog water-table depth (Mitchell et al. 2008). Calibration-in-time (i.e., time series of biota calibrated against environmental data from the same time period) has been developed for chironomids and temperature (Larocque-Tobler et al. 2011) and forms a special case of intra-site calibration

data sets. Most intra-lake transfer functions have been used to infer environmental variables from sediments cored in the transfer function lake (the focal lake) (Kurek and Cwynar 2009b; Larocque-Tobler et al. 2011; Luoto and Helama 2010; Mackay et al. 2005). However, Luoto (2010) extended the methodology by using intra-site transfer function to reconstruct three environment variables (lotic index, water depth, distance from littoral vegetation) by use of one proxy in a single core in a non-focal lake.

Intra-site calibration data sets have never been rigorously tested, although Kurek and Cwynar (2009b) highlight several of the challenges that exist with their downcore application. It is not known if the cross-validation performance of an intra-site transfer function is a realistic guide to the quality of a resulting reconstruction. Nor has it been tested if meaningful reconstructions can be expected when intra-site transfer functions are applied to non-focal sites. We aim to validate intra-site transfer functions by testing their performance under cross-validations and when applied to assemblages from known conditions in non-focal sites as external validation. We use five intra-lake chironomid water depth transfer functions from Southern Norway and four from sub-Arctic Alaska to infer known water depth from multiple samples in each test lake. The transfer functions are also applied to four Holocene sequences and the reconstructions from the different models are compared and tested for statistical significance of the inferred variations in water depth. Our results are

expected to provide insights that are applicable for proxies and environmental variables other than chironomids and water depth.

## Materials and methods

### Data sets

We used chironomid larval head capsules from within-lake surface samples collected in five lakes in Southern Norway (Holebudalen, Histøl, Lisletjønn, Råtåsjøen, and Vestre Økjamyrtjern; Heiri 2004; Heiri et al. 2003) and four lakes in sub-Arctic Alaska (No Itch, Puyuk, Zagoskin, and Red Cliff; Kurek and Cwynar 2009a, b) (Table 1). Intra-site transfer functions were produced from each of these lakes and tested at the other lakes in the same region. In addition, the intra-site transfer functions were applied downcore to Holocene sequences of fossil chironomids from Råtåsjøen, Holebudalen, Vestre Økjamyrtjern, and Zagoskin. The Holocene cores were retrieved from the deepest part of the lakes and have lengths of 197, 250, 301 and 306 cm, respectively. For details on sampling procedure, lake morphology, and taxonomy, see Heiri (2004) and Heiri et al. (2003) for the Norwegian intra-lake data; Kurek and Cwynar (2009a, b) for the Alaskan intra-lake data; Velle et al. (2005a, b) for the Norwegian downcore data; and Kurek et al. (2009) for the downcore Alaskan data. The taxonomy of the data sets was harmonized within each region prior to

**Table 1** Study lake characteristics

Lake	Code	Region	Elevation (m a.s.l.)	Vegetation	July air T (°C)	Area (km <sup>2</sup> )	<i>n</i>	Max/min depth (m)	S/mean N2
Histøl	HIS	Norway	245	c, d	14.3	0.048	20	9.2/0.4	70/13.2
Holebudalen	HOL	Norway	1,144	a	8.2	0.03	21	7.5/1.2	34/6.7
Lisletjønn	LIS	Norway	518	c	12.9	0.044	20	7.3/0.6	52/11.8
Råtåsjøen	RAT	Norway	1,169	a	8.6	0.125	20	7.5/0.8	51/9.3
V. Økjamyrtjern	OYK	Norway	570	h, a	11.0	0.02	20	7.8/1.0	56/9.2
No Itch	NI	Alaska	22	st	~12	0.11	20	4.0/0.8	37/11.3
Puyuk	PL	Alaska	23	st	~12	0.24	33	18.5/1.3	34/6.3
Zagoskin	ZL	Alaska	5	st	~12	0.24	33	19.7/0.2	38/6.1
Red Cliff	RC	Alaska	62	st	~12	0.35	25	34.5/3.0	26/4.8

Vegetation type: *c* coniferous, *d* deciduous, *a* alpine, *h* heathland, *st* shrub tundra. *n* is the number of intra-site observations, Max/min depth are the maximum depth and the minimum depth of the intra-site observations, S is the number of taxa and N2 is the effective number of taxa in the sample (Hill's N2)

numerical analysis. Only taxa that occurred in two or more observations were included in analyses. The relative abundance of taxa in the assemblages was square-root transformed prior to analysis.

### Numerical analyses

For each lake, we used redundancy analysis (RDA) to assess the statistical significance of lake depth as an explanatory variable of the chironomid assemblages. We used the ratio of the first, constrained, eigenvalue ( $\lambda_1$ ) of the RDA to the second, unconstrained, eigenvalue ( $\lambda_2$ ) as an indicator of the importance of depth. When the  $\lambda_1/\lambda_2$  ratio is greater than one, the transfer functions are likely to be useful (ter Braak 1987). Dixit et al. (1993) suggest that variables which do not fulfil ter Braak's criterion may still be useful for calibration purposes, but with greater uncertainty. The  $\lambda_1/\lambda_2$  ratio only indicates the importance of the environmental variable of interest in the calibration data set: a high ratio does not imply that the palaeoenvironmental reconstruction is correct. At sites where depth was a statistically significant predictor of assemblage composition, intra-lake transfer functions were developed using weighted averaging with inverse deshrinking ( $WA_{inv}$ ).  $WA_{inv}$  was chosen as this transfer function method is fairly robust to the spatial autocorrelation that is inevitable in an intra-lake calibration data set (Telford and Birks 2005). The performance of the resulting transfer functions was assessed with leave-one-out cross-validation. The utility of each transfer function when applied to external sites was established with external validation by comparing the measured and predicted depth of observations in the other lakes.

Spatial coordinates of the sampling locations were only readily available from the Alaskan calibration data sets. We checked for the influence of spatial autocorrelation on the transfer functions by plotting variograms of the detrended  $WA_{inv}$  residuals (Legendre and Legendre 1998; Telford and Birks 2009).

The intra-lake transfer functions developed here were used to reconstruct water depth from three Norwegian and one Alaskan Holocene stratigraphies. The reconstructions from the different models were compared, and their statistical significance was evaluated using both methods developed by Telford and

Birks (2011). Under the first test, a reconstruction is statistically significant if it explains more of the variance in the fossil data than 95 % of reconstructions derived from random environmental variables. Under the second test, a reconstruction is statistically significant if the weighted correlation between the environmental optima of the taxa in the calibration data set and their axis 1 scores in a RDA of fossil data constrained by the reconstruction is higher than it is for 95 % of trials using random environmental data. For both tests we use 999 trials.

Transfer functions are less likely to produce reliable results if the assemblages at the site being reconstructed do not have good analogues in the calibration data set (Birks 1998). We evaluated dissimilarities between the assemblages with two approaches. First, observations lacking analogues in the calibration data set were identified using squared chord distance as a measure of dissimilarity and a 50 % threshold, a much more liberal threshold than usually used (Simpson 2007). Second, the taxonomic distance among the assemblages was visualized using non-metric multi-dimensional scaling (NMDS) with the Bray-Curtis dissimilarity.

All computations were performed using the statistical language R version 2.11.1 (R Development Core Team 2010). We used the rioja package version 0.5-6 (Juggins 2009) for transfer function development; the vegan package version 1.17-2 (Oksanen et al. 2011) for ordinations, and palaeoSig version 1.0 (Telford 2011) to estimate the statistical significance of reconstructions.

## Results

### Transfer functions

Water depth was a significant environmental variable explaining the variation in the chironomid assemblages in all intra-site data sets. Depth explained between 9.4 % (NI) and 31.9 % (RAT) of the variance in the chironomid relative abundance data (Table 2). The two lakes with the best fit between depth and fauna within each region were also the largest in area. For several of the lakes, the  $\lambda_1/\lambda_2$  ratio was high. For the Norwegian lakes, all but OYK produced transfer functions with reasonable cross-validation performance with RAT

**Table 2** The importance of depth as predictor for the chironomid assemblages at five Norwegian lakes (HIS, HOL, LIS, RAT, OYK) and four Alaska lakes (NI, PL, ZL, RC) as analyzed by RDA

	Explained (%)	$\lambda_1/\lambda_2$	<i>p</i>
HIS	21.7	0.99	0.005
HOL	24.9	1.49	0.005
LIS	13.7	0.82	0.005
RAT	31.9	2.16	0.005
OYK	11.1	0.50	0.005
NI	9.4	0.38	0.030
PL	12.6	0.60	0.005
ZL	21.6	0.67	0.005
RC	30.3	1.68	0.005

Explained (%) is the proportion of variance explained,  $\lambda_1/\lambda_2$  is the ratio of the constrained eigenvalue to the first unconstrained eigenvalue, and *p* denotes the significance level as assessed by a Monte Carlo permutation test

performing best (Table 3). In Alaska, ZL, RC, and possibly PL produced apparently useful transfer functions with RC the best (Table 4).

External validation

When the intra-site depth transfer functions were validated externally against the test lakes, the results were variable. The NMDS (Fig. 1) suggests that calibration data sets from the Alaska lakes ZL, RC, and PL and the Norwegian lakes HIS and LIS can be used in at least some of the test lakes within their region. The assemblages from HOL and NI are taxonomically dissimilar from the other lakes and inferences involving these calibration data sets should be interpreted with caution. The model from HIS had good performance statistics under cross-validation and when validated at LIS (Table 3), as did the model from RAT. RC seemingly performed best of the Alaska transfer functions. When tested externally, this model’s  $r^2$  was only slightly worse in PL than under cross-validation, but the RMSEP is almost double. Somewhat surprisingly, the inference model from LIS had better statistical performance when validated against HIS and RAT than it did under cross-validation (Table 3, Fig. 2). Likewise, the inference model from PL seems inadequate when cross-validated

**Table 3** Calibration data set performance ( $r^2$ /RMSEP/proportion of observations with analogues in the calibration data set lakes at a 50 % threshold) using  $WA_{inv}$  when the Norwegian intra-site transfer functions are used to infer depths in test lakes

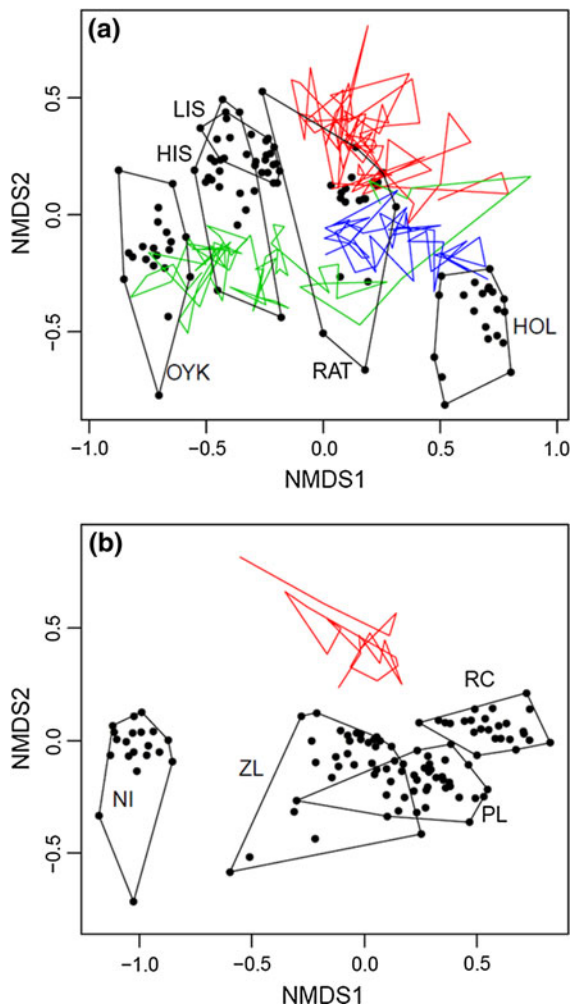
	Test lake ( $r^2$ /RMSEP/proportion of lakes with good analogues)				
	HIS	HOL	LIS	RAT	OYK
Transfer function lake					
HIS	<b>0.78/1.6/–</b>	0.16/5.8/0	0.58/3.2/0.75	0.08/2.7/0	0.18/4/0
HOL	0.43/4.3/0	<b>0.64/1.4/–</b>	0.00/2.5/0	0.22/2.2/0	0.24/4.4/0
LIS	0.8/1.6/0.05	0.15/3.0/0	<b>0.52/1.7/–</b>	0.81/1.9/0	0.39/7.3/0
RAT	0.15/4.1/0	0.64/2.9/0	0.31/2/0	<b>0.8/1.1/–</b>	0.02/2.5/0
OYK	0.43/2.7/0	0.13/2.7/0	0.00/2.9/0	0.06/3.1/0	<b>0.01/2.9/–</b>

Numbers in bold indicate the leave-one-out cross-validation performance. RMSEP are in m

**Table 4** Calibration data set performance ( $r^2$ /RMSEP/proportion of observations with analogues in the calibration data set lakes at a 50 % threshold) using  $WA_{inv}$  when the Alaska intra-site transfer functions are used to infer depths in the test lakes

	Test lake ( $r^2$ /RMSEP/proportion of lakes with good analogues)			
	NI	PL	ZL	RC
Transfer function lake				
NI	<b>0.05/1.2/–</b>	0.1/9.4/0	0.14/9.7/0	0/16.9/0
PL	0.03/14.7/0	<b>0.13/5.1/–</b>	0.46/4.6/0.24	0.57/11.7/0.36
ZL	0/2.8/0	0.44/4.1/0.67	<b>0.49/4.3/–</b>	0.47/11.6/0.04
RC	0.31/18.9/0	0.3/4.7/0.06	0.61/4.2/0	<b>0.64/5.8/–</b>

Numbers in italics indicate the model’s leave-one-out cross-validation performance. RMSEP are in m



**Fig. 1** Non-metric multidimensional scaling (NMDS) of the modern taxon assemblages with fossil samples included as trajectories; **a** the Norwegian sites Holebudalen (HOL), Histøl (HIS), Lisletjønn (LIS), Råtåsjøen (RAT), and Vestre Økjamyrjern (OYK), with the fossil samples from RAT (*red*), OYK (*green*), and HOL (*blue*). **b** The Alaska sites No Itch (NI), Puyuk (PL), Zagoskin (ZL), and Red Cliff (RC), with fossil samples from ZL (*red*). (Color figure online)

(Table 4), but performed considerably better at ZL and RC (Fig. 3). Most other external validations show poor performance.

#### Spatial autocorrelation

Variograms of the detrended WA residuals for the Alaskan sites show little evidence of spatial autocorrelation (not shown), suggesting that spatial autocorrelation is not a major problem at these sites for depth

transfer functions with the current sampling density. If the sampling density was greater, the environmental gradient less ecologically important, or the lakes more complex, spatial autocorrelation might have been more of a problem.

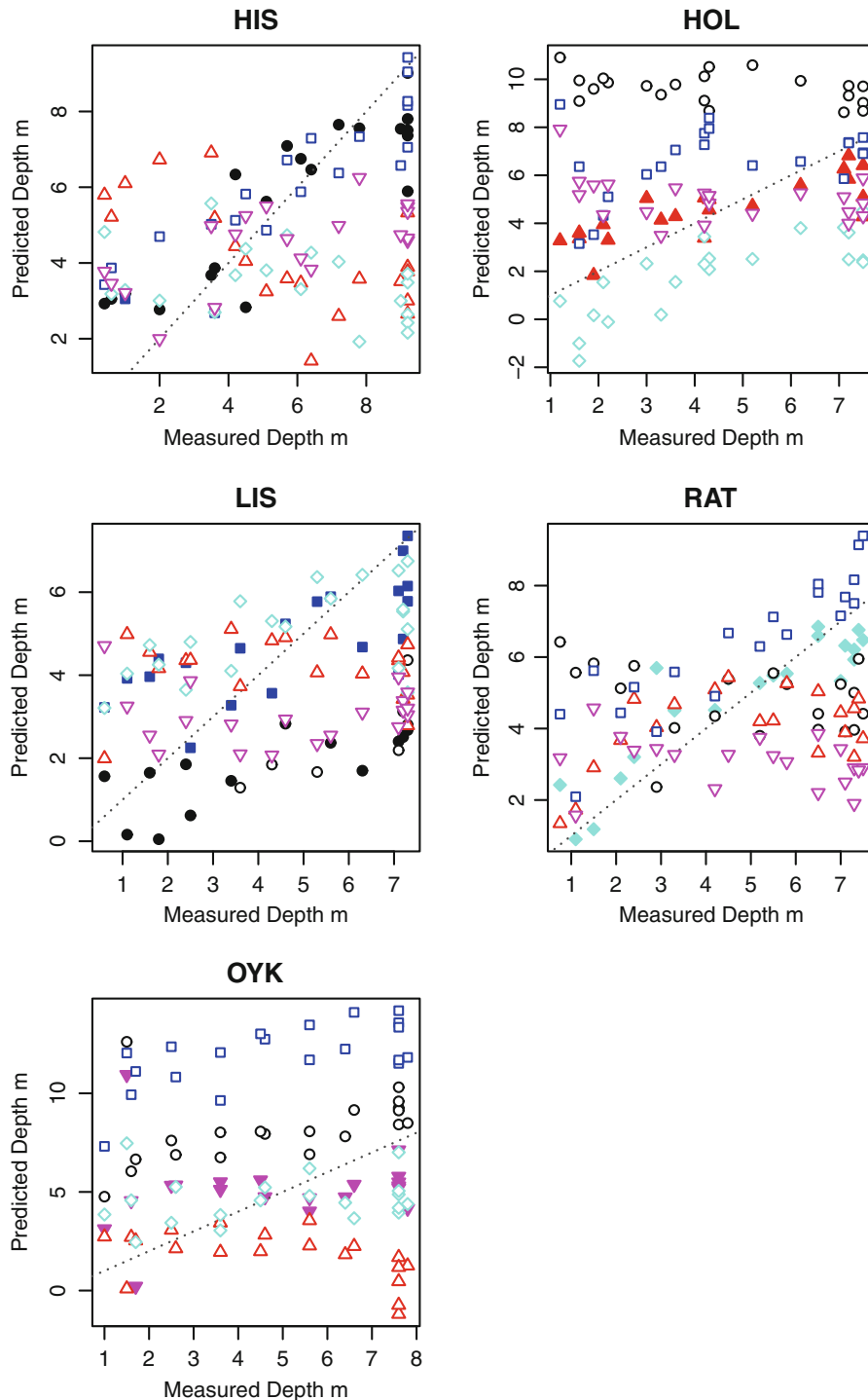
#### Holocene depth reconstructions

Ignoring the poorly performing transfer functions from OYK, NI, and PL, a total of 14 downcore water depth reconstructions were performed for fossil stratigraphies from RAT, HOL, OYK, and ZL. No downcore observations have analogues in the calibration data sets using a dissimilarity threshold of 10 % and few have analogues using a dissimilarity threshold of 50 % (Fig. 4). Only three of the reconstructions are statistically significant (Table 5). Significant downcore reconstructions were obtained when the inference model from HIS was used at HOL and OYK, and when HOL was employed at OYK. Reconstructions are shown in Fig. 4. With the exception of ZL, reconstructions within any one lake are inconsistent; they differ by more than the RMSEP and several have opposite trends. Some reconstructions generate physically impossible values with depths less than zero or greater than the lake's threshold. Many depths are under-estimated (see discussion on the edge effect below).

#### Discussion

##### Can intra-site transfer functions be used at non-focal sites?

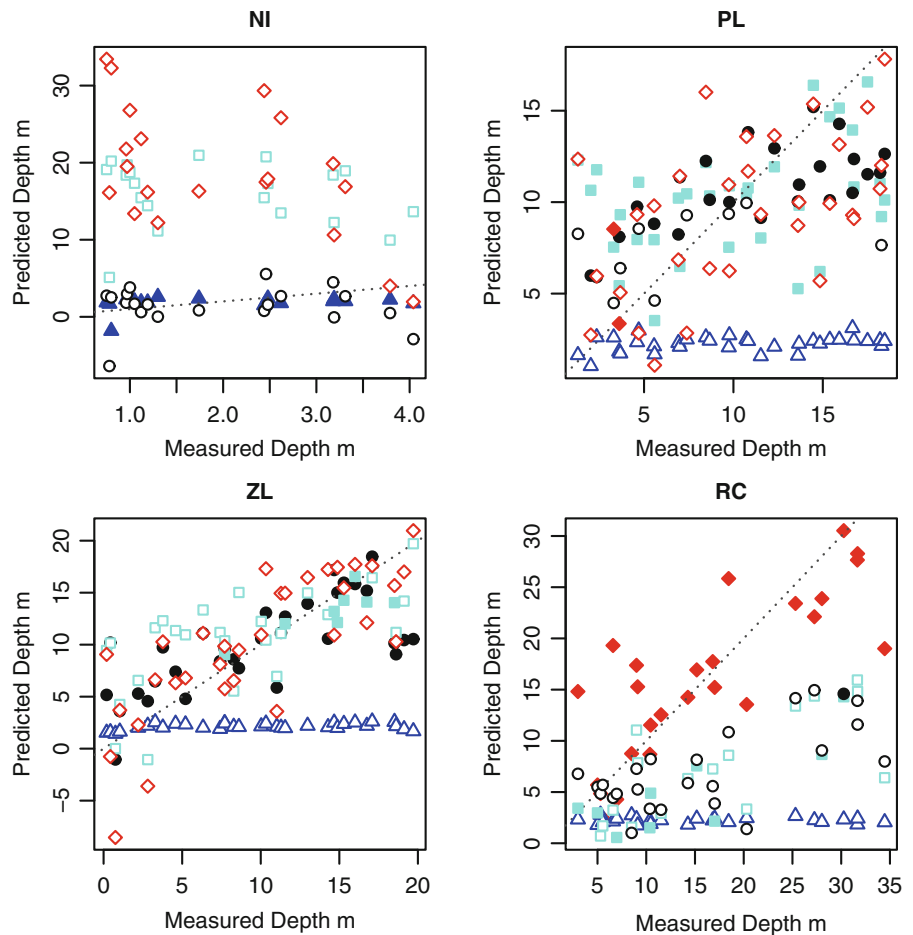
The intra-site transfer functions in this study perform well in some lakes and poorly at others. It is not straightforward to predict at which sites a model is likely to return reliable inferences. The strength of the relationships between water depth and fauna in a lake, estimated either by constrained ordination or by the cross-validation performance of the transfer function does not give any indication of how well the model performs externally. Taxonomic distance between sites can be used to predict where a calibration data set is not likely to return reliable results. However, closeness in taxonomic space does not guarantee reliable inferences. The model from RAT outperforms published multiple-lake regional depth transfer



**Fig. 2** Predicted against measured lake water depth using intra-site transfer functions in Norway for Histøl (HIS: black circles), Holebudalen (HOL: red triangles), Lisletjønn (LIS: dark blue squares), Råtåsjøen (RAT: pale blue diamonds), and Vestre

Økjamyr tjern (OYK: purple inverted triangles). Observations without analogues in the calibration data set are shown as hollow symbols. The 1:1 line is shown as a dotted line. (Color figure online)

**Fig. 3** Predicted against measured lake water depth using intra-site transfer functions for Alaska for No Itch (NI: dark blue triangles), Puyuk (PL: pale blue squares), Zagoskin (ZL: black circles), and Red Cliff (RC: red diamonds). Observations without analogues in the calibration data set are shown as hollow symbols. The 1:1 line is shown as a dotted line. (Color figure online)

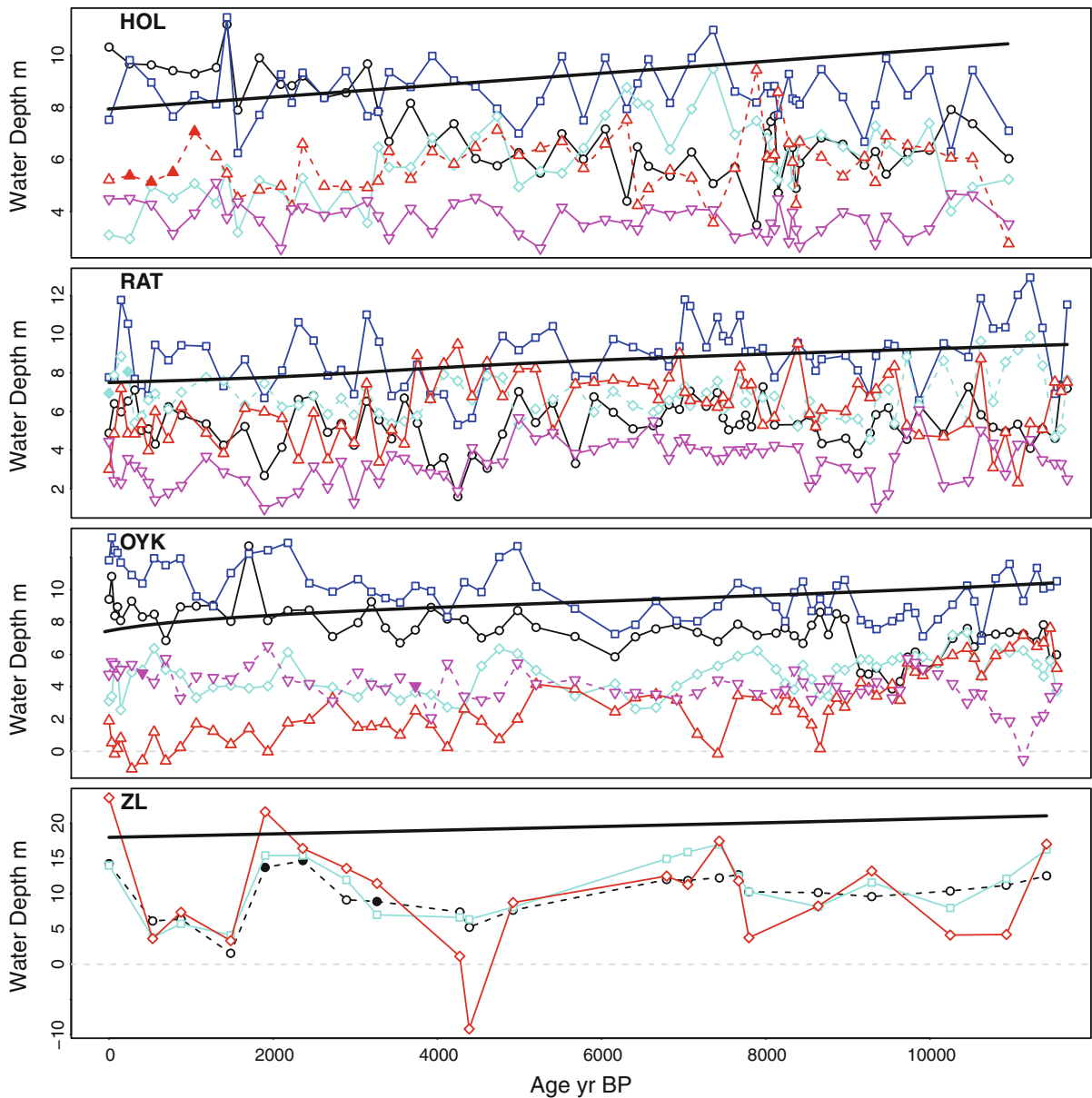


functions, such as those developed in Finland (Korhola et al. 2000; Luoto 2009) and north-west North America (Barley et al. 2006). Yet, when evaluated externally against independent test-sites, the performance of the intra-site model decreases considerably. Even where external validation suggests reasonable numerical performance, such as the models from LIS and HIS do at RAT, application downcore returns inconsistent reconstructions that are not statistically significantly different from random (Table 5, Fig. 4). There are several possible reasons why the reconstructions are mostly not statistically significant, including the lack of good modern analogues, assemblages changes that might be driven by other environmental variables, and Holocene depth changes that might have been small in these lakes. In these lakes, the expected minimum change in lake depth over the Holocene is about equal to the depth of the sediment profile (Fig. 4). The most likely reason for the lack of analogues is the small amount of dispersion in the intra-lake calibration data

sets: the 50th percentile of squared chord distances range between 0.2 and 0.5 in the intra-lake calibration set, in contrast the 5th percentile of distances in a Norwegian multi-lake calibration data set (Brooks and Birks 2001) is 0.63. Consequently, even a small amount of taxonomic turnover downcore or among lakes will result in non-analogue assemblages.

In short, the results of our analyses of intra-site transfer functions at non-focal sites are not encouraging: validation of the transfer functions at locations with known depth at external sites gives poor performance in most cases. Few reconstructions downcore are statistically significant and several are physically implausible. Even though the downcore reconstructions at ZL are consistent using the intra-site transfer functions, the reconstructions are not consistent with the reconstructions derived from a local (three lakes and 87 samples) or a regional (136 lakes) transfer function (Kurek and Cwynar 2009b). This poor performance is not unexpected given the lack of good





**Fig. 4** Holocene changes in lake water-depth inferred for Holebudalen (HOL), Råtåsjøen (RAT), Vestre Øykjamyrtjern (OYK), and Zagoskin (ZL) using the intra-site transfer functions from Råtåsjøen (pale blue diamonds), LIS (dark blue squares), Histøl (black circles), Holebudalen (red triangles), Vestre Økjamyrtjern (purple inverted triangles), Zagoskin (black circles with dashed line), Red Cliff (red diamonds), and Puyuk

(pale blue squares). Observations with analogues at a 50 % threshold using squared chord distance are shown with filled symbols. The black solid line represents the present maximum depth and the theoretical maximum depth backwards in time, estimated using sedimentation rates from Kurek et al. (2009) and Velle et al. (2005a) and assuming the present lake bedrock threshold has remained constant. (Color figure online)

analogues, suggesting that the environment in the non-focal lakes was insufficiently similar to the calibration site. Another potential reason for the poor performance of intra-sites transfer functions when applied to non-focal sites is that any benefit that an intra-site

transfer function gains from being adapted to fit the intrinsic morphometric features of a lake (Birks 1998) is lost when it is applied to another lake. If intra-lake transfer functions are to be used in non-focal lakes, it is imperative that their performance is verified in this

**Table 5** Proportion of the variance in the fossil data explained by the depth reconstruction and the correlation between RDA axis 1 scores and species optima

	Norway	Test core lake				
		HOL	RAT	OYK	Alaska	ZL
Transfer function lake	HIS	0.24*/0.20	0.04/0.29	0.12/0.45*	NI	0.12/0.20
	HOL	0.05/0.17	0.06/0.36	0.25*/0.09	PL	0.16/0.12
	LIS	0.02/0.19	0.04/0.00	0.09/0.41	ZL	0.17/0.18
	RAT	0.16/0.30	0.08/0.54 <sup>+</sup>	0.11/0.05	RC	0.12/0.57
	OYK	0.06/0.04	0.06/0.33	0.10/0.01	–	–

\*  $p < 0.05$ , <sup>+</sup>  $p < 0.10$

non-focal lake by comparing predicted and measured environmental variables from modern surface samples and checking the fossil observations have good modern analogues. Validation with multiple proxies is also important and key to robust studies. This need for validation largely negates the advantage of using a transfer function derived from another lake as it would be little extra effort to construct a new intra-site calibration data set. Reconstruction derived using intra-site transfer functions in non-focal lakes without validation should be treated with caution.

Can intra-site transfer functions be used at focal sites?

Given the poor performances downcore, it appears that the cross-validation performance statistics for the intra-site transfer functions were overly-optimistic. This may indicate that some of the assumptions of transfer functions have been violated. Birks et al. (2010) list six assumptions; we argue below that these are at greater risk of being violated by intra-site transfer functions than by multi-site transfer functions.

1. The taxa in the calibration data set are systematically related to the environment

When creating a multi-lake calibration data set, it is often assumed that a single sediment sample from the centre of a lake integrates biota from the whole lake and represents an overall environmental feature of the lake. Conversely, an intra-lake calibration requires the assumption that the majority of biotic remains are deposited in situ, for if taphonomic processes redistribute fossils around the lake, the relationship between species and the local environment will be corrupted. The extent to which the fossil

assemblage at the sampling point represents the biota living at that point could be assessed by comparing living assemblages with fossil assemblages. Depending on lake bathymetry, fossil chironomid assemblages may be similar across the entire lake basin (Brodersen and Lindegaard 1999; Hofmann 1998; Lehman 1975; van Hardenbroek et al. 2011), or vary across the lake basin (Heiri 2004; Heiri et al. 2003; Kurek and Cwynar 2009a, b; Luoto 2010). The varying importance of taphonomic effects from lake to lake suggests that these processes should be understood before applying an intra-site calibration data set to a non-focal site. Also, if the degree to which assemblages are re-deposited has changed over time, perhaps due to changes in sediment focusing as the lake's morphology changes, this may bias reconstructions.

2. The environmental variable to be reconstructed is, or is linearly, or at least monotonically, related to an ecologically important determinant in the system of interest

The causal link between the biota and environmental variables being reconstructed is not always straightforward. Some environmental variables, such as oxygen concentration, directly affect the biota (Brodersen et al. 2004) whereas others, like lake depth, do not affect the biota directly, but have a relationship with other variables that do (Kernan et al. 2009). If this relationship is persistent in time and space, transfer functions will be robust. If, however, this relationship is transient, and at risk of breaking down, transfer functions with good cross-validation performance may be over-optimistic.

While not an inherent property of intra-site calibration data sets, many are used to reconstruct variables that do not have a direct effect on the biota. Chironomids, for example, respond to changes in depth

(Engels and Cwynar 2011; Lindegaard 1992), but are unlikely to be directly sensitive to the changes in water pressure that accompany changes in depth. Instead, they respond to a combination of variables that may correlate with depth, for example macrophyte and algal growth, water temperature, oxygen concentration, substrate, food quality or stability (Brodersen et al. 2001; Eggermont et al. 2007; Korhola et al. 2000; Larocque et al. 2001). Several studies conclude that the concentration of dissolved organic carbon (DOC) in lakes is a major driver determining light transparency and the depth of the euphotic zone and the epilimnion (Laird et al. 2011). If the relationship between variables that correlate with depth is not stable over time, either driven by external drivers or lake ontology, reconstructions may be erroneous (see assumption 5).

3. The taxa in the calibration data set are the same biological entities as in the fossil data and their ecological responses have not changed

In both the calibration data set and in the fossil samples, many taxonomic entities aggregate multiple taxa that cannot be distinguished based on morphological characters (Brooks et al. 2007), so-called cryptic species. A multi-site calibration data set might contain several cryptic species in a species aggregate, most likely resulting in a higher RMSEP than would have been achieved had a species level determination been possible (Heiri and Lotter 2010). An intra-lake calibration data set might contain only a single cryptic species from an aggregate, increasing the risk that a different cryptic species from the aggregate, with a different ecological niche, occurs in the fossil data. This could render the reconstructions erroneous.

4. The mathematical methods used adequately model the species responses and have sufficient predictive power to allow accurate and unbiased reconstructions

Intra-lake transfer functions may be at greater risk of statistical edge-effects than multi-lake calibration data sets. The ‘edge effect’ is a problem inherent in weighted averaging with an inverse regression, caused by the truncation of the taxa’s responses to the environmental gradient when a finite gradient is sampled. This moves the apparent optima towards the mean of the gradient in the calibration data set (Velle et al. 2011). Hence, taxa that occur near the end of the gradient are likely to have poorly estimated

optima. For intra-site depth calibration data sets, the implication is that lake-level fluctuations near the current maximum depth may be systematically underestimated, and extrapolation to greater depths is problematic. This problem can partly account for the underestimated Holocene depths in the present study (Fig. 4). Noise still predominates the depth reconstructions and under-estimation also occurs when training sets from deeper lakes are being used, such as when HIS is used at HOL. To some extent, the edge effect could be overcome by coring at intermediate depth in the lake, which is strongly recommended by Laird and Cumming (2008, 2009) who inferred depth from multiple cores within a lake.

5. Environmental variables other than the one of interest had negligible influence during the time window of interest, the joint distribution of these variables of interest in the past was the same as today, or their effect did not lead to past changes in assemblage states resembling shifts indicative in the modern environment of changes in the variable of interest

In intra-site calibration data sets, the influence of the environmental variable of interest on the species assemblages is assessed within one site. The advantage is that the effects of nuisance variables that vary between sites, such as air temperature and catchment characteristics, will be minimised, optimising the cross-validation performance of the transfer function. If, however, ecological important nuisance variables have changed over time, the environmental conditions that ensued when the fossil assemblages were deposited will not be found in the calibration data set, and non-analogue assemblages and poor reconstructions will be generated. In multi-site transfer functions, the influence of the environmental variables on the species assemblages is assessed across many sites, each with a different configuration of nuisance variables. The cross-validation performance is then likely to be worse (low precision), but reconstructions should be more robust to changes in nuisance variables (high accuracy). If it is reasonable *a priori* to assume that the changes in the environmental variables other than the one of interest will have either been small or maintained their joint distribution, it may be reasonable to use intra-site calibration data sets. This assumption is unlikely to be valid for any reconstruction in systems where there are strong external drivers of environmental change,

such as Holocene-scale reconstructions where there have been substantial changes in productivity (Velle et al. 2005b) and climate (Mayewski et al. 2004).

In addition to these general problems with intra-site calibration data sets, there are specific problems with attempts to reconstruct water depth. As discussed above, depth appears possible to reconstruct because of its correlation with variables such as temperature, oxygen availability, and food quality. It is not easy to identify or control for changes in these variables that are independent of depth changes. Many chironomid taxa, for example, are associated with macrophytes or feed on diatoms that are depending on light (Armitage et al. 1995). Macrophytes may be linked to water clarity and to water depth. If the macrophytes or euphotic zone migrate vertically through time according to altered water clarity independent of water depth, then some chironomids will follow. Because of these confounding factors, Korhola et al. (2000) and Brodersen and Quinlan (2006) called for caution when interpreting changes in the deep water/littoral fauna over time, and Kurek and Cwynar (2009b) suggest that past lake level fluctuations need to be of the order of metres to be reconstructed satisfactorily.

6. The test data in cross-validation are statistically independent of the remainder of the calibration data set

Spatial autocorrelation in the calibration data set can violate this assumption (Telford and Birks 2005, 2009). Positive spatial autocorrelation is the tendency of neighbouring observations to resemble one another more than expected by chance. It can be caused by spatial structure in environmental variables or through contagious biological processes such as dispersal. Spatial autocorrelation is not normally considered a problem for multi-site palaeolimnological calibration sets, although it can be for multi-site pollen calibrations data sets (Telford and Birks 2009). It is potentially a problem in intra-lake calibration sets since both depth and sediment re-deposition in small lakes is inevitably spatially autocorrelated, and many potential nuisance variables will be too. This may account for some of the higher model performance of cross-validation compared to external validation. However, our results show that spatial autocorrelation appeared to be a negligible problem for the Alaskan intra-lake WA transfer functions analysed here (Kurek and Cwynar 2009b). It might become a larger problem if there were more

observations from each lake, reducing the distance between adjacent observations; in more structurally complex lakes with stronger nuisance variables; or if WA-PLS with several components has been used as WA-PLS is less robust to spatial autocorrelation.

#### Violation of assumptions

Whichever of these violations of the assumptions of transfer functions is most critical for explaining the poor performance downcore may vary for different systems. For example, lateral homogenisation of assemblages is probably much less important in bogs than in lakes, and the importance of spatial autocorrelation will depend on both site and sampling characteristics. We used a liberal threshold for defining non-analogue assemblages and still found a prevalence of non-analogues. Given the frequency of non-analogues, even when intra-site calibration data sets are applied to their focal lake, we suspect that the assumption that other environmental variables have no influence is most critically violated. Also, the prevalence of non-analogue assemblages may suggest that the temporal community structure is not likely to follow a pattern similar to the spatial structure found today. Kurek and Cwynar (2009b) discuss how intra-site models may suffer from many potential limitations, such as (1) lower applicability to other lakes, (2) shorter environmental gradients, (3) lake-specific discrepancies in the modern distribution, and (4) transport, mixing, and re-suspension of head capsules. We concur with all these concerns.

#### Conclusions

Intra-site calibration data sets remain a useful tool for examining local patterns in biotic assemblages. However, intra-site transfer functions should only be used with great care outside the system where the model was calibrated. Good model performance within one site does not guarantee good model performance or valid palaeoenvironmental reconstructions at other sites. The model should be validated against modern samples with known environmental data at the site of interest and attention paid to the quality of analogues. Reconstructions from the focal site need to be treated with caution if large changes in environmental variables other than the variables of

interest are expected. All reconstructions should be critically assessed using multiple approaches. These conclusions are valid for relatively small lakes using chironomids and water depth as example. The general implication for other systems, proxies, and environmental variables is that the usefulness of intra-site transfer functions outside the system where the model was calibrated needs to be demonstrated rather than assumed.

**Acknowledgments** We wish to thank Stephen J. Brooks for providing fossil data from Holebudalen and for valuable comments on the manuscript. Les Cwynar is acknowledged for help generating the Alaska data. We also thank Steve Juggins and an anonymous reviewer for valuable comments. This work was supported by the Norwegian Research Council projects AQUACLIM and ARCTREC.

## References

- Anderson NJ (2000) Diatoms, temperature and climatic change. *Eur J Phycol* 35:307–314
- Armitage PD, Cranston PS, Pinder LCV (1995) *The Chironomidae: biology and ecology of non-biting midges*. Chapman & Hall, London
- Barley EM, Walker IR, Kurek J, Cwynar LC, Mathewes RW, Gajewski K, Finney BP (2006) A northwest North American training set: distribution of freshwater midges in relation to air temperature and lake depth. *J Paleolimnol* 36:295–314
- Birks HJB (1995) Quantitative palaeoenvironmental reconstructions. In: Maddy D, Brew JS (eds) *Statistical modelling of Quaternary science data*. Quaternary Research Association, Cambridge, pp 161–254
- Birks HJB (1998) Numerical tools in palaeolimnology—progress, potentialities, and problems. *J Paleolimnol* 20:307–332
- Birks HJB, Heiri O, Seppä H, Bjune A (2010) Strengths and weaknesses of quantitative climate reconstructions based on late-quaternary biological proxies. *Open Ecol J* 3:68–110
- Brodersen KP, Anderson NJ (2002) Distribution of chironomids (Diptera) in low arctic West Greenland lakes: trophic conditions, temperature and environmental reconstruction. *Freshw Biol* 47:1137–1157
- Brodersen KP, Lindegaard C (1999) Classification, assessment and trophic reconstruction of Danish lakes using chironomids. *Freshw Biol* 42:143–157
- Brodersen KP, Quinlan R (2006) Midges as palaeoindicators of lake productivity, eutrophication and hypolimnetic oxygen. *Quat Sci Rev* 25:1995–2012
- Brodersen KP, Odgaard BV, Vestergaard O, Anderson NJ (2001) Chironomid stratigraphy in the shallow and eutrophic Lake Sobygaard, Denmark: chironomid-macrophyte co-occurrence. *Freshw Biol* 46:253–267
- Brodersen KP, Pedersen O, Lindegaard C, Hamburger K (2004) Chironomids (Diptera) and oxy-regulatory capacity: an experimental approach to paleolimnological interpretation. *Limnol Oceanogr* 49:1549–1559
- Brooks SJ, Birks HJB (2001) Chironomid-inferred air temperatures from late-glacial and Holocene sites in north-west Europe: progress and problems. *Quat Sci Rev* 20:1723–1741
- Brooks SJ, Langdon PG, Heiri O (2007) *The identification and use of palaeo-arctic chironomidae larvae in palaeoecology*. QRA Technical Guide No. 10, Quaternary Research Association, London
- Brugam RB, McKeever K, Kolesa L (1998) A diatom-inferred water depth reconstruction for an Upper Peninsula, Michigan, lake. *J Paleolimnol* 20:267–276
- Dixit SS, Cumming BF, Birks HJB, Smol JP, Kingston JC, Uutala AJ, Charles DF, Camburn KE (1993) Diatom assemblages from Adirondack lakes (New York, USA) and the development of inference models for retrospective environmental assessment. *J Paleolimnol* 8:27–47
- Eggermont H, De Deyne P, Verschuren D (2007) Spatial variability of chironomid death assemblages in the surface sediments of a fluctuating tropical lake (Lake Naivasha, Kenya). *J Paleolimnol* 38:309–328
- Engels S, Cwynar L (2011) Changes in fossil chironomid remains along a depth gradient: evidence for common faunal thresholds within lakes. *Hydrobiologia* 665:15–38
- Heiri O (2004) Within-lake variability of subfossil chironomid assemblages in shallow Norwegian lakes. *J Paleolimnol* 32:67–84
- Heiri O, Lotter AF (2010) How does taxonomic resolution affect chironomid-based temperature reconstruction? *J Paleolimnol* 44:589–601
- Heiri O, Birks HJB, Brooks SJ, Velle G, Willassen E (2003) Effects of within-lake variability of fossil assemblages on quantitative chironomid-inferred temperature reconstruction. *Palaeogeogr Palaeoclimatol Palaeoecol* 199:95–106
- Hofmann W (1998) Cladocerans and chironomids as indicators of lake level changes in north temperate lakes. *J Paleolimnol* 19:55–62
- Juggins S (2009) rioja: an R package for the analysis of Quaternary science data. Department of Geography, University of Newcastle, Newcastle upon Tyne
- Kernan M, Ventura M, Bitušik P, Brancelj A, Clarke G, Velle G, Raddum GG, Stuchlík E, Catalan J (2009) Regionalisation of remote European mountain lake ecosystems according to their biota: environmental versus geographical patterns. *Freshw Biol* 54:2470–2493
- Korhola A, Olander H, Blom T (2000) Cladoceran and chironomid assemblages as quantitative indicators of water depth in subarctic Fennoscandian lakes. *J Paleolimnol* 24:43–54
- Kurek J, Cwynar LC (2009a) Effects of within-lake gradients on the distribution of fossil chironomids from maar lakes in western Alaska: implications for environmental reconstructions. *Hydrobiologia* 623:37–52
- Kurek J, Cwynar LC (2009b) The potential of site-specific and local chironomid-based inference models for reconstructing past lake levels. *J Paleolimnol* 42:37–50
- Kurek J, Cwynar LC, Ager TA, Abbott MB, Edwards ME (2009) Late Quaternary paleoclimate of western Alaska inferred from fossil chironomids and its relation to vegetation histories. *Quat Sci Rev* 28:799–811
- Laird KR, Cumming BF (2008) Reconstruction of Holocene lake level from diatoms, chrysophytes and organic matter in a drainage lake from the Experimental Lakes Area (northwestern Ontario, Canada). *Quat Res* 69:292–305

- Laird KR, Cumming BF (2009) Diatom-inferred lake level from near-shore cores in a drainage lake from the Experimental Lakes Area, northwestern Ontario, Canada. *J Paleolimnol* 42:65–80
- Laird KR, Kingsbury MV, Lewis CFM, Cumming BF (2011) Diatom-inferred depth models in 8 Canadian boreal lakes: inferred changes in the benthic: planktonic depth boundary and implications for assessment of past droughts. *Quat Sci Rev* 30:1201–1217
- Larocque I, Hall RI, Grahn E (2001) Chironomids as indicators of climate change: a 100-lake training set from a subarctic region of northern Sweden (Lapland). *J Paleolimnol* 26:307–322
- Larocque-Tobler I, Grosjean M, Kamenik C (2011) Calibration-in-time versus calibration-in-space (transfer function) to quantitatively infer July air temperature using biological indicators (chironomids) preserved in lake sediments. *Palaeogeogr Palaeoclimatol Palaeoecol* 299:281–288
- Legendre P, Legendre L (1998) *Numerical ecology*. Elsevier, Amsterdam
- Lehman JT (1975) Reconstructing the rate of accumulation of lake sediment: the effect of sediment focusing. *Quat Res* 5:541–550
- Lindegaard C (1992) Zoobenthos ecology of Thingvallavatn—vertical-distribution, abundance, population-dynamics and production. *Oikos* 64:257–304
- Luoto TP (2009) A Finnish chironomid- and chaoborid-based inference model for reconstructing past lake levels. *Quat Sci Rev* 28:1481–1489
- Luoto TP (2010) Hydrological change in lakes inferred from midge assemblages through use of an intralake calibration set. *Ecol Monogr* 80:303–329
- Luoto TP, Helama S (2010) Palaeoclimatological and palaeolimnological records from fossil midges and tree-rings: the role of the North Atlantic oscillation in eastern Finland through the medieval climate anomaly and little ice age. *Quat Sci Rev* 29:2411–2423
- Mackay AW, Battarbee RW, Flower RJ, Granin NG, Jewson DH, Ryves DB, Sturm M (2003) Assessing the potential for developing internal diatom-based transfer functions for Lake Baikal. *Limnol Oceanogr* 48:1183–1192
- Mackay AW, Ryves DB, Battarbee RW, Flower RJ, Jewson D, Rioual P, Sturm M (2005) 1000 years of climate variability in central Asia: assessing the evidence using Lake Baikal (Russia) diatom assemblages and the application of a diatom-inferred model of snow cover on the lake. *Global Planet Change* 46:281–297
- Mayewski PA, Rohling EE, Stager JC, Karlen W, Maasch KA, Meeker LD, Meyerson EA, Gasse F, van Kreveld S, Holmgren K, Lee-Thorp J, Rosqvist G, Rack F, Staubwasser M, Schneider RR, Steig EJ (2004) Holocene climate variability. *Quat Res* 62:243–255
- Mitchell EAD, Charman DJ, Warner BG (2008) Testate amoebae analysis in ecological and paleoecological studies of wetlands: past, present and future. *Biodivers Conserv* 17:2115–2137
- Moser KA, Korhola A, Weckstrom J, Blom T, Pienitz R, Smol JP, Douglas MSV, Hay MB (2000) Paleohydrology inferred from diatoms in northern latitude regions. *J Paleolimnol* 24:93–107
- Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2011) Community ecology package 'vegan' for R. University of Oulu, Oulu
- R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Simpson GL (2007) Analogue methods in palaeoecology: using the analogue package. *J Stat Softw* 22:1–29
- Telford RJ (2006) Limitations of dinoflagellate cyst transfer functions. *Quat Sci Rev* 25:1375–1382
- Telford R (2011) palaeoSig: significance tests of quantitative palaeoenvironmental reconstructions. R package version 1.0
- Telford RJ, Birks HJB (2005) The secret assumption of transfer functions: problems with spatial autocorrelation in evaluating model performance. *Quat Sci Rev* 24:2173–2179
- Telford RJ, Birks HJB (2009) Evaluation of transfer functions in spatially structured environments. *Quat Sci Rev* 28:1309–1316
- Telford R, Birks HJB (2011) A novel method for assessing the significance of quantitative reconstructions inferred from biotic assemblages. *Quat Sci Rev* 30:1272–1278
- ter Braak CJF (1987) CANOCO—a FORTRAN program for canonical community ordination by (partial) (detrended) (canonical) correspondence analysis (version 2.1). TNO Institute for Applied Computer Science, Statistics Department, Wageningen, NL
- van Hardenbroek M, Heiri O, Wilhelm MF, Lotter AF (2011) How representative are subfossil assemblages of Chironomidae and common benthic invertebrates for the living fauna of Lake De Waay, The Netherlands? *Aquat Sci* 73:247–259
- Velle G, Brooks SJ, Birks HJB, Willassen E (2005a) Chironomids as a tool for inferring Holocene climate: an assessment based on six sites in southern Scandinavia. *Quat Sci Rev* 24:1429–1462
- Velle G, Larsen J, Eide W, Peglar SM, Birks HJB (2005b) Holocene environmental history and climate of Råtåsjøen, a low-alpine lake in south-central Norway. *J Paleolimnol* 33:129–153
- Velle G, Bjune AE, Larsen J, Birks HJB (2010a) Holocene climate and environmental history of Brurskardstjøni, a lake in the catchment of Øvre Heimdalsvatn, south-central Norway. *Hydrobiologia* 642:13–34
- Velle G, Brodersen KP, Birks HJB, Willassen E (2010b) Midges as quantitative temperature indicator species: lessons for palaeoecology. *Holocene* 20:989–1002
- Velle G, Kongshavn K, Birks HJB (2011) Minimizing the edge-effect in environmental reconstructions by trimming the calibration set: chironomid-inferred temperatures from Spitsbergen. *Holocene* 21:417–430
- Velle G, Brodersen KP, Birks HJB, Willassen E (in press) Inconsistent results should not be overlooked: A reply to Brooks et al. (2012). *The Holocene*. doi:10.1177/0959683612449765