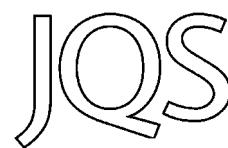


# Statistical testing of a new testate amoeba-based transfer function for water-table depth reconstruction on ombrotrophic peatlands in north-eastern Canada and Maine, United States



MATTHEW J. AMESBURY,<sup>1\*</sup> GUNNAR MALLON,<sup>2</sup> DAN J. CHARMAN,<sup>1</sup> PAUL D. M. HUGHES,<sup>2</sup> ROBERT K. BOOTH,<sup>3</sup> TIMOTHY J. DALEY<sup>4</sup> and MICHELLE GARNEAU<sup>5</sup>

<sup>1</sup>Geography, College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4RJ, UK

<sup>2</sup>Geography and Environment, University of Southampton, Southampton, UK

<sup>3</sup>Earth and Environmental Sciences Department, Lehigh University, Bethlehem, PA, USA

<sup>4</sup>School of Geography, Earth and Environmental Sciences, Plymouth University, Plymouth, UK

<sup>5</sup>Département de Géographie and GEOTOP, Université du Québec à Montréal, Montréal, Québec, Canada

Received 26 March 2012; Revised 13 August 2012; Accepted 14 August 2012

**ABSTRACT:** Proxy reconstructions of climatic parameters developed using transfer functions are central to the testing of many palaeoclimatic hypotheses on Holocene timescales. However, recent work shows that the mathematical models underpinning many existing transfer functions are susceptible to spatial autocorrelation, clustered training set design and the uneven sampling of environmental gradients. This may result in over-optimistic performance statistics or, in extreme cases, a lack of predictive power. A new testate amoeba-based transfer function is presented that fully incorporates the new recommended statistical tests to address these issues. Leave-one-out cross-validation, the most commonly applied method in recent studies to assess model performance, produced over-optimistic performance statistics for all models tested. However, the preferred model, developed using weighted averaging with tolerance downweighting, retained a predictive capacity equivalent to other published models even when less optimistic performance statistics were chosen. Application of the new statistical tests in the development of transfer functions provides a more thorough assessment of performance and greater confidence in reconstructions based on them. Only when the wider research community have sufficient confidence in transfer function-based proxy reconstructions will they be commonly used in data comparison and palaeoclimate modelling studies of broader scientific relevance.

Copyright © 2012 John Wiley & Sons, Ltd.

**KEYWORDS:** cross-validation; North America; peat; testate amoebae; transfer function.

## Introduction

Testate amoebae are a group of Protozoa that construct a shell, or test, that is readily preserved in peat, allowing identification to species level in many cases (Mitchell *et al.*, 2008). They have been extensively studied in recent years as a proxy-indicator of Holocene palaeohydrological and palaeoclimatic change in ombrotrophic peatlands, predominantly in the northern hemisphere (e.g. Charman *et al.*, 2006; Hughes *et al.*, 2006; Sillasoo *et al.*, 2007; Amesbury *et al.*, 2008; Blundell *et al.*, 2008; van Bellen *et al.*, 2011; Lamarre *et al.*, 2012). The value of testate amoebae in such studies is due largely to an understanding of their ecology, particularly in relation to hydrological factors, of which water-table depth (used as a proxy for the actual living conditions experienced) is the most important. Ecological knowledge of testate amoebae has been used to develop transfer functions that now provide quantitative palaeohydrological reconstructions from an increasing range of locations and from regional to continental scales (e.g. Charman and Warner, 1997; Woodland *et al.*, 1998; Bobrov *et al.*, 1999; Charman *et al.*, 2007; Payne and Mitchell, 2007; Booth, 2008; Lamentowicz *et al.*, 2008; Payne *et al.*, 2008; Swindles *et al.*, 2009; Markel *et al.*, 2010). Reconstructions are commonly performed using a range of transfer function models and assessed using leave-one-out cross-validation. They are often presented with error margins commonly derived from 1000 bootstrap cycles. However, recent research (Telford and Birks, 2005, 2009, 2011a,b; Payne *et al.*, 2012) has highlighted a

range of statistical issues that have not been addressed as a matter of course in transfer function development, but which may have a significant effect on the reliability of model output and on any subsequent palaeoclimatic interpretation. The statistical challenges to developing reliable transfer functions which have received recent attention are:

1. **Spatial autocorrelation** (the tendency of proximal sites to resemble one another more than randomly selected sites; Telford and Birks, 2005, 2009). Performance statistics (such as  $r^2$ ) may be high in the presence of strong spatial autocorrelation, even for simulated variables with no ecological relevance. Telford and Birks (2005) applied foraminifera-based models for sea surface temperature from the North Atlantic to a spatially independent training set from the South Atlantic and found that models based on a unimodal species–environment response (i.e. weighted averaging-based models) were more robust to spatial autocorrelation than other model types, particularly the modern analogue technique (MAT). Telford and Birks (2009) developed a method to test for spatial autocorrelation by comparing the reduction in  $r^2$  when sites were removed from the training set either randomly or by geographical proximity to the test site.
2. **Uneven sampling of an environmental gradient** (Telford and Birks, 2011a). This can lead to the overestimation of root mean square error of prediction (RMSEP). Telford and Birks (2011a) showed that some model types were more susceptible to this problem than others (again, MAT was the most susceptible) and presented a method to evaluate this potential bias by using a segment-wise RMSEP procedure. This

\*Correspondence: M. J. Amesbury, as above.  
E-mail: m.jAMESBURY@exeter.ac.uk

divides the environmental gradient into a series of segments and calculates the RMSEP by weighting each segment equally.

3. **The use of clustered data sets.** Payne *et al.* (2012) found that the use of clustered datasets (i.e. many samples per site) can also lead to overly optimistic RMSEP values. They described a method of leave-one-site-out (LOSO) cross-validation to overcome this problem and made recommendations for training set development to minimize loss of model performance.

The ecological relevance and statistical validity of many transfer function reconstructions has also been questioned (Telford and Birks, 2011b). Telford and Birks (2011b) stated that it is reasonable to expect statistical significance when there is a clear underlying ecological process for the species-environment relationship, sensitive sites are used in the training set and the range of the environmental variable is at least as large as the RMSEP. They proposed a test for this assumption of statistical significance, where any single reconstruction should explain more of the variance in the fossil data than 95% of reconstructions from transfer functions trained on randomly generated environmental data to be considered statistically significant.

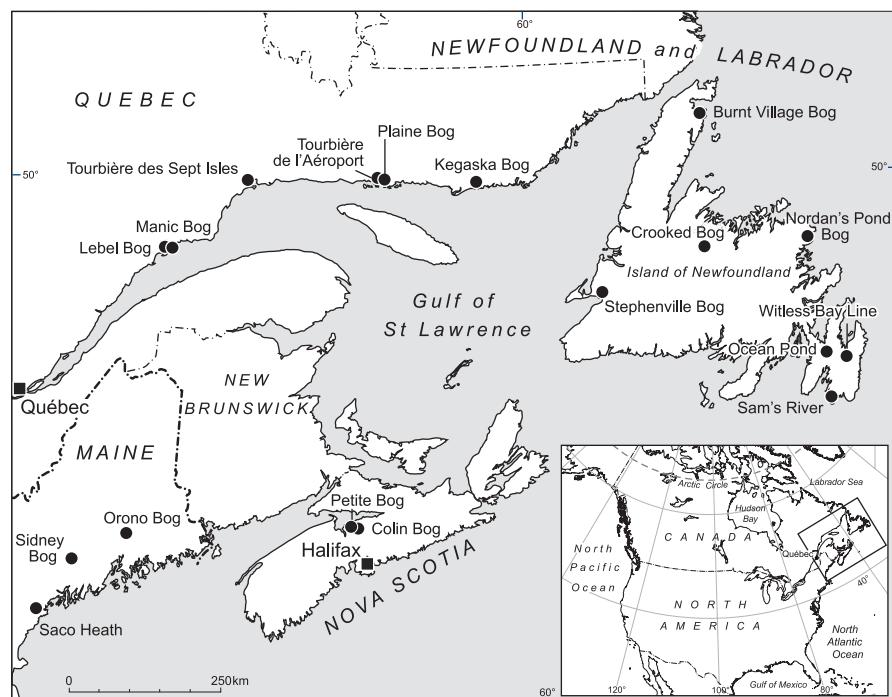
It is crucial that potentially problematic statistical issues are not overlooked in transfer function development so that errors can be realistically modelled (e.g. Salonen *et al.*, 2012) and carried through to applications such as understanding of the mechanisms driving centennial–millennial-scale Holocene climate change (e.g. Hughes *et al.*, 2006; Charman, 2010) and validation of results by comparison with instrumental climate data (e.g. Charman *et al.*, 2009; Booth, 2010; Amesbury *et al.*, 2012). All transfer function development should start with the null hypothesis that the model has no predictive power, but this is rarely, if ever, explicitly tested (Telford and Birks, 2009).

Here, a new regional testate amoeba-based transfer function is presented for water-table depth from the north-eastern seaboard of North America, covering provinces of north-

eastern Canada (Quebec, Nova Scotia and Newfoundland) and the state of Maine, USA, as part of a wider multi-proxy project investigating Holocene climate change in the region. Transfer function models are subjected to thorough statistical testing before being applied to an existing fossil testate amoeba profile (Hughes *et al.*, 2006) so that new reconstructions can be compared with those from existing models. Peatland and palaeoclimate research in this region is expanding (e.g. Hughes *et al.*, 2006; Loisel and Garneau, 2010; van Bellen *et al.*, 2011; Lamarre *et al.*, 2012) but existing transfer functions (Charman and Warner, 1997; Booth, 2008) do not cover the whole region or include site types different from those from which new palaeoclimate records will be developed within the wider project of which the work presented here is a part. The results also provide the first case study of palaeoecological transfer function development using the full suite of recently developed statistical tools.

## Sites and methods

Twelve new sites spanning the latitudinal and longitudinal range of the study region (Fig. 1; Table 1) were sampled in June/July 2010 (Saco Heath and Sidney Bog, Maine; Petite Bog, Nova Scotia; and Burnt Village Bog, Newfoundland) and May/June 2011 (Colin Bog, Nova Scotia; Nordan's Pond Bog, Newfoundland; and all six sites in Quebec) and combined with existing data from Maine, USA (one site; Booth, 2008), and Newfoundland, Canada (five sites; Charman and Warner, 1997), to form the complete dataset for transfer function modelling. All sites are ombrotrophic bogs with vegetation dominated by *Sphagnum* mosses and vascular and ericaceous plants typical of the region such as *Eriophorum* spp., *Vaccinium* spp., *Kalmia* spp. and *Picea mariana* trees (for more detailed information on regional climate, vegetation and bog types see National Wetlands Working Group, 1988). Charman and Warner (1997) described a variety of peatland types in their dataset and here we selected only those sites with  $\text{pH} < 5$  to avoid introducing non-oligotrophic sites or strong pH or conductivity gradients into our dataset.



**Figure 1.** Locations of sites included in the transfer function.

**Table 1.** Details of sites included in the transfer functions.

Site	Latitude (°N)	Longitude (°W)	Elevation (m ASL)	n*	Depth to water-table		Reference (including sampling date for new data)
					Range (cm)	Mean ± SD (cm)	
Saco Heath (SCB), MN	43°33.194	70°28.265	44	26	3 to 65	25 ± 16.7	This study/June 2010
Sidney Bog (SDY), MN	44°23.274	69°47.260	90	24	3 to 58	32 ± 17.6	This study/June 2010
Orono Bog (ORO), MN	44°52.286	68°43.630	40	18	8 to 36	22 ± 11.1	Booth (2008)
Petite Bog (PTB), NS	45°8.659	63°56.219	50	16	0 to 46	16 ± 16.1	This study/June 2010
Colin Bog (CLB), NS	45°6.459	63°53.683	28	27	0 to 56	13 ± 15.9	This study/May 2011
Manic Bog (MNC), QC	49°7.388	68°17.519	20	5	3 to 32	13 ± 12.3	This study/June 2011
Lebel Bog (LBL), QC	49°6.457	68°14.425	13	9	-8 to 51	16 ± 18.6	This study/June 2011
Tourbière des Sept Isles (TSI), QC	50°12.095	66°36.081	64	6	1.5 to 40	15 ± 16.2	This study/June 2011
Tourbière de l'Aéroport (TDA), QC	50°16.501	63°36.841	33	6	-2 to 38	18 ± 16.5	This study/June 2011
Plaine Bog (PLB), QC	50°16.172	63°32.428	19	7	-4 to 28	13 ± 12.9	This study/June 2011
Kegaska Bog (KGK), QC	50°12.640	61°15.008	24	4	1.5 to 40	21 ± 17.2	This study/June 2011
Burnt Village Bog (BVB), NFL	51°7.562	55°55.645	28	25	0 to 43	12 ± 14.0	This study/June 2010
Nordan's Pond Bog (NDN), NFL	49°9.646	53°35.903	27	20	0 to 40	14 ± 13.5	This study/May 2011
Stephenville Bog (STE), NFL	48°32.974	58°26.064	34	6	0 to 46	20 ± 18.9	Charman and Warner (1997)
Crooked Bog A-C (CRO), NFL	49°8.361	56°6.685	190	14	-4 to 26	9 ± 9.2	Charman and Warner (1997)
Ocean Pond (OCE), NFL	47°25.642	53°26.275	92	9	2 to 34	20 ± 10.1	Charman and Warner (1997)
Witless Bay Line (WIT), NFL	47°20.207	53°0.994	240	4	0 to 32	15 ± 13.2	Charman and Warner (1997)
Sam's River A (SAM), NFL	46°43.859	53°31.112	150	6	0 to 31	11 ± 11.6	Charman and Warner (1997)

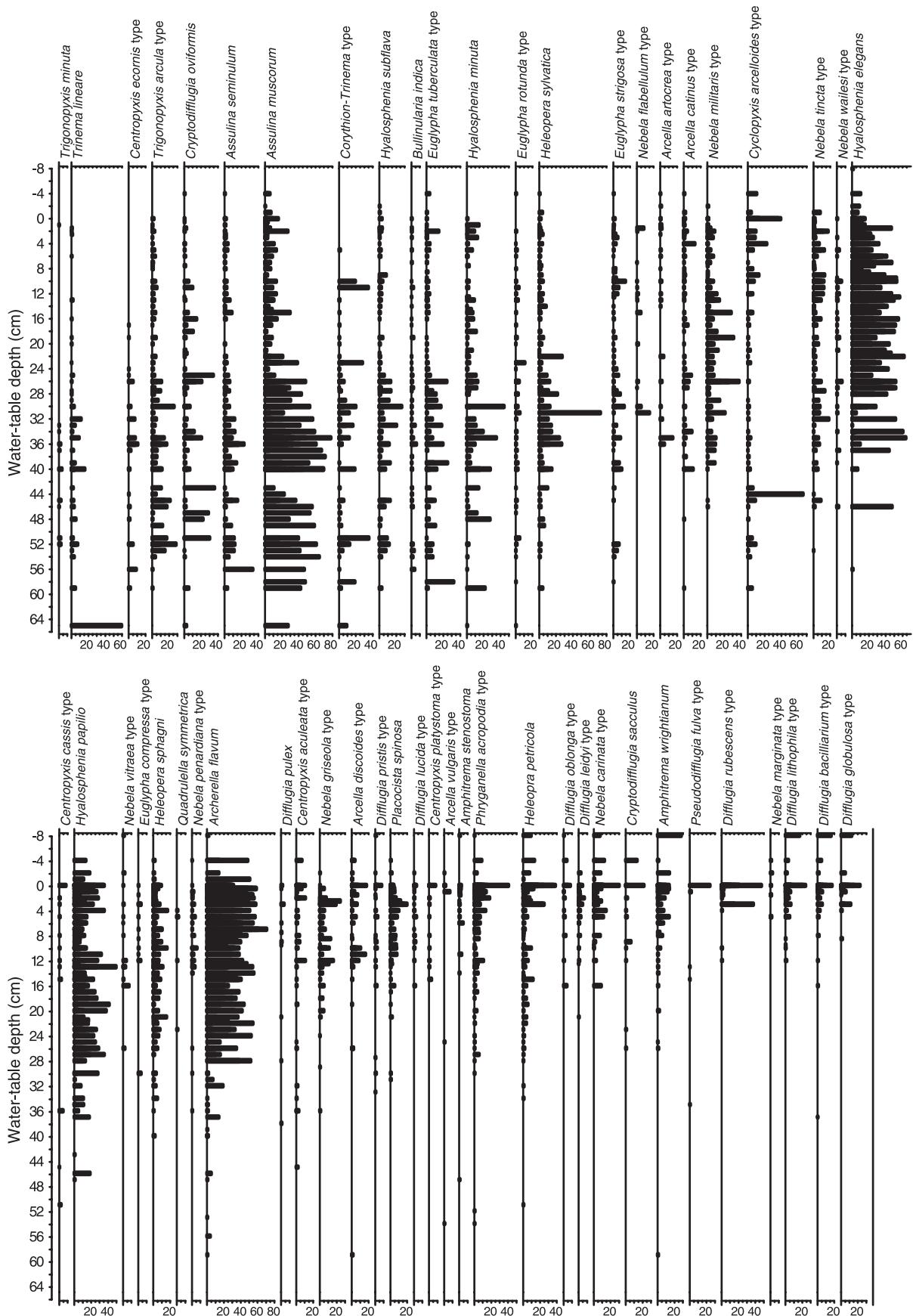
\*n=number of samples analysed (total n=232).

Field sampling methods followed the protocols of previous similar studies and the procedure outlined in Booth *et al.* (2010). At each site, one to six transects resulting in four to 27 samples were taken to sample the microtopographical gradient in water-table depth from hummock to pool. Numbers of samples at sites in Quebec were limited by the period of time available at each site. At each sampling location, a 5 × 5 × 5-cm surface *Sphagnum* moss sample was retrieved and water-table depth measured relative to the surface such that high values represent dry sites with deep water-tables and zero and/or negative values represent wet sites with surface water or submerged *Sphagnum*. At all four sites sampled in June/July 2010, we deployed PVC-lined stakes (Belyea, 1999; Booth *et al.*, 2005; Booth, 2008) and after 11 months returned to these sites and measured the discolouration of the tape to estimate mean annual water-table depth.

Laboratory methods for the preparation of moss samples for microscopic analysis followed standard procedures (Hendon and Charman, 1997; Booth *et al.*, 2010) with the size fraction between 300 and 15 µm being retained for analysis. A minimum of 150 tests were counted in all samples and raw counts converted to percentage data for transfer function modelling. Nomenclature followed Charman *et al.* (2000) except where modified by Sullivan and Booth (2007); the latter guide was used to ensure taxonomic harmonization between new data and that of Booth (2008). *Nebela penardiana* type includes *Nebela collaris* and *Nebela bohemica* as defined by Charman *et al.* (2000) and Sullivan and Booth (2007), and *Difflugia lithophila* type is defined as per Ogden and Hedley (1980). Following the taxonomic work of Meisterfeld (2002) and a number of recent publications (e.g. Lamentowicz *et al.*, 2010; Markel *et al.*, 2010; Sullivan and Booth, 2011), we use the nomenclature *Archerella flavum* for the taxon previously known as *Amphitrema flavum*. *Cyclopyxis arcelloides* type (as defined by Charman *et al.*, 2000) is split into *Cyclopyxis arcelloides*, *Phryganella acropodia* and *Difflugia globulosa* based on the overall dimensions and aperture size of the test, because Booth (2002) showed that these three taxa favour different hydrological niches along a water-table gradient. As a

result, a taxonomic difference exists between the new data and that of Booth (2008) from Orono Bog with the data of Charman and Warner (1997), which does not distinguish between these three taxa. However, we believe combining the two taxonomies for this one group of taxa is more beneficial to overall model performance than either (1) omitting the Charman and Warner (1997) data or (2) grouping the three taxa in all new samples. Weighted-average models were developed for each alternative approach to test their effects. It was found that combining taxonomies maintained similar optima for each of the three taxa to those calculated by omitting the Charman and Warner (1997) data. The wider tolerance for *C. arcelloides* resulting from inclusion of the Charman and Warner (1997) data means that any error resulting from this approach will be incorporated into subsequent fossil reconstructions. The inclusion of the wider range of sites is also advantageous in developing improved optima and tolerance estimates for the other taxa.

Ordination analyses were carried out using the programme Canoco, version 4.51 (ter Braak and Šmilauer, 2002), with Monte Carlo permutation tests (using 499 random permutations) used to determine statistical significance. Transfer functions were developed in R version 2.14.0 (R Development Core Team, 2011) using the rioja package (Juggins, 2009) and applying five commonly used model types, namely: weighted averaging (WA), weighted averaging with tolerance down-weighting (WA-Tol), weighted average partial least squares (WA-PLS), maximum likelihood (ML) and MAT. Due to the gradient length of the primary ordination axis (see Results), linear response models such as partial least squares were not considered (Birks, 1995). The mathematical basis and usage of the models employed is discussed in more detail by ter Braak and Prentice (1988), ter Braak and Juggins (1993), Birks (1995, 1998), Birks *et al.* (2010), Juggins and Birks (2012), and ter Braak and van Dam (1989). The 'best' version of each model was chosen based predominately on the  $r^2$  and RMSEP, but the average and maximum bias of each model was also considered. The most appropriate number of components for the WA-PLS model was chosen by the percentage improvement in RMSEP



**Figure 2.** Percentage distribution of taxa in all samples plotted against water-table depth. Taxa are ordered from 'dry' on the left to 'wet' on the right by jack-knifed water-table optima derived from the WA-Tol (inv) model with rare taxa and high-value residual samples removed.

(we accepted >5% *sensu* Birks, 1998) as each new component was added and the significance of a randomization *t*-test on each model. Other statistical methods such as LOSO cross-validation (Payne *et al.*, 2012) and spatial autocorrelation tests (Telford and Birks, 2009) were also performed in R using rioja and palaeoSig (Telford, 2011).

To test the applicability of the regional transfer function models developed here, we applied them to an 8500-year fossil testate amoeba record from Nordan's Pond Bog, Newfoundland (Hughes *et al.*, 2006), and compared results with the Newfoundland transfer function of Charman and Warner (1997) and the eastern North American transfer function of Booth (2008). Standard error estimates were derived from 1000 bootstrap cycles. We tested the significance of these reconstructions (Telford and Birks, 2011b) using the R package palaeoSig (Telford, 2011).

## Results

Only ~50% of the PVC-lined stakes returned useful results, mainly because of loss, movement or damage over the intervening year, probably due to wildlife and other site users. Poorly discernible discolouration and stakes remaining frozen within hummocks were also problematic (see also Markel *et al.*, 2010). Of 101 locations, 62 provided data with 10 results being questionable due to difficulty in determining the colour change. The correlation between the 2010 single readings and the 2011 annual averages from the PVC stakes was  $r^2 = 0.93$  ( $P < 0.0001$ ,  $n = 62$ ), indicating that the single-point measurements are a reliable measure of the relative wetness of the locations (Supporting Information, Fig. S1). To make use of the complete data set and the single-point measurements recorded at all pre-existing sites, including those from Booth (2008) and Charman and Warner (1997), we use only single-point measurements here, in common with the majority of other published testate amoeba transfer functions. In addition, we also concentrate only on the development of transfer functions for water-table depth. Previous research has consistently shown, on a global scale, that local hydrology, more so than other environmental variables such as pH or conductivity, is the key influence on testate amoeba distribution on bogs (e.g. Charman, 1997; Booth, 2008; Swindles *et al.*, 2009). Water-table depth provides a more reliable measure for palaeoclimatic studies than percentage moisture content, which is subject to more short-term variability and the potential for evaporative loss (Charman *et al.*, 2007).

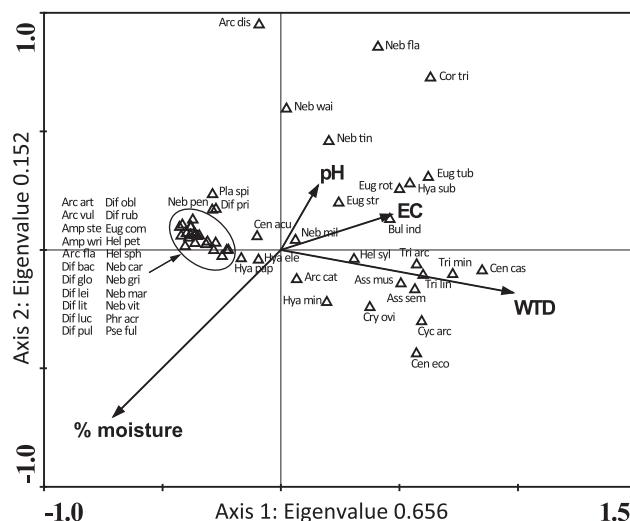
A total of 232 samples with water-table data were analysed from 18 sites. A total of 59 taxa were recorded, with *Assulina muscorum*, *Archerella flavum*, *Hyalosphenia elegans* and *Hyalosphenia papilio* being the most abundant (Fig. 2). To reduce the influence of rare taxa, *Diffugia oviformis*, *Heleopera rosea*, *Nebela tubulosa* type, *Sphenoderia lenta* and *Tracheleuglypha dentata* (these were all present in fewer than five samples at <5% abundance) were removed from all ordinations and transfer function modelling, leaving a total of 54 taxa.

Detrended Correspondence Analysis (DCA) was carried out using the complete dataset (supporting Fig. S2). A biplot of Axis 1 and 2 sample scores shows a high degree of overlap between all sites and no one site or group of sites could be identified in an outlying position (*sensu* Charman *et al.*, 2007). A biplot of Axis 1 and 2 species scores shows taxa arranged in a typical water-table gradient along Axis 1 (eigenvalue 0.651), which has a gradient length of  $>5\sigma$  units, suggesting good separation of taxa at either end, with typically 'wet' taxa such as *Diffugia* spp. and *Archerella wrightianum* having low values and typically 'dry' taxa such as *Assulina* spp., *Trinema lineare* and

*Hyalosphenia subflava* having higher values. This pattern is reflected in the Axis 1 gradient of sample scores, with samples on the left (low values) having higher water-table measurements (i.e. wetter microsites) and samples on the right (high values) having lower water-table measurements (i.e. drier microsites).

To further investigate the influence of a range of environmental variables on species distribution, Canonical Correspondence Analysis (CCA) was also performed on a reduced dataset of 126 samples (from 10 sites) for which water-table depth, percentage moisture content, pH and electrical conductivity data were available (Fig. 3). CCA Axis 1 (eigenvalue 0.656) explained 19.8% of the variance in the testate amoeba data and 75% of the species–environment relationship. Figure 3 illustrates that the principal axis of variation is related to water-table depth, with a correlation of  $r = 0.92$  ( $P < 0.002$ ). The four environmental variables explained 26.4% of the total variance in the dataset. Variance partitioning of this explained variance was undertaken using a series of partial CCAs and showed that water-table depth explained 42.9% ( $P < 0.002$ ), percentage moisture content 17.8% ( $P < 0.002$ ), electrical conductivity 5.5% ( $P < 0.006$ ) and pH 3.3% ( $P < 0.116$ ). The remaining 30.5% of explained variance is related to the inter-correlations of the environmental variables. In a CCA with water-table depth as the only constraining environmental variable, 13.5% of the species data was explained with a species–environment correlation of  $r = 0.89$  ( $P < 0.002$ ).

Performance statistics for the five transfer function models using the complete dataset are shown in Table 2. All models show reasonable performance with  $r^2 > 0.78$  and RMSEP < 7.5 cm, which is equivalent to or better than many published models (e.g. Charman *et al.*, 2007; Booth, 2008; Swindles *et al.*, 2009; Markel *et al.*, 2010). In general, this suggests that using one time measurements of water-table depth rather than a longer term average has not been problematic except in very extreme conditions; Booth (2008) analysed one time measurements in an anomalously dry year which resulted in an RMSEP of 19.1 and maximum bias of >40 cm. Following the initial model run, outliers and residual values were examined and samples with residual values of more than 20% of the overall range of recorded water-table depths (overall range 73 cm, cut-off value  $\geq 14.6$  cm) were removed, in keeping with many other studies (e.g. Payne *et al.*, 2006; Charman *et al.*, 2007; Booth, 2008; Swindles *et al.*, 2009). The number of samples removed therefore varied between models applied. The complete dataset contained 232 samples and 7–15 samples were omitted depending on the model type used (see Tables 2 and 3). This procedure resulted in the improvement of the majority of performance statistics for all models; in particular RMSEP values were reduced by 0.80–1.45 cm, depending on the model used. The effect of removing outlying samples is illustrated in Fig. 4 for the three best performing models: WMAT (weighted mean MAT), WA-Tol (inv) and ML. Figure 4A illustrates that the majority of samples with high residual values were from the dry end of the water-table gradient and Table 3 shows that 15 out of 20 samples that were removed from various models as outliers had water-table depths  $> 25$  cm. Only two out of the 20 samples could be considered 'wet' (water-table depth  $\leq 2$  cm) and their high residual values are likely to result from the dominance of typically 'dry' taxa in the assemblages. Figures 5 and 6 both suggest that taxa typical of drier conditions also tend to have wider tolerances, although several taxa do not conform to this interpretation; for example, *Centropyxis ecornis* type has a relatively narrow, and *Pseudodiffugia fulva* type a relatively wide tolerance when compared with those taxa with similar



**Figure 3.** Canonical Correspondence Analysis using 126 samples from 10 sites (see text for details). Environmental variables are water-table depth (WTD), percentage moisture content (% moisture), pH and electrical conductivity (EC). Full taxa names are: AMP STE, *Amphitrema stenostoma*; AMP WRI, *Archerella wrightianum*; ARC ART, *Arcella artocrea*; ARC CAT, *Arcella catinus* type; ARC DIS, *Arcella discoides* type; ARC VUL, *Arcella vulgaris*; ARC FLA, *Archerella flavum*; ASS MUS, *Assulina muscorum*; ASS SEM, *Assulina seminulum*; BUL IND, *Bullinularia indica*; CEN ACU, *Centropyxis aculeata* type; CEN CAS, *Centropyxis cassis* type; CEN ECO, *Centropyxis ecornis* type; CEN PLA, *Centropyxis platystoma* type; COR TRI, *Corythion-Trinema* type; CRY OVI, *Cryptodifflugia oviformis*; CRY SAC, *Cryptodifflugia sacculus*; CYC ARC, *Cyclopyxis arcelloides*; DIF BAC, *Difflugia bacillarium* type; DIF GLO, *Difflugia globulosa*; DIF LEI, *Difflugia leidyi*; DIF LIT, *Difflugia lithophila* type; DIF LUC, *Difflugia lucida* type; DIF OBL, *Difflugia oblonga* type; DIF PRI, *Difflugia pristis* type; DIF PUL, *Difflugia pulex*; DIF RUB, *Difflugia rubescens*; EUG COM, *Euglypha compressa* type; EUG ROT, *Euglypha rotunda* type; EUG STR, *Euglypha strigosa*; EUG TUB, *Euglypha tuberculata* type; HEL PET, *Heleopera petricola*; HEL SPH, *Heleopera sphagni*; HEL SYL, *Heleopera sylvatica*; HYA ELE, *Hyalosphenia elegans*; HYA MIN, *Hyalosphenia minuta*; HYA PAP, *Hyalosphenia papilio*; HYA SUB, *Hyalosphenia subflava*; NEB CAR, *Nebela carinata*; NEB FLA, *Nebela flabellulum*; NEB GRI, *Nebela griseola*; NEB MAR, *Nebela marginata*; NEB MIL, *Nebela militaris*; NEB PEN, *Nebela penardiiana* type; NEB TIN, *Nebela tincta*; NEB VIT, *Nebela vitraea*; NEB WAL, *Nebela walesi* type; PHR ACR, *Phryganella acropodia*; PLA SPI, *Placocista spinosa*; PSE FUL, *Pseudodifflugia fulva* type; QUA SYM, *Quadrilella symmetrica*; TRI ARC, *Trigonopyxis arcula* type; TRI MIN, *Trigonopyxis minuta* type; TRI LIN, *Trigonopyxis minuta*.

water-table optima (Fig. 5). Charman *et al.* (2007) also found that wet-indicating taxa rarely occurred in dry locations but dry-indicating taxa often occurred at relatively low abundance in wet locations (see Fig. 2). The order of the water-table depth

optima for taxa (Fig. 5) is very similar to the arrangement of taxa along DCA Axis 1 (supporting Fig. S2) and of particular interest is the close association of *Archerella flavum* and *Diffugia pulex* in an intermediate–wet location, discussed further below.

Performance statistics for WA (inv) were generally worse than for WA-Tol (inv) and a WA-PLS component 2 model provided only a 5.9% improvement in RMSEP over simple WA and a randomization *t*-test of the significance of this improvement gave only  $P=0.101$ , meaning that of the weighted-averaging-based models available, WA-Tol (inv) performed best. Given the differences in the underlying assumptions between models and their susceptibility to different statistical problems, particularly spatial autocorrelation and uneven sampling, the WMAT, WA-Tol (inv) and ML models were selected for further statistical testing.

$\text{RMSEP}_{(\text{LOSO})}$  (Payne *et al.*, 2012) and  $\text{RMSEP}_{(\text{segment-wise})}$  (Telford and Birks, 2011a; Fig. 7) were calculated for each of these three models. All results are shown in Table 4. All three models were also tested for spatial autocorrelation (Telford and Birks, 2005, 2009; Fig. 8). All  $\text{RMSEP}_{(\text{LOSO})}$  values were higher than jack-knifed  $\text{RMSEP}$ , suggesting that the clustered nature of the training set does result in over-optimistic  $\text{RMSEP}$  values when leave-one-out cross validation is used, although it is important to note that this may be an artefact of the lower number of observations used in  $\text{RMSEP}_{(\text{LOSO})}$  cross validation (Payne *et al.*, 2012). The relative decrease in model performance, or increase in  $\text{RMSEP}$  (calculated by  $(\text{RMSE}_{\text{P}(\text{LOSO})} - \text{RMSE}_{(\text{jack})})/\text{RMSE}_{(\text{jack})}$ ), was equivalent to that in Payne *et al.* (2012; mean relative decrease in performance of 0.141 from 14 training sets) and for all models  $\text{RMSEP}_{(\text{LOSO})}$  remains considerably lower than the standard deviation of all water-table measurements ( $\text{SD} = 15.85 \text{ cm}$ ,  $n = 232$ ) suggesting that all models have predictive capacity.

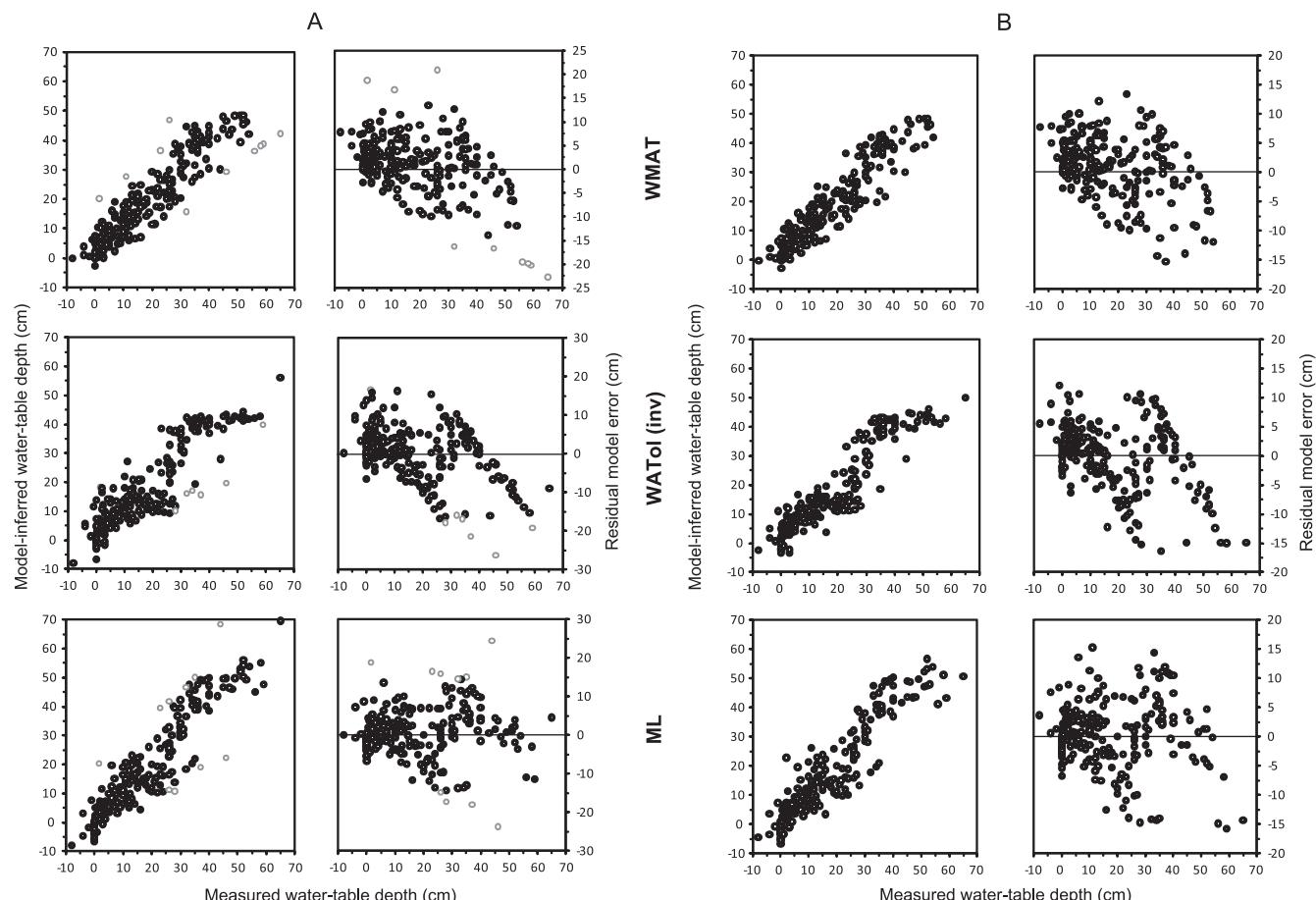
Calculating RMSEP<sub>(segment-wise)</sub> (Fig. 7) resulted in a decrease in model performance (i.e. higher RMSEP) for the WMAT and WA-Tol (inv) models but an improvement in model performance for the ML model when compared with RMSEP<sub>(jack)</sub>. This supports the findings of Telford and Birks (2011a) who found that ML models outperformed MAT- and WA-based models along unevenly sampled gradients. A high proportion of samples in the dataset are clustered at water-table depths of < 15 cm with a broadly decreasing number of samples per segment as water-table depth increases (Fig. 7). In particular, the segments 41–45 cm and 46–50 cm contain only four and five samples, respectively. The RMSEPs for each segment of all models generally follow the expected trend shown in Telford and Birks (2011a, their Fig. 6) where a higher frequency of samples equates to a lower RMSEP. However, in the ML model in particular, the RMSEPs for the low-frequency segments from 41 to 50 cm are anomalously low and result in the apparent improvement in model performance using RMSEP<sub>(segment-wise)</sub> when compared with RMSEP<sub>(jack)</sub>. The standard deviation of

**Table 2.** Performance statistics for all transfer function models for water-table depth based on leave-one-out (jack-knifing) cross validation. Results are in order of performance as assessed by RMSEP<sub>(jack)</sub>. Rare taxa occurring in fewer than five samples at < 5% abundance have been omitted (see text for details). Figures in parenthesis show performance statistics after the removal of outlier samples (see text and Table 3 for details). SD is the standard deviation of all water-table measurements included in each model after the removal of outliers.

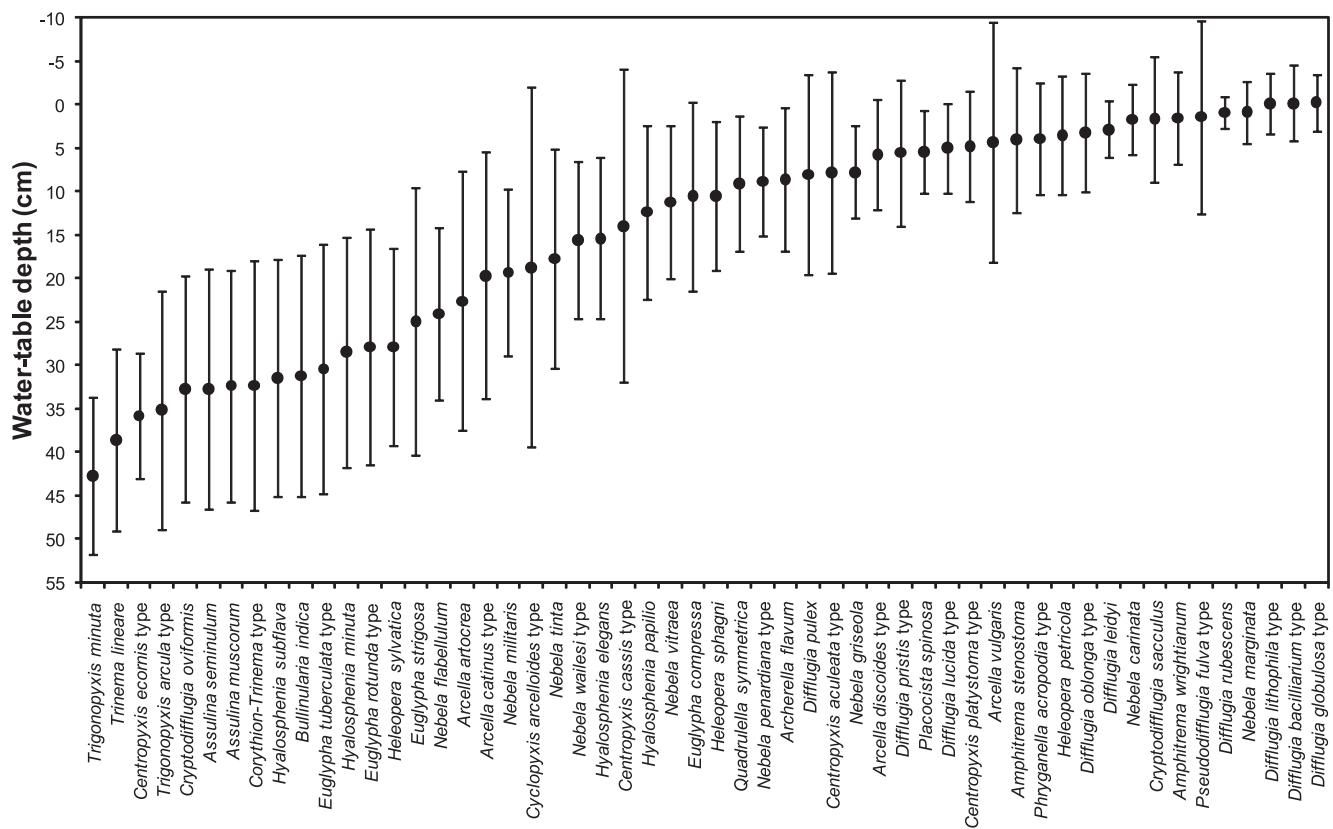
Model	$R^2_{(jack)}$	Average bias <sub>(jack)</sub>	Maximum bias <sub>(jack)</sub>	RMSEP <sub>(jack)</sub>	No. of outlier samples removed	SD
WMAT	0.84 (0.88)	-1.04 (-0.98)	20.95 (6.38)	6.32 (5.27)	9	14.9
WA-Tol (inv)	0.83 (0.87)	0.20 (0.21)	17.03 (14.98)	6.44 (5.66)	7	15.6
ML	0.83 (0.87)	-0.73 (-0.48)	12.78 (12.27)	6.89 (5.91)	10	15.9
WA-PLS Component 2	0.80 (0.84)	0.13 (0.08)	18.32 (14.19)	7.03 (6.03)	7	15.3
WA (inv)	0.78 (0.85)	0.05 (0.04)	17.81 (17.39)	7.47 (6.02)	15	15.4

**Table 3.** Details of samples removed from transfer function models when filtering for residuals >20% of total measured water-table depth (WTD) range (14.6 cm). Site codes as per Table 1.

Site	Sample	Water-table depth (cm)	Residual >20% limit in which model?					Possible explanation of anomaly
			WMAT	WA-Tol (inv)	ML	WA-PLS C2	WA (inv)	
SCB	A1	65	X					~60% <i>Trinema lineare</i> , deepest WTD measured
SCB	A4	28					X	Typical moderate-wet assemblage, but relatively high WTD?
SCB	E1	59	X	X		X	X	>20% <i>Hyalosphenia minuta</i> , very high WTD
SCB	E3	35				X	X	>60% <i>Hyalosphenia elegans</i> , with high WTD
SYB	C1	58	X			X	X	>30% <i>Euglypha tuberculata</i> , very high WTD
SYB	E4	11	X			X		~35% <i>Corythion-Trinema</i> type in relatively wet sample
CLB	C1	35			X			~80% <i>Assulina</i> spp.
CLB	D1	56	X			X	X	~80% <i>Assulina</i> spp.
CLB	E1	44				X	X	>65% <i>Cyclopyxis arcelloides</i>
CLB	E2	26				X	X	Typical moderate-wet assemblage, but relatively high WTD?
MNC	A5	32			X			~ 55% <i>Assulina muscorum</i>
STI	A4	1.5	X	X	X	X	X	Typical moderate-dry assemblage, but WTD of 1.5 cm
NDN	A2	28		X	X	X	X	Typical moderate-wet assemblage, but relatively high WTD?
STE	3	37		X	X	X	X	Typical moderate-wet assemblage, but relatively high WTD?
STE	6	46	X	X	X	X	X	Typical moderate-wet assemblage, but relatively high WTD?
CRO	9	23			X		X	Typical dry assemblage, but moderate WTD?
CRO	14	26	X		X		X	Typical dry assemblage, but moderate WTD?
OCE	4	2					X	Diverse assemblage, no taxa >15%.
OCE	7	34		X			X	Some 'dry' taxa but WTD of 2 cm
WIT	4	32	X	X			X	>60% <i>Hyalosphenia elegans</i> , with high WTD
							X	>50% <i>Hyalosphenia elegans</i> , with high WTD



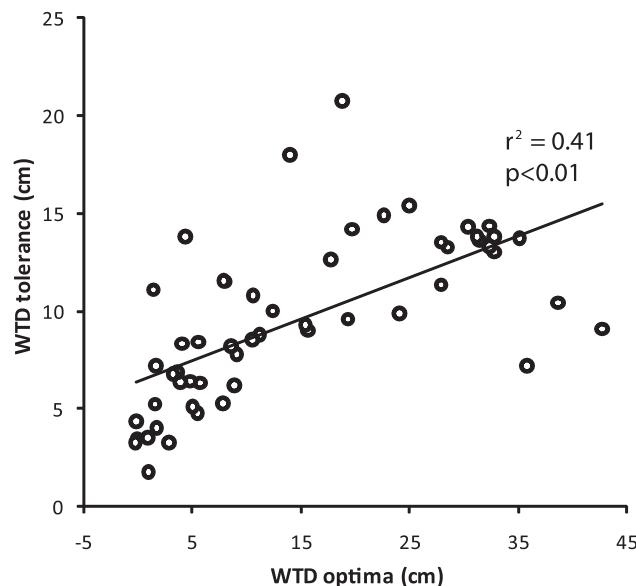
**Figure 4.** Biplots of jack-knifed predicted and observed water-table depth and residual error plots for the three best performing models [as measured by RMSEP<sub>jack</sub>: WMAT, WA-Tol (inv) and ML] before (A) and after (B) the filtering of samples with high residual values. All model were runs with rare taxa removed (see text for details). Samples shown in grey in (A) are those removed from model runs shown in (B).



**Figure 5.** Water-table depth optima and tolerances for 54 taxa (five rare taxa omitted; see text for details) based on the WA-Tol (inv) model after removal of samples with high residual values ( $n=225$ ).

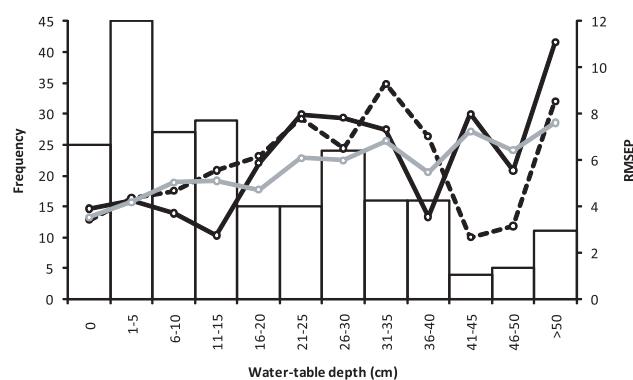
the number of samples in each segment divided by the total number of observations in the model, used by Telford and Birks (2011a) as a metric to describe the unevenness of datasets, was 0.049.

Figure 8 suggests that all models are spatially autocorrelated to some degree as removing sites in the geographical



**Figure 6.** Biplot of water-table depth (WTD) optima and tolerances for 54 taxa (five rare taxa omitted; see text for details) based on the WA-Tol (inv) model after removal of samples with high residual values ( $n=225$ ).

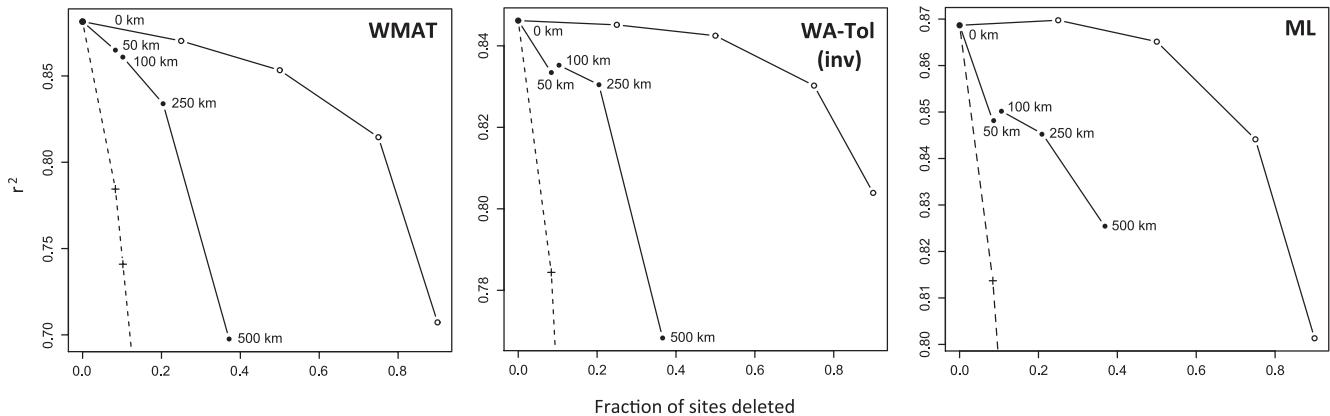
neighbourhood of the test site results in a decline in  $r^2$  more similar to that shown if the most environmentally similar samples are removed as opposed to randomly selected samples. The overall reduction in  $r^2$  is greatest in the WMAT model but the initial reduction when sites within 100 km of the test site are removed is steepest in the ML model. Broadly speaking, if spatial autocorrelation were not a problem, the curves showing a decline in  $r^2$  as sites are deleted in cross-validation should follow similar trajectories, regardless of whether the deleted sites are geographically proximal to the test site or random (see Telford and Birks, 2009, Fig. 1a). However, this interpretation is complicated by the clustered nature of the dataset, which introduces a degree of spatial autocorrelation into the dataset that may not necessarily be problematic if the



**Figure 7.** Sampling distribution of the complete dataset divided into 12 segments. Lines show segment-wise RMSEP for the three best performing models [as measured by RMSEP<sub>(jack)</sub>: grey = WMAT, black = WA-Tol (inv) and dashed black = ML].

**Table 4.** Comparison of jack-knifed (leave-one-out) RMSEP, LOSO (leave-one-site-out) RMSEP, segment-wise RMSEP and palaeoSig values for the three best performing models [WMAT, WA-Tol (inv) and ML]. Figures are based on model runs with rare taxa and samples with high residual values omitted (see text for details). palaeoSig *P*-value is the *P*-value of a test of the significance of each model against randomly generated models (mean of 10 model runs, see text for details). Figures in parentheses for RMSEP<sub>(LOSO)</sub> and RMSEP<sub>(segment-wise)</sub> show the relative decrease (positive value, i.e. higher RMSEP) or increase (negative value, i.e. lower RMSEP) in model performance from RMSEP<sub>(jack)</sub>.

Model	RMSEP <sub>(jack)</sub>	RMSEP <sub>(LOSO)</sub>	RMSEP <sub>(segment-wise)</sub>	palaeoSig <i>P</i> -value
WMAT	5.27	5.79 (0.098)	5.69 (0.080)	0.0062
WA-Tol (inv)	5.66	6.37 (0.125)	5.99 (0.053)	0.0577
ML	5.91	6.20 (0.049)	5.77 (-0.052)	0.0736



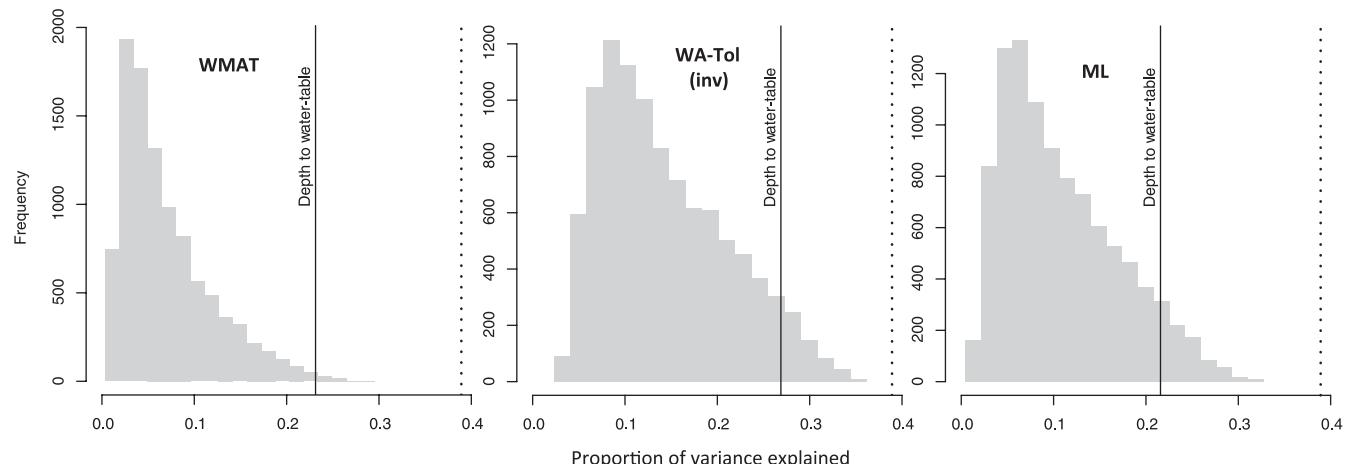
**Figure 8.** Plot showing effect of spatial autocorrelation on  $r^2$  on all model types by deleting sites at random (open circles), from the geographical neighbourhood of the test site (filled circles) or that are most environmentally similar (crosses) during cross-validation (cf. Telford and Birks 2009). Note y-axes are on different scales.

measured environmental variable (i.e. water-table depth) is ecologically important. For this reason these results are used as a qualitative and comparative measure of spatial autocorrelation in the three model types that can be used to help determine the final model choice.

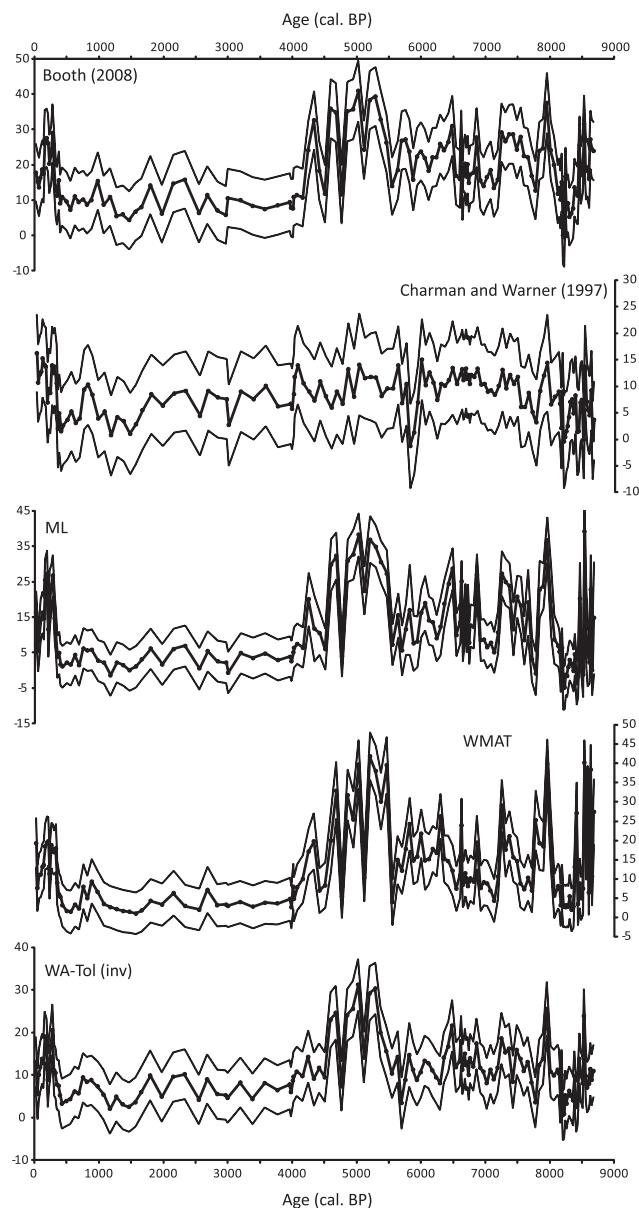
To assess the models, water-table depth reconstructions from an 8500-year profile from Nordan's Pond Bog, Newfoundland (Hughes *et al.*, 2006), were plotted using each model. The significance of these reconstructions was tested against reconstructions from 9999 transfer functions trained on randomly generated environmental data (Telford and Birks,

2011b; Fig. 9; Table 4). Only the WMAT reconstruction fell within the cut-off accepted by Telford and Birks (2011b; 95%), although the WA-Tol (inv) reconstruction explained more of the variance in the fossil data than 94.23% of reconstructions generated by transfer functions trained on random data. The ML reconstruction performed the worst, with a *P* value of 0.0736.

The new reconstructions were also compared with those using the transfer functions of Charman and Warner (1997) from Newfoundland (the original published reconstruction) and Booth (2008) from North America (Fig. 10). These



**Figure 9.** Histograms showing the proportion of variance in the full dataset explained by water-table depth (shown by the black line) in the three best performing models [as measured by RMSEP<sub>(jack)</sub>: WMAT, WA-Tol (inv) and ML]. Grey histograms show the variance explained by 999 reconstructions derived from transfer functions trained on random data. Dashed lines show the proportion of variance explained by first axis of PCA.



**Figure 10.** Comparison of water-table depth reconstructions from Nordan's Pond Bog, Newfoundland, using the three best performing models [as measured by RMSEP<sub>(jack)</sub>: WMAT, WA-Tol (inv) and ML] and the transfer functions of Charman and Warner (1997) and Booth (2008). Standard error shown for each reconstruction is derived from 1000 bootstrap cycles. All y-axis scales are depth to water table (cm).

published transfer functions both contain data used in the new models presented here (Table 1) and use weighted averaging-based models. Although there is obvious co-variation within the same broad range of depth to water-table values between all new models, differences in both the magnitude and the direction of change are evident. The WMAT and ML models exhibit a greater range of water-table values than the WA-Tol (inv) model, reconstructing drier values when the profile is dominated by taxa such as *Hyalosphenia subflava* and *Assulina muscorum* (see Hughes *et al.*, 2006, Fig. 6, for the testate amoeba assemblage), for example from ca. 5500 to 4500 cal a BP, and wetter values when the profile is dominated by taxa such as *Archerella flavum*, for example from ca. 4000 to 1000 cal a BP. Between the new models developed here and the existing models, correlations are higher with the reconstruction based on data from Booth (2008) than with the Charman and Warner (1997) data (Table 5), although the former model generally predicts drier water-tables throughout

the profile as a result of the higher number of continental sites included in that model, which provide better modern analogues for deep water-tables.

## Discussion

Judged by the statistics presented in Table 4, the WMAT model is the best performing, having the lowest RMSEP value of all models (however RMSEP is calculated). The reconstruction using the WMAT model also has the lowest palaeoSig P-value. However, previous research has shown that MAT-type models are the most likely to result in over-optimistic performance statistics and potentially misleading reconstructions as a result of problems such as spatial autocorrelation (Telford and Birks, 2005, 2009; Telford *et al.*, 2004) and the uneven sampling of environmental gradients (Telford and Birks, 2011a). Therefore, careful consideration of the evidence is necessary before a final choice of model is made.

The small relative decrease in model performance for the WMAT and WA-Tol (inv) models and improvement in model performance for the ML model when RMSEP<sub>(segment-wise)</sub> is calculated suggests that the unevenness present in the sampling of the water-table gradient (Fig. 7) does not have a significant adverse effect on the performance of any model type; for the WMAT and WA-Tol (inv) models the relative decrease in model performance is equivalent to that in the most evenly distributed datasets in Telford and Birks (2011a; SWAP and Adirondack). We can therefore be confident that all models will reproduce both wet (more samples) and dry (fewer samples) periods with equal reliability. RMSEP<sub>(LOSO)</sub> is the least optimistic method of calculating performance for all three models, but again the decrease in model performance is similar to published data (Payne *et al.*, 2012). In addition, DCA results showed overlap of all sites (supporting Fig. S2), with none assuming outlying positions (such as Bissendorfer Moor, Germany, in the transfer function of Charman *et al.*, 2007). This suggests that while RMSEP<sub>(jack)</sub> is not an appropriate method for assessing the performance of the clustered data, the limited extent of clustering should not have a detrimental effect on the model's ability to reconstruct relative shifts in peatland palaeohydrology, although absolute values are more uncertain, as shown by the higher RMSEP<sub>(LOSO)</sub>.

The tests for spatial autocorrelation in all models suggest that an effect is present (but not to the extreme extent of the datasets shown in Telford and Birks, 2009, Fig. 1b–f), although at least some of this effect may be attributed to the clustered dataset design. Of the three models, the WMAT showed the most extreme effect with a decrease in  $r^2$  in the neighbourhood analysis of 0.184, compared with 0.077 for WA-Tol (inv) and 0.043 for ML. This result, coupled with the findings of Telford and Birks (2005, 2009), challenges the assumption that the WMAT model is the best performing. Both of the alternative models, WA-Tol (inv) and ML, failed to explain more of the variance in the fossil data than 95% of reconstructions from randomly generated models, although for the WA-Tol (inv) model this failure was very marginal. Considering the similarities between fossil reconstructions produced by the three new models (Fig. 10) and the relatively robust performance of each model in a range of cross validation tests, it is considered likely that all three will reliably reproduce shifts in water-table from a fossil assemblage in the target region (correlation coefficients of the new model reconstructions in Fig. 10 are all  $>0.7$ ,  $P<0.01$ ; see Table 5). However, given the level of spatial autocorrelation in the WMAT model and the relatively poor performance of the ML model in the palaeoSig tests, the WA-Tol (inv) model appears the most reliable of the

**Table 5.** Correlation coefficients between the various transfer function reconstructions shown in Fig. 10. All values are significant at  $P < 0.01$ .

	Booth (2008)	Charman and Warner (1997)	WMAT (this study)	WA-Tol (this study)	ML (this study)
Booth (2008)	–				
Charman and Warner (1997)	0.460	–			
WMAT (this study)	0.807	0.400	–		
WA-Tol (this study)	0.846	0.623	0.733	–	
ML (this study)	0.880	0.636	0.790	0.941	–

three overall and is the favoured model. WA-Tol (inv) has a marginally higher RMSEP<sub>(LOSO)</sub> and RMSEP<sub>(segment-wise)</sub> than ML, but a lower RMSEP<sub>(jack)</sub> and has been commonly applied in published transfer functions (e.g. Charman and Warner, 1997; Lamentowicz *et al.*, 2008; Swindles *et al.*, 2009). The least optimistic RMSEP value for the WA-Tol (inv) model of 6.37 cm (RMSEP<sub>(LOSO)</sub>) is equivalent to or better than many published models (e.g. Charman *et al.*, 2007; Booth, 2008) and has additional statistical reliability because it performs well in the tests applied above.

Further analysis of results relating to the WA-Tol (inv) model provides insight into its reconstructive ability. One problem widely identified in previous testate amoeba transfer function development is the poor modern analogue status of certain taxa, in particular *Difflugia pulex*, which is common in Holocene fossil assemblages but often rare in modern training sets. Booth (2008) identified this problem in continental USA and while Charman *et al.* (2007) considered it largely addressed for their European transfer function, it remains a weakness in the WA-Tol (inv) model. Only 18 *D. pulex* individuals were counted in nine out of 232 samples, with their abundance in any one sample always remaining below 4% of the total assemblage. Figure 5 shows that *D. pulex* assumes an intermediate-wet position along the water-table gradient, which is inconsistent with other studies (e.g. Charman *et al.*, 2000, 2007), where *D. pulex* is typically an intermediate-dry indicator. In our model *D. pulex* has a water-table optimum that is almost identical to that of *Archerella flavum* (Fig. 5), which is commonly regarded as an indicator of wet-intermediate conditions and is exceptionally well represented (occurring in 137/232 samples, maximum abundance 70%, >8500 individuals counted). These two taxa can occur in antiphase in fossil assemblages, representing what may be considered alternating wetter and drier phases of bog surface wetness (seen in unpublished data from the study area), but these would not be reconstructed by our model. Two possible hypotheses may be suggested in such a case: (1) that the water-table optimum of *D. pulex* is poorly modelled and therefore the reconstruction is missing important palaeohydrological shifts, or that (2) the water-table optimum of *D. pulex* is reliably modelled and its past fluctuations are a response to some other factor. The most likely explanation suggested from monitoring of humidity changes in peatlands is that *D. pulex* is more abundant in conditions that show greater short-term fluctuations in hydrological status, whereas *A. flavum* is dependent on more stable hydrological conditions (Sullivan and Booth, 2011). Additionally, recent studies have highlighted the extent of genetic diversity in eukaryotes (e.g. Parfrey *et al.*, 2008), including the testate amoebae (e.g. Gomaa *et al.*, 2012) and it remains a possibility that cryptic diversity exists within *D. pulex*, meaning that individuals found in modern and fossil samples are different taxa.

The WA-Tol (inv) reconstruction is highly and significantly correlated with the ML and WMAT curves as well as with that of Booth (2008; all  $r > 0.7$ ,  $P < 0.01$ ), but the correlation with the

original published reconstruction (Hughes *et al.*, 2006) using the transfer function of Charman and Warner (1997) is weaker ( $0.62$ ,  $P < 0.01$ ). Indeed, while the new model reconstructions are visually similar and agree well with that of Booth (2008), the Charman and Warner (1997) curve appears relatively flat, with only one major increase in bog surface wetness occurring at ca. 5750 cal a BP, associated with an increase in *D. pulex* in the raw data (see Hughes *et al.*, 2006, Fig. 10) and suggesting this taxon is poorly modelled in this earlier and much smaller data set. However, other major shifts in the new reconstructions such as those at ca. 8250 and ca. 5500–4500 cal a BP are less evident in the Charman and Warner (1997) curve, suggesting a significant improvement in the hydrological modelling of many key taxa such as *Hyalosphenia subflava*, *Hyalosphenia papilio* and *Trigonopyxis arcula* type as a result of the much higher number of peatlands and microsites sampled. Comparison of the WA-Tol (inv) reconstruction to the Nordan's Pond Bog plant macrofossil DCA curve (see Hughes *et al.*, 2006, Fig. 8) also shows plausible ecological relationships between plant species/groups and testate amoebae and, by proxy, between phases of wetter and drier conditions, further supporting the reliability of our new model.

## Conclusions

Testate amoeba-based transfer functions to reconstruct the palaeohydrology of ombrotrophic peatlands have been developed in many geographical regions over the past 20 years. At the same time, research into the statistical assumptions underpinning these models has identified a number of potential problems that suggest, at best, that the performance statistics for existing models based on a range of proxies are over-optimistic and, at worst, that existing models may have no credible predictive power. Here, a new testate amoeba-based transfer function for water-table depth was developed and various model types were subjected to rigorous statistical analysis designed to directly address key problems such as spatial autocorrelation, clustered sampling design and uneven sampling gradients. Leave-one-out cross-validation, the most commonly applied method in recent studies, produced over-optimistic performance statistics for all models tested. However, by applying the methods suggested by Telford and Birks (2009, 2011a,b) and Payne *et al.* (2012) it was shown that a preferred model (developed using weighted averaging with tolerance downweighting) could be identified which retained a predictive capacity equivalent to other published models, even when less optimistic performance statistics were chosen. Qualitative assessment of new reconstructions on a fossil assemblage from Newfoundland supported this premise.

The recommendations of Telford and Birks (2009, 2011a,b) and Payne *et al.* (2012) that the statistical tests applied here should become standard practice in the development of new transfer functions are reiterated. For peatland testate amoeba, it is expected in most cases that the outcome will confirm that

transfer functions are performing well and that palaeohydrological reconstructions based on such data identify the major changes reliably. However, the additional rigour of these new approaches will help to differentiate between underlying models, identify any problematic datasets and enable improved error estimates. The application of the newly available tests will increase confidence in the results being produced, especially in data compilation and comparison efforts that tackle broader palaeoclimatic questions.

### Supporting information

Additional supporting information can be found in the online version of this article:

Figure S1. Correlation between single-point water-table depth measurements taken at 52 sampling locations in 2010 with annual average measurements determined by PVC stake discoloration (recorded in 2011).

Figure S2. Axis 1 and 2 biplots of sample and taxa scores from detrended correspondence analysis.

Please note: This supporting information is supplied by the authors, and may be re-organized for online delivery, but is not copy-edited or typeset by Wiley-Blackwell. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

**Acknowledgements.** This work was carried out as part of the NERC-funded project PRECIP (Palaeo REconstructions of ocean-atmosphere Coupling In Peat; grant numbers NE/G019851/1, NE/G020272/1, NE/G019673/1 and NE/G02006X/1). Thanks to Sue Rouillard in the Geography drawing office at the University of Exeter for assistance with Fig. 1. Richard Telford provided valuable advice and assistance in the application of various statistics in R. The comments of two anonymous reviewers improved an earlier version of the manuscript.

**Abbreviations.** CCA, Canonical Correspondence Analysis; DCA, Detrended Correspondence Analysis; LOSO, leave-one-site-out; MAT, modern analogue technique; ML, maximum likelihood; RMSEP, root mean square error of prediction; WA, weighted averaging; WA-PLS, weighted average partial least squares; WA-Tol, weighted averaging with tolerance down-weighting; WMAT, weighted mean MAT.

### References

- Amesbury MJ, Charman DJ, Fyfe RM, et al. 2008. Bronze Age upland settlement decline in southwest England: testing the climate change hypothesis. *Journal of Archaeological Science* **35**: 87–98.
- Amesbury MJ, Barber KE, Hughes PDM. 2012. The relationship of fine-resolution, multi-proxy palaeoclimate records to meteorological data at Fågelmossen, Värmland, Sweden and the implications for the debate on climate drivers of the peat-based record. *Quaternary International* **268**: 77–86.
- Belyea LR. 1999. A novel indicator of reducing conditions and water-table depth in mires. *Functional Ecology* **13**: 431–434.
- Birks HJB. 1995. Quantitative palaeoenvironmental reconstructions. In *Statistical Modelling of Quaternary Science Data*, Maddy D, Brew JS (eds.). Cambridge: Quaternary Research Association: Technical Guide 5, 161–254.
- Birks HJB. 1998. Numerical tools in palaeolimnology – progress, potentialities and problems. *Journal of Paleolimnology* **20**: 307–332.
- Birks HJB, Heiri O, Seppä H, et al. 2010. Strengths and weaknesses of quantitative climate reconstructions based on late-Quaternary biological proxies. *The Open Ecology Journal* **3**: 68–110.
- Blundell AC, Charman DJ, Barber KE. 2008. Multiproxy late Holocene peat records from Ireland: towards a regional palaeoclimatic curve. *Journal of Quaternary Science* **23**: 59–71.
- Bobrov AA, Charman DJ, Warner BG. 1999. Ecology of testate amoebae (Protozoa: Rhizopoda) on peatlands in western Russia with special attention to niche separation in closely related taxa. *Protist* **150**: 125–136.
- Booth RK. 2002. Testate amoebae as paleoindicators of surface-moisture changes on Michigan peatlands: modern ecology and hydrological calibration. *Journal of Paleolimnology* **28**: 329–348.
- Booth RK. 2008. Testate amoebae as proxies for mean annual water-table depth in Sphagnum-dominated peatlands of North America. *Journal of Quaternary Science* **23**: 43–57.
- Booth RK. 2010. Testing the climate sensitivity of peat-based paleoclimate reconstructions in mid-continental North America. *Quaternary Science Reviews* **29**: 720–731.
- Booth RK, Lamentowicz M, Charman DJ. 2010. Preparation and analysis of testate amoebae in peatland paleoenvironmental studies. *Mires and Peat* **7**: 1–7.
- Charman DJ. 1997. Modelling hydrological relationships of testate amoebae (Protozoa: Rhizopoda) on New Zealand peatlands. *Journal of the Royal Society of New Zealand* **27**: 465–483.
- Charman DJ. 2010. Centennial climate variability in the British Isles during the mid-late Holocene. *Quaternary Science Reviews* **29**: 1539–1554.
- Charman DJ, Warner BG. 1997. The ecology of testate amoebae (Protozoa: Rhizopoda) in oceanic peatlands in Newfoundland, Canada: modelling hydrological relationships for palaeoenvironmental reconstruction. *Écoscience* **4**: 555–562.
- Charman DJ, Hendon D, Woodland WA. 2000. *The Identification of Testate Amoebae (Protozoa: Rhizopoda) in Peats*. London: Quaternary Research Association: Technical Guide 9.
- Charman DJ, Blundell A, Chiverrell RC, et al. 2006. Compilation of non-annually resolved Holocene proxy climate records: stacked Holocene peatland palaeo-water table reconstructions from northern Britain. *Quaternary Science Reviews* **25**: 336–350.
- Charman DJ, Blundell A, and ACCROTEL MBERS 2007. A new European testate amoebae transfer function for palaeohydrological reconstruction on ombrotrophic peatlands. *Journal of Quaternary Science* **22**: 209–221.
- Charman DJ, Barber KE, Blaauw M, et al. 2009. Climate drivers for peatland palaeoclimate records. *Quaternary Science Reviews* **28**: 1811–1819.
- Gomaa F, Todorov M, Heger TJ, et al. 2012. SSU rRNA phylogeny of Arcellinida (Amoebozoa) reveals that the largest Arcellinid genus, *Difflugia* Leclerc 1815, is not monophyletic. *Protist* **163**: 389–399.
- Hendon D, Charman DJ. 1997. The preparation of testate amoebae (Protozoa: Rhizopoda) samples from peat. *The Holocene* **7**: 199–205.
- Hughes PDM, Blundell A, Charman DJ, et al. 2006. An 8500 cal. year multi-proxy climate record from a bog in eastern Newfoundland: contributions of meltwater discharge and solar forcing. *Quaternary Science Reviews* **25**: 1208–1227.
- Juggins S. 2009. rioja: Analysis of Quaternary Science Data. R package version 0.5–6.
- Juggins S, Birks HJB. 2012. Quantitative environmental reconstructions from biological data. In *Tracking Environmental Change Using Lake Sediments. Data Handling and Numerical Techniques*, Birks HJB, Lotter AF, Juggins S, Smol JP (eds) London: Springer: 431–494.
- Lamarre A, Garneau M, Asnong H. 2012. Holocene palaeohydrological reconstruction of a permafrost peatland using testate amoebae and macrofossil analyses, Kuujjuarapik, subarctic Québec, Canada. *Review of Palaeobotany and Palynology*, in press
- Lamentowicz L, Lamentowicz M, Gabka M. 2008. Testate amoebae ecology and a local transfer function from a peatland in western Poland. *Wetlands* **28**: 164–175.
- Lamentowicz M, van der Knapp W, Lamentowicz L, et al. 2010. A near-annual palaeohydrological study based on testate amoebae from a sub-alpine mire: surface wetness and the role of climate during the instrumental period. *Journal of Quaternary Science* **25**: 190–202.
- Loisel J, Garneau M. 2010. Late Holocene paleoecohydrology and carbon accumulation estimates from two boreal peat bogs in eastern Canada: potential and limits of multi-proxy archives. *Palaeogeography, Palaeoclimatology, Palaeoecology* **291**: 493–533.
- Markel E, Booth RK, Qin Y. 2010. Testate amoebae and  $\delta^{13}\text{C}$  of Sphagnum as surface-moisture proxies in Alaskan peatlands. *The Holocene* **20**: 463–475.
- Meisterfeld R. 2002. Testate amoebae with filopodia. In *The Illustrated Guide to the Protozoa*, Lee J, Leedale JJ, Bradbury P (eds). Society of Protozoologists: Lawrence, KS; 1054–1084.

- Mitchell EAD, Charman DJ, Warner BG. 2008. Testate amoebae analysis in ecological and palaeoecological studies of wetlands: past, present and future. *Biodiversity and Conservation* **17**: 2115–2137.
- National Wetlands Working Group. 1988. *Wetlands of Canada*. Ecological Land Classification Series, No. 24. Sustainable Development Branch, Environment Canada: Ottawa, and Polyscience Publications Inc.: Montreal.
- Ogden CG, Hedley RH. 1980. *An Atlas of Freshwater Testate Amoebae*. Oxford; Oxford University Press.
- Parfrey LW, Lahr DJG, Katz LA. 2008. The dynamic nature of Eukaryotic genomes. *Molecular Biology and Evolution* **25**: 787–794.
- Payne RJ, Kishaba K, Blackford JJ, et al. 2006. Ecology of testate amoebae (Protista) in south-central Alaska peatlands: building transfer-function models for paleoenvironmental studies. *The Holocene* **16**: 403–414.
- Payne RJ, Mitchell EAD. 2007. Ecology of testate amoebae from mires in the central Rhodope Mountains, Greece and development of a transfer function for palaeohydrological reconstruction. *Protist* **158**: 159–171.
- Payne RJ, Charman DJ, Matthews S, et al. 2008. Testate amoebae as palaeohydrological proxies in Sürmeme Ağacbaşı Yaylası peatland (northeast Turkey). *Wetlands* **28**: 311–323.
- Payne RJ, Telford RJ, Blackford JJ, et al. 2012. Testing peatland testate amoeba transfer functions: appropriate methods for clustered training-sets. *The Holocene* **22**: 818–825.
- R Development Core Team. 2011. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna.
- Salonen JS, Ilvonen L, Seppä H, et al. 2012. Comparing different calibration methods (WA/WA-PLS regression and Bayesian modeling) and different-sized calibration sets in pollen-based quantitative climate reconstruction. *The Holocene* **22**: 413–424.
- Sillasoo U, Mauquoy D, Blundell A, et al. 2007. Peat multi-proxy data from Männikjärve bog as indicators of late Holocene climatic changes in Estonia. *Boreas* **36**: 20–37.
- Sullivan M, Booth RK. 2007. *Key to Testate Amoebae Inhabiting Sphagnum-Dominated Peatlands with an Emphasis on Taxa Preserved in Holocene Sediments*. Earth and Environmental Science Department: Lehigh University: Bethlehem, PA.
- Sullivan ME, Booth RK. 2011. The potential influence of short-term environmental variability on the composition of testate amoeba communities in Sphagnum peatlands. *Microbial Ecology* **62**: 80–93.
- Swindles GT, Charman DJ, Roe HM, et al. 2009. Environmental controls on peatland testate amoebae (Protozoa: Rhizopoda) in the North of Ireland: implications for Holocene palaeoclimate studies. *Journal of Paleolimnology* **42**: 123–140.
- Telford RJ. 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. *Quaternary Science Reviews* **24**: 2173–2179.
- Telford RJ, Birks HJB. 2009. Evaluation of transfer functions in spatially structured environments. *Quaternary Science Reviews* **28**: 1309–1316.
- Telford RJ, Birks HJB. 2011a. Effect of uneven sampling along an environmental gradient on transfer function performance. *Journal of Paleolimnology* **46**: 99–106.
- Telford RJ, Birks HJB. 2011b. A novel method for assessing the statistical significance of quantitative reconstructions inferred from biotic assemblages. *Quaternary Science Reviews* **30**: 1272–1278.
- Telford RJ, Andersson C, Birks HJB, et al. 2004. Biases in the estimation of transfer function prediction errors. *Paleoceanography* **19**: DOI: 10.1029/2004PA001072.
- ter Braak CJF, Prentice IC. 1988. A theory of gradient analysis. *Advances in Ecological Research* **18**: 271–317.
- ter Braak CJF, van Dam H. 1989. Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologica* **178**: 209–223.
- ter Braak CJF, Juggins S. 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologica* **269/270**: 485–502.
- ter Braak CJF, Šmilauer P. 2002. *CANOCO Reference Manual and CanoDraw for Windows User's guide: Software for Canonical Community Ordination (version 4.5)*. Microcomputer Power: Ithaca, NY.
- van Bellen S, Garneau M, Booth RK. 2011. Holocene carbon accumulation rates from three ombrotrophic peatlands in boreal Quebec, Canada: impact of climate-driven ecohydrological change. *The Holocene* **21**: 1217–1231.
- Woodland WA, Charman DJ, Sims PC. 1998. Quantitative estimates of water tables and soil moisture in Holocene peatlands from testate amoebae. *The Holocene* **8**: 261–273.