Significance testing testate amoeba water table reconstructions

Richard J Payne[1,2], Kirill V Babeshko[2], Simon van Bellen[3], Jeffrey J Blackford[4], Robert K Booth[5], Dan J Charman[6], Megan R Ellershaw[7], Daniel Gilbert[8], Paul D M Hughes[9], Vincent EJ Jassey[8,10], Łukasz Lamentowicz[11], Mariusz Lamentowicz[11], Elena A Malysheva[2], Dmitri Mauquoy[12], Yuri Mazei[2,13], Edward A D Mitchell[14,15], Graeme T Swindles[16], Andrey N Tsyganov[2], T Edward Turner[16], Richard J Telford[17]

1 Environment, University of York, Heslington, York, YO105DD, United Kingdom.

2 Department of Zoology and Ecology, Penza State University, Krasnaya str. 40, 440026 Penza, Russia.

3 GEOTOP-Université du Québec à Montréal, 201 Avenue Président-Kennedy, Montréal, Québec, H2X 3Y7, Canada.

4 Department of Geography, Environment and Earth Sciences, University of Hull, Cottingham Road, Hull, HU6 7RX, United Kingdom.

5 Department of Earth and Environmental Sciences, Lehigh University, 1 West Packer Avenue, Bethlehem, PA 18015-3001, USA.

6 Geography, College of Life and Environmental Sciences, University of Exeter, Amory Building, Rennes Drive, Exeter, EX4 4RJ, United Kingdom.

7 Natural England, Northminster House, Peterborough, PE1 1UA, United Kingdom.

8 Laboratoire Chrono-Environnement, UMR 6249 Université de Franche-Comté/CNRS, UFR Sciences et Techniques, 16 route de Gray, 25030 Besançon, France.

9 Geography and Environment, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom.

10 Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Site Lausanne, Station 2, 1015 Lausanne, Switzerland.

11 Department of Biogeography and Palaeoecology, Faculty of Geographical and Geological Science, Adam Mickiewicz University, Poznan, ul. Dziegielowa 27, 61-680 Poznan, Poland.

12 School of Geosciences, University of Aberdeen, Elphinstone Road, Aberdeen AB24 3UF, United Kingdom.

13 Department of Hydrobiology, Lomonosov Moscow State University, Leninskiye gory, 1, Moscow 119991, Russia.

14 Laboratory of Soil Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland.

15 Jardin Botanique de Neuchâtel, Pertuis-du-Sault 58, CH- 2000 Neuchâtel, Switzerland

16 School of Geography, University of Leeds, Leeds, LS2 9JT, United Kingdom.

17 EECRG, Department of Biology, University of Bergen, Allegaten 41, N-5007 Bergen, Norway.

ABSTRACT

Transfer functions are valuable tools in palaeoecology, but their output may not always be meaningful. A recently-developed statistical test (‘randomTF’) offers the potential to distinguish among reconstructions which are more likely to be useful, and those less so. We applied this test to a large number of reconstructions of peatland water table depth based on testate amoebae. Contrary to our expectations, a substantial majority (25 of 30) of these reconstructions gave non-significant results (P>0.05). The underlying reasons for this outcome are unclear. We found no significant correlation between randomTF P-value and transfer function performance, the properties of the training set and reconstruction, or measures of transfer function fit. These results give cause for concern but we believe it would be extremely premature to discount the results of non-significant reconstructions. We stress the need for more critical assessment of transfer function output, replication of results and ecologically-informed interpretation of palaeoecological data.

KEYWORDS: Testate amoeba; Protist; Palaeoecology; Palaeohydrology; Transfer function; randomTF

INTRODUCTION

Testate amoebae are widely-used proxies in palaeoecological studies; in particular for the reconstruction of water table depth in peatlands (Charman, 2001; Mitchell et al., 2008). Over the last 25 years palaeoecology has been revolutionised by the use of statistical models (transfer functions) to quantitatively reconstruct environmental variables. However, questions are increasingly being raised about the reliability and robustness of transfer function results (Belyea, 2007; Juggins, 2013).

A transfer function will always give an output but that output may not always be meaningful. The only way to establish whether the output of a transfer function is ‘true’ is by comparing the results to independent data, but such data are not always available and even in such cases correlations are complicated by temporal autocorrelation and the limitations of the chronology.

Although we cannot realistically assess whether all reconstructions are correct we can conceivably test whether they are potentially useful. Telford and Birks (2011) propose a pragmatic solution: that a reconstruction can be considered statistically significant if it explains more of the variance in the fossil data than those of transfer functions trained on randomly-generated data. Telford and Birks (2011) propose a method, ‘randomTF’, in which:

1. The transfer function is applied to the fossil data to derive a reconstruction (using any commonly-applied method).
2. The proportion of variance in the fossil data explained by the reconstruction is determined using constrained ordination.
3. Multiple new transfer functions are derived using the established modern species data but with the environmental data replaced by uniformly distributed random variables.
4. These transfer functions are applied in turn to the fossil data and the variance they explain tested. This is repeated a large number of times, typically 999.
5. A reconstruction is considered statistically significant when the proportion of variance explained is greater than that of 95% of the transfer functions based on randomly-generated data.

We would expect reliable reconstructions to explain more variance in the fossil data than transfer functions trained on random data, and therefore to give significant results. However, a significant randomTF value is not proof of accuracy and a non-significant result does not necessarily imply inaccuracy. Non-significant results do however give cause for concern and suggest that transfer function output should be treated with caution. randomTF tests can potentially tell us which reconstructions we should trust more, which less, and whether we can predict more than one environmental variable from the same fossil dataset. Telford and Birks (2011) also propose an alternative test (‘obs.cor’) based on the correlation of optima values with axis species scores from a constrained ordination of the fossil data. This test is not applicable to all transfer functions methods and is not considered here. The randomTF test has been applied in a few studies (Amesbury et al., 2013; Lamarre et al., 2013; Swindles et al., 2015a) but is not yet routinely used in testate amoeba palaeoecology. Here we apply this test to a large number of published and unpublished records with the aim to identify the characteristics which are likely to lead to better reconstructions, giving better randomTF results.

METHODS

We identified 30 published and unpublished testate amoeba palaeoecological records (Table 1). These records span a large range of regions, mire types, analysts, time periods, and sampling resolutions, and form a large and reasonably representative sample of testate amoeba palaeoecological research. Reconstructions of water table depth were produced using either the transfer function used in the original study, the most geographically-appropriate model where a transfer function was not previously applied, or in a few cases transfer functions which have been produced since the data were originally published. Taxonomy was harmonised between the fossil data and training set, which in many instances required the grouping or deletion of some taxa (performance statistics may therefore differ slightly from those previously published). Transfer functions were applied based on the model selected by the original authors with sample specific errors calculated by bootstrapping (1000 cycles). All transfer functions were based on either weighted averaging, weighted averaging with tolerance downweighting or weighted average-partial least squares (Birks, 1995). We applied randomTF using 999 permutations with redundancy analysis as the ordination method. Analyses were conducted in R3.1.2 (R Development Core Team, 2014) using the packages analogue (Simpson, 2007), rioja (Juggins, 2009) and palaeoSig (Telford, 2011).

RESULTS and DISCUSSION

Only five of the 30 tests yielded a significant P-value ($P<0.05$; Table 1). While we expected that some reconstructions would give non-significant results this proportion is much higher than we anticipated. While a few reconstructions fail to reach $P=0.05$ by a relatively narrow margin (Tørvesø 1, Staroselsky Moch, Dot Lake B), many more have P-values which substantially exceed this value.

Another two records published in the literature have given significant P-values: those of Swindles et al. (2015a) for Stordalen, Sweden and Lamarre et al. (2013) for Lac Le Caron, Canada. Amesbury et al. (2013) found a significant result for the Nordans Pond site of Hughes et al. (2006) using an extended transfer function, whereas here we find a non-significant result using the transfer function used in the original study (Charman and Warner, 1997). In the latter three cases multiple model structures were tested with some producing significant reconstructions, and some not. We note that in these instances a correction for multiple comparisons (such as a Bonferroni correction) would probably have meant that the reconstructions did not reach significance. However, even if these results are included, eight significant P-values out of 32 reconstructions remains a surprisingly low proportion.

Telford and Birks (2011) identify four factors which might make the randomTF test prone to type II error (“false negative”): low numbers of effective species; small numbers of fossil samples; limited variability in the reconstruction and poorly-performing or poorly-fitting transfer functions. All of these factors apply to some of the reconstructions we examine but it is not clear that any are a consistent cause of non-significant P-values. Overall, P-value was not significantly correlated with properties of the training set (mean/standard deviation/range of WTD) or fossil data (species richness, Hill’s N2 or number of samples), performance metrics of the transfer function (leave one out RMSEP or R^2), properties of the reconstruction (mean/standard deviation/range of predications, ratio of range to RMSEP or training set range, mean boot-strapped error estimates) or measures of transfer function fit (proportion of shared taxa, proportion of fossil samples with poor modern analogues, squared residual length)(Spearman Rs; $P>0.05$). P-value was strongly correlated with the proportion of variance in the fossil data explained by the reconstruction (Spearman Rs=-0.89, $P<0.001$), suggesting (unsurprisingly) that where a high proportion of variance is explained this is unlikely to be exceeded by transfer functions trained on random data.

The five reconstructions yielding significant results were three short records from the Elatia Forest of northern Greece (Dexameni; Krya Vrissi 1&2; Payne and Pates (2009)), the high-resolution Mauntschas record from the Swiss Alps (Lamentowicz et al., 2010; van der Knaap et al., 2011) and a record from Frasne in the Jura Mountains of eastern France (Jassey et al. unpublished). These five records have little obvious similarity. The transfer functions used for the Dexameni, Krya Vrissi and Mauntschas reconstructions all included samples from the same sites and for Frasne the closest training set site was only c.10 km distant. However, ten of the sites with non-significant reconstructions were also included in their respective training sets. The three short records from Greece (Payne and Pates, 2009) are all characterised by a single large change –a shift to drier conditions in the recent past but this is not a feature of the Frasne or Mauntschas records and some non-significant reconstructions are similar (e.g. Andorra: van Bellen et al., in press).

Non-significant reconstructions include studies where it is difficult to see any *a priori* reason to suspect problems: sites with high resolution, good numbers of samples and species, transfer functions which perform well in cross-validation, include samples from the same site with good overlap in assemblage and with modern and fossil samples counted by the same analyst (e.g. Dead Island: (Swindles et al., 2010), Minden: Booth and Jackson (2003)). Non-significant results in these instances are a real surprise.

Our results provide some evidence that where there is a choice of transfer function this can affect randomTF significance level. With the Nordans Pond record (Hughes et al., 2006) the transfer function used in the original publication yields a non-significant P-value (P=0.83; Table 1) while a more recent transfer function with a larger training set and better performance statistics gives a significant P-value (Amesbury et al., 2013). With the Frasne record (Jassey and Gilbert unpublished) marginally better results are found with the smaller and weaker-performing, but geographically closer, Jura transfer function (Mitchell et al., 1999) than the larger, better-performing, north-west Europe model (Charman and Blundell, 2007) although both are $P<0.05$. It is also probable that in borderline cases difference in selection of model or samples included in the training set or fossil data might make the difference between a P-value above or below the usual threshold of 0.05 (*cf*. Amesbury et al., 2013; Swindles et al., 2015).

So what should we take from these results?

Interpreting these findings is a challenge and among the authors of this paper there is a considerable range of viewpoints. In the original paper Telford and Birks (2011) state that '*reconstructions that fail this test have limited credibility and should be treated with considerable caution*'. On this basis these results could be taken to question the reliability of a substantial proportion of published testate amoeba water table reconstructions and thus raise questions about the approach as a whole.

However there are also strong arguments for a more cautious interpretation. Unlike many proxies the ecological underpinnings of testate amoeba palaeoecology are strong. The thickness of water films, for which water table depth is a surrogate, determine an amoeba's ability to move and feed, (although seasonal variability is an important area of uncertainty: (Marcisz et al., 2014b)). Numerous modern studies have found significant correlations between amoeba communities and water table depth (Mitchell et al., 2008) and studies have begun to support this link experimentally (Marcisz et al., 2014a; Mulot et al., 2014). While there undoubtedly are both practical and fundamental issues which can complicate palaeoecological reconstruction, many of these are common to other proxies and archives. Most testate amoeba analysts would expect our reconstructions to satisfy the fundamental requirements for quantitative palaeoecology laid out by Birks (1995).

A substantial proportion of all applications of the randomTF test, with a variety of proxies in a variety of settings, have produced non-significant results (Cwynar et al., 2012; Luoto et al., 2014; Salonen et al., 2013; Wooller et al., 2012). If non-significant results are so common it is arguable that the test may be overly pessimistic. In the case of the records we consider here it can be argued that there are reasons to accept many of the non-significant reconstructions based on correlations between proxies and with independent data sources. There are also an increasing number of studies which show transfer functions to have acceptable performance when tested with independent data (Payne et al., 2012; Swindles et al., 2015b). The increasing number of pitfalls and caveats which have been identified in transfer functions over recent decades should be a warning to palaeoecologists of the dangers of uncritical acceptance of new statistical methods.

It is important to reiterate that even from the most sceptical viewpoint a non-significant randomTF P-value does not *prove* that a reconstruction is invalid. It would be very premature to discount the results of reconstructions identified as non-significant here; even non-significant reconstructions may still be useful. However the unexpected finding that many reconstructions fail this test clearly shows the requirement for a more detailed and critical assessment of reconstructions and a better understanding of the factors which cause non-significant results. Transfer function reconstructions should always be accompanied by thorough ecological interpretation of the record. For instance where a dry shift is reconstructed by the transfer function on the basis of a switch in dominance from *Archerella flavum* to *Trigonopyxis arcula* this can probably be considered robust given the well-understood hydrological preferences of these two taxa. However a similar reconstructed change should be treated with much greater caution if it is based on a change in dominance from (for instance) *Heleopera petricola* to *Cryptodifflugia sacculus*; taxa with much less well-understood ecological preferences.

Replication of results among cores, sites, proxies and archives, and informed ecological interpretation of the primary data remain the best ways for palaeoecologists to ensure that our results are valid and useful.

Acknowledgements

Elements of this project were supported by various funding agencies including: grant PSPB-013/2010 through the Swiss Contribution to the enlarged EU; project 017008 through EU FP6 (Millenium); grant 14-14-00891 from the Russian Scientific Fund; grant IE150173 from the Royal Society; NERC doctoral training grant NE/G52398X/1; NERC research grants NE/I022809/1, NE/I022981/1, NE/I022833/1 and NE/I023104/1; a DEL-NI doctoral training grant and the Worldwide Universities Network (project: Arctic Environments, Vulnerabilities and Opportunities). RJP is supported by the Leverhulme Trust through project RPG-2015-162.

Author contributions: The study was conceived by RJP and RJT. RJP conducted statistical analyses and wrote the first draft of the manuscript. Other authors contributed data, advised on taxonomy, and provided interpretation.

References

Amesbury, M.J., Mallon, G., Charman, D.J., Hughes, P.D.M., Booth, R.K., Daley, T.J., Garneau, M., 2013. 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. Journal of Quaternary Science 28, 27-39.

Belyea, L., 2007. Revealing the Emperor's new clothes: niche-based environmental reconstruction in the light of recent ecological theory. The Holocene 17, 683-688.

Birks, H.J.B., 1995. Quantitative palaeoenvironmental reconstructions, in: Maddy, D., Brew, J.S. (Eds.), Statistical modelling of Quaternary science data. Quaternary Research Association, Cambridge, pp. 161–254.

Booth, R.K., 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, R.K., Jackson, S.T., 2003. A high-resolution record of late-Holocene moisture variability from a Michigan raised bog, USA. The Holocene 13, 863-876.

Charman, D.J., 2001. Biostratigraphic and palaeoenvironmental applications of testate amoebae. Quaternary Science Reviews 20, 1753-1764.

Charman, D.J., Blundell, A., 2007. A new European testate amoebae transfer function for palaeohydrological reconstruction on ombrotrophic peatlands. Journal of Quaternary Science 22, 209-221.

Charman, D.J., Warner, B.G., 1997. ecology of testate amoebae (Protozoa: Rhizopoda) in oceanic peatlands in Newfoundland, Canada: modelling hydrological relationships for palaeoenvironmental reconstruction. Ecoscience.

Cwynar, L., Rees, A.H., Pedersen, C., Engels, S., 2012. Depth distribution of chironomids and an evaluation of site-specific and regional lake-depth inference models: a good model gone bad? J Paleolimnol 48, 517-533.

Ellershaw, M., 2004. Holocene climate change in the North Atlantic region: evidence from peat deposits. University of London, London.

Hendon, D., Charman, D., Kent, M., 2001. Palaeohydrological records derived from testate amoebae analysis from peatlands in northern England: within-site variability, between-site comparability and palaeoclimatic implications. The Holocene 11, 127-148.

Hughes, P., Blundell, A., Charman, D., Bartlett, S., Daniell, J., Wojatschke, A., Chambers, F.M., 2006. An 8500cal. 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. URL http://cran.r-project.org/package= rioja.

Juggins, S., 2013. Quantitative reconstructions in palaeolimnology: new paradigm or sick science? Quaternary Science Reviews 64, 20-32.

Lamarre, A., Magnan, G., Garneau, M., Boucher, É., 2013. A testate amoeba-based transfer function for paleohydrological reconstruction from boreal and subarctic peatlands in northeastern Canada. Quaternary International 306, 88-96.

Lamentowicz, M., Gałka, M., Milecka, K., Tobolski, K., Lamentowicz, Ł., FIAŁKIEWICZ-KOZIEŁ, B., Blaauw, M., 2013. A 1300-year multi-proxy, high-resolution record from a rich fen in northern Poland: reconstructing hydrology, land use and climate change. Journal of Quaternary Science 28, 582-594.

Lamentowicz, M., Obremska, M., 2010. A rapid response of testate amoebae and vegetation to inundation of a kettle hole mire. J Paleolimnol 43, 499-511.

Lamentowicz, M., Obremska, M., Mitchell, E.A., 2008. Autogenic succession, land-use change, and climatic influences on the Holocene development of a kettle-hole mire in Northern Poland. Review of Palaeobotany and Palynology 151, 21-40.

Lamentowicz, M., Tobolski, K., Mitchell, E.A., 2007. Palaeoecological evidence for anthropogenic acidification of a kettle-hole peatland in northern Poland. The Holocene 17, 1185-1196.

Lamentowicz, M., Van Der Knaap, W., Lamentowicz, Ł., Van Leeuwen, J.F., Mitchell, E.A., Goslar, T., Kamenik, C., 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.

Luoto, T.P., Kaukolehto, M., Weckström, J., Korhola, A., Väliranta, M., 2014. New evidence of warm early-Holocene summers in subarctic Finland based on an enhanced regional chironomid-based temperature calibration model. Quaternary Research 81, 50-62.

Marcisz, K., Fournier, B., Gilbert, D., Lamentowicz, M., Mitchell, E.A., 2014a. Response of Sphagnum peatland testate amoebae to a 1-year transplantation experiment along an artificial hydrological gradient. Microbial ecology 67, 810-818.

Marcisz, K., Lamentowicz, Ł., Słowińska, S., Słowiński, M., Muszak, W., Lamentowicz, M., 2014b. Seasonal changes in Sphagnum peatland testate amoeba communities along a hydrological gradient. European journal of protistology 50, 445-455.

Mitchell, E.A., Buttler, A.J., Warner, B.G., Gobat, J.-M., 1999. Ecology of testate amoebae (Protozoa: Rhizopoda) in Sphagnum peatlands in the Jura mountains, Switzerland and France. Ecoscience, 565-576.

Mitchell, E.A., van Der Knaap, W., van Leeuwen, J.F., Buttler, A., Warner, B.G., Gobat, J.-M., 2001. The palaeoecological history of the Praz-Rodet bog (Swiss Jura) based on pollen, plant macrofossils and testate amoebae (Protozoa). The Holocene 11, 65-80.

Mitchell, E.A.D., Charman, D.J., Warner, B.G., 2008. Testate amoebae analysis in ecological and paleoecological studies of wetlands: past, present and future. Biodiversity and Conservation 17, 2115-2137.

Mitchell, E.A.D., Payne, R.J., van der Knaap, W.O., Lamentowicz, Ł., Gąbka, M., Lamentowicz, M., 2013. The performance of single-and multi-proxy transfer functions (testate amoebae, bryophytes, vascular plants) for reconstructing mire surface wetness and pH. Quaternary Research 79, 6-13.

Mulot, M., Villard, A., Mitchell, E.A., 2014. Adaptation of *Sphagnum* peatland testate amoebae communities to an artificial water table depth gradient over a 19 month period, in: Marcisz, K. (Ed.), 7th International Symposium on Testate Amoebae. Adam Mickiewicz University, Poznan, Poland, p. 17.

Novenko, E.Y., Tsyganov, A.N., Volkova, E.M., Babeshko, K.V., Lavrentiev, N.V., Payne, R.J., Mazei, Y.A., 2015. The Holocene paleoenvironmental history of central European Russia reconstructed from pollen, plant macrofossil, and testate amoeba analyses of the Klukva peatland, Tula region. Quaternary Research 83, 459-468.

Payne, R., Telford, R., Blackford, J., Blundell, A., Booth, R., Charman, D., Lamentowicz, Ł., Lamentowicz, M., Mitchell, E., Potts, G., Swindles, G., Warner, B., W, W., 2012. Testing peatland testate amoeba transfer functions: appropriate methods for clustered training-sets. Holocene 22, 819-825.

Payne, R.J., 2011. Can testate amoeba-based palaeohydrology be extended to fens? Journal of Quaternary Science 26, 15-27.

Payne, R.J., 2014. A natural experiment suggests little direct temperature forcing of the peatland palaeoclimate record. Journal of Quaternary Science 29, 509-514.

Payne, R.J., Charman, D.J., Matthews, S., Eastwood, W.J., 2008. Testate amoebae as palaeohydrological proxies in sürmene ağaçbaşi yaylasi peatland (northeast Turkey). Wetlands 28, 311-323.

Payne, R.J., Kishaba, K., Blackford, J.J., Mitchell, E.A., 2006. Ecology of testate amoebae (Protista) in south-central Alaska peatlands: building transfer-function models for palaeoenvironmental studies. The Holocene 16, 403-414.

Payne, R.J., Malysheva, E., Tsyganov, A., Pampura, T., Novenko, E., Volkova, E., Babeshko, K., Mazei, Y., in press. A multi-proxy record of Holocene environmental change, peatland development and carbon accumulation from Staroselsky Moch peatland, Russia. . The Holocene.

Payne, R.J., Mitchell, E.A., 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, R.J., Mitchell, E.A., 2009. How many is enough? Determining optimal count totals for ecological and palaeoecological studies of testate amoebae. J Paleolimnol 42, 483-495.

Payne, R.J., Pates, J.M., 2009. Vertical stratification of testate amoebae in the Elatia Mires, northern Greece: palaeoecological evidence for a wetland response to recent climatic change, or autogenic processes? Wetlands Ecology and Management 17, 355-364.

R Development Core Team, 2014. R: A language and environment for statistical computing. R foundation for Statistical Computing.

Salonen, J.S., Helmens, K.F., Seppä, H., Birks, H.J.B., 2013. Pollen-based palaeoclimate reconstructions over long glacial–interglacial timescales: methodological tests based on the Holocene and MIS 5d–c deposits at Sokli, northern Finland. Journal of Quaternary Science 28, 271-282.

Simpson, G.L., 2007. Analogue methods in palaeoecology: using the analogue package. Journal of Statistical Software 22, 1-29.

Swindles, G., Charman, D., Roe, H., Sansum, P., 2009. Environmental controls on peatland testate amoebae (Protozoa: Rhizopoda) in the North of Ireland: implications for Holocene palaeoclimate studies. J Paleolimnol 42, 123-140.

Swindles, G.T., Amesbury, M.J., Turner, T.E., Carrivick, J.L., Woulds, C., Raby, C., Mullan, D., Roland, T.P., Galloway, J.M., Parry, L., Kokfelt, U., Garneau, M., Charman, D.J., Holden, J., 2015a. Evaluating the use of testate amoebae for palaeohydrological reconstruction in permafrost peatlands. Palaeogeography, Palaeoclimatology, Palaeoecology 424, 111-122.

Swindles, G.T., Blundell, A., Roe, H.M., Hall, V.A., 2010. A 4500-year proxy climate record from peatlands in the North of Ireland: the identification of widespread summer ‘drought phases’? Quaternary Science Reviews 29, 1577-1589.

Swindles, G.T., Holden, J., Raby, C.L., Turner, T.E., Blundell, A., Charman, D.J., Menberu, M.W., Kløve, B., 2015b. Testing peatland water-table depth transfer functions using high-resolution hydrological monitoring data. Quaternary Science Reviews 120, 107-117.

Telford, R., 2011. palaeoSig: significance tests of quantitative palaeoenvironmental reconstructions. R package version 1.

Telford, R.J., Birks, H.J.B., 2011. A novel method for assessing the statistical significance of quantitative reconstructions inferred from biotic assemblages. Quaternary Science Reviews 30, 1272-1278.

Turner, T.E., Swindles, G.T., Charman, D.J., Blundell, A., 2013. Comparing regional and supra-regional transfer functions for palaeohydrological reconstruction from Holocene peatlands. Palaeogeography, Palaeoclimatology, Palaeoecology 369, 395-408.

Turner, T.E., Swindles, G.T., Roucoux, K.H., 2014. Late Holocene ecohydrological and carbon dynamics of a UK raised bog: impact of human activity and climate change. Quaternary Science Reviews 84, 65-85.

Van Bellen, S., Mauquoy, D., Payne, R.J., Roland, T.P., Daley, T.J., Hughes, P.D.M., Loader, N.J., Street-Perrott, F.A., Rice, E.M., Pancotto, V.A., 2014. Testate amoebae as a proxy for reconstructing Holocene water table dynamics in southern Patagonian peat bogs. Journal of Quaternary Science 29, 463-474.

van der Knaap, W., Lamentowicz, M., van Leeuwen, J., Hangartner, S., Leuenberger, M., Mauquoy, D., Goslar, T., Mitchell, E.A., Kamenik, C., 2011. A multi-proxy, high-resolution record of peatland development and its drivers during the last millennium from the subalpine Swiss Alps. Quaternary Science Reviews 30, 3467-3480.

Wooller, M.J., Kurek, J., Gaglioti, B.V., Cwynar, L.C., Bigelow, N., Reuther, J.D., Gelvin-Reymiller, C., Smol, J.P., 2012. An~ 11,200 year paleolimnological perspective for emerging archaeological findings at Quartz Lake, Alaska. J Paleolimnol 48, 83-99.

Palaeoecological data	**Contemporary data**	**Transfer function method**	**Leave-one-out**		**Reconstruction range (cm)**	**Mean boostrap SE (cm)**	**Proportion explained variance**	**Random TF p**
			R^2	**RMSEP (cm)**				
Krya Vrissi-1 (Payne and Pates, 2009)	Greece (Payne and Mitchell, 2007)	WA-PLS(2)	0.75	2.02	11.20	0.70	0.58	**<0.01**
Krya Vrissi-2 (Payne and Pates, 2009)	Greece (Payne and Mitchell, 2007)	WA-PLS(2)	0.75	2.02	10.00	2.00	0.50	**0.01**
Dexameni (Payne and Pates, 2009)	Greece (Payne and Mitchell, 2007)	WA-PLS(2)	0.75	2.02	6.00	2.00	0.35	**0.05**
Jigsaw Lake (Mitchell and Kishaba unpublished)	Alaska (Payne et al., 2006)	WA-PLS(2)	0.55	9.82	28.60	10.60	0.27	0.17
Dot Lake B (Payne and Mitchell, 2009)	Alaska (Payne et al., 2006)	WA-PLS(2)	0.55	9.82	53.40	11.10	0.26	0.10
Sürmene Ağaçbaşi Yaylasi (Payne et al., 2008)	Turkey (Payne et al., 2008)	WA-PLS(2)	0.57	7.22	65.10	12.08	0.43	0.62
Mauntschas (van der Knaap et al., 2011)	Engadine (Lamentowicz et al., 2010; Mitchell et al., 2013)	WA-tol(Inv)	0.70	9.80	48.94	10.80	0.39	**0.02**
Mukzra (Lamentowicz and Obremska, 2010)	Poland (Lamentowicz et al., 2008)	WA-tol(Inv)	0.60	8.90	20.46	9.24	0.16	0.65
Jelenia Wyspa (Lamentowicz et al., 2007)	Poland (Lamentowicz et al., 2008)	WA-tol(Inv)	0.60	8.90	10.96	9.18	0.12	0.50
Tuchola (Lamentowicz et al., 2008)	Poland (Lamentowicz et al., 2008)	WA-tol(Inv)	0.60	8.90	19.53	9.11	0.10	0.80
Stążki (Lamentowicz et al., 2013)	Poland (Lamentowicz et al., 2008)	WA-tol(Inv)	0.60	8.90	54.35	9.14	0.03	0.98
Praz Rodet (Mitchell et al., 2001)	Jura (Mitchell et al., 1999)	WA-PLS(2)	0.61	8.13	50.10	11.38	0.09	0.97
Minden (Booth and Jackson, 2003)	North America (Booth, 2008)	WA-tol(Inv)	0.75	7.79	39.38	7.92	0.24	0.20
Dead Island (Swindles et al., 2010)	Northern Ireland (Swindles et al., 2009)	WA-tol(Inv)	0.83	4.99	38.66	5.91	0.19	0.42
Nordans Pond Bog (Hughes et al., 2006)	Newfoundland (Charman and Warner, 1997)	WA-tol(Inv)	0.65	6.98	16.25	7.68	0.07	0.83[1]
Butterburn Flow A (Hendon et al., 2001)	Northwest Europe (Charman and Blundell, 2007)	WA-PLS(2)	0.71	5.63	27.84	5.81	0.31	0.14
Butterburn Flow B (Hendon et al., 2001)	Northwest Europe (Charman and Blundell, 2007)	WA-PLS(2)	0.71	5.63	12.42	5.87	0.12	0.86
Creusate (Jassey et al.	Jura (Mitchell et al., 1999)	WA-PLS(2)	0.61	8.13	31.75	10.76	0.09	0.93

unpublished)								
Frasne (Jassey et al. unpublished)	Jura (Mitchell et al., 1999)	WA-PLS(2)	0.61	8.13	42.80	9.58	0.41	**0.01**[2]
Tørvesø 1 (Ellershaw, 2004)	Northwest Europe (Charman and Blundell, 2007)	WA-PLS(2)	0.71	5.63	40.11	11.58	0.56	0.09
Gjótarholt (Ellershaw, 2004)	Northwest Europe (Charman and Blundell, 2007)	WA-PLS(2)	0.71	5.63	12.85	7.16	0.19	0.75
Hill of Shurton (Ellershaw, 2004)	Northwest Europe (Charman and Blundell, 2007)	WA-PLS(2)	0.71	5.63	18.77	7.05	0.16	0.91
Andorra (Van Bellen et al. submitted)[3]	Patagonia (Van Bellen et al., 2014)	WA-PLS(2)	0.72	13.49	66.98	14.99	0.13	0.90
Tierra Australis (Van Bellen et al. submitted)[3]	Patagonia (Van Bellen et al., 2014)	WA-PLS(2)	0.72	13.49	71.43	14.86	0.17	0.82
Karukinka (Van Bellen et al. submitted)[3]	Patagonia (Van Bellen et al., 2014)	WA-PLS(2)	0.72	13.49	37.99	15.36	0.24	0.66
Imnati (Payne, 2014)	Eastern Mediterranean (Payne, 2011)	WA-PLS(2)	0.75	9.34	44.42	9.94	0.08	0.80
Didadjara (Mazei et al. unpublished)	Eastern Mediterranean (Payne, 2011)	WA-PLS(2)	0.75	9.34	15.59	10.44	0.09	0.98
Staroselsky Moch (Payne et al., in press)	Russia (Tsygnov et al. submitted)	WA(Inv)	0.73	5.64	29.05	5.89	0.30	0.10
Kluvka Mire (Novenko et al., 2015)	Russia (Tsygnov et al. submitted)	WA(Inv)	0.73	5.64	22.67	6.04	0.32	0.28
Malham Tarn Moss (Turner et al., 2014)	Northern England (Turner et al., 2013)	WA-tol(Inv)	0.70	7.56	30.36	8.42	0.18	0.64

1 Note that using a larger alternative training set and a WMAT model (but not an ML or WA-tol(Inv) model) Amesbury et al. (2013) did find a significant result for this record, suggesting that choice of training set can have a strong influence on significance level.

2 A significant result but with slightly lower significance level (P=0.04) is produced if using the Northwest Europe transfer function (Charman and Blundell, 2007).

3 randomTF results previously presented by van Bellen et al. (submitted), recalculated here.