

This is a repository copy of *Significance testing testate amoeba water table reconstructions*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/100146/>

Version: Accepted Version

Article:

Payne, Richard J., Babeshko, Kirill V., van Bellen, Simon et al. (17 more authors) (2016) Significance testing testate amoeba water table reconstructions. *Quaternary Science Reviews*. pp. 131-135. ISSN 0277-3791

<https://doi.org/10.1016/j.quascirev.2016.01.030>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

1 **Significance testing testate amoeba water table reconstructions**

2 Richard J Payne^{1,2}, Kirill V Babeshko², Simon van Bellen³, Jeffrey J Blackford⁴, Robert K Booth⁵, Dan J
3 Charman⁶, Megan R Ellershaw⁷, Daniel Gilbert⁸, Paul D M Hughes⁹, Vincent EJ Jassey^{8,10}, Łukasz
4 Lamentowicz¹¹, Mariusz Lamentowicz¹¹, Elena A Malysheva², Dmitri Mauquoy¹², Yuri Mazei^{2,13},
5 Edward A D Mitchell^{14,15}, Graeme T Swindles¹⁶, Andrey N Tsyganov², T Edward Turner¹⁶, Richard J
6 Telford¹⁷

7 1 Environment, University of York, Heslington, York, YO105DD, United Kingdom.

8 2 Department of Zoology and Ecology, Penza State University, Krasnaya str. 40, 440026 Penza,
9 Russia.

10 3 GEOTOP-Université du Québec à Montréal, 201 Avenue Président-Kennedy, Montréal, Québec,
11 H2X 3Y7, Canada.

12 4 Department of Geography, Environment and Earth Sciences, University of Hull, Cottingham Road,
13 Hull, HU6 7RX, United Kingdom.

14 5 Department of Earth and Environmental Sciences, Lehigh University, 1 West Packer Avenue,
15 Bethlehem, PA 18015-3001, USA.

16 6 Geography, College of Life and Environmental Sciences, University of Exeter, Amory Building,
17 Rennes Drive, Exeter, EX4 4RJ, United Kingdom.

18 7 Natural England, Northminster House, Peterborough, PE1 1UA, United Kingdom.

19 8 Laboratoire Chrono-Environnement, UMR 6249 Université de Franche-Comté/CNRS, UFR Sciences
20 et Techniques, 16 route de Gray, 25030 Besançon, France.

21 9 Geography and Environment, University of Southampton, University Road, Southampton, SO17
22 1BJ, United Kingdom.

23 10 Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Site Lausanne, Station 2,
24 1015 Lausanne, Switzerland.

25 11 Department of Biogeography and Palaeoecology, Faculty of Geographical and Geological Science,
26 Adam Mickiewicz University, Poznan, ul. Dziegielowa 27, 61-680 Poznan, Poland.

27 12 School of Geosciences, University of Aberdeen, Elphinstone Road, Aberdeen AB24 3UF, United
28 Kingdom.

29 13 Department of Hydrobiology, Lomonosov Moscow State University, Leninskiye gory, 1, Moscow
30 119991, Russia.

31 14 Laboratory of Soil Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel,
32 Switzerland.

33 15 Jardin Botanique de Neuchâtel, Pertuis-du-Sault 58, CH- 2000 Neuchâtel, Switzerland

34 16 School of Geography, University of Leeds, Leeds, LS2 9JT, United Kingdom.

35 17 EECRG, Department of Biology, University of Bergen, Allegaten 41, N-5007 Bergen, Norway.

36 ABSTRACT

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

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

49 INTRODUCTION

50 Testate amoebae are widely-used proxies in palaeoecological studies; in particular for the
51 reconstruction of water table depth in peatlands (Charman, 2001; Mitchell et al., 2008). Over the last
52 25 years palaeoecology has been revolutionised by the use of statistical models (transfer functions)
53 to quantitatively reconstruct environmental variables. However, questions are increasingly being
54 raised about the reliability and robustness of transfer function results (Belyea, 2007; Juggins, 2013).

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

59 Although we cannot realistically assess whether all reconstructions are correct we can conceivably
60 test whether they are potentially useful. Telford and Birks (2011) propose a pragmatic solution: that
61 a reconstruction can be considered statistically significant if it explains more of the variance in the
62 fossil data than those of transfer functions trained on randomly-generated data. Telford and Birks
63 (2011) propose a method, 'randomTF', in which:

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

75 We would expect reliable reconstructions to explain more variance in the fossil data than transfer
76 functions trained on random data, and therefore to give significant results. However, a significant
77 randomTF value is not proof of accuracy and a non-significant result does not necessarily imply
78 inaccuracy. Non-significant results do however give cause for concern and suggest that transfer
79 function output should be treated with caution. randomTF tests can potentially tell us which
80 reconstructions we should trust more, which less, and whether we can predict more than one
81 environmental variable from the same fossil dataset. Telford and Birks (2011) also propose an
82 alternative test ('obs.cor') based on the correlation of optima values with axis species scores from a
83 constrained ordination of the fossil data. This test is not applicable to all transfer functions methods
84 and is not considered here. The randomTF test has been applied in a few studies (Amesbury et al.,
85 2013; Lamarre et al., 2013; Swindles et al., 2015a) but is not yet routinely used in testate amoeba
86 palaeoecology. Here we apply this test to a large number of published and unpublished records with
87 the aim to identify the characteristics which are likely to lead to better reconstructions, giving better
88 randomTF results.

89 METHODS

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

105 RESULTS and DISCUSSION

106 Only five of the 30 tests yielded a significant P-value ($P < 0.05$; Table 1). While we expected that some
107 reconstructions would give non-significant results this proportion is much higher than we
108 anticipated. While a few reconstructions fail to reach $P = 0.05$ by a relatively narrow margin (Tørvesø
109 1, Staroselsky Moch, Dot Lake B), many more have P-values which substantially exceed this value.

110 Another two records published in the literature have given significant P-values: those of Swindles et
111 al. (2015a) for Stordalen, Sweden and Lamarre et al. (2013) for Lac Le Caron, Canada. Amesbury et
112 al. (2013) found a significant result for the Nordans Pond site of Hughes et al. (2006) using an
113 extended transfer function, whereas here we find a non-significant result using the transfer function
114 used in the original study (Charman and Warner, 1997). In the latter three cases multiple model
115 structures were tested with some producing significant reconstructions, and some not. We note that
116 in these instances a correction for multiple comparisons (such as a Bonferroni correction) would

117 probably have meant that the reconstructions did not reach significance. However, even if these
118 results are included, eight significant P-values out of 32 reconstructions remains a surprisingly low
119 proportion.

120 Telford and Birks (2011) identify four factors which might make the randomTF test prone to type II
121 error (“false negative”): low numbers of effective species; small numbers of fossil samples; limited
122 variability in the reconstruction and poorly-performing or poorly-fitting transfer functions. All of
123 these factors apply to some of the reconstructions we examine but it is not clear that any are a
124 consistent cause of non-significant P-values. Overall, P-value was not significantly correlated with
125 properties of the training set (mean/standard deviation/range of WTD) or fossil data (species
126 richness, Hill’s N2 or number of samples), performance metrics of the transfer function (leave one
127 out RMSEP or R²), properties of the reconstruction (mean/standard deviation/range of predications,
128 ratio of range to RMSEP or training set range, mean boot-strapped error estimates) or measures of
129 transfer function fit (proportion of shared taxa, proportion of fossil samples with poor modern
130 analogues, squared residual length)(Spearman Rs; P>0.05). P-value was strongly correlated with the
131 proportion of variance in the fossil data explained by the reconstruction (Spearman Rs=-0.89,
132 P<0.001), suggesting (unsurprisingly) that where a high proportion of variance is explained this is
133 unlikely to be exceeded by transfer functions trained on random data.

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

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

151 Our results provide some evidence that where there is a choice of transfer function this can affect
152 randomTF significance level. With the Nordans Pond record (Hughes et al., 2006) the transfer
153 function used in the original publication yields a non-significant P-value (P=0.83; Table 1) while a
154 more recent transfer function with a larger training set and better performance statistics gives a
155 significant P-value (Amesbury et al., 2013). With the Frasné record (Jassey and Gilbert unpublished)
156 marginally better results are found with the smaller and weaker-performing, but geographically
157 closer, Jura transfer function (Mitchell et al., 1999) than the larger, better-performing, north-west
158 Europe model (Charman and Blundell, 2007) although both are P<0.05. It is also probable that in

159 borderline cases difference in selection of model or samples included in the training set or fossil data
160 might make the difference between a P-value above or below the usual threshold of 0.05 (*cf.*
161 Amesbury et al., 2013; Swindles et al., 2015).

162 So what should we take from these results?

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

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

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

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

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

204 Acknowledgements

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

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

215 References

- 216 Amesbury, M.J., Mallon, G., Charman, D.J., Hughes, P.D.M., Booth, R.K., Daley, T.J., Garneau, M.,
217 2013. Statistical testing of a new testate amoeba-based transfer function for water-table depth
218 reconstruction on ombrotrophic peatlands in north-eastern Canada and Maine, United States.
219 *Journal of Quaternary Science* 28, 27-39.
- 220 Belyea, L., 2007. Revealing the Emperor's new clothes: niche-based environmental reconstruction in
221 the light of recent ecological theory. *The Holocene* 17, 683-688.
- 222 Birks, H.J.B., 1995. Quantitative palaeoenvironmental reconstructions, in: Maddy, D., Brew, J.S.
223 (Eds.), *Statistical modelling of Quaternary science data*. Quaternary Research Association,
224 Cambridge, pp. 161-254.
- 225 Booth, R.K., 2008. Testate amoebae as proxies for mean annual water-table depth in Sphagnum-
226 dominated peatlands of North America. *Journal of Quaternary Science* 23, 43-57.
- 227 Booth, R.K., Jackson, S.T., 2003. A high-resolution record of late-Holocene moisture variability from a
228 Michigan raised bog, USA. *The Holocene* 13, 863-876.
- 229 Charman, D.J., 2001. Biostratigraphic and palaeoenvironmental applications of testate amoebae.
230 *Quaternary Science Reviews* 20, 1753-1764.
- 231 Charman, D.J., Blundell, A., 2007. A new European testate amoebae transfer function for
232 palaeohydrological reconstruction on ombrotrophic peatlands. *Journal of Quaternary Science* 22,
233 209-221.
- 234 Charman, D.J., Warner, B.G., 1997. ecology of testate amoebae (Protozoa: Rhizopoda) in oceanic
235 peatlands in Newfoundland, Canada: modelling hydrological relationships for palaeoenvironmental
236 reconstruction. *Ecoscience*.
- 237 Cwynar, L., Rees, A.H., Pedersen, C., Engels, S., 2012. Depth distribution of chironomids and an
238 evaluation of site-specific and regional lake-depth inference models: a good model gone bad? *J*
239 *Paleolimnol* 48, 517-533.
- 240 Ellershaw, M., 2004. *Holocene climate change in the North Atlantic region: evidence from peat*
241 *deposits*. University of London, London.
- 242 Hendon, D., Charman, D., Kent, M., 2001. Palaeohydrological records derived from testate amoebae
243 analysis from peatlands in northern England: within-site variability, between-site comparability and
244 palaeoclimatic implications. *The Holocene* 11, 127-148.

245 Hughes, P., Blundell, A., Charman, D., Bartlett, S., Daniell, J., Wojatschke, A., Chambers, F.M., 2006.
246 An 8500cal. year multi-proxy climate record from a bog in eastern Newfoundland: contributions of
247 meltwater discharge and solar forcing. *Quaternary Science Reviews* 25, 1208-1227.
248 Juggins, S., 2009. rioja: analysis of Quaternary science data. R package version 0.5-6. URL
249 <http://cran.r-project.org/package=rioja>.
250 Juggins, S., 2013. Quantitative reconstructions in palaeolimnology: new paradigm or sick science?
251 *Quaternary Science Reviews* 64, 20-32.
252 Lamarre, A., Magnan, G., Garneau, M., Boucher, É., 2013. A testate amoeba-based transfer function
253 for paleohydrological reconstruction from boreal and subarctic peatlands in northeastern Canada.
254 *Quaternary International* 306, 88-96.
255 Lamentowicz, M., Gałka, M., Milecka, K., Tobolski, K., Lamentowicz, Ł., FIAŁKIEWICZ-KOZIEŁ, B.,
256 Blaauw, M., 2013. A 1300-year multi-proxy, high-resolution record from a rich fen in northern
257 Poland: reconstructing hydrology, land use and climate change. *Journal of Quaternary Science* 28,
258 582-594.
259 Lamentowicz, M., Obremska, M., 2010. A rapid response of testate amoebae and vegetation to
260 inundation of a kettle hole mire. *J Paleolimnol* 43, 499-511.
261 Lamentowicz, M., Obremska, M., Mitchell, E.A., 2008. Autogenic succession, land-use change, and
262 climatic influences on the Holocene development of a kettle-hole mire in Northern Poland. *Review*
263 *of Palaeobotany and Palynology* 151, 21-40.
264 Lamentowicz, M., Tobolski, K., Mitchell, E.A., 2007. Palaeoecological evidence for anthropogenic
265 acidification of a kettle-hole peatland in northern Poland. *The Holocene* 17, 1185-1196.
266 Lamentowicz, M., Van Der Knaap, W., Lamentowicz, Ł., Van Leeuwen, J.F., Mitchell, E.A., Goslar, T.,
267 Kamenik, C., 2010. A near-annual palaeohydrological study based on testate amoebae from a sub-
268 alpine mire: surface wetness and the role of climate during the instrumental period. *Journal of*
269 *Quaternary Science* 25, 190-202.
270 Luoto, T.P., Kaukolehto, M., Weckström, J., Korhola, A., Väliiranta, M., 2014. New evidence of warm
271 early-Holocene summers in subarctic Finland based on an enhanced regional chironomid-based
272 temperature calibration model. *Quaternary Research* 81, 50-62.
273 Marcisz, K., Fournier, B., Gilbert, D., Lamentowicz, M., Mitchell, E.A., 2014a. Response of Sphagnum
274 peatland testate amoebae to a 1-year transplantation experiment along an artificial hydrological
275 gradient. *Microbial ecology* 67, 810-818.
276 Marcisz, K., Lamentowicz, Ł., Słowińska, S., Słowiński, M., Muszak, W., Lamentowicz, M., 2014b.
277 Seasonal changes in Sphagnum peatland testate amoeba communities along a hydrological gradient.
278 *European journal of protistology* 50, 445-455.
279 Mitchell, E.A., Buttler, A.J., Warner, B.G., Gobat, J.-M., 1999. Ecology of testate amoebae (Protozoa:
280 Rhizopoda) in Sphagnum peatlands in the Jura mountains, Switzerland and France. *Ecoscience*, 565-
281 576.
282 Mitchell, E.A., van Der Knaap, W., van Leeuwen, J.F., Buttler, A., Warner, B.G., Gobat, J.-M., 2001.
283 The palaeoecological history of the Praz-Rodet bog (Swiss Jura) based on pollen, plant macrofossils
284 and testate amoebae (Protozoa). *The Holocene* 11, 65-80.
285 Mitchell, E.A.D., Charman, D.J., Warner, B.G., 2008. Testate amoebae analysis in ecological and
286 paleoecological studies of wetlands: past, present and future. *Biodiversity and Conservation* 17,
287 2115-2137.
288 Mitchell, E.A.D., Payne, R.J., van der Knaap, W.O., Lamentowicz, Ł., Gałka, M., Lamentowicz, M.,
289 2013. The performance of single-and multi-proxy transfer functions (testate amoebae, bryophytes,
290 vascular plants) for reconstructing mire surface wetness and pH. *Quaternary Research* 79, 6-13.
291 Mulot, M., Villard, A., Mitchell, E.A., 2014. Adaptation of *Sphagnum* peatland testate amoebae
292 communities to an artificial water table depth gradient over a 19 month period, in: Marcisz, K. (Ed.),
293 7th International Symposium on Testate Amoebae. Adam Mickiewicz University, Poznan, Poland, p.
294 17.

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

299 Payne, R., Telford, R., Blackford, J., Blundell, A., Booth, R., Charman, D., Lamentowicz, Ł.,
300 Lamentowicz, M., Mitchell, E., Potts, G., Swindles, G., Warner, B., W, W., 2012. Testing peatland
301 testate amoeba transfer functions: appropriate methods for clustered training-sets. *Holocene* 22,
302 819-825.

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

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

307 Payne, R.J., Charman, D.J., Matthews, S., Eastwood, W.J., 2008. Testate amoebae as
308 palaeohydrological proxies in sürmene ağaçbaşı yaylasi peatland (northeast Turkey). *Wetlands* 28,
309 311-323.

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

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

316 Payne, R.J., Mitchell, E.A., 2007. Ecology of testate amoebae from mires in the Central Rhodope
317 Mountains, Greece and development of a transfer function for palaeohydrological reconstruction.
318 *Protist* 158, 159-171.

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

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

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

326 Salonen, J.S., Helmens, K.F., Seppä, H., Birks, H.J.B., 2013. Pollen-based palaeoclimate
327 reconstructions over long glacial–interglacial timescales: methodological tests based on the
328 Holocene and MIS 5d–c deposits at Sokli, northern Finland. *Journal of Quaternary Science* 28, 271-
329 282.

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

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

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

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

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

345 Telford, R., 2011. palaeoSig: significance tests of quantitative palaeoenvironmental reconstructions.
346 R package version 1.
347 Telford, R.J., Birks, H.J.B., 2011. A novel method for assessing the statistical significance of
348 quantitative reconstructions inferred from biotic assemblages. *Quaternary Science Reviews* 30,
349 1272-1278.
350 Turner, T.E., Swindles, G.T., Charman, D.J., Blundell, A., 2013. Comparing regional and supra-regional
351 transfer functions for palaeohydrological reconstruction from Holocene peatlands. *Palaeogeography,*
352 *Palaeoclimatology, Palaeoecology* 369, 395-408.
353 Turner, T.E., Swindles, G.T., Roucoux, K.H., 2014. Late Holocene ecohydrological and carbon
354 dynamics of a UK raised bog: impact of human activity and climate change. *Quaternary Science*
355 *Reviews* 84, 65-85.
356 Van Bellen, S., Mauquoy, D., Payne, R.J., Roland, T.P., Daley, T.J., Hughes, P.D.M., Loader, N.J.,
357 Street-Perrott, F.A., Rice, E.M., Pancotto, V.A., 2014. Testate amoebae as a proxy for reconstructing
358 Holocene water table dynamics in southern Patagonian peat bogs. *Journal of Quaternary Science* 29,
359 463-474.
360 van der Knaap, W., Lamentowicz, M., van Leeuwen, J., Hangartner, S., Leuenberger, M., Mauquoy,
361 D., Goslar, T., Mitchell, E.A., Kamenik, C., 2011. A multi-proxy, high-resolution record of peatland
362 development and its drivers during the last millennium from the subalpine Swiss Alps. *Quaternary*
363 *Science Reviews* 30, 3467-3480.
364 Wooller, M.J., Kurek, J., Gaglioti, B.V., Cwynar, L.C., Bigelow, N., Reuther, J.D., Gelvin-Reymiller, C.,
365 Smol, J.P., 2012. An ~ 11,200 year paleolimnological perspective for emerging archaeological findings
366 at Quartz Lake, Alaska. *J Paleolimnol* 48, 83-99.

367

368

369

370

371

372

373

374

375

376

Palaeoecological data	Contemporary data	Transfer function method	Leave-one-out		Reconstruction range (cm)	Mean bootstrap 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 Ağaçaşası 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
Stażki (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
Gjótárholt (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
Klůvka 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.