Article:

Payne, Richard J., Lamentowicz, Mariusz and Mitchell, Edward A. D. (2011) The perils of taxonomic inconsistency in quantitative palaeoecology: experiments with testate amoeba data. Boreas. pp. 15-27. ISSN: 0300-9483

https://doi.org/10.1111/j.1502-3885.2010.00174.x

The perils of taxonomic inconsistency in quantitative palaeoecology: experiments with testate amoeba data

RICHARD J. PAYNE, MARIUSZ LAMENTOWICZ AND EDWARD A.D. MITCHELL

Payne, R.J., Lamentowicz, M. & Mitchell, E.A.D.: The perils of taxonomic inconsistency in quantitative palaeoecology: experiments with testate amoeba data

A fundamental requirement of quantitative palaeoecology is consistent taxonomy between a modern training set and palaeoecological data. In this study we assess the possible consequences of violation of this requirement by simulating taxonomic errors in testate amoeba data. Combinations of easily-confused taxa were selected and data manipulated to reflect confusion of these taxa, transfer functions based on unmodified data were then applied to these modified data sets. Initially these experiments were carried out one error at a time using four modern training sets, subsequently multiple errors were separately simulated in both four modern training sets and four palaeoecological datasets. Some plausible taxonomic confusions caused major biases in reconstructed values. In the case of two palaeoecological datasets a single consistent taxonomic error was capable of changing the pattern of environmental reconstruction beyond all recognition, totally removing any real palaeoenvironmental signal. The issue of taxonomic consistency is one which many researchers would rather ignore; our results show that the consequences of this may ultimately be severe.

Keywords: Testate amoebae; Palaeoecology; Transfer Functions; Peatlands; Palaeohydrology; Palaeoclimatology

Richard J. Payne (e-mail: r.payne@mmu.ac.uk), Department of Environmental & Geographical Sciences, Manchester Metropolitan University, Chester Street, Manchester, M1 5GD, United Kingdom, and Geography, School of Environment and Development, The University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom; Mariusz Lamentowicz, Department of Biogeography and

Palaeoecology, Adam Mickiewicz University, Dzięgielowa 27, 61-680 Poznań, Poland; Edward Mitchell, WSL, Swiss Federal Institute for Forest, Snow and Landscape Research, Ecosystem Boundaries Research Unit, Wetlands Research Group, Station 2, CH-1015 Lausanne, Switzerland, École Polytechnique Fédérale de Lausanne (EPFL), Laboratory of Ecological Systems, Station 2, CH - 1015 Lausanne, Switzerland, and Laboratory of Soil Biology, University of Neuchâtel, Rue Emile Argand 11, CH-2009 Neuchâtel, Switzerland

Quantitative palaeoecology generally proceeds by modelling the relationship between species and an environmental variable in modern environments and then applying this model to palaeoenvironmental data to produce quantitative estimates of environmental changes through time. Among the basic requirements of this 'transfer function' approach is that 'the fossil data-sets used for reconstruction purposes should be of comparable taxonomy and nomenclature... as the modern training set' (Birks 1995) i.e. that individuals of the same species are identified consistently and called the same name in both the modern and palaeoecological data (Belyea 2007). However, there are good reasons to suppose that this assumption is sometimes violated; human error is inevitable and in some microfossil groups there is considerable uncertainty regarding the underlying taxonomy. Such a microfossil group is the testate amoebae, a group of protists which are abundant in many aquatic to terrestrial ecosystems and whose solid shells ('tests') may be preserved long after death (Fig. 1), allowing community changes to be tracked through time. Testate amoebae are increasingly used in palaeoecology, in particular as proxies for hydrological change, and therefore palaeoeclimate, in peatlands (Charman 2001; Mitchell *et al*. 2008).

The taxonomy of testate amoebae is not straightforward. Difficulties start with the problem of applying a biological species concept to micro-organisms which, as far as we know, overwhelmingly reproduce asexually and for which there are little genetic data (Schlegel & Meisterfeld 2003). Testate amoeba taxonomy is built around the concept of morphospecies, that consistent morphological forms represent valid taxonomic units, at least in the absence of any superior approach (Finlay *et al.* 1996; Finlay 1998). However there are no biometric data for many morphospecies, leaving considerable room for personal interpretation of what degree of difference justifies the erection of new morphospecies and what can simply be considered intraspecific variability (Medioli *et al.* 1987; Odgen & Meisterfeld 1989). Delineation of species is further complicated by considerable morphological variability in tests (Heal 1963; Wanner 1999; Bobrov & Mazei 2004). Testate amoebae can show marked phenotypic plasticity (Lüftnegger *et al.* 1988; Wanner & Meisterfeld 1994; Wanner 1999) and in some taxa (adaptive) polymorphism (Schönborn 1992). The test morphology of taxa which build their shells from particles in their environment (xenosomes) depends on the available material; large particles may obscure the underlying test morphology (Ogden 1983). It is probable that many described taxa may just represent extreme forms of this morphological variability. A difference in taxonomies between 'lumpers' and 'splitters' is highly apparent in the literature. For instance the *Centropyxis constricta* of Medioli & Scott (1983) would probably include 20 or more species and subspecies considered separable by Chardez (1967).

Issues with the differentiation of morphospecies are common to other micro-organisms (e.g. Mann & Droop 1996; Pawlowski *et al.* 2002). However in the case of testate amoebae these issues are particularly acute due to the inadequacies of the taxonomic literature. Unlike for instance freshwater diatom analysis, where the floras of Krammer & Lange-Bertalot (1986, 1988, 1991a, b) are widely used (at least as a baseline), there is no 'standard text' for testate amoeba taxonomy. The obscurity of testate amoebae to many biologists, combined with the general decline in morphological taxonomic research over recent decades (Lee 2000; Wheeler 2004) have contributed to the poor state of testate amoeba taxonomy. Those attempting to apply testate amoeba analysis in ecology and palaeoecology are forced to use a fragmented body of literature, much of which dates back to the early part of the last century, and much of which is mutually-contradictory. There are no clear rules for separating many taxa and few taxonomic keys are available (none of which are comprehensive and few of which are in English, the *de facto* language of modern science).

In environmental studies using testate amoebae these problems are particularly serious because of the large number of tests which must be counted; typically at least 100 individuals per sample and 40-50 samples (Payne & Mitchell 2009). This number of tests pragmatically requires that all identification and counting be carried out using light microscopy under normal (200x to 400x) magnifications. Many fine taxonomic distinctions rest on very subtle features which are simply not practicable under these conditions (e.g. in *Euglypha*: Wylezich *et al.* 2002, *Cyphoderia*: Todorov *et al.* 2009; Heger *et al.* in press, and *Difflugia*: Ogden 1983). In palaeoecology problems are compounded by the loss of diagnostic features. The division between taxa with lobose and filose pseudopodia is the most fundamental in testate amoebae taxonomy but is not applicable in palaeoecology. Diagnostic features of the test such as spines may be lost through taphonomic processes or in sample preparation and tests may become compressed (Charman *et al.* 2000). Taxonomic schemes used in palaeoecology are therefore a compromise between practical simplicity and loss of palaeoenvironmental discernment (Charman *et al.* 2000). Given all these problems it would be little surprise if there were considerable taxonomic differences among researchers. In the absence of a formal inter-comparison exercise it is impossible to know to what extent different researchers apply the same name to different taxa or different names to the same taxon. We can however make observations that: i) The taxonomic literature lacks clarity. ii) There are considerable differences in the taxonomic resolution adopted by different studies. iii) Inter-comparison exercises for other microfossil groups used in Quaternary palaeoecology have shown considerable variability between different analysts and research groups (Munro *et al.* 1990; Pederson & Moseholm 1993; Kelly *et al.* 2002; Prygiel *et al.* 2002). For instance, in the diatom inter-comparison exercise of Kelly *et al.* (2002) some taxa were identified correctly less than 20% of the time. iv) When researchers are learning testate amoeba taxonomy several mistakes are consistently made.

On the basis of these observations we feel it would be naïve to assume that taxonomies are identical among all researchers. In this study we attempt to gain an understanding of the possible implications of taxonomic variability for environmental reconstructing by simulating possible errors in previously established modern and palaeoecological datasets.

Methods

Four modern training sets and four palaeoecological datasets were used in our experiments. The four modern training sets are all derived from *Sphagnum*-dominated, mostly ombrotrophic mires and span a considerable region from North America to western Asia (Table 1). They are: i) Poland, from peatlands of Poland (Lamentowicz *et al.* 2005, 2007, 2008); ii) Jura, from peatlands in the Jura Mountains of France and Switzerland (Mitchell *et al.* 1999, 2001); iii) Turkey, from the Sürmene Ağaçbaşı Yaylası peatland in north-eastern Turkey (Payne *et al.* 2008); and iv) Alaska, from peatlands in south-central Alaska (Payne *et al.* 2006). The final selected transfer function models were used in our experiments to infer depth to water table (DWT; Table 1). The four palaeoecological datasets are: 1. 'Site DLB', a peatland in sub-Arctic Alaska (Payne *et al.* unpublished, but see Payne & Mitchell 2009); 2. Praz-Rodet, a peatland in Switzerland (Mitchell *et al.* 2001); 3. Tuchola, a peatland in Poland (Lamentowicz *et al.* 2008), and 4. Jelenia Wyspa, another peatland in Poland (Lamentowicz *et al.* 2007). All of these palaeoecological datasets have an applicable transfer function from the same area (i.e. the Alaska, Jura and Poland training sets,

Table 1) which was produced by the same analysts. We are as confident as possible that these palaeoecological datasets and their respective transfer functions have consistent taxonomic schemes.

A first step in our experiments was to select pairs of species which we considered could be confused (Table 2). Our combinations were based on three sources of evidence: i) Our assessment of the distinctiveness of the taxon based upon the literature, in particular where taxa have been considered inseparable by some authors. ii) Our observations of the mistakes made by undergraduate and postgraduate students in learning testate amoeba taxonomy. iii) Our own experience of learning testate amoeba taxonomy. We produced separate lists of taxon combinations for each of our training sets, reflecting the differing communities encountered in those studies and the slightly different taxonomic schemes adopted by the analysts. For simplicity we refer to each of these taxon combinations as an 'error combination', however with some of these pairings we note that the distinction between the taxa may not always be clear. We would not claim that our taxon combinations reflect all possible errors or that all of these errors have a high probability. However, we do feel that our taxon combinations include all of the most common confusions. Three sets of experiments were conducted:

Individual errors

The first group of experiments used only the modern training sets and was designed to quantitatively investigate the impacts of individual errors on transfer function predictions. We identified three possible ways in which each pair of species could be confused: 1. All of taxon A could be recorded as taxon B. 2. All of taxon B could be recorded as taxon A. 3. The taxa could be switched. The training set data were then transformed to reflect each of these three types of error for each of the taxon pairs identified. So for instance with the Alaska data we identified 15 taxon pairs (Table 2), which could each be transformed in three different ways giving a total of 45 possible individual modifications to the data. We then applied the transfer function derived from the original, unmodified training set to each of these modified data-sets in turn to predict depth to water table (DWT). This approach of applying a transfer function based on a training set to the same training set but with simulated taxonomic errors is not representative of any real-world situation but is a useful tool to investigate the impact that these errors might have on transfer function results.

Inferred depth to water table values (termed 'testate amoeba-inferred depth to water table': TI-DWT) were compared to predictions based on the unmodified data set and residuals calculated ($\text{TI-DWT}_{\text{original}}$ - $\text{TI-DWT}_{\text{modified}}$). Differences between predictions based on the original and modified data were calculated in terms of root mean square error (RMSE), R^2 and the maximum difference between predictions for any one sample (Maximum Bias). All transfer function analyses were carried out using C^2 (Juggins 2003).

Multiple errors

To investigate the cumulative impact of more than one error we also carried out experiments simulating multiple errors in our modern training sets. The same taxon combinations were used as in the individual errors experiments. A random numbers system was used to select a taxon pair, with each pair assigned an equal probability of selection. Where more than two taxa could be confused with each other only one taxon pair could be selected at a time (where more than one pair were selected the data were kept unchanged). Each taxon pair could be transformed in one of the three ways described above with each of these three modifications given an equal probability of being selected. The number of errors in the data was steadily increased up to the maximum number of possible changes, with fifteen repetitions for each error total. The transfer function based on the unmodified training set was then applied to this modified training set and RMSE, R^2 and Maximum Bias calculated as above.

A related possible source of bias in inferred values is that taxonomic errors in a training set lead to selection of a different transfer function model structure which may, in itself, lead to differences in model output. To investigate the potential implications of this issue alternative model structures (WA, WA-Tol, WA-PLS, ML) were tested using the maximum number of simulated errors in each training set and 15 replicates. The best performing model was selected based on $\text{RMSEP}_{\text{jack}}$ with no penalty for model complexity.

Errors in palaeoecological sequences

To see how the simulated errors might affect palaeoenvironmental inference we also manipulated the four palaeoecological data-sets and then applied transfer functions based on unmodified training sets. The same taxon combinations were used when simulating errors in the palaeoecological data-sets as were used in the two experiments simulating errors in training sets described above. The number of errors was successively increased from one to ten. Transfer functions based on the unmodified training set data were applied and TI-DWT values calculated for each modified palaeoecological data-set.

Results

Individual errors

Results of individual error experiments are shown in Table 2. With all training sets a few error combinations have a great deal more impact on predictions than most others. With the Poland data much the most significant error combination is *Difflugia globulosa/Cyclopyxis arcelloides*, introducing a mean error of up to 2.5 cm (7% of the total measured DWT range) depending on which of the three permutations is considered, the next most important error combination is *Arcella vulgaris/Arcella discoides* (RMSE≤0.55 cm, 1.5% measured range). With the Jura data the two most important error combinations are *Cyclopyxis arcelloides/Phryganella acropodia*, leading to a mean error of up to 1.95 cm (4% measured range) and *Centropyxis aerophila/Centropyxis platystoma*, leading to a mean error of up to 1.1 cm (2% measured range). With the Turkey data the most important error combination is *Corythion dubium/Trinema lineare*, leading to a mean error of up to 1.7 cm (2% measured range). With the Alaska data the most important error combinations are *Euglypha ciliata/Euglypha strigosa* (RMSE≤3.06 cm, 5% measured range), *Nebela tincta/Nebela penardiana* (RMSE≤2.78 cm, 4.6% measured range) and *Heleopera petricola/Heleopera sphagni* (RMSE≤2.13 cm, 3.5% measured range). Maximum bias data show that many of these single errors lead to the predicted TI-DWT values of some samples changing by more than 10 cm, and in some cases more than 20 cm. These are highly significant changes; 20 cm represents the DWT difference between a lawn and a low hummock.

Multiple errors

When multiple errors are simulated there is a steady increase in the deviation of predictions from those based on the unmodified data (Fig. 2). With the Alaska data there is an approximately equal division between samples with TI-DWT over- and under-predicted relative to the original data. However with the other three data-sets there is a trend in one direction; with the Poland data this is towards under-prediction of TI-DWT while with the Jura and Turkey data this is towards over-prediction of TI-DWT. This directional bias is most apparent with the Jura data with the TI-DWT values of the majority of samples being over-predicted relative to the unmodified data. These directional biases are largely driven by just a few errors, so with the Jura data the trend is mostly due to the *N. tincta/N. parvula* combination, with the Poland data the trend is mostly due to the *C. arcelloides/D. globulosa* combination and with the Turkey data the trend is mostly attributable to the *C. dubium/T. lineare* and *H. petricola /H. rosea* combinations.

If alternative transfer function model structures are tested using the training sets with simulated errors a different model structure is selected with 93% of replicates with the Jura data, 60% of replicates with the Poland data, 40% of replicates with the Turkey data and in no replicates with the Alaska data.

Errors in palaeoecological sequences

The consequences of these errors for palaeoecological reconstruction are shown in Fig. 3A-D. With the Site DLB data (Fig. 3A) the most distinct features of the reconstruction based on unmodified data are pronounced wet phases at the base of the profile, from 52-56 cm and from 25-28 cm. These wet phases generally remain apparent even when taxonomic errors are introduced, although with increasing number of errors the phases become less distinct in some experiments. A notable change with even one error is a period of higher values between 11 and 15 cm due to counting *Centropyxis ecornis* as *Centropyxis laevigata*. With the Praz Rodet data (Fig. 3B) simulated errors make relatively little difference to reconstructed values. The maximum deviation is 7.6 cm but in none of these experiments is the TI-DWT reconstruction different enough to change interpretation of the record. With the Tuchola data (Fig. 3C) even a single error can drastically change the pattern of the reconstruction: If *Cyclopyxis arcelloides* is recorded as *Difflugia globulosa* it fundamentally changes the reconstruction giving an overall reduction in predicted values, introducing a period of rapidly fluctuating values between 20 and 120 cm depth and adding a trough at 360 cm. Interpretation of these data with and without this error would be utterly different. Increasing error load slightly increases the variability of predictions, but the overall pattern is largely determined by whether or not *C. arcelloides* and *D. globulosa* are confused.

With the Jelenia Wyspa data (Fig. 3D) the difference that even a single error can make is even more marked. Again the most important error is recording *C. arcelloides* as *D. globulosa*. This error leads to a general under-prediction of TI-DWT by 5 cm or more and an almost total difference in the pattern of change. Introducing this error leads to the reconstruction of major TI-DWT peaks at 42, 95 and 110 cm, features which are totally absent in the reconstruction based on unmodified data. One of the most distinctive features of the TI-DWT reconstruction based on the unmodified data is a period of high values between 50 and 65 cm. However in several experiments with one or more errors this feature is less distinct or not apparent at all. In these experiments *Centropyxis cassis* has been recorded as either *Centropyxis platystoma* or *Centropyxis aerophila*. With increasing number of errors there is an increasing variability in the pattern of reconstructed change, although reconstructions group around two basic patterns determined by whether *C. arcelloides*/*D. globulosa* are confused or not. In some experiments where both *C. arcelloides*/*D. globulosa,* and *C. cassis* and *C. aerophila* or *C. platystoma* are confused TI-DWT values deviate from the unmodified data by more than 17 cm.

Discussion

All of our experiments make several important assumptions: they assume that mistakes are made consistently, that these are all possible errors and all have an equal probability, and they do not account for tests simply over-looked or mistaken for taxa not included in the transfer function and therefore excluded. While we acknowledge that our experiments represent a considerable simplification of the real way in which taxonomic errors may affect transfer function output the results are undeniably revealing. While many possible errors make very little difference to predicted values some possible errors can change predicted values drastically, giving reconstructions which bear little apparent resemblance to those based on full data.

The specific errors which produce major impacts in our experiments seem by no means improbable. For instance the confusion of *C. dubium* with *T. lineare* (important in the Turkey training set) and *E. ciliata* with *E. strigosa* (important in the Alaska training set) are both common mistakes among our students. The most dramatic illustration of the possible impacts of taxonomic errors in our experiments is provided by the experiments simulating errors in palaeoecological data sets from Tuchola and Jelenia Wyspa. Major differences in reconstructions are produced by confusing *D. globulosa* and *C. arcelloides*, two taxa that have a similar overall morphology and would probably be grouped by Charman *et al.* (2000) or Medioli & Scott (1983). The drastic impact that this error makes is particularly notable given the relative scarcity of these taxa in the Tuchola data, constituting only 2.7% of total tests and only exceeding 5% of count in 5 samples. In the Jelenia Wyspa data the taxa are slightly more abundant, constituting 10.1% of total tests. The difference that this single change makes to the reconstructions highlights the extent to which the pattern of palaeoenvironmental reconstruction may be determined by just a few important taxa. It is worryingly easy to envisage a scenario where somebody, perhaps relatively new to testate amoebae palaeoecology and using one of the more agglomerative taxonomies as their main guide, could make such an error to produce an environmental reconstruction which is substantially biased, or in the worst case entirely an artefact of taxonomic inconsistency. Taxonomic errors in a training set may change the transfer function model structure selected, but it is likely that this change alone would have limited impact on model output (cf. Booth 2007).

The large impacts of some of the simulated errors may suggest the need to group these potentially problematic taxa in our transfer functions. However these taxa frequently have significantly differing hydrological optima, therefore a corollary of the impacts of these errors is that if these taxa are grouped considerable ecological information will be lost. In the worst case grouping may considerably bias reconstructions. If one of a pair of taxa is well represented in a training set and the other not, the ecological optima of the group will mostly match that of the first taxon, however if the second taxon is more abundant in palaeoecological samples then reconstructed values will be biased.

In the absence of any formal taxonomic inter-comparison it is not possible to make any definitive assessment of how much of a problem taxonomic inconsistency may be *in praxis*. We would suggest that these errors are far from implausible. However, whether or not these specific taxonomic errors are very likely, our results suggest a wider point, that it is possible for taxonomic errors to radically distort environmental reconstructions. Taxonomic errors will not necessarily make any significant difference to environmental reconstruction; indeed, most errors will probably make very little difference. However, there is the potential for a single taxonomic mistake made consistently to so change an environmental reconstruction that the real palaeoecological signal is totally masked. Although our experiments only consider water table reconstruction in peatlands it is likely that similar results would be found when considering reconstruction of other variables and in other environments. Problems may be particularly acute in minerotrophic peatlands where there may be a greater abundance of 'difficult' taxa (e.g. genera *Difflugia* and *Centropyxis*).

Taxonomic comparability is critical; what a palynomorph used in palaeoecology is called matters little as long as the name is used *consistently*. For instance, non-pollen palynomorphs are commonly referred to as simply a numbered 'type' as the origin of the palynomorph may not be known (van Geel 2001). Given the taxonomic limitations imposed by palaeoecological counting some authors have considered it necessary to use a parallel naming system, for instance Joosten & de Klerk (2002) have suggested the differentiation of fossil pollen from plant species (and indeed modern pollen) by referring to the former in SMALL CAPITALS. While we do not feel that such a system is necessarily required for testate amoebae we would appeal for clarity in the description of taxonomies used in palaeoecological studies of testate amoebae. Until a revised taxonomic framework with clear identification criteria and keys is available and consistently used, researchers publishing training sets should clearly state identification criteria and the taxa included in groupings where these are not obvious.

Extreme caution should be used when applying transfer functions, particularly when using training sets counted by different analysts. Researchers attempting to use a transfer function derived by other analysts should work in close cooperation to ensure the same identification criteria are consistently employed. In our experience this is best done by close communication during counting, rather than trying to post-hoc adjust the taxonomy of a palaeoecological data-set to fit the taxonomy of a transfer function. Comparison of photographs of difficult taxa between analysts is a useful approach to ensure this consistency. Where there is any doubt at all over the criteria for differentiating taxa these taxa should be grouped or excluded from the data-sets. The fact that extremely large reconstruction errors can be introduced by relatively modest taxonomic errors adds to the case for comparing testate amoeba-based records with other data in a multi-proxy approach, and ideally replicating records with multiple cores. All palaeoecological techniques are imperfect, testate amoeba analysis is no exception.

There appears to be a tendency in testate amoeba-based palaeoecological reconstruction to use boot-strapping to derive estimates of standard errors and consider any changes which exceed these error bars (or even do not: Hendon & Charman 2004) to be a palaeoecological 'signal'. However, these standard errors only provide an estimate of the error inherent in the model, additional errors may well be introduced if the transfer function does not provide an adequate fit to the palaeoecological data (cf. Wilmshurst *et al.* 2003) or taxonomic errors are made. In our experiments even quite minor taxonomic errors produced a bias that significantly exceeded the boot-strapped standard errors. Boot-strapped standard errors should be used with caution as other sources of error can produce biases which considerably exceed these estimates.

To ensure taxonomic consistency there is a need for a common standard taxonomy which can be applied uniformly among analysts given the constraints imposed by counting large numbers of sub-fossil tests using optical microscopy. The guide of Charman *et al.* (2000) is the best attempt at this and is widely used (79 citations in 'Google Scholar' at the time of writing). However, the taxonomic scheme set out has not met with uniform acceptance with many authors either not adopting this scheme or adapting it to varying extents. Major reasons for this lack of consistent use may include the exclusion of some relatively common peatland taxa (e.g. *Euglypha cristata, Tracheleuglypha dentata*) and the broad 'types' adopted for some groups of taxa (perhaps most notably the '*Cyclopyxis arcelloides* type'). The guide of Charman *et al.* (2000) provides a first attempt at a difficult task and is a very useful contribution. However we would argue that now, ten years after publication, is the time for a reconsideration and refinement of the scheme in an attempt to achieve a broad consensus. A consistent taxonomy is essential given increasing attempts to compare and combine modern data-sets while the more widespread use of testate amoebae in palaeoecology means that more environmental reconstructions are being produced using transfer functions derived by other researchers. Taxonomic inconsistency is a neglected issue in biological sciences, but its consequences may ultimately be very severe (Bortolus 2008).

Conclusions

- Errors of taxonomy and enumeration are inevitable in palaeoecology. Testate amoeba analysis is likely to be particularly susceptible to such errors due to the inadequacies of the taxonomy.
- Our experiments suggest that some likely confusions can produce significant biases in quantitative environmental reconstructions.
- These results call for improvement of the taxonomic baseline. For now,

extreme caution should be used when applying transfer functions and especially interpreting small changes.

- There are many possible causes of bias in environmental reconstructions. Taxononomic inconsistency is but one of these.

Acknowledgements- RJP was supported by fellowships from the University of Manchester and British Institute at Ankara. Funding to EM by Swiss NSF projects no. 205321-109709/1, 205321-109709/2 and CCES projects RECORD and BigLink is kindly acknowledged. ML was supported by the Foundation for Polish Science (FNP) (Outgoing Fellowship KOLUMB). We thank Pim van der Knaap and John Birks for constructive comments on a previous version of the manuscript.

‘*Truth is mighty and will prevail. There is nothing the matter with this, except that it ain't so.*’ (Mark Twain)

References

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

Birks, H.J.B. 1995: Quantitative palaeoecological reconstructions. *In* Maddy, D. & Brew, S. (eds.): Statistical modelling of Quaternary science data, 161-54. Quaternary Research Association, Cambridge.

Bobrov, A.A. & Mazei, Y. 2004: Morphological variability of testate amoebae (Rhizopoda: Testacealobosea) in natural populations. *Acta Protozoologica 43*, 133–146.

Booth, R.K. 2007: Testate amoebae as proxies for mean annual water-table depth in Sphagnum-dominated peatlands of North America. *Journal of Quaternary Science 23*, 43-57.

Bortolus, A. 2008: Error cascades in the biological sciences: The unwanted consequences of using bad taxonomy in ecology. *Ambio 37*, 114-118.

Chardez, D. 1967: Histoire naturelle des Protozoaires Thecamoebiens. *Les Naturalistes Belges 48*, 484-586.

Charman, D., Hendon, D. & Woodland, W. 2000: *The identification of testate amoebae (Protozoa: Rhizopoda) in peats*. 147pp. Quaternary Research Association Technical Guide Series, Cambridge.

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

Finlay, B.J. 1998: The global diversity of protozoa and other small species. *International Journal of Parasitology 28***,** 29-48.

Finlay, B.J., Corliss, J.O., Estban, G. & Fenchel, T. 1996: Biodiversity at the microbial level: the number of free-living ciliates in the biosphere. *Quarterly Review of Biology 71***,** 221-237.

Heal, O.W. 1963: Morphological variation in certain Testacea (Protozoa: Rhizopoda). *Archiv für Protistenkunde 106*, 351-368.

Heger, T.J., Mitchell, E.A.D., Todorov, M., Golemansky, V., Lara, E. & Pawlowski, J. In press: Phylogenies of euglyphid testate amoebae (Cercozoa: Euglyphida) reveal rare transitions between the marine supralittoral and the freshwater or terrestrial environments. *Molecular Phylogenetics and Evolution.*

Hendon, D. & Charman, D.J. 2004: High-resolution peatland water-table changes for the past 200 years: the influence of climate and implications for management. *The Holocene 14*, 125-134.

Joosten, H. & de Klerk, P. 2002: What's in a name? Some thoughts on pollen classification, identification and nomenclature in Quaternary palynology. *Review of Paleobotany and Palynology 122*, 29-45.

Juggins, S. 2003: C2 user guide. *Software for ecological and palaeoecological data analysis and visualisation*. University of Newcastle, Newcastle Upon Tyne.

Kelly, M.G., Bayer, M.M., Hürlimann, J. & Telford, R.J. 2002: Human error and quality assurance in diatom analysis. In: du Buf, H. & Bayer, M.M. (eds.): Automatic diatom identification, 75-92. World Scientific Publishing, Singapore.

Krammer, K. & Lange-Bertalot, H. 1986: *Bacillariophyceae. i: Teil Naviculaceae.* 230pp. Gustav Fischer Verlag, Stuttgart.

Krammer, K. & Lange-Bertalot, H. 1988: *Bacillariophyceae. ii Teil: Bacillariaceae, Epithemiaceae, Surirellaceae.* 596pp. Gustav Fischer Verlag, Stuttgart.

Krammer, K. & Lange-Bertalot, H. 1991a: *Bacillariophyceae. iii Teil: Centrales, Fragilariaceae, Eunotiaceae.* 576pp. Gustav Fischer Verlag, Stuttgart.

Krammer, K. & Lange-Bertalot, H. 1991b: *Bacillariophyceae. iv Teil: Achnanthaceae, kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema Gesamtliteraturverzeichnis.* 437pp. Gustav Fischer Verlag, Stuttgart.

Lamentowicz, M. & Mitchell, E.A.D. 2005: The ecology of testate amoebae (Protists) in Sphagnum in north-west Poland in relation to peatland ecology. *Microbial Ecology 50*, 48-63.

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

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

Lee, M.S.Y. 2000: A worrying systematic decline. *Trends in Ecology and Evolution 15*, 346.

Lüftnegger, G., Petz, W., Berger, H., Foissner, W. & Adam, H. 1988: Morphologic and biometric characterization of twenty-four soil testate amoebae (Protozoa, Rhizopoda). *Archiv für Protistenkunde 136*, 153-189.

Mann, D.G. & Droop, S.J.M. 1996: Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia 336*, 19-32.

Medioli, F.S. & Scott, D.B. 1983: Holocene Arcellacea (Thecamoebians) from Eastern Canada. 63pp. Cushman Foundation for Foraminiferal Research, Special Publication 21, Washington DC.

Medioli, F.S., Scott, D.B. & Abbott, B.H. 1987: A case study of protozoan intraclonal variability: Taxonomic implications. *Journal of Foraminiferal Research 17*, 28-47.

Mitchell, E.A.D., 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 6*, 565-576.

Mitchell, E.A.D., van der Knaap, W.O., van Leeuwen, J.F.N., 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.

Munro, M.A.R., Kreiser, A.M., Battarbee, R.W., Juggins, S., Stevenson, A.C., Anderson, D.S., Anderson, N.J., Berge, F., Birks, H.J.B., Davis, R.B., Flower, R.J., Fritz, S.C., Haworth, E.J., Jones, V.J., Kingston, J.C. & Renberg, I. 1990: Diatom quality control and data handling. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 327*, 257- 261.

Ogden, C.G. & Meisterfeld, R. 1989: The taxonomy and systematics of some species of Cucurbitella, Difflugia and Netzelia (Protozoa: Rhizopoda); with an evaluation of diagnostic characters. *European Journal of Protistology 27*, 109-128.

Ogden, C.G. 1983: Observations on the systematics of the genus Difflugia in Britain (Rhizopoda, Protozoa). *Bulletin of the British Museum of Natural History (Zoology) 44*, 1-73.

Pawlowski, J., Fahrni, J.F., Brykczynska, U., Habura, A. & Bowser, S.S. 2002: Molecular data reveal high taxonomic diversity of allogromiid foraminifera in Explorers Cove (McMurdo Sound, Antarctica). *Polar Biology 25*, 96-105.

Payne, R., Kishaba, K., Blackford, J. & Mitchell, E. 2006: The ecology of testate amoebae in southcentral Alaskan peatlands: Building transfer function models for palaeoenvironmental inference. *The Holocene 16*, 403-414.

Payne, R., Charman, D., Matthews, S. & Eastwood, W. 2008: Testate amoebae as palaeoclimate proxies in Sürmene Ağaçbaşi Yaylasi peatland (Northeast Turkey). *Wetlands 28*, 311-323.

Payne R.J. & Mitchell, E.A.D. 2009: How many is enough? Determining adequate count totals for ecological and palaeoecological studies of testate amoebae. *Journal of Paleolimnology 42*, 483-495.

Pedersen, B. & Moseholm, L. 1993: Precision of the daily pollen count. Identifying sources of variation using variance component models. *Aerobiologia 9*, 15-26.

Prygiel, J., Carpentier, P., Almeida, S., Coste, M., Druart, J.-C., Ector, L., Guillard, D., Honore, M-A., Iserentant, R., Ledeganck, P., Lalanne-Cassou, C., Lesniak, C., Mercier, I., Moncaut, P., Nazart, M., Nouchet, N., Peres, F., Peeters, V., Rimet, F., Rumeau, A., Sabater, S., Straub, F., Torrisi, M., Tudesque, L., Van de Vijver, B., Vidal, H., Vizinet, J. & Zydek, N. 2002: Determination of the biological diatom index (IBD NF T 90–354): Results of an intercomparison exercise. *Journal of Applied Phycology 14*, 27–39.

Schönborn, W. 1992: Adaptive polymorphism in soil-inhabiting testate amebas (Rhizopoda) - its importance for delimitation and evolution of asexual species. *Archiv für Protistenkunde 142*, 139-155.

Schlegel, M. & Meisterfeld, R. 2003: The species problem in protozoa revisited. *European Journal of Prostistology 39*, 349-355.

Todorov, M., Golemanski, V., Mitchell, E.A.D. & Heger, T.J. 2009: Morphology, biometry and taxonomy of five freshwater and marine interstitial testate amoebae of the genus Cyphoderia (Testaceafilosia: Cyphoderiidae). *Journal of Eukaryotic Microbiology 56*, 279-289.

Van Geel, B. 2001: Non-pollen palynomorphs, *In* Smol, J.P., Birks, H.J.B. & Last, W.M. (eds.): *Tracking Environmental Change Using Lake Sediments. Volume 3: Terrestrial, Algal, and Siliceous Indicators*, 99-119. Kluwer Academic Publishers, Dordrecht.

Wanner, M. 1999: A review on the variability of testate amoebae: Methodological approaches, environmental influences and taxonomical implications. *Acta Protozoologica 38*, 15–29.

Wanner, M. & Meisterfeld, R. 1994: Effects of some environmental factors on the shell morphology of testate amoebae (Rhizopoda, Protozoa). *European Journal of Protistology 30*, 191–195.

Wheeler, Q.D. 2004: Taxonomic triage and the poverty of phylogeny. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 359*, 571-583.

Wilmshurst, J., Wiser, S. & Charman, D. 2003: Reconstructing Holocene water tables in New Zealand using testate amoebae: differential preservation of tests and implications for the use of transfer functions. *The Holocene 13*, 61-72.

Wylezich, C., Meisterfeld, R., Meisterfeld, S. & Schlegel, M. 2002: Phylogenetic analyses of small subunit ribosomal RNA coding regions reveal a monophyletic lineage of euglyphid testate amoebae (order Euglyphida). *Journal of Eukaryotic Microbiology 49*, 108-118.

FIGURES and TABLES

Figure 1. Illustrations of selected testate amoeba taxa discussed in this paper. A. *Nebela tincta* var. *major*. B. *N. tincta* var. *major* and *N. tincta*. C. *N. marginata*. D. *N. carinata*. E. *N. tincta* var. *major*. F. *N. flabellulum*. G. *N. penardiana*. H. *N. militaris*. I. *Centropyxis aerophila*. J. *C. aerophila* var. *sphagnicola*. K. *C. ecornis*. L. *C. laevigata*. M. *C. platystoma*. N. *Phryganella acropodia*. O. *Difflugia globulosa*. P. *Corythion dubium*. Q & R. *Trinema lineare*. S. *Euglypha ciliata*. T. *E. compressa*. U. *E. strigosa*. Scale bar is 20µm for P,Q and R, 50µm for others.

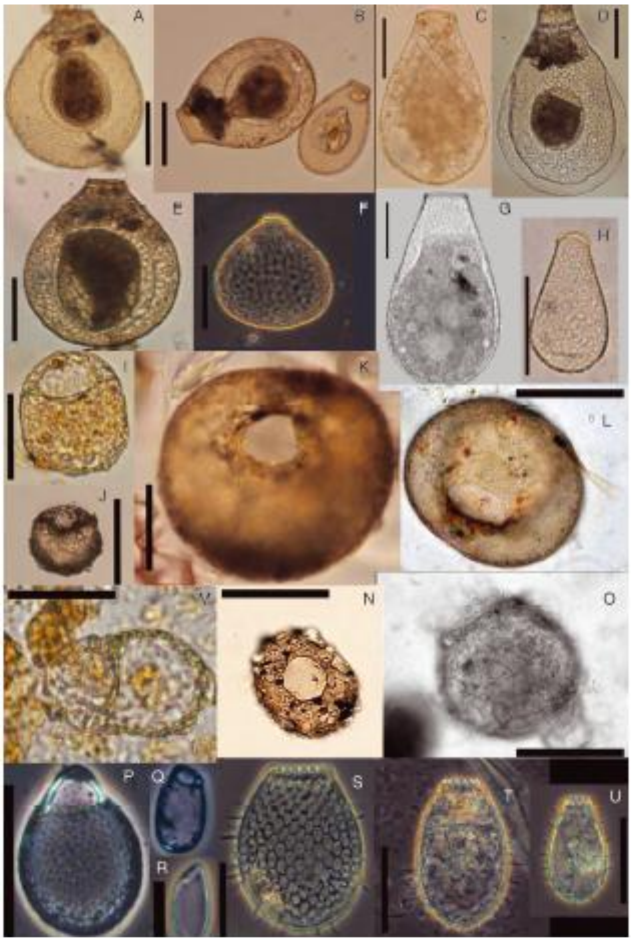


Figure 2. Results of multiple error experiments (see Methods) with four modern training sets. Plots A-D show residuals (TI-DWT$_{\text{original}}$ - TI-DWT$_{\text{modified}}$), plots E-H show the same data presented as an overall mean TI-DWT deviation. Box plots show median (central line), first and third quartiles (grey box), tenth and ninetieth percentiles ('whiskers') and fifth and ninety-fifth percentiles (dots).

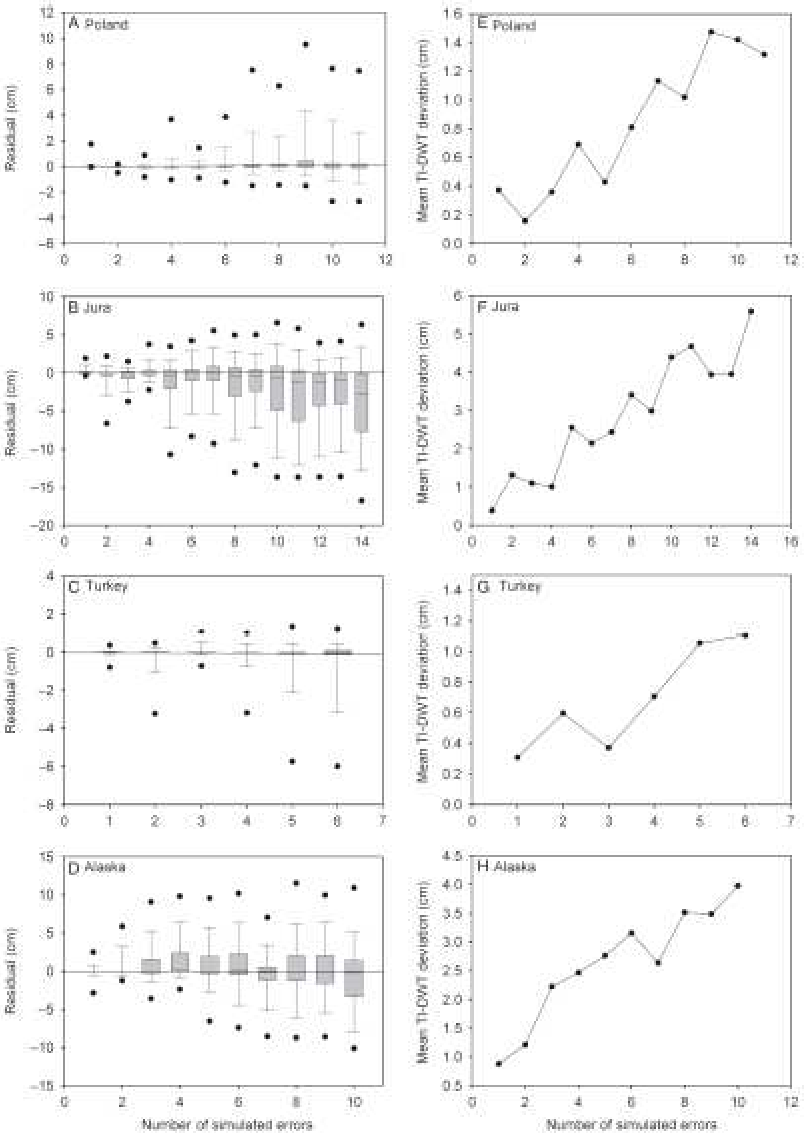
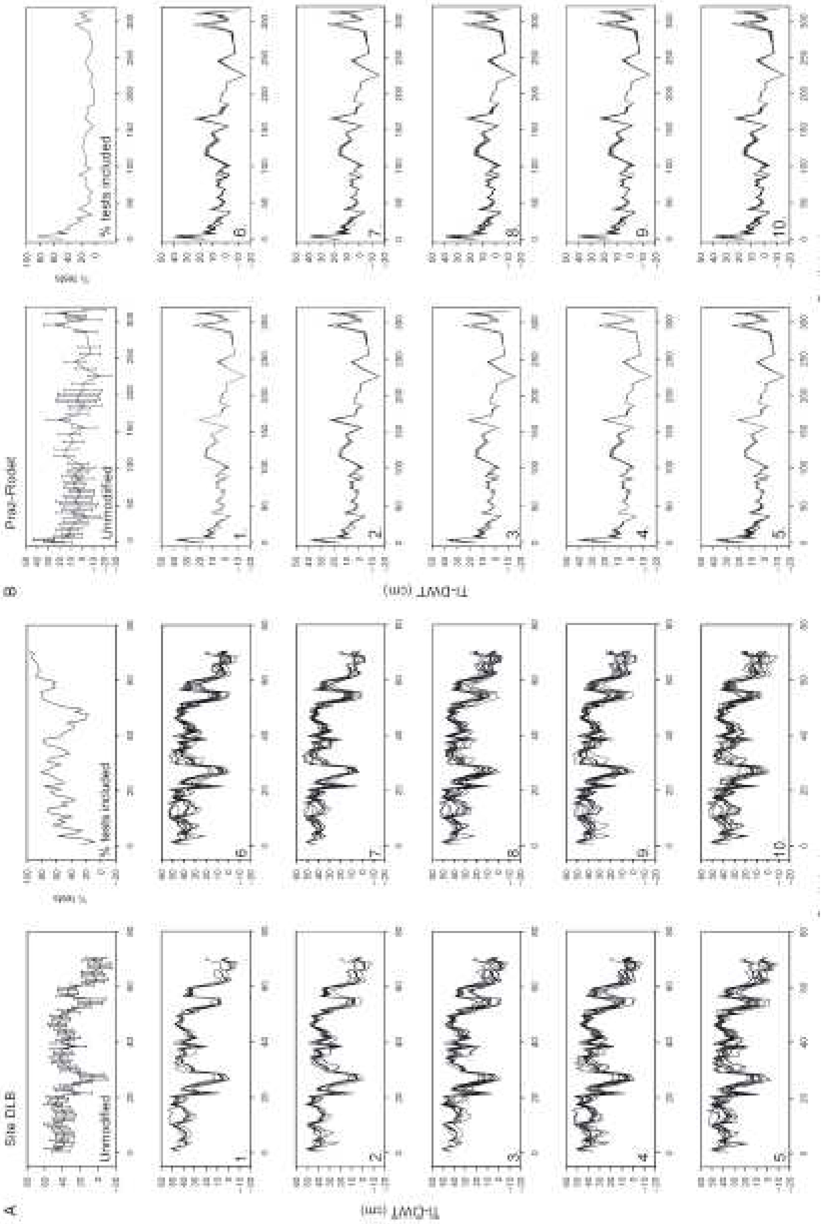


Figure 3. Results of errors in palaeoecological sequences experiments (see Methods) with palaeoecological data from A) 'Site DLB', Alaska, B) Praz-Rodet, Swiss Jura, C) Tuchola, Poland, and D) Jelenia Wyspa, Poland. For each dataset the plot on the upper left shows reconstruction based on unmodified data and the adjacent plot shows percentage of tests contributed by the taxa which could be confused. Other plots show reconstructions for increasing number of errors from 1-10 with fifteen cycles of random re-selection for each error total.


A
Site DLB
Unmodified
% tests included
% tests
TI-DWT (cm)
B
Praz-Rodet
Unmodified
% tests included
% tests
TI-DWT (cm)
1
2
3
4
5
6
7
8
9
10

174

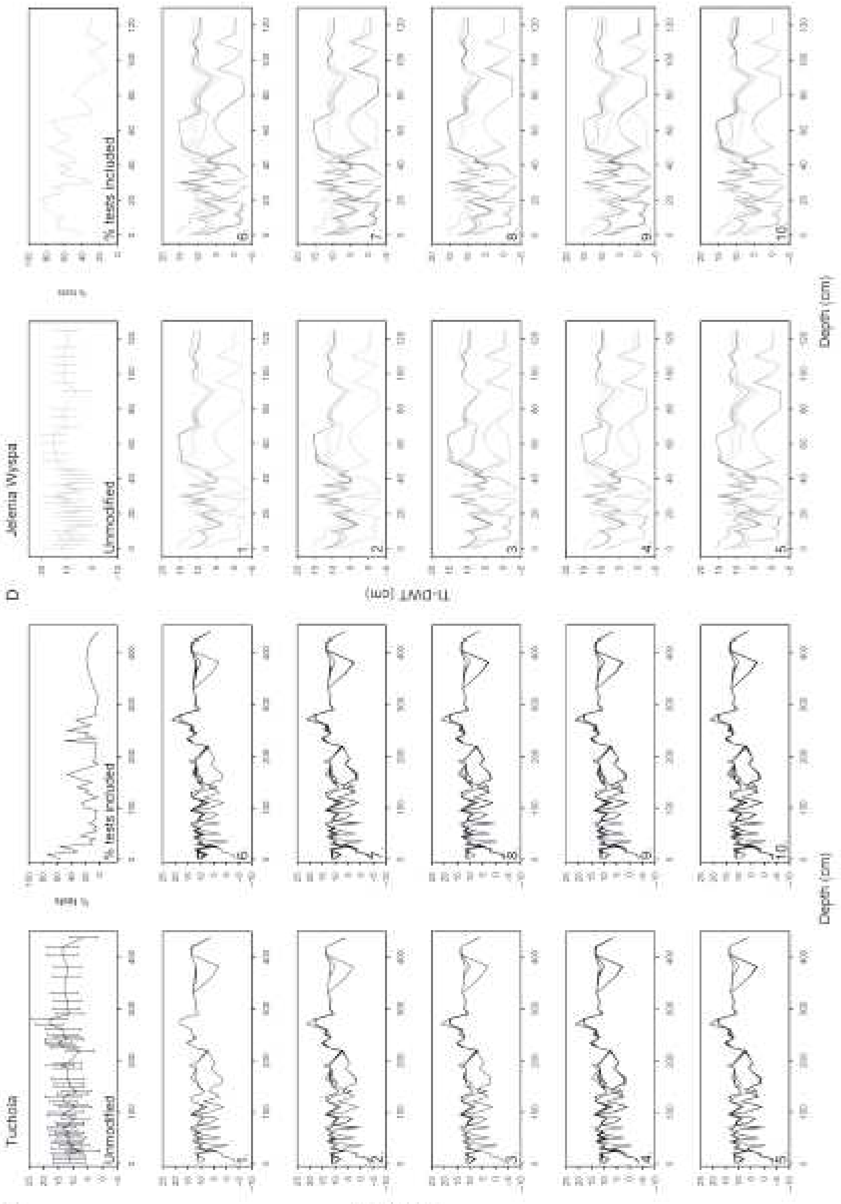


Table 1.

Attributes of the datasets used in this study showing number of samples (n), and for modern training sets: transfer function model structure, jack-knifed root mean square error of prediction (RMSEP), Maximum Bias and R^2. Location given in parentheses after palaeoecological data set name indicates applicable transfer function.

Location	n	Model structure	$RMSEP_{jack}$ (cm)	Max $Bias_{jack}$ (cm)	R^2_{jack}	Reference
Modern training sets:						
Poland	84	WA-Tol, Inverse deshrinking	4.6	9.0	0.71	Lamentowicz *et al.* (2007)*
Jura	37	WA-PLS (2 component)	8.0	21	0.62	Mitchell *et al.* (1999, 2001)†
Turkey	42	ML	7.1	21	0.81	Payne *et al.* (2008)

Alaska	91	WA-PLS (2 component)	9.7	14	0.55	Payne *et al.* (2006)
Palaeoecological data sets:						
Site DLB (Alaska)	71					Payne *et al.* (unpublished)
Praz Rodet (Jura)	57					Mitchell *et al.* (2001)
Tuchola (Poland)	50					Lamentowicz *et al.* (2008)
Jelenia Wyspa (Poland)	38					Lamentowicz *et al.* (2007)

*Values slightly different from published due to re-calculation of percentages. †Re-calculated using WA-PLS, see Payne and Mitchell (2009).

Table 2. Results of individual error experiments (Methods section 1) for A) Poland (Lamentowicz *et al.* 2007), B) Jura (Mitchell *et al.* 1999, 2001), C) Turkey (Payne *et al.* 2008), D) Alaska (Payne *et al.* 2006). Showing, taxon pair (A and B), percentage of total tests these taxa represent, number of occurrences of each taxon (N), DWT optima estimated by weighted averaging ('WA Optima') and impact of simulated errors in terms of RMSE, maximum bias and R^2 between TI-DWT based on original and modified datasets. Each taxon pair could be changed in three ways: all of taxon A could be counted as taxon B (A→B), all of taxon B could be counted as taxon A (A←B), and the two taxa could be switched (A↔B).

A) Poland

Taxon A	Taxon B	% total		N		WA optima		RMSE		
		A	B	A	B	A	B	A→B	A←B	A↔B
Corythion dubium	*Corythion-Trinema* type	0.80	0.03	13	4	23.08	20.90	0.03	0.00	0.03
Cyclopyxis arcelloides	*Difflugia globulosa*	3.63	1.74	33	6	4.36	-0.18	2.33	0.28	2.49
Nebela parvula	*Nebela tincta*	1.37	2.40	32	33	19.04	21.59	0.04	0.08	0.08
Nebela bohemica	*Nebela collaris*	2.49	0.12	24	6	11.60	19.72	0.19	0.02	0.20
Nebela militaris	*Nebela collaris*	1.21	0.12	15	6	25.11	19.72	0.12	0.01	0.11
Heleopera sphagni	*Heleopera petricola*	0.42	1.56	15	31	13.29	13.02	0.00	0.01	0.01
Heleopera sylvatica	*Heleopera petricola*	0.16	1.56	5	31	20.10	13.02	0.01	0.05	0.06
Euglypha strigosa	*Euglypha compressa*	0.25	0.43	10	11	19.75	6.92	0.11	0.06	0.17
Euglypha compressa	*Euglypha ciliata*	0.43	0.41	11	8	6.92	6.51	0.02	0.02	0.02
Euglypha ciliata	*Euglypha strigosa*	0.41	0.25	8	10	6.51	19.75	0.40	0.05	0.07
Centropyxis cassis	*Centropyxis aerophila*	0.27	0.07	5	3	13.98	7.41	0.03	0.03	0.07
Centropyxis aerophila	*Centropyxis platystoma*	0.07	0.03	3	2	7.41	8.68	0.00	0.00	0.00
Centropyxis cassis	*Centropyxis platystoma*	0.27	0.03	5	2	13.98	8.68	0.05	0.01	0.05
Amphitrema stenostoma	*Amphitrema wrightianum*	0.11	0.65	5	5	0.08	0.06	0.01	0.06	0.06
Arcella artocrea	*Arcella catinus*	0.03	3.05	4	35	11.64	15.08	0.00	0.15	0.15
Arcella discoides	*Arcella vulgaris*	7.58	2.20	33	17	1.36	3.15	0.43	0.16	0.55
Arcella gibbosa	*Arcella hemispherica*	0.59	0.59	6	5	0.77	-0.23	0.02	0.02	0.05

B) Jura

Taxon A	Taxon B	% total		N		WA optima		RMSE		
		A	B	A	B	A	B	A→B	A←B	A↔B
Arcella artocrea	*Arcella catinus*	0.10	1.64	7	19	13.16	26.33	0.06	0.88	0.92
Centropyxis aerophila	*Centropyxis platystoma*	2.10	0.95	17	8	17.17	23.31	1.10	0.50	1.07
Corythion dubium	*Trinema* type	5.31	3.70	33	20	24.97	26.38	0.49	0.34	0.36
Cyclopyxis arcelloides	*Difflugia globulosa*	0.55	0.24	7	1	11.12	3.00	0.02	0.01	0.03
Cyclopyxis arcelloides	*Phryganella acropodia*	0.55	2.99	7	28	11.12	28.25	0.32	1.76	1.95
Difflugia longicollis	*Difflugia oblonga*	0.37	0.02	3	1	27.35	16.00	0.26	0.01	0.27
Euglypha alveolata	*Euglypha tuberculata*	0.01	0.01	1	1	41.00	8.00	0.02	0.01	0.03
Euglypha ciliata	*Euglypha compressa*	2.08	0.29	31	8	21.66	26.25	0.72	0.10	0.69
Euglypha ciliata	*Euglypha strigosa*	2.08	1.04	31	19	21.66	25.78	0.30	0.15	0.27
Euglypha laevis	*Euglypha rounda*	1.66	2.62	22	24	24.24	24.75	0.27	0.42	0.47
Euglypha strigosa	*Euglypha compressa*	1.04	0.29	19	8	25.78	26.25	0.21	0.06	0.22
Heleopera petricola	*Heleopera rosea*	2.47	2.82	27	22	26.90	26.04	0.29	0.33	0.52
Nebela bohemica	*Nebela collaris*	0.72	0.23	6	5	20.68	23.20	0.13	0.04	0.09
Nebela carinata	*Nebela marginata*	0.18	0.91	5	9	8.82	9.59	0.01	0.05	0.05
Nebela militaris	*Nebela collaris*	6.62	0.23	30	5	27.85	23.20	0.81	0.03	0.83
Nebela parvula	*Nebela tincta*	0.04	14.68	2	37	29.35	29.29	0.01	5.87	5.86
Nebela penardiana	*Nebela tubulosa*	0.42	0.69	8	8	19.12	16.41	0.12	0.20	0.23
Phryganella acropodia	*Difflugia globulosa*	2.99	0.24	28	1	28.25	3.00	1.88	0.15	2.00
Sphenoderia lenta	*Tracheleuglypha dentata*	0.13	0.81	5	13	17.01	23.01	0.04	0.25	0.21

C) Turkey

Species A	Species B	% total		N		WA optima		RMSEP		
		A	B	A	B	A	B	A→B	A←B	A↔B
Phryganella acropodia	*Cyclopyxis arcelloides*	1.04	0.27	22	3	39.74	9.34	0.03	0.00	0.03
Cyclopyxis eurystoma	*Phryganella acropodia*	0.84	1.04	8	22	68.28	39.74	0.22	0.35	0.37
Cyclopyxis arcelloides	*Cyclopyxis eurystoma*	0.27	0.84	3	8	9.34	68.28	0.55	0.17	0.72
Corythion dubium	*Trinema lineare*	8.24	1.41	31	13	47.40	63.76	1.65	0.35	1.59
Euglypha compressa	*Euglypha ciliata*	0.12	0.49	5	15	25.39	48.87	0.01	0.13	0.12
Euglypha strigosa	*Euglypha compressa*	0.07	0.12	4	5	30.29	25.39	0.01	0.01	0.01
Euglypha strigosa	*Euglypha ciliata*	0.07	0.49	4	15	30.29	48.87	0.00	0.03	0.03
Heleopera rosea	*Heleopera petricola*	3.45	0.08	27	2	41.03	28.59	0.90	0.01	0.90
Nebela penardiana	*Nebela tubulosa*	0.03	0.03	2	2	29.63	29.46	0.00	0.00	0.00
Nebela tincta	*Nebela penardiana*	0.47	0.03	14	2	43.69	29.63	0.01	0.00	0.01
Centropyxis aerophila type	*Plagiopyxis* cf. *callida*	2.33	0.06	20	2	57.28	12.62	0.38	0.01	0.38

D) Alaska

Taxon A	Taxon B	% total		N		WA optima		RMSE		
		A	B	A	B	A	B	A→B	A←B	A↔B
Arcella arenaria	*Arcella artocrea*	2.02	0.10	58	4	35.79	30.92	0.44	0.02	0.46
Centropyxis ecornis	*Centropyxis laevigata*	0.76	1.26	19	20	28.35	44.19	0.48	0.80	1.28
Centropyxis aerophila	*Centropyxis platystoma*	3.05	0.12	38	5	26.43	28.06	0.95	0.04	0.93
Corythion dubium	*Trinema* spp.	4.81	0.96	48	33	31.44	29.41	1.32	0.26	1.10
Difflugia globulosa	*Phryganella acropodia* type	0.15	6.89	3	85	19.59	34.72	0.01	0.29	0.29
Euglypha ciliata	*Euglypha compressa*	4.95	0.83	67	28	35.76	37.60	0.78	0.13	0.75
Euglypha ciliata	*Euglypha strigosa*	4.95	0.23	67	11	35.76	23.47	3.06	0.14	2.97
Euglypha strigosa	*Euglypha compressa*	0.23	0.83	11	28	23.47	37.60	0.18	0.64	0.82
Heleopera petricola	*Heleopera sylvatica*	3.84	0.31	43	12	32.45	33.42	0.57	0.05	0.58
Heleopera petricola	*Heleopera sphagni*	3.84	3.74	43	33	32.45	24.39	1.17	1.14	2.13
Nebela penardiana	*Nebela marginata*	0.06	0.33	3	6	18.27	18.35	0.02	0.10	0.09
Nebela tincta	*Nebela penardiana*	3.25	0.06	60	3	42.25	18.27	2.74	0.05	2.78
Hyalosphenia elegans	*Nebela militaris*	3.98	1.76	47	40	32.03	46.80	2.59	1.15	2.71
Euglypha rotunda	*Tracheleuglypha dentata*	1.15	0.03	32	3	31.69	14.52	0.74	0.02	0.73
Tracheleuglypha dentata	*Sphenoderia lenta*	0.03	0.35	3	12	14.52	20.68	0.00	0.04	0.04