Testate amoeba transfer function performance along localised hydrological gradients

Andrey N. Tsyganov[a,*], Olga A. Mityaeva[a], Yuri A. Mazei[a,c], Richard J. Payne[a,b]

[a]Department of Zoology and Ecology, Penza State University, Krasnaya str. 40, 440026 Penza, Russia
[b]Environment Department, University of York, Heslington, York YO10 5DD, United Kingdom
[c]Department of Hydrobiology, Lomonosov Moscow State University, Leninskiye gory, 1, Moscow 119991, Russia

Abstract

Testate amoeba transfer functions are widely used for reconstruction of palaeo-hydrological regime in peatlands. However, the limitations of this approach have become apparent with increasing attention to validation and assessing sources of uncertainty. This paper investigates effects of peatland type and sampling depth on the performance of a transfer function using an independent test-set from four *Sphagnum*-dominated sites in European Russia (Penza Region). We focus on transfer function performance along localised hydrological gradients, which is a useful analogue for predictive ability through time. The performance of the transfer function with the independent test-set was generally weaker than for the leave-one-out or bootstrap cross-validations. However, the transfer function was robust for the reconstruction of relative changes in water-table depth, provided the presence of good modern analogues and overlap in water-table depth ranges. When applied to subsurface samples, the performance of the transfer function was reduced due to selective decomposition, the presence of deep-dwelling taxa or vertical transfer of shells. Our results stress the importance of thorough testing of transfer functions, and highlight the role of taphonomic processes in determining results. Further studies of stratification, taxonomy and taphonomy of testate amoebae will be needed to improve the robustness of transfer function output.

Introduction

One of the most widespread applications of testate amoebae is as proxies in palaeoenvironmental studies. The sub-fossil remains of testate amoebae are widely used to indicate sea level in coastal sediments (Gehrels et al. 2006) and water-table depth in peatlands (Mitchell et al. 2008a). Most of these studies have used a quantitative approach involving the development of a statistical model – a transfer function – based on contemporary data, and then the application of this model to palaeoecological data to infer changes in the variable of interest over time. In peatlands more than thirty transfer functions have been developed for prediction of water-table depth (WTD) in locations ranging from Amazonia (Swindles et al. 2014) to China (Qin et al. 2013) (although the distribution is biased towards the mid-latitudes). The development of transfer functions has had a huge impact on environmental reconstruction and the availability of these methods has played an important role in the popularisation of testate amoebae as a proxy over the last twenty years. However, in recent years the limitations of the transfer function

approach have become increasingly apparent across palaeoecology with increasing attention to validation and assessing sources of uncertainty in the results (Juggins 2013). The uncertainties in peatland testate amoeba-based environmental reconstructions arise from both fundamental issues with confounding variables, and from more practical issues such as differential decomposition (Mitchell et al. 2008b; Payne et al. 2012).

An essential first step in evaluating the results of quantitative palaeoecological studies is establishing the performance of the transfer function with contemporary data. Cross-validation methods provide an estimate of performance by successively applying models based on a large sub-set of the training set to predict environmental values for a smaller sub-set of the training set. Three techniques have been widely applied in testate amoeba studies:

(1) In *leave-one-out* (LOO) cross-validation (also called jack-knifing) a transfer function based on all but one of the samples is applied to the remaining sample; predictions for that sample are compared to the measured value and the process is repeated for all samples (Birks 1995; Frey and Deevey 1998).

(2) In *boot-strap* cross-validation a dataset of the same size as the full training set is randomly selected but with the possibility for each sample to be selected more than once; the remaining unselected samples form a test-set (Birks 1995; Frey and Deevey 1998). Because the number of independent samples in the training set is smaller, boot-strapping tends to produce more conservative performance estimates than LOO.

(3) Both boot-strapping and LOO are affected by the spatial structure of the training set. In development of peatland testate amoeba transfer functions it has been usual to extract samples from several bogs but with several samples per bog. Even nearby bogs can differ considerably in terms of factors which are important to testate amoeba assemblages (e.g. nutrient status, vegetation and the history of fire and drainage) and ordinations of testate amoeba data frequently show within-site clustering. One consequence of this clustering is that LOO and boot-strapping give overly optimistic performance estimates because samples from the same site remain in the training set. To account for this issue (Payne et al. 2012) proposed an additional approach – *leave one site out* cross-validation (LOSO) – in which each site is successively excluded from the training set.

Cross-validation has been widely-used to establish the performance of transfer functions because it makes efficient use of the data collected; all the samples analysed are used in developing the model. This is sensible because counting testate amoebae is time-consuming and a training set with more samples is likely to better characterise the optima of species, to include more rare species and to incorporate better analogues for assemblages in the past. However, in reality, there are factors which may affect transfer function performance which are impossible to account for with cross-validation. For instance, it is reasonably common for transfer functions to be applied outside of the immediate region of their development (e.g. Borgmark and Schoning (2006), Willis et al. (2015)), to data produced by different analysts *cf*. Mitchell et al. (2014), Payne et al. (2011), samples prepared using somewhat different methods *cf*. Avel and Pensa (2013) and to phases in the history of a site where other environmental variables may be important e.g. Payne (2011). Recent studies have demonstrated how cross-validation statistics may be skewed by previously unconsidered statistical factors (Payne et al. 2012; Telford and Birks 2005). It is clear that although cross-validation is a pragmatic option to establish the performance of a transfer function, a better indication of true model performance will require the model to be applied to a truly *independent* test-set of samples not used in model development.

Only a minority of testate amoeba transfer functions have been tested with independent data. The largest attempt has been by Payne et al. (2012) who used 18 independent datasets to assess the performance of five published transfer functions. Independent tests have also been conducted in the process of developing some transfer functions e.g. (Payne 2011; Swindles et al. 2009; Swindles et al. 2015a; Van Bellen et al. 2014). The majority (although not all) of these tests have suggested weaker performance than implied by cross-validation. Most recently (Swindles et al. 2015b) used sites with logged water table records to suggest that absolute values of transfer function predictions should be treated with considerable caution, although relative trends may be more robust.

The sampling structure of the test-sets used in these studies is generally similar to those of the transfer function training sets with several samples scattered across a comparatively small number of sites. An alternative approach is to consider the predictive ability of a transfer function for samples spanning a localised hydrological gradient within a single site, such as from a hollow to a hummock. Theoretically, a range of water tables in a small area may be a better analogue for a change in water table in a single spot over time than a range of water tables over a much broader area. Such localised hydrological gradients also offer the potential to investigate the ability of the proxy to discern small differences in wetness. If a transfer function is able to accurately reproduce the water table profile across a localised spatial gradient this is a very positive indicator that the transfer function will accurately reconstruct temporal change. This approach has been taken in some previous studies. Hummock-hollow transects have been studied in the process of transfer function development by Van Bellen et al. (2014) and Amesbury et al. (2012) and near surface stratigraphies from sites along such a transect have been studied by (Avel-Niinemets et al. 2011).

Here we use localised hydrological gradients both to test the performance of an individual transfer function and to explore the value of this approach more generally. Our study forms part of the process of developing and establishing the