Diatom-based models for inferring past water chemistry in western Ugandan crater lakes

Keely Mills (author for correspondence)
Centre for Environmental Management,
School of Science, Information Technology & Engineering, University of Ballarat, VIC 3353, Australia.
Phone: +61 (0) 3 5327 9941
Fax: +61 (0) 3 5327 9602
Email: k.mills@ballarat.edu.au

David B. Ryves
Centre for Hydrological and Ecosystem Science (CHES),
Department of Geography, Loughborough University, Loughborough, Leics LE 11 3TU, UK.
Phone: +44 (0) 1509 22 8192
Fax: +44 (0) 1509 22 3930
Email: d.b.ryves@lboro.ac.uk

Keywords: Eastern Africa; Diatoms; Conductivity transfer function; Weighted averaging; Palaeoclimate; EDDI (European Diatom Database Initiative).

Abstract

Diatom surface sediment samples and corresponding water chemistry were collected from 56 lakes across a natural conductivity gradient in western Uganda (reflecting a regional climatic gradient of effective moisture) to explore factors controlling diatom distribution. Here we develop a regional training set from these crater lakes to test the hypothesis that this approach, by providing more appropriate and closer analogues, can improve the accuracy of palaeo-conductivity reconstructions, and so environmental inferences in these lake systems compared to larger training sets. We compare this output to models based on larger, but geographically and limnologically diverse training sets, using the European Diatom Database Initiative (EDDI) database. The relationships between water chemistry and diatom distributions were explored using canonical correspondence analysis (CCA) and partial CCA. Variance partitioning indicated that conductivity accounted for a significant and independent portion of this variation. A transfer function was developed for conductivity (r^2_{jack} = 0.74). Prediction errors, estimated using jack-knifing, are low for the conductivity model (0.256 $\log_{10}$ units). The resulting model was applied to a sedimentary sequence from Lake Kasenda, western Uganda. Comparison of conductivity reconstructions using the Ugandan crater lake training set and the East Africa training set (EDDI) highlighted a number of differences in the optima of key diatom taxa, which lead to differences in reconstructed values and could lead to misinterpretation of the fossil record. This study highlights issues of how far transfer functions based on continental-scale lake datasets such as the EDDI pan-African models should be used and the benefits that may be obtained from regional training sets.

Introduction

Recent research has highlighted the complexity of climatic fluctuations in eastern Africa over the last 2000 years (Russell et al. 2007; Russell et al. 2009; Stager et al. 2009), especially in terms of fluctuations associated with the Mediæval Warm Period and the Little Ice Age. Biotic proxies in lake sediments have been extensively used to understand climate and environmental change in many different lake types and regions. Diatom analysis is one of the most powerful and successful proxies used in reconstructing past water chemistry parameters (Battarbee 2000; Battarbee et al. 2001). Calibrating the present day distributions of diatoms in lakes against a known environmental gradient of ecological and environmental significance (e.g. water chemistry, such as conductivity) provides a tool for inferring changes from sedimentary records in lake basins. In particular, diatoms are known to respond to changes in conductivity even below the true saline threshold (3 g L^{-1} total dissolved solids; Hammer 1986; Wilson et al. 1996; Ryves et al. 2002), while a direct control by salinity (osmotic pressure) on diatom metabolism is supported on strong physiological grounds.

In eastern Africa, quantitative diatom models have been developed to trace changes in past water chemistry (Gasse and Tekaia 1983; Gasse et al. 1995). The broad geographic area, wide species coverage and statistical strength of the EDDI East Africa transfer function (http://craticula.ncl.ac.uk/Eddi) has led to its use in numerous studies that seek to reconstruct former African hydrological and climatic variability, especially in larger lakes (Gasse and Van Campo, 1994; Barker et al. 2002; Chalié and Gasse, 2002; Stager et al. 2005; Stager et al. 2009, Stager et al. 2011). Such calibration sets are continental in scale, and often result through the merging of regional training sets. However, these may not be appropriate in regions dominated by lake types or taxa rare in these datasets, and may lead to unreliable inferences. There are analogue issues in their application to geographic areas and lake types not well covered in the models, and inevitably taxonomic concerns arise from merging training sets analysed by different workers at different times using a range of sampling methods. This is particularly the case for taxonomically difficult, but ecologically diverse, diatom groups. Whilst some of these aspects can be explored using analogue-matching techniques within these large datasets or developing sample-specific training sets and moving-window models (Hübener et al. 2008), these essentially statistical approaches do not address fundamental issues about regional and local representation within these datasets that other studies have shown to be important (Reed 1998; Avery et al. 2009; Mills 2009).

Volcanic crater lakes form a distinct limnological group (Hutchinson 1957) and are numerous and widespread in western Uganda (Melack 1978; Kizito et al. 1993), where they provide a large and representative sample of tropical mid-altitude crater lakes, which have the advantage of proximity and accessibility. Here, the crater lakes extend over a natural conductivity gradient, vary in their trophic status and exhibit differing degrees of human impact within their catchments. We sought to establish

which measured limnological and catchment variables exerted a strong control on diatom communities and could therefore be used to create models for quantitative inference.

A training set from crater lakes in western Uganda was developed to test the hypothesis that regional datasets, by providing more appropriate and closer analogues, can improve the accuracy of palaeo-conductivity reconstructions compared to larger training sets. The environmental inferences from the regional dataset were compared to models based on larger, but geographically and limnologically diverse training sets. We applied both the EDDI and crater lake model to a fossil diatom assemblage from Lake Kasenda, within the study region, to compare how each performed in inferring water chemistry, in the context of other independent proxy records of environmental change from the sediment record over the last ca. 1000 years.

Study area

In western Uganda, there are over 80 crater lakes in four distinct lake districts (Fig. 1) formed in association with the western branch of the East African rift valley system, and straddle the equator at 30°E from 0°42'N to 0°19'S. Most have been formed as a direct result of volcanic activity and comprise maars and phreatomagmatic explosion craters (lake types 11 and 12, respectively; Hutchinson 1957). The four lake clusters of Fort Portal, Kasenda, Katwe-Kikorongo and Bunyaruguru differ in landscape setting, climate and hydrology, and hence water chemistry. The northern lakes, near Fort Portal, lie slightly higher in altitude (1520 m asl) than those in the Kasenda cluster (1220-1400 m asl). The Katwe-Kikorongo cluster lie on the rift valley floor (895-925 m asl), whilst those in the Bunyaruguru cluster extend into the southern uplands (975-1250 m asl). The crater lakes are spread along a strong ecotonal gradient between the moist shoulders north and south, and the dry floor of the rift valley itself (Supplementary Table 1 summarises the physical and chemical data from the four lake clusters).

Rainfall in equatorial eastern Africa is seasonally bimodal, driven by the biannual migration of the Inter Tropical Convergence Zone. In the northern and southern crater regions, annual rainfall varies from 1300-1600 mm yr^{-1}. By contrast, the rift floor averages 750-1000 mm yr^{-1}, with extreme dry seasons (Atlas of Uganda 1962). The natural vegetation of the region is low-altitude montane forest (Langdale-Brown et al., 1964; Ssemmanda et al. 2005), which, outside of the national parks has been subjected to widespread clearance to make way for small-scale farming.

Although the crater lakes of western Uganda share similar modes of formation, age and geology, they display a great variety of physical, chemical and biological characteristics (Melack 1978) arising from variations in climate, land-use and morphometry, and as such are natural laboratories for studying limnological responses to a range of drivers in one (bio)geographic area (Mills 2009). There is a strong conductivity gradient amongst the lakes, which vary from dilute to hypersaline (<100 to

135,400 µS cm^{-1}), and from shallow and polymictic to deep and permanently stratified (Melack 1978, Kizito et al. 1993; Eggermont and Verschuren 2004; Mills 2009).

Lake Kasenda was chosen for the application of the transfer function. Lake Kasenda (0° 26'N, 30° 17'E, 1260 m asl) is a closed crater lakes located close to the Kibale Forest National Park (Ssemmanda et al. 2005). Lake Kasenda is a 0.13 km^2, 13 m deep, freshwater lake (380 µS cm^{-1}) in a crater with a topographic catchment area of c. 2 km^2. This lake has been the subject of multiproxy palaeoecological studies including pollen, diatoms, macrofossils, chironomids and authigenic carbonates (Ssemmanda et al. 2005, Ryves et al., 2011).

Methods

Field sampling

The collection of contemporary diatom samples was carried out during several independent field campaigns between March 2000 and January 2007. In all cases, the sampling was undertaken during the dry season. Sampling yielded 64 surface-sediment samples, surface-water conductivity measurements (a common proxy for salinity in closed basins in semi-arid regions; Fritz et al. 1999) and other water chemistry parameters from 56 lakes in western Uganda (with some sites having multiple samples from different sampling years). Surface sediment samples were either from the intact sediment-water interface in short cores using a HON-Kajak corer (Renberg 1991) or, in lakes <1 m deep, samples were taken by hand (D. Verschuren, pers. comm.). Basic water chemistry data (e.g. temperature, pH and conductivity) were collected using Hydrolab Quanta and Datasonde probes.

Water chemistry

Filtered water samples were analysed for cations (Ca, Mg, Na, K, Li, Ba, Sr, Fe, Mn), anions (F, Cl, NO_3, SO_4), dissolved silica (Si), total dissolved solids (TDS), dissolved phosphorus (PO_4-P) and dissolved inorganic carbon (DIC). Samples for all water chemistry analyses were stored in a portable refrigeration unit at <4°C and frozen (anion and cation samples were refrigerated only) on return to the UK before being sent for analysis.

Filtrates for cation, PO_4-P and TDS analysis were stabilized with ultra-concentrated nitric acid and stored in 60 ml acid pre-washed bottles. Filtered samples for analysis of anions, Si and DIC were stored untreated. Cations and Si were analysed by Inductively Coupled Plasma Atom Emission Spectrometry (ICP-AES; IRIS, Thermo Elemental). F, Cl, NO_3 and SO_4 were analysed by ion-exchange chromatography (DX100, Dionex).

Total phosphorus (TP) and total nitrogen (TN) samples were unfiltered and fixed by adding concentrated sulphuric acid (reducing pH to <2). TP was determined by wet oxidation in an acid

persulphate solution (120°C, 30 min) and TN (as nitrate plus nitrite) by wet oxidation in an alkaline persulphate solution following the methods of Grasshoff et al. (1983). Pre-treatment for chlorophyll-*a* determinations involved filtering a known volume of water through Whatmann GF/F glass-fibre filters (0.8 µm). Chl-*a* was determined by high-pressure liquid chromatography following the protocol of Wright et al. (1997). Samples for dissolved organic carbon (DOC) were filtered and analysed using a Shimadzu Total Organic Carbon analyser (TOC-VCSN).

Water samples stored in acid pre-washed bottles were taken for isotopic analyses ($\delta^{13}C_{TDIC}$, $\delta^{18}O$ and δD). Total dissolved inorganic carbon (TDIC) for $\delta^{13}C$ analysis was precipitated by the addition of $BaCl_2$ + NaOH solution, filtered and washed with de-ionised water. A smaller, untreated sample was used for the analysis of $\delta^{18}O$ and δD. The isotope ratios are reported in parts per mille (‰) versus V-SMOW.

Samples for diatom analysis were prepared following the procedure of Renberg (1990). Strewn slides were mounted in Naphrax and at least 500 valves per sample were counted under oil-immersion phase-contrast light microscopy (x1000) using a Leica DME microscope. A variety of general (Krammer and Lange-Bertalot 1986, 1988, 1991a, 1991b) and regional floras (Gasse 1986; Cocquyt 1998) were consulted, and valves identified to species level where possible. Diatom species abundances are expressed as percentages, calculated using total number of valves recorded for each sample.

Numerical methods

Prior to statistical analyses, the water chemistry data were transformed to remove skewed distributions and to reduce the influence of extreme values (Supplementary Table 2). The major water chemistry gradients within the data set were analysed using principal components analysis (PCA). Detrended correspondence analysis (DCA; Hill and Gauch 1980) of the % transformed diatom data was used to determine patterns in the assemblage data. Axis 1 gradient lengths derived through DCA were used to determine whether unimodal or linear ordination techniques were the most appropriate means to examine species response to environmental gradients (ter Braak 1995).

Canonical correspondence analysis (CCA) was used to explore the relationships between diatom assemblages and environmental variables. Forward selection, using a Bonferroni adjustment, identified significant variables (p <0.05) suitable for model development. The significance of the first axis and the sum of all canonical axes (forward selected variables) was tested using a Monte Carlo permutation test (n = 999 unrestricted permutations).

Variance partitioning (Borcard et al. 1992) by a series of CCAs was performed on the forward selected variables. The significance and unique contribution of each variable to the variance in the diatom data was determined by removing the effects of other significant variables. Partial CCAs were performed on all combinations of variables to establish the interaction between them. The importance

of each forward selected environmental variable was determined by comparing the first constrained axis (λ_1) to the unconstrained axis (λ_2; Kingston et al. 1992).

Transfer functions were derived using C2 (Juggins 2003). A variety of models were tested. All models were internally validated using the leave-one-out jack-knifing cross-validation method and were compared on the basis of their root mean squared error of prediction (RMSEP). Potential outliers in the dataset were explored by assessing the jack-knifed residual (inferred-observed; Gasse et al. 1995). "True" outliers were identified using the method of Racca and Prairie (2004) which reduces the apparent bias in the residual data, instilling greater confidence in the predictive capability of the model. The influence of the identified samples on the model was assessed using Cook's *D* (Cook and Weisberg 1982) and compared against the critical value of $n/4$ (where n = number of sites in the model). As the surface sediment from the Lake Kasenda core is included in the crater lake training set, the corresponding sample was removed from the final model prior to application to the fossil data. This had negligible effect on model performance (as estimated under internal validation) or on the inferred values.

Datasets

Three calibration datasets (UG-40, UG-58 and UG-76; coding refers to number of sites in each training set) were formed on the basis of water chemistry data availability. All samples from the 2006-2007 field work have diatom samples with full limnological and water chemistry data sampled simultaneously. Samples collected during 2000-2001 have only basic limnological data to complement the diatom samples. The samples collected by Verschuren and colleagues (2000-2002) have temporally 'offset' water chemistries (Gasse et al. 1983; Gasse 1986). Although sediment samples and water chemistry data were taken simultaneously, a change in the laboratories used for water chemistry analyses led to discrepancies in the output and an incompatibility between the results (D. Verschuren, pers. comm.). As a result, the lakes from this dataset were re-sampled for water chemistry (but not sediments) during 2007, thus creating an offset between the sediment samples and water chemistry. Comparison of the old (2001-2002) and new (2007) measured conductivities of these samples show little variation in the majority of the lakes. The largest discrepancies are noted in the hypersaline lakes, known to be the most variable on inter-annual scales (Hammer 1986; Mills 2009).

The full dataset consists of 64 sediment samples from 56 lakes, although only 58 samples contained diatoms (UG-58). The six sites without sedimentary diatoms are hypersaline (96,100 to 217,000 μS cm^{-1}). Of the 58 samples with diatoms (from 53 lakes), 40 have full water chemistry data (UG-40). Thus two crater lake datasets were created: UG-40, representing the lakes with full water chemistry and fossil diatom data and UG-58, representing all lakes with fossil diatom data and basic limnological data (e.g. conductivity and depth).

In an attempt to fill the gap between the fresh water and hypersaline sites in the UG-58 dataset, 18 additional lake sites were integrated into the dataset using sites and samples from eastern African samples within the European Diatom Database (EDDI; Supplementary Table 3). The use of these sites aimed to overcome this issue and to enhance the representation of some key taxa often found in fossil assemblages (Ryves et al. 2011; Mills 2009) but not in the training set. Only lakes from the EDDI East Africa dataset were reviewed for inclusion. This was to ensure that there was an overlap in the diatom flora and for biogeographical consistency. Sites from EDDI were chosen primarily to bridge the conductivity range from 1050 to 16,300 µS cm^{-1}. Hypersaline sites from EDDI were avoided to prevent the inclusion of unique flora associated with these unusual sites. Following this process, sites were also examined to ensure that the dominant flora were similar to that of the crater lake training set in order to create an ecologically continuous combined training set of 76 samples (UG-76). All datasets were screened to exclude taxa which had a maximum abundance of less than 0.5%, to allow consistency with the EDDI training sets.

Results

Exploratory analyses of the training set

Two hundred and twenty seven taxa were identified during diatom analysis, seven of which could not be resolved to species level. For all analyses any taxon with a maximum abundance of <0.5% was excluded, coinciding with the cut-off used within EDDI. Exploratory ordination analyses were carried out on the UG-40 dataset, as this dataset had full corresponding water chemistry. The hypersaline Lake Kitigata and freshwater Lake Kyogo were outliers. These lakes were removed and the analyses rerun. DCA of the UG-40 dataset, using untransformed (none), square-root (√) and log-transformed (log) percentage data (species >0.5%, rare species down-weighted), revealed axis 1 gradient lengths of 3.34 (√) – 5.66 (none) SD units. The resultant gradient lengths for all data transformations were greater than 2 SD suggesting unimodal methods were the most appropriate for further analysis (Hill and Gauch 1980; ter Braak 1995).

CA indicated an arch in the data; therefore DCA (with detrending by segments and down-weighting or rare taxa) was used to explore the major patterns of variation within the diatom data. Between 10.2% (none) and 13.5% (log) of species variance was explained on axis 1, and a further 6.6 (√) to 7.3% (log) on axis 2 in the DCA ordinations. Broken stick analyses revealed that axes 1 and 2 were both significant in all cases ($p = 0.05$).

PCA of the transformed environmental variables of UG-40 (Fig. 2a) indicated axis 1 (explaining 35.8%) was highly correlated with conductivity and the major ion concentrations, as well as TP and TN, whilst axis 2 (20.1%) contrasts lakes with no human impact in their catchments with

those that have partially impacted catchments. Broken stick analyses revealed that axes 1 and 2 were both significant at $p = 0.05$.

Constrained ordinations

Sites were originally chosen for the construction of a conductivity model and DCCA was used to estimate conductivity gradient length (Birks 1995). Results of the DCCA analysis on UG-40 (with 2 outliers removed, as above) revealed an axis 1 gradient length of 3.649 and CCA produced eigenvalues of 0.531 (λ_1) and 0.296 (λ_2). The first two ordination axes were significant ($p < 0.05$), accounting for 11.8% of the total variation in the diatom data. Whilst a low percentage of explained variance is relatively typical of noisy datasets with large numbers of taxa and many zero values in the species matrix, the variance explained here is higher than several other published studies (e.g. 6.8% Gasse et al. 1995; 7.3%, Reed 1998; 9.9%, Davies et al. 2002). The three constrained ordination axes explained 15.3% of the total variance. The patterns of species and sites under CCA were similar regardless of the species transformation used and showed close correspondence to the unconstrained ordinations.

Forward selection and variance partitioning

A subset of three significant environmental variables (conductivity, TN and depth) was determined using forward selection with a Bonferroni adjustment with both the first axis and full model significant at $p < 0.001$ (n = 999 permutations). The relationship between the three significant variables and the UG-40 training set is shown in Fig. 2b. These variables account for a total of 15.8% of the variance in the diatom data (Fig. 3). The total explained variance is composed of unique contributions (each significant at p = 0.001) from conductivity (6.65%), total nitrogen (TN; 3.87%) and depth (4.36%), with only 1% variance shared. These results indicate that conductivity is a dominant driver of diatom distribution (explaining 7.46% as a sole variable; Fig. 3), and therefore a diatom-conductivity model can be explored with the larger dataset, taking advantage of more samples with observed conductivity measurements.

Transfer function development

After exploration of the UG-40 dataset, CCA of the UG-58 and UG-76 were carried out to test the significance of conductivity within these datasets. Under DCCA, log-transformed conductivity as the sole explanatory variable, the ratio between the eigenvalues of the first (λ_1; constrained) DCCA axis and the second (λ_2; unconstrained) DCA axis on all three datasets (UG-40, UG-58 and UG-76) was ≥ 0.5, suggesting that the relationship between diatom species distribution and conductivity was

sufficiently strong to generate a diatom-conductivity calibration dataset (Kingston et al. 1992; ter Braak and Juggins 1993). The gradient length for conductivity (DCCA axis 1) in all datasets was >4 SD units, suggesting unimodal models (WA and WAPLS; ter Braak and Juggins 1993) were the most appropriate. Weighted Averaging (WA) models with and without tolerance downweighting were tested using classical and inverse deshrinking (Birks 1995).

For all models two clear outliers were identified: the hypersaline Lakes Mahega and Kitigata. These lakes both support unique flora, with Lake Mahega dominated by a form of *Nitzschia latens* v. *etoshensis* (Choln.) (97%) and Lake Kitigata dominated by *Navicula gawaniensis* (Gasse) and *N. latens* (Hust.) (35% and 21%, respectively). Other potential outliers in the training set were identified as those with a jack-knifed residual (inferred-observed) greater than the standard deviation of the environmental parameter in question (Gasse et al. 1995) after the removal of Lakes Mahega and Kitigata (log conductivity SD = 0.52; Fig. 4a). This highlighted six additional sites that were potential outliers in the dataset, four from the original crater lake dataset and two from the EDDI East Africa dataset.

An alternative method exists for the identification of 'true outliers' (Racca and Prairie 2004), where residuals (observed-inferred) are plotted against predicted conductivity values (Fig. 4b). This method (with Kitigata and Mahega omitted) again indicates that the same six lakes in the dataset are potential outliers. Five were removed after critically inspecting their flora and sample type from the model based on the following criteria: (1) whether removal of the outliers actually improved the model performance (r^2_{jack}); (2) whether samples had a low Cook's *D* and (3) whether removal had an impact on species optima. Although one site (Lake Kikorongo) passed conditions (1) and (2), it was kept in the model, as the deletion of this sample significantly altered the conductivity optima of some taxa in the UG-76 model. Several conductivity models were developed, based on the three datasets (UG-40, UG-58 and UG-76, with and without outliers, the latter coded "b") and using different techniques (WA, WAPLS and MAT). This was carried out to highlight the importance of the additional EDDI sites to the original UG dataset. In the majority of cases, for all datasets, WA models with tolerance down-weighting proved the strongest statistically.

Simple weighted averaging (WA) and weighted averaging-partial least squares (WA-PLS) models out performed MAT techniques (both simple MAT and weighted averaging MAT [WMAT]). MAT and WMAT resulted in models with low r^2 for the UG-40b and UG-58b training sets; RMSE and maximum bias values were large. Although the UG-76b model performed much better (r^2 = 0.62; WMAT), the RMSE and maximum bias are large (>0.3 and >0.8 log conductivity units respectively). Weighted averaging and WAPLS performed well on all three training sets, with r^2 >0.75 in all cases. However, WA performed much better when leave-one-out cross-validation techniques (jack-knifing) were employed. Values of r^2_{jack} for the WAPLS components in many cases were much less than 0.5, although high for the UG-76b training set (r^2_{jack} = 0.67). However, this improvement was related to

WAPLS component 1 (identical to simple WA with inverse deshrinking). As a result weighted averaging models were investigated further (Table 1).

Weighted averaging with tolerance down weighting (WA_{tol}) models produced lower jack-knifed RMSEP for the UG-76b model, whereas WAPLS showed no improvement at all. In all cases, the UG-76 dataset performed better than either of the smaller datasets (UG-40, UG-58). The model chosen was the WA_{tol} with classical deshrinking; r^2_{jack} = 0.74, RMSEP = 0.256 log conductivity units, based on the UG-76 dataset after removal of 7 outliers (UG-76b). This performed well (in terms of the highest r^2_{jack} and lowest RMSEP), although the WA_{tol} with inverse deshrinking gave an almost identical performance (r^2_{jack} = 0.737, RMSEP = 0.257). The model using classical deshrinking was finally selected as it has a lower maximum bias than that using inverse deshrinking (0.36 compared to 0.49 log conductivity units). The results of the diatom conductivity model of EDDI East Africa dataset and Gasse et al. (1995) are displayed in Table 1 to highlight the comparable performance of the crater lake training set. Values of r^2 and r^2_{jack} for both the UG-76b dataset and that of Gasse et al. (1995) are very similar, whilst the RMSEP and maximum bias are substantially lower in the UG-76b model; RMSEP in particular is 35-50% smaller in UG-76b compared to EDDI models. The model based on the UG-76b jack-knife WA_{tol} (classical deshrinking; hereafter referred to as the crater lake transfer function) was applied to the sediment fossil diatom data to produce conductivity reconstructions at Lake Kasenda (the deshrinking coefficients for this transfer function are given in Supplementary Table 4).

Comparison of the crater lake and EDDI transfer functions

The diatom-conductivity model was applied to sedimentary diatom assemblages from a sediment sequence collected from Lake Kasenda, a crater lake in western Uganda. The record from Lake Kasenda spans 1200 years (Ssemmanda et al. 2005; Ryves et al., 2011; Fig. 5). Conductivity reconstructions were carried out using the crater lake transfer function and the East Africa conductivity training set (EDDI). The diatom assemblages at Kasenda are dominated by *Nitzschia bacillum* (Hust.), with important contributions in certain periods by *Aulacoseira granulata* (Ehr. Simonsen), *Synedra acus* (Kütz.), *Gomphonema* spp., *Amphora copulata* (Kütz.) and *Cocconeis placentula* (Ehrenb.). Full details of the Kasenda sequence (biotic and abiotic records) are provided in Ssemmanda et al. (2005) and Ryves et al. (2011). All fossil samples from Kasenda had an average of 90% of species data covered by the crater lake model and >80% were included in the EDDI model.

The two conductivity reconstructions for Lake Kasenda show a strong inverse correlation (Fig. 5), with the EDDI model implying a saline phase early on in the record, giving way to lower conductivity for much of the record from c. AD 1000, and the opposite inferred by the crater lake model. Evidence from other proxies in this sediment sequence, and from other lakes in the wider region, strongly support the crater lake model as providing the more accurate conductivity history at

this site. Low conductivity from c. AD 700 to AD 1000 is supported by pollen and other proxy data, although conditions were likely relatively dry regionally, with freshwater conditions sustained by water inflow and salt loss via groundwater connections. Generally elevated conductivities at Kasenda from c. AD 1200 to 1600 agree with high Poaceae pollen %, low tree % and isotopic data (Ssemmanda et al. 2005) that this was a dry interval, while fresh events at c. AD 1000, 1400 and 1700 stand out (in agreement with isotopic data and regional palaeorecords; Ryves et al. 2011, Fig. 5).

Discussion

Species-environment relationships

Unimodal models assume that diatom taxa respond with a distinct single peak (optimum), with symmetric distributions (tolerance) along the environmental gradient under investigation. Whilst the gradient length of the first unconstrained and constrained axes (DCA = 5.89 SD; CCA = 8.88 SD of untransformed % data) of the diatom data indicates that diatom species responses were largely unimodal for the crater lake (UG-76) dataset, the form of response of individual species was tested using the maximum-likelihood method of Huisman-Olff- Fresco (HOF; Huisman et al. 1993). Statistical analyses of species response, using HOF for species present in at least 10 samples indicated a range of responses, from symmetric unimodal responses (e.g. *Amphora copulata, Fragilaria tenera* (W. Sm.) and *Gomphonema pumilum* (Grun. Reichardt & Lange-Bertalot) to skewed unimodal responses (e.g. *Amphora pediculus* (Kütz. Grun.) and *Aulacoseira granulata*) and sigmoidal responses (e.g. *Amphora veneta* (Kütz.) and *Aulacoseira ambigua* (Grun.)). The HOF analyses included 57 taxa from the crater lake training set. Almost 70% (39) of the 57 taxa included have a unimodal response (HOF types IV and V) to conductivity, whereas only 7% (4) show a monotonic decrease or increase; 24% (14) do not show any relationship.

There is a good relationship between the distribution of diatom taxa and conductivity in the crater lake training set. Whilst the assumption of a unimodal response for all taxa is an oversimplification of a complex system, there is a turnover of species as conductivity increases (Fig. 6). In addition to this, many species demonstrate distinct optima across this gradient, although some taxa appear indifferent or eurytopic in their response to conductivity across much of the gradient (e.g. *Nitzschia fonticola* (Hust.)). For such taxa, their distributions are thus controlled by other variables than conductivity in contemporary crater lake systems (such as *Cyclotella meneghiniana* (Kütz.); see below). Samples with the highest and lowest conductivity waters support distinct diatom floras.

Amalgamating regional and continental-scale transfer functions

One of the most significant problems encountered whilst developing the crater lake training set was the paucity of modern lakes with 'intermediate salinities' (i.e. in the range 10^3-10^4 µS cm^{-1}; Verschuren 2003; Eggermont et al. 2006). This is a phenomenon that occurs throughout eastern Africa (Verschuren 2003). Lakes across the study area (from Fort Portal in the north to Bunyaruguru in the south) are either fresh (typically <1500 µS cm^{-1}) or (hyper)saline (>20,000 µS cm^{-1}), thus few lakes lie close to the biologically important freshwater-saline transition (3000 µS cm^{-1}; Hammer 1986). This is perhaps not unexpected from the perspective of lake water balance and the potential role of local groundwater. For this reason, selected sites from the EDDI database were amalgamated with the crater lake training set to fill this gap.

However, the merging of different datasets is not without problems, one of which is the taxonomic and methodological inconsistencies between datasets due to different analysts (Birks 1995). In this instance, the EDDI diatom taxonomy was updated to match that of the crater lakes (e.g. *Amphora libyca* (Ehr.) renamed *Amphora copulata*), but this was completed using the images and morphologic descriptions available online, as the original EDDI slides could not be consulted. Furthermore it may cause secondary environmental gradients to become more prominent (Eggermont et al. 2006).

Including sites of intermediate conductivities into the transfer function not only improved the model performance, but also its utility, as intermediate conductivities may have existed in the past in western Uganda under different climatic regimes (Fig. 5; Mills 2009, Ryves et al. 2011). Given the current pattern of lake conductivity across eastern Africa (Verschuren 2003), however, it remains possible that some of these lakes have never experienced intermediate conductivities, with only abrupt changes characterising the fossil record, and thus rendering the reconstruction of intermediate conductivities misleading. The sediment record from Lake Kasenda, however, suggests otherwise, at this lake at least (Fig. 5; Ryves et al. 2011).

The distributions of some well-represented taxa from the crater lake training set can be compared with those from the EDDI East Africa training set (n = 167; Gasse et al. 1995). As demonstrated in Fig. 6, while there is generally good agreement between the two training sets, there are significant differences between the optima of some key taxa in the UG-76b training set (such as *Amphora acutiuscula, Nitzschia inconspicua* and *Cyclotella meneghiniana*).

There are a number of potential reasons why this may be the case. Firstly, and perhaps the most important reason for the differences in the species optima, is the size of the training set. Gasse et al. (1995) use a 276-lake dataset from across Africa (167 of which are in the East Africa dataset). The crater lake training set here is relatively small in size (69 after deletions), and the lakes sampled cover a smaller conductivity gradient than the African training set. The minimum recorded conductivity for the African training set is 40 µS cm^{-1} (55 µS cm^{-1} in the crater lake training set), with a maximum of 99,060 µS cm^{-1} (16,300 µS cm^{-1} in the crater lake training set after the removal of the two hypersaline outliers) and the median conductivity for the African training set is 925 µS cm^{-1} (520 µS cm^{-1} in the

crater lake training set). This highlights the fact that the crater lake transfer function contains more samples from, and is therefore biased towards, the fresher end of the conductivity gradient. It is also likely that the true optima for some species may not have been identified in the crater lake training set, given the shorter conductivity gradient (for example, *Amphora acutiuscula*, *Anomoeoneis sphaerophora* and *Navicula elkab* at the upper end of the conductivity gradient). Some taxa show very different optima despite good coverage within the crater lake training set (such as *Gomphonema intricatum*, *Nitzschia inconspicua* and *Cyclotella meneghiniana*). This is especially apparent for *Cyclotella meneghiniana*, an important diatom in many stratigraphic sequences from western Uganda (e.g. Mills 2009, Ryves et al. 2011), where there is a particularly large discrepancy between optima (Fig. 6). *Cyclotella meneghiniana*, a cosmopolitan facultatively-planktonic species is described as euryhaline (Caljon and Cocquyt 1992), but its salinity (conductivity) preference varies amongst studies (Tuchman et al. 1984; Roubiex and Lancelot 2008).

Cyclotella meneghiniana is a species most often associated with waters of higher conductivity in eastern Africa (Gasse 1986), but elsewhere has also been shown to bloom in freshwater eutrophic lakes (Dong et al. 2008) and freshwater lakes with a high water temperature (Mitrovic et al. 2008). Despite this, the optimum generated for the crater lake training set is more ecologically plausible than the EDDI African dataset optimum. There are only a handful of sites where *C. meneghiniana* occurs in significant abundance within the EDDI African dataset with extremely high conductivities, and these clearly exert a large influence on the dataset and calculated conductivity optimum (Gasse 1986; EDDI website). Furthermore, all but one of these sites included in the EDDI African dataset are either hot springs or salt swamps. The discrepancy between the optima calculated for *C. meneghiniana* between the crater lake transfer function and the EDDI East Africa dataset is more plausibly attributed to an overestimation of optimum in the latter dataset, caused by the inclusion of a small number of highly saline swamps. Alternatively, other factors which influence the presence/abundance of *C. meneghiniana* (eutrophication and high water temperatures; Dong et al. 2008; Mitrovic et al. 2008) and the robust nature of *C. meneghiniana* valves allowing them to be preferentially preserved and identified (Barker 1990; Ryves et al. 2009) may be confounding factors in the distribution of *C. meneghiniana* within EDDI.

It is possible that, in some cases, there is a mismatch between the 0-0.5 cm surface sediment and the measured water chemistry, with the sediment sample covering either a shorter or longer time period than the water sample (where sedimentation rates are very fast or very slow, respectively). Although one of the more stable aspects of water chemistry, conductivity will change seasonally corresponding to wet or dry seasons, while the upper 0.5 cm of sediment may represent a single season, or up to a year or more, and so integrate fluctuations in diatom species assemblage composition over different periods (and potentially mismatched to measured conductivity). For many smaller lakes in the region, present-day sedimentation rates are not known, but recent work across a range of crater

lakes in Uganda suggest current rates range between 0.2-2 cm yr^{-1} (e.g. rates of 2 cm yr^{-1} calculated in Lakes Wandakara and Kyasanduka; Ssemmanda et al. 2005; Mills 2009).

There are a number of limitations to quantitative palaeoecological reconstructions in western Uganda using the crater lake training set (even with additional EDDI sites included). For example, a number of important fossil species in Ugandan lake sediments are poorly represented. *Thalassiosira rudolfi* (Bach.), *Amphora coffeaeformis* (Ag.) and *A. veneta* are abundant in analysed core sections from some lakes (Mills 2009), but they are rarely found in the modern regional environment. This can result in problems when applying the transfer function to palaeoecological data. Another potential issue when applying the transfer function to core sediments is the potential modification of modern lakes and their respective surface sediment samples due to human impact. Many lake basins in western Uganda have experienced significant human impact in the recent past (last 50-100 years; Ssemmanda et al. 2005; Mills 2009). While TN is important within the Ugandan crater lakes training set, there is little statistical overlap with conductivity (although this does not preclude such an overlap in fossil assemblages). Clearly, however, in contemporary lakes, conductivity is not the dominant driver for all diatom taxa, or important at all sites, and care is needed in choosing suitable lakes if palaeoclimate is the principal research interest. Even in this climatically-sensitive region, diatom response may be driven by non-climatic (or only indirectly climatic) factors, such as nutrient concentrations or turbidity, underlining the need for a multi-disciplinary and multiproxy approach to corroborate environmental inference from any single signal. Nonetheless, we suggest that carefully constructed regional training sets and models can improve inferences made from larger, merged multi-lake type and multi-analyst datasets, especially for such smaller, limnologically distinct lakes. Such an approach can help unlock the palaeoenvironmental and palaeoclimatic potential of the rich sedimentary archives of these systems.

Conclusion

The diatom-conductivity transfer function is the first of its kind focussing on crater lakes and highlights issues of how far transfer functions based on continental-scale lake datasets such as the EDDI pan-African models should be used and the benefits that may be obtained from regional training sets. The training set assembled for this study has enabled the development of a robust transfer function for reconstructing past conductivity changes. Other important gradients were present in the data set (i.e. TN and depth), and whilst this has implications for the performance of the conductivity transfer functions, it may also facilitate the development of a robust transfer function for TN from other lakes in the region. As a result it should be used in concert with other independent biological proxies, both quantitative (e.g. chironomid conductivity model of Eggermont et al. 2006) and qualitative (e.g. pollen, macrofossils, stable isotopes).

There was a need for a new transfer function to address changes in the western Ugandan crater lakes as pre-existing transfers function tend to over-estimate the conductivity optima of several key species (as observed from modern-day diatom distributions), causing erroneously high values in reconstructions. However, significant environmental changes associated with climatic fluctuations and anthropogenic disturbance in many lakes in western Uganda has meant that a number of important species in other regional palaeolimnological records do not have modern analogues. The lakes in western Uganda today are either fresh (shoulder of the rift valley), or very saline (floor of the rift valley), and the training set perhaps bears similarities to the Spanish dataset of Reed (1998). This problem was overcome by the addition of extra sites of intermediate salinities from the EDDI East Africa training set. Whilst this undoubtedly improves the model's performance, it may not be ecologically realistic if these systems are subject to abrupt transitions between fresh and saline conditions. Multiproxy analyses can generally determine if this has been the case, however, in particular lakes.

Models are best judged by how useful they are (rather than how accurate), and by this test the crater lake model is more useful for such lakes than the larger, but heterogeneous, EDDI models. The transfer function developed from the crater lakes of western Uganda improves on the regional EDDI models which have been widely, and often uncritically, applied across a range of lake types in eastern Africa. Nonetheless, the EDDI training set remains useful, especially for larger lakes, and can be refined by applying models to planktonic taxa only for example (Stager et al. 2005, 2009, 2011). We argue that developing more regionally- and limnologically-specific models, as here, may provide the appropriate key to unlock the archives of eastern African crater lakes. These lakes house as yet under-exploited records of palaeoenvironmental information, but their value is being increasingly recognised to explore and answer key questions of regional and local-scale change over the Holocene at high temporal resolution.

Acknowledgements

This work was completed by KM as part of a Ph.D. carried out at and funded by Loughborough University. KM is currently a Research Fellow working within the University of Ballarat's Collaborative Research Network. Financial support for the fieldwork was provided through NERC (UK) within a New Investigators' Competition award (NE/D000157/1) to DBR, and grants from the Quaternary Research Association and the British Institute in Eastern Africa to KM. Additional support for isotopic analyses of water samples by the NERC Isotope Geosciences Laboratory (NIGL award code IP/884/1105). Fieldwork in Uganda was supported by the Department of Geology at Makerere University, Makerere University Biological Field Station (Kibale) and Loughborough University. We thank the Uganda National Council for Science and Technology (permit EC482), Uganda Wildlife Authority and the Office of the President for fieldwork permission. We thank Dirk Verschuren, Bob

Rumes, Hilde Eggermont (Ghent University) for providing sediment samples and lake data and Georg Schettler (GeoForschungsZentrum), Lei Chou (Université Libre de Bruxelles) and Renaat Dasseville (Ghent University) for the processing of the water chemistry samples. Many sincere thanks go to John Anderson, Sergi Pla, Immaculate Ssemmanda, Julius Lejju and (especially) Richard Nyakoojo for their invaluable help in the field.

References

Atlas of Uganda (1962) Ugandan Department of Lands and Surveys. The Government Printer, Entebbe

Avery L, Shinneman C, Edlund MB, Almendinger JE, Sonunkhishig N (2009) Diatoms as indicators of water quality in Western Mongolian lakes: a 54-site calibration set. J Paleolimnol 42: 373-389

Barker P (1990) Diatoms as palaeolimnological indicators: a reconstruction of Late Quaternary environments in two East African salt lakes. Unpublished PhD Thesis, Loughborough University of Technology

Barker P, Telford R, Gasse F, Florian T (2002) Late Pleistocene and Holocene palaeohydrology of Lake Rukwa, Tanzania, inferred from diatom analysis. Palaeogeogr Palaeoclimatol Palaeoecol 187: 295-305

Battarbee RW (2000) Palaeolimnolgical approaches to climate change, with special regard to the biological record. Quat Sci Rev 19: 107-204

Battarbee R, Jones VJ, Flower RJ, Cameron NG, Bennion H, Carvalo L, Juggins S (2001) Diatoms. In: Smol JP, Birks HJB, Last WM (eds) Tracking Environmental Change Using Lake Sediments. Volume 3: Terrestrial, Algal, and Siliceous Indicators. Kluwer Academic Publishers, Dordrecht, pp 155-202

Birks HJB (1995) Quantitative palaeoenvironmental reconstructions. In: Maddy D, Brew JS (eds) Statistical Modelling of Quaternary Science Data. Quaternary Research Association Technical Guide 5, Cambridge, pp 161-254

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73: 1045-1055

Caljon AG, Cocquyt CZ (1992) Diatoms from surface sediments of the northern part of Lake Tanganyika. Hydrobiologia 230: 135-156

Chalié F, Gasse F (2002) Late Glacial-Holocene diatom record of water chemistry and lake level change from the tropical East African Rift Lake Abiyata (Ethiopia). Palaeogeogr Palaeoclimatol Palaeoecol 187: 259-283

Cocquyt C (1998) Diatoms from the Northern Basin of Lake Tanganyika. Bibliotheca Diatomologica 39. Crammer, Berlin/Stuttgart.

Cook RD, Weisberg S (1982) Residuals and Influence in Regression. Chapman and Hall, New York

Davies SJ, Metcalfe SE, Caballero ME, Juggins S (2002) Developing diatom-based transfer functions for Central Mexican lakes. Hydrobiologia 467: 199-213

Dong X, Bennion H, Battarbee R, Yang X, Yang H, Liu E (2008) Tracking eutrophication in Taihu Lake using the diatom record: potential and problems. J Paleolimnol 40: 413-429

Eggermont H, Verschuren D (2004) Subfossil Chironomidae from East African low- and mid-elevation lakes. 1. Tanypodinae and Orthocladiinae. J Paleolimnol 32: 383-412

Eggermont H, Heiri O, Verschuren D (2006) Subfossil Chironomidae (Insecta: Diptera) as quantitative indicators for past salinity variation in African lakes. Quat Sci Rev 25: 1966-1994

Fritz SC, Cumming BF, Gasse F, Laird K (1999). Diatoms as indicators of hydrologic and climate change in saline lakes. In: Stoermer EF, Smol JP (eds) The Diatoms: Applications for Environmental and Earth Sciences. Cambridge University Press, Cambridge, pp 41-72

Gasse F, Van Campo E (1994) Abrupt post-glacial climate events in West Asia and North Africa monsoon domains. Earth Planet Sci Lett 126: 435-456

Gasse F (1986) East African diatoms; Taxonomy, ecological distribution. Bibliotheca Diatomologica 11. Crammer, Berlin/Stuttgart

Gasse F, Tekaia F (1983) Transfer functions for estimating paleoecological conditions (pH) from East African diatoms. Hydrobiologia 103: 85-90

Gasse F, Talling JF, Kilham P (1983) Diatom assemblages in East Africa: classification, distribution and ecology. Review Hydrobiology Tropical 116: 3-34

Gasse F, Juggins S, Khelifa LB (1995) Diatom-based transfer functions for inferring past hydrochemical characteristics of African lakes. Palaeogeogr Palaeoclimatol Palaeoecol 117: 31-54

Grasshoff K, Ehrhardt M, Kremling K (1983) Methods of seawater analysis. Verlag Chemie, Weinheim

Hammer UT (1986) Saline lake ecosystems of the world. In: Dumont HJ (ed) Monographiae Biologicae. Junk, Dordrecht

Hill MO, Gauch HG (1980) Detrended correspondence analysis: An improved ordination technique. Plant Ecol 42: 47-58

Hübener T, Dreßler M, Schwarz A, Langner K, Adler S (2008) Dynamic adjustment of training sets (‘moving-window’ reconstruction) by using transfer functions in palaeolimnology – a new approach. J Paleolimnol 40: 79-95

Huisman J, Olff H, Fresco LFM (1993) A hierarchical set of models for species response analysis. J Veg Sci 4: 37–46

Hutchison GE (1957) A treatise on limnology. John Wiley & Sons, New York

Juggins S (2003) C2, user Guide; Software for Ecological and Palaeoecological Data Analysis and Visualisation. University of Newcastle, Newcastle upon Tyne

Kingston JC, Birks HJB, Uutala AJ, Cumming BF, Smol JP (1992) Assessing the trends in fishery resources and lake water aluminium from palaeolimnological analyses of siliceous algae. Can J Fish Aquat Sci 49: 127–138

Kizito YS, Nauwerck A, Chapman LJ, Koste W (1993) A limnological survey of western Uganda crater lakes. Limnologica 23: 335-347

Krammer K, Lange-Bertalot H (1986) Bacillariophyceae. 1. Teil. Naviculaceae. Süßwasserflora von Mitteleuropa, Band 2/1. Gustav Fischer Verlag, Stuttgart

Krammer K, Lange-Bertalot H (1988) Bacillariophyceae. 2. Teil. Bacillariaceae, Epithemiaceae, Surirellaceae. Süßwasserflora von Mitteleuropa, Band 2/2. Gustav Fischer Verlag, Stuttgart

Krammer K, Lange-Bertalot H (1991a) Bacillariophyceae. 3. Teil. Centrales, Fragilariaceae, Eunotiaceae. Süßwasserflora von Mitteleuropa, Band 2/3. Gustav Fischer Verlag, Stuttgart

Krammer K, Lange-Bertalot H (1991b) Bacillariophyceae. 4. Teil. Achnanthaceae, kritsche erganzungen zu Navicula (Lineolatae) und Gomphonema Gesamliteraturverzeichnis. Süßwasserflora von Mitteleuropa, Band 2/4. Gustav Fischer Verlag, Stuttgart

Langdale-Brown I, Osmaston HA, Wilson JG (1964) The Vegetation of Uganda and Its Bearing on Land-Use. The Government Printer, Entebbe

Melack JM (1978) Morphometric, physical and chemical features of the volcanic crater lakes of western Uganda. Arch Hydrobiol 84: 430-453

Mills K (2009) Ugandan crater lakes: Limnology, palaeolimnology and palaeoenvironmental history. Unpublished Ph.D. Thesis, Loughborough University

Mitrovic SM, Chessman BC, Davie A, Avery EL, Ryan N (2008) Development of blooms of *Cyclotella meneghiniana* and *Nitzschia* spp. (Bacillariophyceae) in a shallow river and estimation of effective suppression flows. Hydrobiologia 596: 173-185

Racca JMJ, Prairie YT (2004) Apparent and real bias in numerical transfer functions in palaeolimnology. J Paleolimnol 31: 117-124

Reed JM (1998) A diatom-conductivity transfer function for Spanish salt lakes. J Paleolimnol 19: 399-416

Renberg I (1990) A procedure for preparing large sets of diatom slides from sediment cores. J Paleolimnol 4: 87-90

Renberg I (1991) The HON-Kajak sediment corer. J Paleolim 6: 167-70

Roubiex V, Lancelot C (2008) Effect of salinity on growth, cell size and silification of an euryhaline freshwater diatom: *Cyclotella meneghiniana* Kütz. Transitional Waters Bulletin 1: 31-38

Russell JM, Eggermont H, Verschuren D (2007) Spatial complexity during the Little Ice Age in tropical East Africa: sedimentary records from contrasting crater lake basins in western Uganda. Holocene 17: 183-193

Russell JM, McCoy SJ, Verschuren D, Bessems I, Huang Y (2009) Human impacts, climate change, and aquatic ecosystem response during the past 2,000 years at Lake Wandakara, Uganda. Quat Res 72: 315-324

Ryves DB, McGowan S, Anderson NJ (2002) Development and evaluation of a diatom-conductivity model from lakes in West Greenland. Freshwater Biol 47: 995-1014

Ryves DB, Mills K, Bennike O, Brodersen KP, Lamb AL, Leng MJ, Russell JM, Ssemmanda I (2011) Environmental change over the last millennium recorded in two contrasting crater lakes in western Uganda, eastern Africa (Lakes Kasenda and Wandakara). Quat Sci Rev 30: 555-569

Ssemmanda I, Ryves DB, Bennike O, Appleby PG (2005) Vegetation history in west Uganda during the last 1200 years: a sediment-based reconstruction from two crater lakes. Holocene 15: 119-132

Stager JC, Ryves D, Cumming BF, Meeker LD, Beer J (2005) Solar variability and the levels of Lake Victoria, East Africa, during the last millennium. J Paleolimnol 33: 243-251

Stager JC, Cocquyt C, Bonnefille R, Weyhenmeyer C, Bowerman N (2009). A late Holocene paleoclimatic history of Lake Tanganyika, East Africa. Quat Res 72: 47-56

Stager JC, Ryves DB, Chase BM, Pausata SR (2011) Catastrophic drought in the Afro-Asian monsoon region during Heinrich Event 1. Science 331: 1299-1302

ter Braak CJF, Juggins S (1993). Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. Hydrobiologia 269/270: 485-502

ter Braak CJF (1995) Non-linear methods for multivariate statistical calibration and their use in palaeoecology: a comparison of inverse (k-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. Chemom Intell Lab Syst 28, 165-180

Tuchman ML, Theriot E, Stoermer EF (1984) Effects of low level salinity concentrations on the growth of *Cyclotella meneghinianana* Kütz (Bacillariophyta). Arch Protistenkd 128: 319-326

Verschuren D (2003) Lake-based climate reconstruction in Africa: progress and challenges. Hydrobiologia 500: 315-330

Wilson SE, Cumming BF, Smol JP (1996) Assessing the reliability of salinity inference models from diatom assemblages: an examination of a 219-lake data set from western North America. Can J Fish Aquat Sci 53: 1580–1594

Wright SW, Wright JW, Jeffrey SW (1997) High resolution system for chlorophylls and carotenoids of marine phytoplankton. In: Jeffrey SW, Mantoura RFC, Wright SW (eds) Phytoplankton pigments in oceanography: a guide to advanced methods. SCOR-UNESCO, Paris, pp 327-360

Table 1 Comparison of methods for various weighted averaging (WA) diatom-conductivity models. Results are shown for the original crater lake models (UG-40 and UG-58) and for the UG-76 model, supplementing the original dataset with sites of 'intermediate salinities' (1000-10,000 μS cm^{-1}) from the European Diatom Database (EDDI). All modelling was carried out on 0.5% screened data (to ensure consistency with the EDDI dataset). The best model in each set is highlighted. The final model is highlighted in **bold.**

Model	Method	N° Sites	N° Species	r^2	RMSE	Max bias	r^2_{jack}	RMSEP	Max bias
UG-40	WA_Inv	40	143	0.90	0.1587	0.3411	0.32	0.4336	2.2218
	WA_Cla	40	143	0.90	0.1670	0.4022	0.37	0.4178	2.1026
	WA_{tol}_Inv	40	143	0.91	0.1552	0.3696	0.45	0.4157	2.1188
	WA_{tol}_Cla	40	143	0.91	0.1629	0.3775	0.50	0.3977	1.9968
UG-40b (1 outlier[#])	WA_Inv	39	138	0.75	0.1577	0.2934	0.33	0.2603	0.5216
	WA_Cla	39	138	0.75	0.1819	0.3151	0.36	0.2709	0.4429
	WA_{tol}_Inv	39	138	0.79	0.1443	0.2899	0.44	0.2394	0.4012
	WA_{tol}_Cla	39	138	0.79	0.1623	0.3862	0.46	0.2352	0.4054
UG-58	WA_Inv	58	165	0.90	0.1873	0.4205	0.58	0.4500	0.4205
	WA_Cla	58	165	0.90	0.1979	0.4188	0.59	0.4335	0.4188
	WA_{tol}_Inv	58	165	0.91	0.1743	0.4280	0.28	0.5227	0.4280
	WA_{tol}_Cla	58	165	0.91	0.1827	0.4273	0.29	0.5169	0.4273
UG-58b (2 outliers*)	WA_Inv	56	159	0.75	0.1776	0.2544	0.14	0.3314	1.4236
	WA_Cla	56	159	0.75	0.2056	0.7462	0.15	0.3461	1.5106
	WA_{tol}_Inv	56	159	0.78	0.1662	0.2507	0.35	0.2843	0.7817
	WA_{tol}_Cla	56	159	0.78	0.1885	0.5261	0.37	0.2816	0.6339
UG-76	WA_Inv	76	206	0.86	0.2383	0.3795	0.61	0.4003	1.5657
	WA_Cla	76	206	0.86	0.2574	0.3954	0.62	0.3927	1.4209
	WA_{tol}_Inv	76	206	0.89	0.2107	0.3227	0.68	0.3687	1.3003
	WA_{tol}_Cla	76	206	0.89	0.2235	0.2886	0.69	0.3572	1.1597
UG -76b (7 outliers[†])	WA_Inv	69	198	0.82	0.2111	0.2733	0.66	0.2897	0.4932
	WA_Cla	69	198	0.82	0.2331	0.3838	0.67	0.3111	0.3410
	WA_{tol}_Inv	69	198	0.87	0.1797	0.2639	0.74	0.2567	0.4937
FINAL MODEL	**WA_{tol}_Cla**	**69**	**198**	**0.87**	**0.1927**	**0.2812**	**0.74**	**0.2560**	**0.3611**
EDDI (East African Dataset)	WA_Inv	179	332	0.86	0.3237	0.3064	0.78	0.4101	0.4609
	WA_{tol}_Inv	179	332	0.86	0.3482	0.3392	0.78	0.4224	0.4789
	WA_Cla	179	332	0.89	0.2847	0.1970	0.71	0.4739	0.4034
	WA_{tol}_Cla	179	332	0.89	0.3010	0.2036	0.71	0.4927	0.3454
Gasse et al. (1995) (African Dataset)	WA_Inv	274	389	0.87	0.32	0.40	0.81	0.39	0.52
	WA_{tol}_Inv	274	389	0.92	0.26	0.32	0.80	0.41	0.48
	WA_Cla	274	389	0.87	0.34	0.35	0.81	0.40	0.46
	WA_{tol}_Cla	274	389	0.92	0.27	0.28	0.80	0.42	0.39

Figure Captions

Fig. 1 a Regional location of Uganda and the EDDI sites; **b** Map of Uganda showing the four crater lake clusters: Fort Portal (FP), Kasenda (Ka), Katwe-Kikorongo (KK) and Bunyaruguru (Bu) as described by Melack (1978); **c** location of the EDDI lakes included in the transfer function (using the coordinates of the sampling points as given in Supplementary Table 4).

Fig. 2 a Principal Components Analysis (PCA) of all the measured environmental variables on the UG-40 dataset (with the outliers Lakes Kitigata and Lake Kyogo removed); **b** Canonical Correspondence Analysis (CCA) of UG-40 showing (i) three forward-selected environmental variables (vectors); (ii) surface sediment samples; (iii) WA optima of 32 diatom taxa (>10% abundance; open triangles). The samples are given as numbers and the diatom species are given as codes (see Supplementary Table 5 for details).

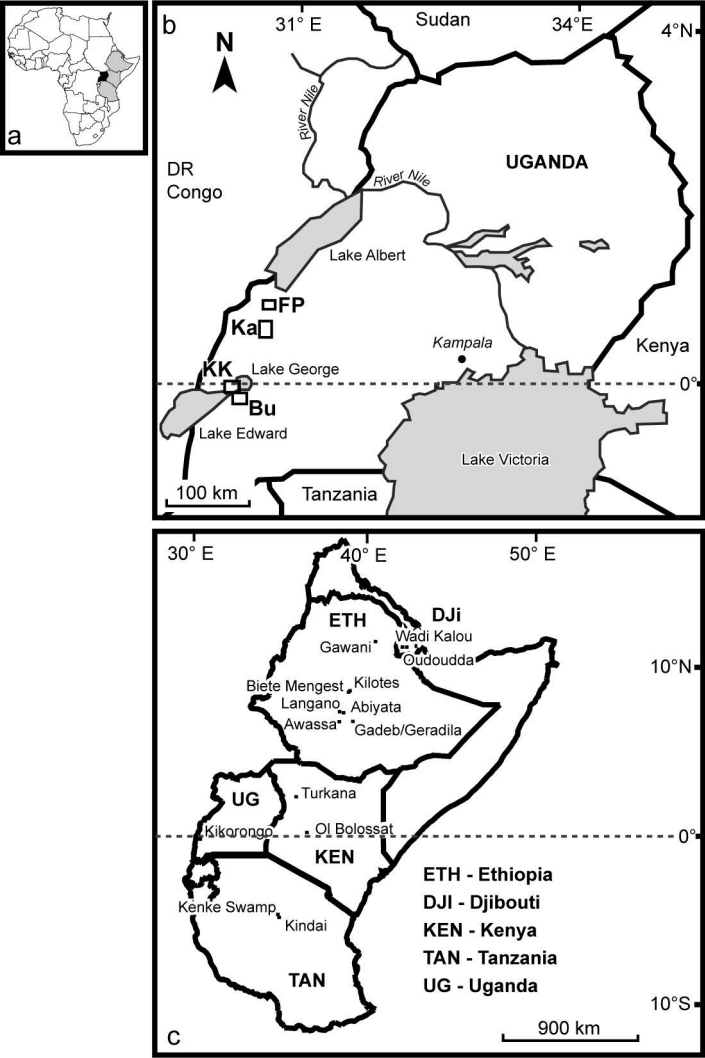
Fig. 3 Results of variance partitioning for the three environmental variables for the UG-40 training set. All of the environmental variables were significant in the partial CCA ($p < 0.05$).

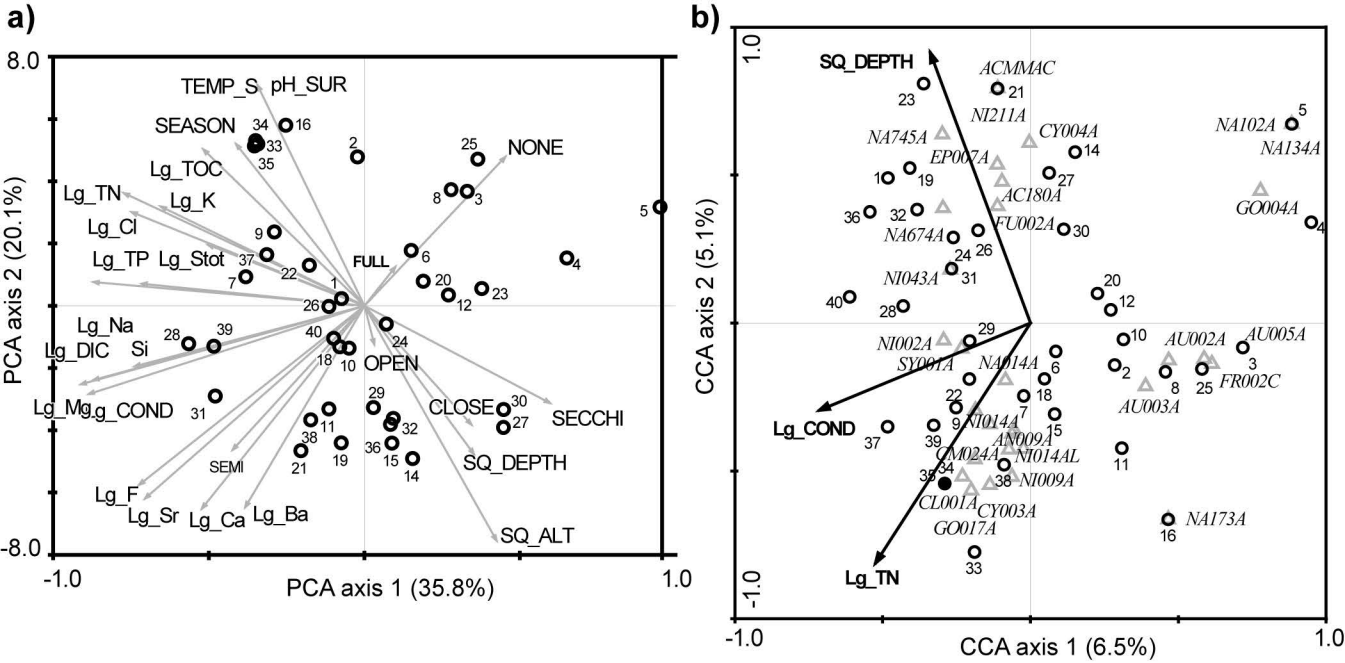
Fig. 4 a: i-iv Performance of the weighted averaging model with tolerance downweighting and classical deshrinking (with leave-one-out jack-knife) for the UG-76 dataset (EDDI sites denoted as open circles). A lowess smooth has been applied to the residual data (ii and iii). Potential outliers are highlighted in were identified using the standard deviation of the conductivity data (0.52 log conductivity units; ii); **b** Potential outliers identified in the UG-76 dataset using the method of Racca and Prairie (2004). The dashed line denotes the standard deviation of the conductivity data. This alternative method aims to identify 'true outliers' whilst reducing the trend in the data. This method highlights the same six outliers identified in ii; **c** Apparent vs. observed conductivity for the final model, UG-76b (the crater lake model).

Fig. 5 Comparison of reconstructed conductivity using the crater lake model and EDDI (East Africa) training sets on a sediment core from Lake Kasenda (see Ryves et al., 2011). Pollen data are from Ssemmanda et al. (2005) and bulk carbonate data from Ryves et al. (2011). % sum included = % of fossil (core) data included in the model. Undiff. = undifferentiated pollen types (by habitat). Excursions to lighter (less positive) values of $\delta^{18}O$ and $\delta^{13}C$ are interpreted as wetter conditions (reduced evaporative impact on the isotope signal).

Fig. 6 Weighted average conductivity optima and tolerances for species >10% in any one sample from the crater lake model dataset (closed circles) compared to the conductivity optima as estimated

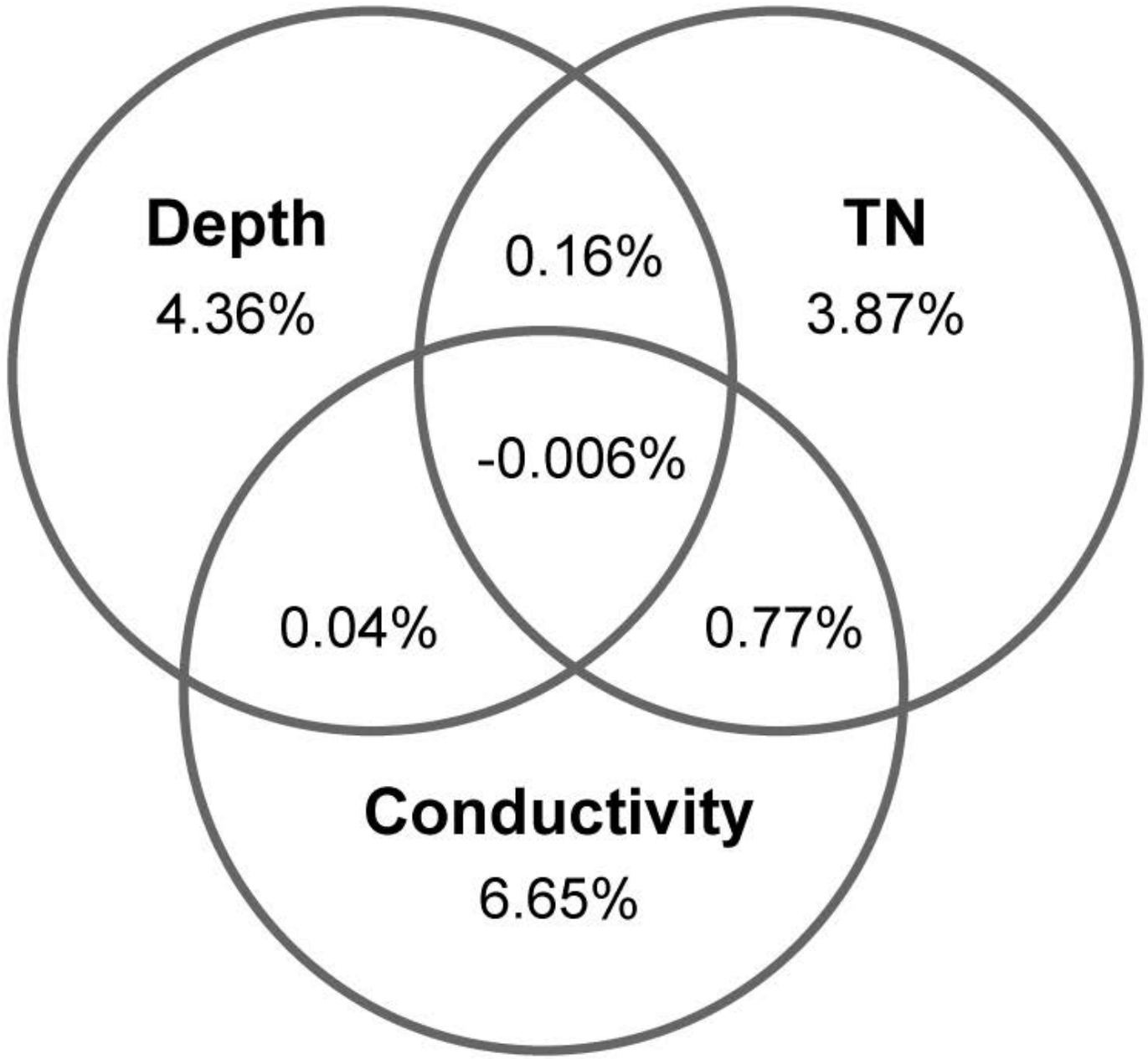
using the EDDI transfer function for conductivity (East Africa dataset; open circles). The species names follow that of EDDI.


a
b
N
31° E
Sudan
34°E
4°N
River Nile
River Nile
DR
Congo
UGANDA
Lake Albert
FP
Ka
KK
Lake George
Bu
Lake Edward
Kampala
Kenya
0°
Lake Victoria
100 km
Tanzania
c
30° E
40° E
50° E
ETH
DJi
Wadi Kalou
Gawani
Oudoudda
10°N
Biete Mengest
Kilotes
Langano
Abiyata
Awassa
Gadeb/Geradila
UG
Turkana
Kikorongo
Ol Bolossat
0°
KEN
ETH - Ethiopia
DJI - Djibouti
KEN - Kenya
TAN - Tanzania
UG - Uganda
Kenke Swamp
Kindai
TAN
10°S
900 km

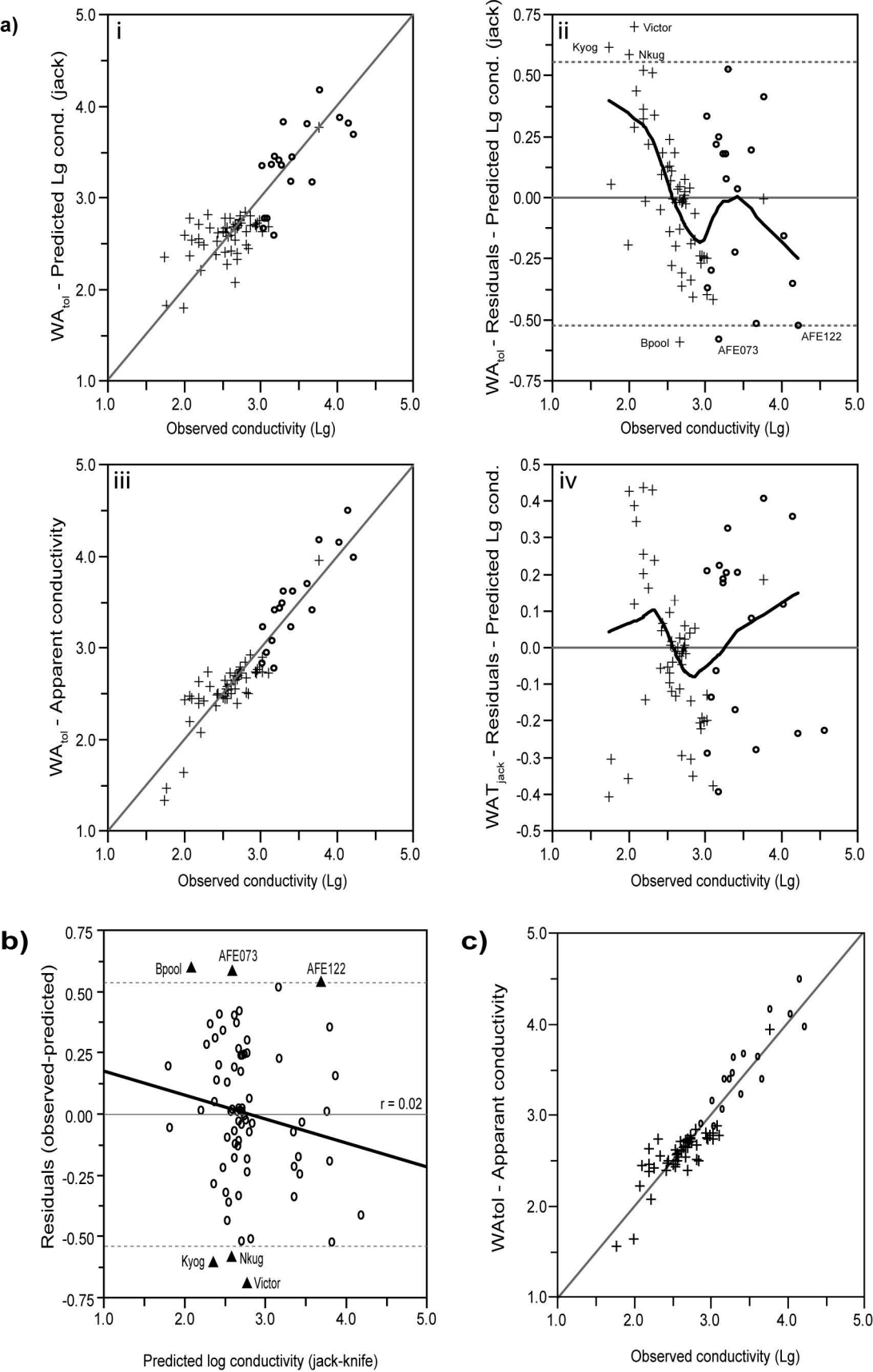

a)
8.0
-8.0
-1.0
1.0
PCA axis 2 (20.1%)
PCA axis 1 (35.8%)
TEMP_S
pH_SUR
SEASON
Lg_TOC
Lg_TN
Lg_K
Lg_Cl
Lg_TP
Lg_Stot
Lg_Na
Lg_DIC
Si
Lg_Mg
Lg_COND
Lg_F
Lg_Sr
Lg_Ca
Lg_Ba
SEMI
FULL
OPEN
NONE
CLOSE
SECCHI
SQ_DEPTH
SQ_ALT
b)
1.0
-1.0
CCA axis 2 (5.1%)
CCA axis 1 (6.5%)
SQ_DEPTH
Lg_COND
Lg_TN
ACMMAC
NI211A
NA745A
EP007A
CY004A
AC180A
FU002A
NA674A
NI043A
NI002A
SY001A
NA014A
NI014A
AN009A
CM024A
NI014AL
NI009A
CL001A
CY003A
GO017A
NA102A
NA134A
GO004A
AU002A
AU005A
FR002C
AU003A
NA173A

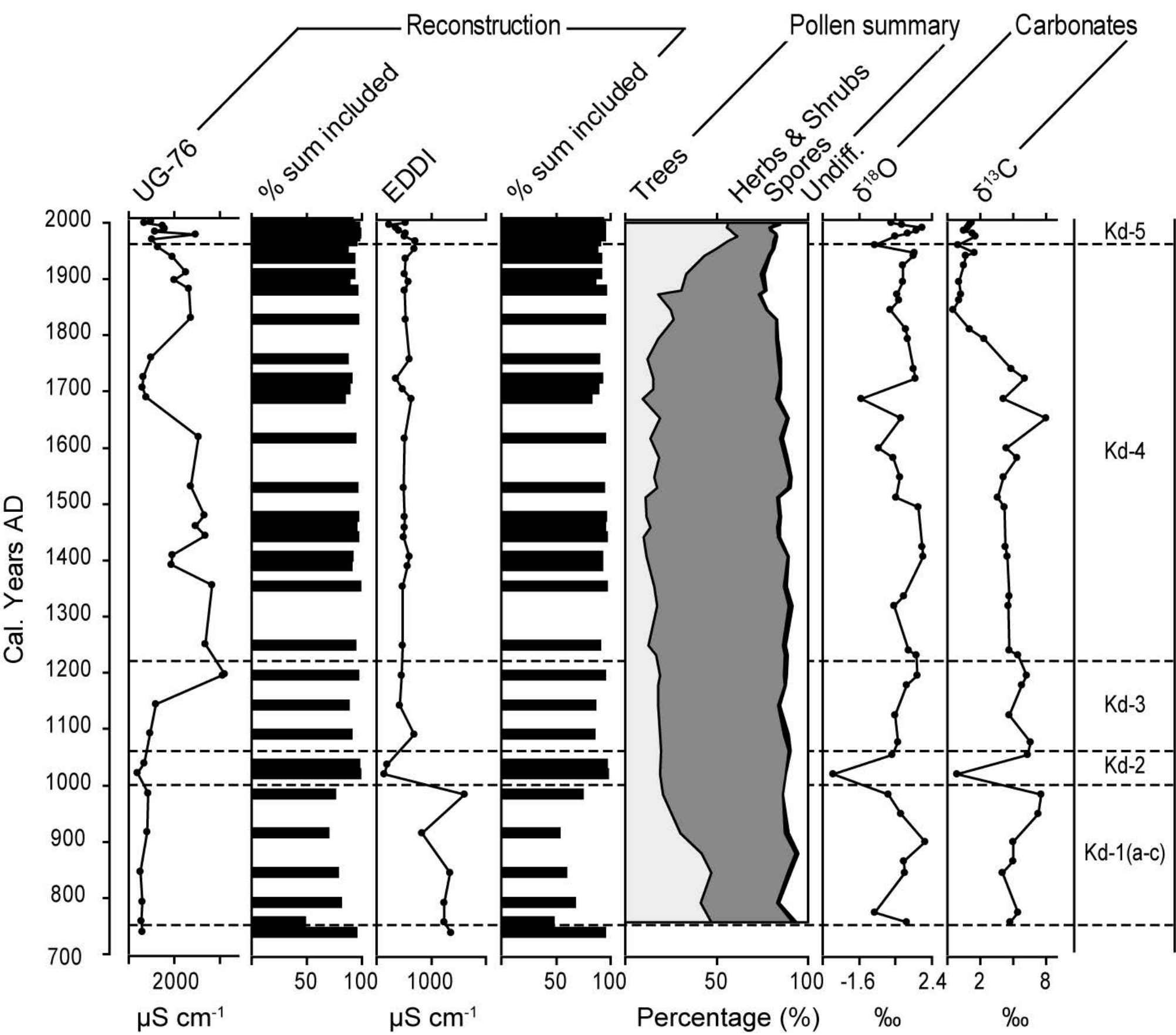
Total variance = 100%
(Total inertia = 4.957)

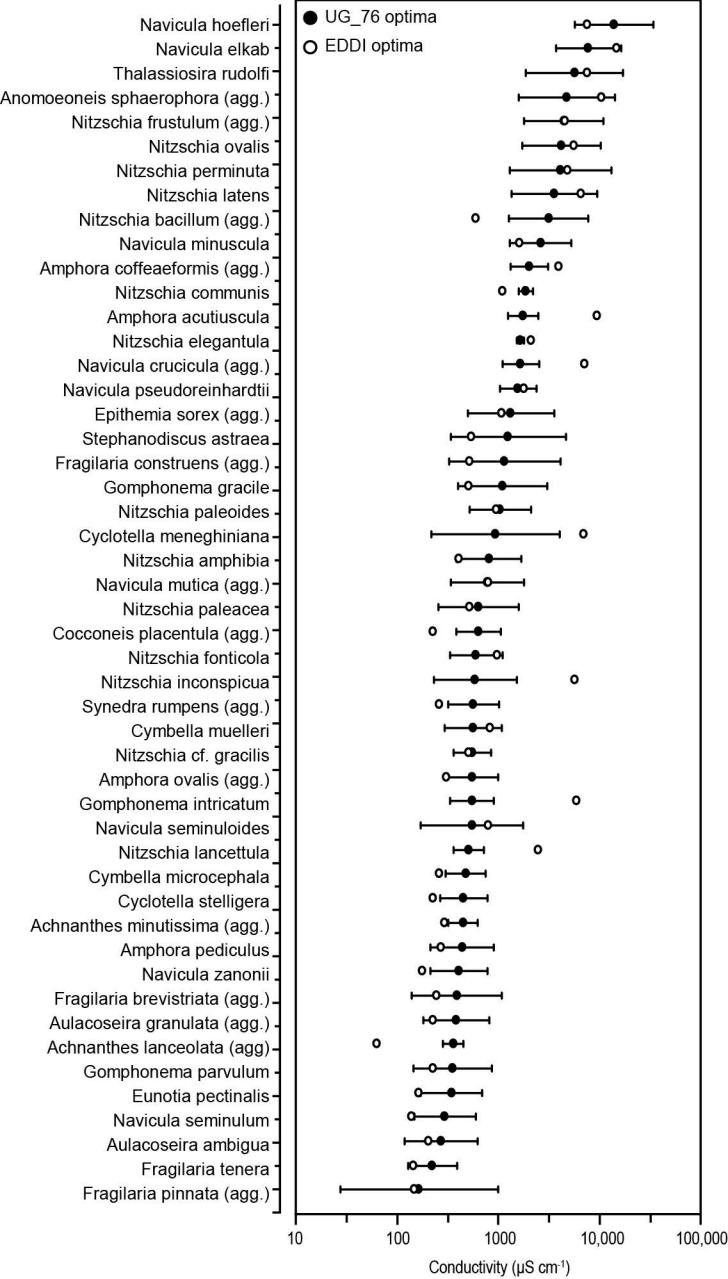
Explained = 15.8% (0.783)
Unexplained = 84.2% (4.174)



Conductivity = 7.46%
TN = 4.8%
Depth = 4.56%


a)
i
WA$_{tol}$ - Predicted Lg cond. (jack)
Observed conductivity (Lg)
ii
Victor
Kyog
Nkug
Bpool
AFE073
AFE122
WA$_{tol}$ - Residuals - Predicted Lg cond. (jack)
Observed conductivity (Lg)
iii
WA$_{tol}$ - Apparent conductivity
Observed conductivity (Lg)
iv
WAT$_{jack}$ - Residuals - Predicted Lg cond.
Observed conductivity (Lg)
b)
Bpool
AFE073
AFE122
r = 0.02
Kyog
Nkug
Victor
Residuals (observed-predicted)
Predicted log conductivity (jack-knife)
c)
WAtol - Apparant conductivity
Observed conductivity (Lg)


Reconstruction
Pollen summary
Carbonates
UG-76
% sum included
EDDI
% sum included
Trees
Herbs & Shrubs
Spores
Undiff.
δ18O
δ13C
Cal. Years AD
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
Kd-5
Kd-4
Kd-3
Kd-2
Kd-1(a-c)
2000
µS cm-1
50
100
1000
µS cm-1
50
100
50
100
Percentage (%)
-1.6
2.4
‰
2
8
‰


UG_76 optima
EDDI optima
Navicula hoefleri
Navicula elkab
Thalassiosira rudolfi
Anomoeoneis sphaerophora (agg.)
Nitzschia frustulum (agg.)
Nitzschia ovalis
Nitzschia perminuta
Nitzschia latens
Nitzschia bacillum (agg.)
Navicula minuscula
Amphora coffeaeformis (agg.)
Nitzschia communis
Amphora acutiuscula
Nitzschia elegantula
Navicula crucicula (agg.)
Navicula pseudoreinhardtii
Epithemia sorex (agg.)
Stephanodiscus astraea
Fragilaria construens (agg.)
Gomphonema gracile
Nitzschia paleoides
Cyclotella meneghiniana
Nitzschia amphibia
Navicula mutica (agg.)
Nitzschia paleacea
Cocconeis placentula (agg.)
Nitzschia fonticola
Nitzschia inconspicua
Synedra rumpens (agg.)
Cymbella muelleri
Nitzschia cf. gracilis
Amphora ovalis (agg.)
Gomphonema intricatum
Navicula seminuloides
Nitzschia lancettula
Cymbella microcephala
Cyclotella stelligera
Achnanthes minutissima (agg.)
Amphora pediculus
Navicula zanonii
Fragilaria brevistriata (agg.)
Aulacoseira granulata (agg.)
Achnanthes lanceolata (agg)
Gomphonema parvulum
Eunotia pectinalis
Navicula seminulum
Aulacoseira ambigua
Fragilaria tenera
Fragilaria pinnata (agg.)
10
100
1000
10,000
100,000
Conductivity (µS cm-1)

Supplementary Table 1 Summary of the main physical and chemical properties of the four crater lake clusters

		$Depth_{max}$ (m)	Secchi (cm)	pH	Temp °C	Cond. µS cm^{-1}	DO_{top} mg l^{-1}	TP µg l^{-1}	TN µg l^{-1}	Na mg l^{-1}	K mg l^{-1}	Ca mg l^{-1}	Mg mg l^{-1}	Cl mg l^{-1}	Si mg l^{-1}
Fort Portal	Ave	48.0	257	8.2	23.7	510	7.9	20.93	498.91	23.95	13.19	40.20	25.20	2.28	11.78
	Min	7.8	60	7.9	22.2	424	6.1	5.32	103.78	10.84	8.08	30.24	15.70	1.27	9.10
	Max.	72.0	660	8.5	24.7	662	10.8	36.54	894.04	39.05	17.57	58.00	36.40	3.90	13.15
	Median	56.1	154	8.2	23.9	478	7.4	20.93	498.91	22.95	13.55	36.28	24.34	1.98	12.43
Kasenda	Ave	42.9	188	8.7	24.1	539	7.6	58.94	430.31	21.99	17.86	32.75	31.99	9.02	14.84
	Min	4.5	31	8.4	18.7	216	6.2	4.06	86.10	1.80	2.90	12.60	7.10	1.42	6.27
	Max.	80.0	660	9.1	27.2	1222	9.4	359.69	835.44	123.98	55.70	61.60	79.40	50.95	22.20
	Median	38.0	135	8.7	24.3	418	7.7	25.03	443.66	9.60	10.76	29.70	26.17	4.83	15.30
Katwe-Kikorongo	Ave	4.7	43	10.0	30.4	86850	11.9	5039.38	5272.07	20045.85	5508.59	9.64	62.12	9374.78	14.00
	Min	0.01	0.02	9.3	26.9	22200	0.7	42.76	1821.39	384.39	172.79	7.70	5.70	70.87	4.97
	Max.	11.0	160	10.6	32.3	138100	31.8	13720.00	11914.00	40719.81	15473.95	15.00	134.00	37992.89	28.00
	Median	4.3	24	10.1	31.0	82150	3.2	1355.38	2080.83	15218.06	2374.07	8.88	50.24	2371.46	11.97
Bunyaruguru	Ave	31.7	166	8.6	26.2	1733	6.9	45.86	589.36	545.89	176.07	14.49	18.19	41.60	12.09
	Min	0.08	2	7.0	24.3	55	3.2	1.68	66.22	1.30	3.10	3.00	1.97	1.15	0.11
	Max.	141.0	770	10.2	29.9	31600	13.7	272.66	2003.01	11602.12	3326.86	27.60	41.30	720.87	36.76
	Median	23.1	135	8.7	25.9	297	6.2	19.36	334.70	5.84	13.19	12.65	17.39	3.35	11.95

Supplementary Table 2 Summary of the 19 environmental variables measured from the lakes in the diatom surface sediment training set, indicating units of measurement, summary statistics (average, minimum, maximum and median), data transformations (trans.) applied prior to numerical analyses (e.g. log or square root) and the abbreviated code

Type	Variable	Unit	Average	Max.	Min	Median	Trans.	Code
Major ion chemistry	Conductivity	µS cm^{-1}	9456	138100	55	464	Log(x)	Lg_COND
	pH		9.81	10.62	7.09	8.75	None	pH
	Sodium	mg l^{-1}	2301	40719.81	1.3	12.5	Log(x)	Lg_Na
	Potassium	mg l^{-1}	650	15473.95	2.9	14.9	Log(x)	Lg_K
	Calcium	mg l^{-1}	22	61.6	3	20.18	Log(x)	Lg_Ca
	Magnesium	mg l^{-1}	28.96	134	1.96	24.18	Log(x)	Lg_Mg
	Fluoride	mg l^{-1}	2.91	40.39	0.04	0.78	Log(x)	Lg_F
	Chloride	mg l^{-1}	979	37991.89	1.15	3.88	Log(x)	Lg_Cl
	Strontium	mg l^{-1}	1.18	6.56	0.06	0.52	Log(x)	Lg_Sr
	Barium	mg l^{-1}	0.39	5.85	0.004	0.08	Log(x)	Lg_Ba
	Total S	mg l^{-1}	276.85	12380	0.091	2.31	Log(x)	Lg_Stot
	Silica	mg l^{-1}	13.25	36.76	0.11	12.58	None	Si
Nutrient	Total P	µg l^{-1}	406.25	13720	1.68	24.61	Log(x)	Lg_TP
	Total N	µg l^{-1}	877.69	11914	66.22	417.2	Log(x)	Lg_TN
Physical	Altitude	m.a.s.l	1186	1568	903	1176	Square root	SQ_ALT
	Max. depth	m	32.05	141	0.01	19.8	Square root	SQ_DEPTH
	Secchi depth	cm	154.87	770	0.02	115	None	SECCHI
	Temperature	°C	25.82	32.3	18.71	25.82	None	TEMP_S
	DIC	mg l^{-1}	497	12221.68	4.89	47.5	Log(x)	Lg_DIC
	TOC	mg l^{-1}	38.92	668.87	0.56	5.48	Log(x)	Lg_TOC
	Catchment impact	Dummy[1]*	n/a	n/a	n/a	n/a	None	NONE / SEMI / FULL
	Lake status	Dummy[2]*	n/a	n/a	n/a	n/a	None	OPEN / SEASON / CLOSE

* Dummy variable – is used to denote the presence or absence of the variable using 0 or 1 codes.

[1] Catchment impact (based on contemporary lake setting): whether the lake catchment was subject to human activity. NONE (protected in forest reserves and national parks), SEMI (signs of human activity, e.g. clearance) and FULL (over 50% of the catchment impacted). Status assigned according to field observations and map data.

[2] Lake status: whether the lake system is OPEN (permanent inflow and outflow), SEASON (outflow following seasonal rains) or closed (CLOSE; no dry season outflow).

Supplementary Table 3 European Diatom Database (EDDI) sites added into the crater lake conductivity transfer function (* excluded from the UG-76b model)

EDDI Code	Site Name	Country	Site Type	Sample Type	Latitude	Longitude	Sampling date	Zm (m)	Cond. ($\mu S\ cm^{-1}$)
AFE004	Oudoudda	Dijibouti	Well	Periphytic	11° 34’ 0”	42° 4’ 0”	Nov-73	n/a	4060
AFE011	Awassa	Ethiopia	Lake	Plankton	6° 58’ 0”	38° 22’ 0”	Mar-80	21.6	1050
AFE012	Langano	Ethiopia	Lake	Plankton	7° 31’ 0”	38° 40’ 0”	Apr-80	47.9	1900
AFE013	Biete Mengest	Ethiopia	Lake	Plankton	8° 46’ 0”	38° 58’ 0”	Apr-80	38	2500
AFE017	Abiyata	Ethiopia	Lake	Bottom mud	7° 32’ 0”	38° 27’ 0”	Apr-80	14.20	10700
AFE021	Kilotes	Ethiopia	Lake	Plankton	8° 48’ 0”	39° 4’ 0”	Mar-80	6.4	5930
AFE035	Wadi Kalou	Dijibouti	River	Mud	11° 34’ 0”	42° 19’ 0”	Nov-75	0.1	1540
AFE036	Wadi Kalou	Dijibouti	River	Mud	11° 34’ 0”	42° 19’ 0”	Nov-75	0.1	1720
AFE037	Wadi Kalou	Dijibouti	River	Mud	11° 34’ 0”	42° 19’ 0”	Nov-75	0.1	1720
AFE055	Gadeb/Geradila	Ethiopia	Lake	Mud	7° 0’ 0”	39° 15’ 0”	Feb-77	0.2	2000
AFE061	Ol Bolossat	Kenya	Swampy lake	Littoral mud	0° 9’ 0”	36° 25’ 0”	Mar-79	0.5	1390
AFE073	Marairs ol Bolossat*	Kenya	Lake	Littoral mud	0° 9’ 0”	36° 25’ 0”	Dec-79	0.1	1500
AFE080	Turkana	Kenya	Lake	Plankton	2° 25’ 0”	35° 49’ 0”	Jan-79	120	2630
AFE090	Awassa	Ethiopia	Lake	Bottom mud	6° 58’ 0”	38° 22’ 0”	Mar-64	21.6	1050
AFE115	Kenke swamp	Tanzania	Lake	Bottom mud	-4° 40’ 0”	34° 40’ 0”	Jul-69	n/a	1200
AFE116	Kindai	Tanzania	Lake	Bottom mud	-4° 50’ 0”	34° 43’ 0”	Jul-69	n/a	4800
AFE122	Kikorongo	Uganda	Lake	Bottom mud	0° 0’ 0”	30° 01’ 0”	Nov-69	0.3	16300
AFE175	Gawani	Ethiopia	Swamp	Littoral mud	11° 40’ 0”	40° 40’ 0”	Nov-72	0.1	14300

Supplementary Table 4 Regression coefficients for the UG-76b transfer function

Coefficient	Inverse deshrinking	Classical deshrinking
WA_b0	-0.9587	1.0948
WA_b1	1.3396	0.6122
WATOL_b0	-0.6176	0.8043
WATOL_b1	1.2254	0.7097

*Classical deshrinking: y = b0+b1*x where,*
y = inferred conductivity; x = observed conductivity

Supplementary Table 5 Species list with assigned EDDI codes (the taxonomy follows that of EDDI). The total number of occurrences (N), effective number of occurrences (Hill's N2), maximum percent abundance *(Max)*, and optima and tolerances for WA and WA_{jack} transfer function for conductivity. Optima and tolerances for conductivity are in $\log_{10}$ (*x*) units where x is measured in µS cm^{-1}

EDDI Code	Species Name	Occurrences	Max %	N2	Optimum	Tolerance	$Optimum_{jack}$	$Tolerance_{jack}$
AC180A	*Achnanthes buccola*	4	2.50	2.75	2.58	0.34	2.58	0.34
AC166A	*Achnanthes chlidanos*	1	0.91	1.00	2.08	0.30	2.08	0.30
AC060A	*Achnanthes curtissima*	3	1.78	1.43	2.37	0.35	2.37	0.36
XXC949	*Achnanthes exigua (agg.)*	27	4.86	11.35	2.68	0.43	2.68	0.43
AC032A	*Achnanthes hungarica*	16	6.04	5.44	2.76	0.22	2.76	0.22
XXC995	*Achnanthes lanceolata (agg)*	7	31.03	2.28	2.56	0.10	2.55	0.10
XXC987	*Achnanthes minutissima (agg.)*	23	92.46	2.37	2.65	0.15	2.65	0.15
ACMMAC	*Achnanthes minutissima v. macrocephala*	1	1.92	1.00	2.80	0.30	2.80	0.30
ACMINSC	*Achnanthes minutissima v. scotica*	3	83.33	1.88	2.23	0.19	2.23	0.19
AC175A	*Achnanthes rupestoides*	1	2.02	1.00	2.33	0.30	2.33	0.30
AC161A	*Achnanthes ventralis*	5	39.33	1.22	2.54	0.09	2.54	0.09
AM002A	*Amphora acutiuscula*	3	10.23	1.84	3.25	0.15	3.25	0.17
XXC948	*Amphora coffaeaformis (agg.)*	3	38.00	2.73	3.31	0.18	3.31	0.19
AF_0315	*Amphora mexicana + v. major*	3	2.00	2.84	3.22	0.03	3.22	0.03
AM084A	*Amphora montana*	18	3.16	6.95	2.65	0.27	2.65	0.27
XXC914	*Amphora ovalis (agg.)*	47	16.86	17.45	2.74	0.26	2.74	0.26
AM012A	*Amphora pediculus*	42	18.27	15.29	2.65	0.31	2.65	0.31
AF_0375	*Amphora somalica*	2	1.40	1.28	3.94	0.53	3.93	0.52
AF_0351	*Amphora submontana*	1	1.57	1.00	2.19	0.30	2.19	0.30
AM110A	*Amphora tenerrima*	2	1.00	1.86	3.24	0.30	3.24	0.30
XXC988	*Amphora veneta (agg.)*	8	2.71	4.76	3.29	0.28	3.29	0.28
XXC947	*Anomoeoneis sphaerophora (agg.)*	21	13.74	5.39	3.68	0.47	3.68	0.47
AF_3501	*Aulacoseira agassizii + v. malayensis*	1	0.51	1.00	3.02	0.30	3.02	0.30
AU002A	*Aulacoseira ambigua*	17	62.98	6.45	2.44	0.36	2.44	0.36
XXC946	*Aulacoseira distans (agg.)*	3	1.17	2.58	2.51	0.53	2.51	0.53
XXC945	*Aulacoseira granulata (agg.)*	32	82.08	8.35	2.59	0.33	2.59	0.33
XXC944	*Aulacoseira nyassensis (agg.)*	1	2.01	1.00	3.28	0.30	3.28	0.30
XXC975	*Brachysira vitrea (agg.)*	2	2.55	1.90	3.02	0.30	3.02	0.30
XXC943	*Caloneis aequatorialis (agg.)*	3	5.13	1.47	3.21	0.30	3.21	0.31
XXC942	*Caloneis bacillum (agg.)*	13	1.99	7.30	2.78	0.40	2.78	0.40
XXC940	*Campylodiscus clypeus (agg.)*	2	3.97	1.20	3.72	0.34	3.73	0.33

EDDI Code	Species Name	Occurrences	Max	N2	Optimum	Tolerance	$Optima_{jack}$	$Tolerance_{jack}$
XXC976	*Cocconeis placentula (agg.)*	33	51.54	7.05	2.80	0.22	2.80	0.22
CO009A	*Cocconeis thumensis*	2	3.68	1.10	2.66	0.04	2.66	0.05
XXC999	*Craticula ambigua (agg.)*	14	9.92	3.44	3.05	0.87	3.05	0.87
CC003A	*Cyclostephanos tholiformis*	2	34.66	1.01	2.42	0.16	2.42	0.16
CY010A	*Cyclotella comensis*	4	7.09	1.54	2.71	0.15	2.71	0.15
XXC973	*Cyclotella distinguenda var. unipunctata (agg.)*	1	0.82	1.00	3.40	0.30	3.40	0.30
XXC974	*Cyclotella glomerata*	3	8.01	2.06	2.32	0.49	2.32	0.49
CY020A	*Cyclotella iris*	3	1.01	2.13	3.83	0.68	3.83	0.68
CY003A	*Cyclotella meneghiniana*	24	58.35	9.27	2.98	0.64	2.98	0.64
AF_1635	*Cyclotella stelligera*	6	47.15	1.39	2.66	0.23	2.66	0.23
CY045A	*Cyclotella stelligeroides*	2	18.06	1.19	2.69	0.06	2.69	0.07
CL001A	*Cymatopleura solea*	9	4.11	2.06	2.72	0.18	2.72	0.18
AF_1939	*Cymbella affinis var. afarensis*	3	6.00	2.82	3.22	0.03	3.22	0.03
CM070A	*Cymbella caespitosa*	2	0.78	1.47	2.87	0.00	2.87	0.01
CM006A	*Cymbella cistula*	7	9.58	2.64	2.62	0.10	2.62	0.10
CM027A	*Cymbella leptoceros*	1	0.73	1.00	2.71	0.30	2.71	0.30
CM004A	*Cymbella microcephala*	9	25.19	1.90	2.68	0.20	2.68	0.20
CM031A	*Cymbella minuta*	15	1.38	11.17	2.65	0.22	2.65	0.22
CM024A	*Cymbella muelleri*	27	11.84	11.20	2.75	0.28	2.75	0.28
CM013A	*Cymbella silesiaca*	10	1.17	6.20	2.57	0.45	2.57	0.45
CMSUBALP	*Cymbella subalpina*	9	26.63	1.60	2.59	0.11	2.59	0.11
CM002A	*Cymbella turgida*	1	1.02	1.00	3.02	0.30	3.02	0.30
CM110A	*Cymbella turgidula*	1	0.91	1.00	2.08	0.30	2.08	0.30
DP053A	*Diploneis pseudovalis*	3	2.08	2.80	2.55	0.12	2.55	0.12
DP061A	*Diploneis subovalis*	1	0.57	1.00	3.24	0.30	3.24	0.30
EP007A	*Epithemia adnata*	14	5.90	5.09	2.68	0.39	2.68	0.39
EP005A	*Epithemia hyndmanii*	1	0.99	1.00	3.30	0.30	3.30	0.30
XXC979	*Epithemia sorex (agg.)*	8	13.97	1.45	3.12	0.43	3.13	0.43
EU017A	*Eunotia flexuosa*	2	0.59	1.53	2.67	0.43	2.68	0.42
EU024A	*Eunotia glacialis*	2	2.56	1.29	2.56	0.09	2.56	0.09
EU110A	*Eunotia minor*	5	2.95	3.64	2.50	0.17	2.50	0.17
EU008A	*Eunotia monodon*	1	1.18	1.00	2.54	0.30	2.54	0.30
EU040A	*Eunotia paludosa*	6	1.73	2.98	2.64	0.35	2.64	0.35
XXC936	*Eunotia pectinalis*	1	25.76	1.00	2.54	0.30	2.54	0.30
FR026A	*Fragilaria bidens*	1	0.77	1.00	2.64	0.30	2.64	0.30

EDDI Code	Species Name	Occurrences	Max	N2	Optimum	Tolerance	$Optima_{jack}$	$Tolerance_{jack}$
XXC989	*Fragilaria brevistriata (agg.)*	9	38.41	2.64	2.59	0.45	2.59	0.45
AF_2604	*Fragilaria capucina*	9	4.54	3.30	2.63	0.34	2.63	0.34
FR009L	*Fragilaria capucina v. amphicephala*	2	1.71	1.22	2.54	0.21	2.54	0.21
FR009H	*Fragilaria capucina v. gracilis*	4	3.77	2.98	2.65	0.12	2.65	0.12
FR009G	*Fragilaria capucina v. rumpens*	2	1.15	1.30	2.59	0.30	2.58	0.30
XXC935	*Fragilaria construens (agg.)*	19	62.04	4.86	3.07	0.55	3.06	0.55
FR011A	*Fragilaria lapponica*	1	1.40	1.00	4.03	0.30	4.03	0.30
NAMICR	*Fragilaria microrhombus*	3	2.52	1.74	2.25	0.18	2.25	0.18
XXC934	*Fragilaria pinnata (agg.)*	12	35.49	1.43	2.22	0.78	2.23	0.78
FR060A	*Fragilaria tenera*	18	27.11	6.71	2.35	0.24	2.35	0.24
XXC917	*Fragilaria ulna (agg.)*	34	7.46	14.95	2.75	0.33	2.75	0.33
AF_2701	*Frustulia rhomboides*	5	0.96	3.14	2.76	0.17	2.76	0.17
AF_2803	*Gomphocymbella beccari*	7	1.36	4.04	2.86	0.22	2.86	0.22
GO020A	*Gomphonema affine*	25	45.33	5.17	2.84	0.19	2.84	0.19
GO019A	*Gomphonema augur*	10	6.96	3.81	2.89	0.24	2.89	0.24
XXC915	*Gomphonema clevei*	1	4.71	1.00	2.72	0.30	2.72	0.30
GOFGRA	*Gomphonema f. gracile*	2	69.78	1.97	1.86	0.16	1.86	0.17
GO004A	*Gomphonema gracile*	25	43.30	3.05	3.04	0.44	3.04	0.44
XXC996	*Gomphonema intricatum*	34	38.25	8.67	2.74	0.22	2.74	0.22
GO017A	*Gomphonema lanceolatum*	10	4.37	5.18	3.05	0.51	3.05	0.51
GO013A	*Gomphonema parvulum*	43	32.02	15.70	2.55	0.39	2.55	0.39
GO083A	*Gomphonema productum*	2	1.17	1.59	2.53	0.02	2.53	0.03
AF_2924	*Gomphonema spp. 1 (Kenya)*	1	0.84	1.00	3.14	0.30	3.14	0.30
AF_3001	*Gomphonitzschia ungeri*	1	2.91	1.00	3.02	0.30	3.02	0.30
XXC964	*Hantzschia amphioxys (agg.)*	27	7.16	10.61	2.92	0.47	2.92	0.47
XXC962	*Mastogloia elliptica (agg.)*	4	3.13	2.29	3.33	0.37	3.33	0.37
NACOARS	*Navicula (Coarse)*	3	2.31	1.69	2.68	0.30	2.68	0.30
NABARB	*Navicula barbarica*	5	2.87	3.06	2.66	0.20	2.66	0.20
NA045A	*Navicula bryophila*	3	0.79	1.97	2.33	0.29	2.33	0.29
NA066B	*Navicula capitata var. hungarica*	1	1.45	1.00	2.71	0.30	2.71	0.30
NA745A	*Navicula capitoradiata*	2	0.73	1.82	2.24	0.35	2.25	0.34
AF_3626	*Navicula cf. fluminicata*	1	0.84	1.00	3.14	0.30	3.14	0.30
XXC959	*Navicula cincta (agg.)*	4	1.63	2.39	3.25	0.27	3.25	0.27
NA050A	*Navicula clementis*	2	0.78	1.56	2.49	0.14	2.49	0.15
NA118A	*Navicula confervacea*	3	2.36	1.32	3.21	0.54	3.21	0.54

EDDI Code	Species Name	Occurrences	Max	N2	Optimum	Tolerance	$Optima_{jack}$	$Tolerance_{jack}$
NA046A	*Navicula contenta*	24	5.96	9.85	2.68	0.35	2.68	0.35
XXC930	*Navicula crucicula (agg.)*	2	10.87	1.51	3.22	0.18	3.22	0.19
AF_3613	*Navicula cryptocephala*	8	13.89	3.53	2.41	0.21	2.41	0.21
NA007D	*Navicula cryptocephala v. veneta*	2	1.00	1.86	3.24	0.30	3.24	0.30
NA751A	*Navicula cryptonella*	27	7.41	13.17	2.58	0.22	2.58	0.22
NA317A	*Navicula descussis*	2	3.02	1.49	3.31	0.10	3.31	0.11
NA730A	*Navicula elkab*	11	40.48	2.84	3.89	0.32	3.89	0.32
NA173A	*Navicula erifuga*	1	0.58	1.00	2.31	0.30	2.31	0.30
AF_3629	*Navicula gastrum*	7	1.34	4.99	2.63	0.11	2.63	0.11
AF_3640	*Navicula gawaniensis*	2	8.70	1.14	4.15	0.09	4.15	0.10
NA776A	*Navicula gregaria*	1	0.58	1.00	2.53	0.30	2.53	0.30
NA415A	*Navicula harderii*	9	0.79	6.76	2.67	0.31	2.67	0.31
NA167A	*Navicula hoefleri*	2	12.45	1.03	4.15	0.39	4.14	0.38
AF_3671	*Navicula iranensis*	3	2.00	2.77	3.22	0.03	3.22	0.03
NA102A	*Navicula laevissima*	1	0.54	1.00	1.76	0.30	1.76	0.30
NA122A	*Navicula miniscula*	1	20.20	1.00	3.42	0.30	3.42	0.30
XXC998	*Navicula monoculata*	12	1.17	7.79	2.46	0.28	2.46	0.28
NA082A	*Navicula muralis*	1	0.51	1.00	3.14	0.30	3.14	0.30
XXC928	*Navicula mutica (agg.)*	38	18.99	9.54	2.89	0.36	2.89	0.36
AF_3716	*Navicula nolens*	1	0.84	1.00	3.14	0.30	3.14	0.30
AF_3717	*Navicula nyassensis*	2	1.36	1.88	3.16	0.27	3.16	0.27
NA058A	*Navicula phyllepta*	1	0.95	1.00	2.51	0.30	2.51	0.30
NA127A	*Navicula pseudoreinhardtii*	3	13.88	1.78	3.20	0.18	3.20	0.18
NA590A	*Navicula pseudoventralis*	1	4.19	1.00	2.08	0.30	2.08	0.30
XXC957	*Navicula pupula (agg.)*	37	4.90	17.03	2.85	0.49	2.85	0.49
NA003A	*Navicula radiosa*	11	7.22	4.06	2.56	0.14	2.56	0.14
NA008A	*Navicula rhyncocephala*	2	7.91	1.58	4.02	0.39	4.02	0.38
XXC916	*Navicula salinicola*	1	0.20	1.00	2.19	0.30	2.19	0.30
NA110A	*Navicula schadei*	1	0.55	1.00	2.08	0.30	2.08	0.30
NA128A	*Navicula schoenfeldii*	1	0.73	1.00	2.08	0.30	2.08	0.30
NA129A	*Navicula seminuloides*	17	11.05	4.57	2.74	0.51	2.74	0.51
NA005A	*Navicula seminulum*	9	13.96	1.76	2.48	0.31	2.48	0.31
NA134A	*Navicula subminuscula*	1	0.72	1.00	1.76	0.30	1.76	0.30
NA166A	*Navicula submuralis*	1	25.47	1.00	2.44	0.30	2.44	0.30
NA114A	*Navicula subrotundata*	1	0.73	1.00	2.71	0.30	2.71	0.30

EDDI Code	Species Name	Occurrences	Max	N2	Optimum	Tolerance	$Optima_{jack}$	$Tolerance_{jack}$
NA674A	*Navicula tenella*	4	1.93	3.40	2.87	0.11	2.87	0.12
NA131A	*Navicula twymania*	1	1.85	1.00	3.14	0.30	3.14	0.30
NA054A	*Navicula veneta*	3	1.98	1.86	3.73	1.00	3.72	0.99
NA017A	*Navicula ventralis*	1	0.96	1.00	2.58	0.30	2.58	0.30
XXC956	*Navicula viridula v. rostellata*	1	1.75	1.00	2.53	0.30	2.53	0.30
AF_3762	*Navicula zanonii*	14	11.03	7.21	2.62	0.28	2.62	0.28
NE030A	*Neidium minutissimum*	2	0.78	1.81	2.50	0.07	2.50	0.08
AF_3904	*Nitzschia adapta*	1	7.82	1.00	3.02	0.30	3.02	0.30
NIAEQU	*Nitzschia aequalis*	3	4.45	2.54	2.72	0.37	2.72	0.37
NI202A	*Nitzschia alpina*	2	1.71	1.86	2.52	0.02	2.52	0.03
NI014A	*Nitzschia amphibia*	38	24.07	12.77	2.91	0.32	2.91	0.32
NI014AL	*Nitzschia amphibia (L)*	17	3.52	9.32	2.68	0.25	2.68	0.25
XXC990	*Nitzschia bacata*	16	9.18	5.69	2.56	0.20	2.56	0.20
XXC955	*Nitzschia bacillum (agg.)*	9	66.86	2.37	3.50	0.39	3.50	0.39
NI082A	*Nitzschia cf. gracilis*	25	69.73	6.97	2.75	0.18	2.75	0.18
NI010A	*Nitzschia communis*	3	16.30	1.18	3.27	0.07	3.27	0.07
NI200A	*Nitzschia compressa*	3	0.78	2.00	3.67	0.32	3.67	0.32
NI204A	*Nitzschia elegantula*	4	14.00	2.96	3.22	0.04	3.22	0.04
AF_3918	*Nitzschia epiphytica*	1	1.96	1.00	3.42	0.30	3.42	0.30
AF_3919	*Nitzschia epiphyticoides*	1	4.59	1.00	3.02	0.30	3.02	0.30
NI002A	*Nitzschia fonticola*	33	49.81	11.20	2.78	0.26	2.78	0.26
XXC913	*Nitzschia frustulum (agg.)*	21	68.70	4.16	3.65	0.39	3.65	0.39
AF_3936	*Nitzschia goetzeana*	2	2.01	1.72	3.86	0.84	3.85	0.83
NI201A	*Nitzschia graciliformis*	7	3.41	2.66	2.52	0.19	2.52	0.19
NI007A	*Nitzschia hungarica*	3	7.65	2.27	3.70	0.24	3.70	0.24
NI043A	*Nitzschia inconspicua*	40	18.15	11.45	2.77	0.41	2.77	0.41
NI044A	*Nitzschia intermedia*	6	1.65	4.56	2.57	0.07	2.57	0.07
XXC929	*Nitzschia lacuum*	4	18.18	1.64	2.41	0.16	2.41	0.16
AF_4003	*Nitzschia lancettula*	26	58.44	8.61	2.71	0.15	2.71	0.15
AF_4004	*Nitzschia latens*	9	24.90	2.87	3.55	0.42	3.55	0.42
NI203A	*Nitzschia leibetruthii*	18	14.17	6.60	2.68	0.30	2.68	0.30
XXC928A	*Nitzschia linearis*	7	4.13	3.48	2.74	0.33	2.74	0.33
AF_4008	*Nitzschia mediocris*	1	6.63	1.00	3.02	0.30	3.02	0.30
NI027A	*Nitzschia microcephala*	1	2.17	1.00	4.16	0.30	4.16	0.30
NI036A	*Nitzschia obtusa*	2	7.00	1.83	3.24	0.30	3.24	0.30

EDDI Code	Species Name	Occurrences	Max	N2	Optimum	Tolerance	$Optima_{jack}$	$Tolerance_{jack}$
NI045A	*Nitzschia ovalis*	2	10.59	1.07	3.63	0.39	3.64	0.38
XXC926	*Nitzschia paleacea*	45	37.45	14.78	2.81	0.40	2.81	0.40
AF_4049	*Nitzschia paleoides*	1	24.25	1.00	3.02	0.30	3.02	0.30
NI193A	*Nitzschia perminuta*	4	17.06	1.16	3.62	0.50	3.62	0.51
XXC911	*Nitzschia pusilla (agg.)*	4	7.85	1.95	3.27	0.46	3.27	0.46
NI006A	*Nitzschia sigma*	5	2.00	3.26	3.54	0.50	3.54	0.50
NI046A	*Nitzschia sigmoidea*	2	0.38	1.82	2.43	0.26	2.43	0.26
NISUBRO	*Nitzschia subrostrata*	1	9.34	1.00	3.02	0.30	3.02	0.30
NI047A	*Nitzschia tropica*	2	1.01	1.38	4.11	0.43	4.11	0.42
XXC3518	*Orthoseira roeseana*	8	1.19	4.83	2.70	0.43	2.70	0.43
AF_4203	*Pinnularia borealis + varieties*	9	2.23	4.29	2.92	0.43	2.92	0.43
PI010A	*Pinnularia fasciata*	1	0.99	1.00	3.30	0.30	3.30	0.30
XXC954	*Pinnularia gibba (agg.)*	1	0.17	1.00	3.14	0.30	3.14	0.30
PI022A	*Pinnularia subcapitata*	1	1.12	1.00	3.08	0.30	3.08	0.30
XXC953	*Rhopalodia gibba (agg.)*	9	6.70	1.86	2.98	0.40	2.98	0.40
XXC952	*Rhopalodia gibberula (agg.)*	11	7.39	4.03	3.39	0.47	3.39	0.47
AF_4608	*Rhopalodia gracilis*	4	3.97	1.96	2.66	0.15	2.66	0.15
RH002A	*Rhopalodia musculus*	2	0.51	1.69	3.18	0.10	3.18	0.10
RHRHOP	*Rhopalodia rhopala*	9	3.13	4.37	2.57	0.28	2.57	0.28
SA006A	*Stauroneis phoenicenteron*	1	1.12	1.00	3.08	0.30	3.08	0.30
AF_5001	*Stephanodiscus astraea + v. minutula*	10	21.65	1.63	3.10	0.57	3.10	0.57
SU004A	*Surirella biseriata*	2	2.75	1.12	3.40	0.28	3.39	0.28
AF_5310	*Surirella engleri*	2	0.98	1.36	2.80	0.61	2.80	0.60
XXC992	*Synedra acus*	4	2.23	2.27	2.75	0.54	2.75	0.54
XXC983	*Synedra rumpens (agg.)*	10	10.13	3.40	2.76	0.25	2.76	0.25
AF_5701	*Thalassiosira faurii*	4	4.43	1.95	3.57	0.62	3.57	0.62
TH003A	*Thalassiosira rudolfi*	11	65.83	3.37	3.76	0.48	3.76	0.48