The use of multivariate statistics to resolve multiple contamination signals in the oxygen isotope analysis of biogenic silica

Katy E. Wilson[1*], Melanie J. Leng[2,3], Anson W. Mackay[4]

[1] Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK.

[2] Centre for Environmental Geochemistry, School of Geography, University of Nottingham, Nottingham NG7 2RD, UK.

[3] NERC Isotope Geosciences Facilities, British Geological Survey, Keyworth, Nottingham NG12 5GG, UK.

[4] Environmental Change Research Centre, Department of Geography, University College London, Gower Street, London WC1E 6BT, UK.

*To whom correspondence should be addressed: k.e.wilson@ucl.ac.uk

Abstract

The analysis of the oxygen isotope composition ($\delta^{18}O$) of diatom silica is a commonly-used tool for palaeoclimate reconstruction that recent studies have demonstrated may be complicated by the presence of non-diatom detrital material. Such contamination can mask any true climate-driven signal, leading to spurious results. Analysis of the 2.6 million year old Barsemoi Diatomites from the East African Rift Valley highlights the presence of both tephra and clay in purified samples. Here we present a new method for assessing the relative contribution and geochemical composition of contamination components where sedimentary samples may be affected by more than one type of contamination. This approach shows that the incorporation of analytical techniques such as x-ray fluorescence spectrometry, coupled with statistical modelling, can be used to develop a three end-member model to successfully resolve climate-driven changes in $\delta^{18}O_{diatom}$. Mass-balance corrections made to $\delta^{18}O_{diatom}$ data demonstrate the importance of adopting quantitative geochemical analysis in tandem with the $\delta^{18}O$ analysis of biogenic silica, in order to obtain accurate and meaningful results for palaeoclimate reconstruction.

Keywords: oxygen isotopes, diatom, contamination, multivariate statistics

Introduction

The oxygen isotope analysis of biogenic silica, principally diatoms ($\delta^{18}O_{diatom}$), has become an increasingly popular proxy for palaeoclimate change (Lamb *et al*., 2005, 2007; Leng *et al*., 2005; Morley *et al*., 2005; Moschen *et al*., 2005; Leng and Barker, 2006; Swann *et al*., 2006; Swann and Leng, 2009; Barker *et al*., 2011; Mackay *et al*., 2011, 2013; Rosqvist *et al*., 2013). Diatoms are unicellular, algae that precipitate siliceous cell walls, which are preserved as rigid frustules within the sediment record after cell death (Round *et al*., 1990). They are abundant in areas with limited carbonate sedimentation, such as in soft-water lakes and the high-latitude oceans, and thus can be utilised as a reliable palaeoenvironmental proxy to complement existing palaeoclimatic records. They are ubiquitous in the photic zones of most aquatic environments (including lakes) where levels of key nutrients such as silicon, nitrogen and phosphorus are sufficient to sustain productivity (Leng and Barker, 2006).

Productivity is largely controlled by seasonal climate patterns that influence habitat conditions and nutrient availability. In large, monomictic tropical lakes, increased mixing and productivity tend to occur during the dry season (Bootsma, 1993). In dimictic, temperate lakes such as those in the mid- to high-latitudes, mixing occurs twice each year, during spring and autumn, when similarities between the temperature and density of the hypolimnion and epilimnion create a strong mixing regime (Wetzel, 2001). The isotopic signature of diatom silica is largely acquired during these growth periods (Leng and Barker, 2006). Within lacustrine environments, $\delta^{18}O_{diatom}$ varies as a function of temperature and the isotopic composition of ambient lake water ($\delta^{18}O_{water}$), which in turn is heavily influenced by some aspect of precipitation (open lakes) or balance between precipitation and evaporation (closed lakes) (Leng and Barker, 2006). A possible additional control is exerted by depth constraints, such as vertical stratification, that limit diatom productivity to the upper part of the water column (the photic zone) and can result in a $\delta^{18}O_{diatom}$ value that is reflective of a localised $\delta^{18}O_{water}$ signal (Raubitschek *et al.*, 1999).

The analysis of $\delta^{18}O_{diatom}$ offers the potential to obtain palaeoenvironmental records, which contain a mineral-water fractionation that is dependent on temperature. Various calibration studies have attempted to define the empirical relationship between $\delta^{18}O_{diatom}$ and temperature; the diatom-temperature coefficient is thought to be –0.2‰/°C but estimates have ranged up to –0.5‰/°C (Labeyrie 1974; Juillet-Leclerc and Labeyrie, 1987; Shemesh *et al.*, 1992; Brandriss *et al.*, 1998; Moschen *et al.*, 2005). Where there is little seasonal variation in temperature (i.e. at low-latitudes) or in open lake systems, changes in $\delta^{18}O_{water}$, the isotopic composition of precipitation and possible changes in atmospheric circulation, become more important. For example, changes in the hydrological balance between precipitation and evaporation have also been invoked as the cause for variations in a $\delta^{18}O_{diatom}$ record from Lake Challa near Kilimanjaro (Barker *et al.*, 2011), while changes in

$\delta^{18}O_{diatom}$ from Laguna Zacapu in Mexico are thought to reflect variations in the $\delta^{18}O$ composition of precipitation driven by salinity, temperature or air-mass and moisture contribution balance between the Pacific Ocean and Gulf of Mexico (Leng *et al.*, 2005).

In some sediments, purification using chemical and physical cleaning steps is sufficient to remove non-diatom material from samples prior to isotope analysis (Rosqvist *et al.*, 2013). However it has become apparent that the precision of $\delta^{18}O_{diatom}$ data can be compromised by the presence of small amounts of tephras, clays and carbonates which can sometimes remain within the purified samples because of difficulties with the process due to similarities of size, specific gravity or chemistry between the diatom and contaminant (Morley *et al.*, 2005; Lamb et al. 2007; Brewer *et al.,* 2008). Because oxygen is liberated from all components of the sample during the analytical procedure, even small proportions of contamination can have a significant effect, causing negative excursions and high-frequency noise in the $\delta^{18}O_{diatom}$ record, masking any true climate signal (Morley *et al.*, 2005; Lamb *et al.,* 2007; Brewer *et al.,* 2008). Silicate minerals tend to have low $\delta^{18}O$, for example silt contained within Lake Baikal sediments has a measured $\delta^{18}O$ value of +12.3‰ (Morley *et al.*, 2005) while tephra remaining in samples from Lake Tilo, Ethiopia had an average $\delta^{18}O$ value of +11.6 ‰ (Lamb *et al.*, 2005).

The application of whole-sample geochemistry to analyse contamination remaining within purified diatom samples was first described by Lamb *et al.* (2007) and later expanded by Brewer *et al.* (2008), Swann and Leng (2009), Mackay *et al.* (2011; 2013) and Chapligin *et al.* (2012). By adopting a chemical-based technique to investigate contamination, it is possible to quantify the type and volume of contaminant material affecting cleaned diatom samples, and thus develop a way of removing its effect on measured $\delta^{18}O_{diatom}$ using mass-balance calculations. With the exception of the FTIR method (Swann and Patwardhan, 2011), the on-going development of techniques used to assess purity has largely concerned the use of x-ray fluorescence (XRF) spectrometry to quantify variations in geochemical composition. To date, these have been applied in the analysis of lacustrine systems containing one non-diatom component (e.g. clay, tephra or carbonates) where the amount of Al_2O_3 is used to quantify the amount of remaining contamination (Brewer *et al*., 2008; Mackay *et al*., 2011, 2013). However, in systems, which may contain two or more types of contamination, more precise assessment of sample geochemistry is required in order to differentiate between components.

Here we analyse a 2.6 million year old (Ma) diatomite sequence from the East African Rift Valley that contains numerous air-fall ash deposits from volcanic activity (Deino *et al*., 2006) and was deposited in a system known to experience high clay and silt influx from the catchment (Tarits *et al*., 2006). We use multivariate statistics to identify the geochemical signatures of the different contamination components and develop a three end-member model based on elemental oxide abundance data to accurately model climate-driven changes in $\delta^{18}O_{diatom}$.

The Barsemoi Diatomites

The Barsemoi Diatomites are a well-dated sequence exposed within the Tugen Hills in the Baringo-Bogoria basin (Fig. 1). The Tugen Hills is a complex, westward-tilting fault block which extends for 75 km between the Kerio Valley and the Baringo-Suguta axial trough and is uplifted along the N-S trending synthetic Saimo fault. The fault block represents a 3,000 m thick sedimentary succession spanning the period between 14 - 1 Ma that was deposited in a down-warped half-graben that has served as a depositional basin since the initiation of rift activity in the region at 16 Ma (Chapman *et al*., 1978; Morley *et al*., 1992). The Barsemoi Diatomites record the rhythmic cycling of a major freshwater lake system in the Baringo-Bogoria basin between 2.55 - 2.68 Ma and offer a unique, high-resolution archive of Late Pliocene climate history (Deino *et al.,* 2006; Kingston *et al.*, 2007).

Modern-day Lake Baringo is situated close to the Tugen Hills in the axial graben of the Central Kenyan Rift (0.33-1° N, 36.08° E) at an altitude of 970 m.a.s.l. (Fig. 1). Depending on the strength and duration of the rainy seasons, the surface area varies between 108-160 km^2 and the lake drains a catchment encompassing a total area of 6,200 km^2 (Tarits *et al*., 2006). The region is semi-arid, with mean annual rainfall rates that range from 600–900 mm on the valley floor to >1000 mm in the adjacent highlands. Potential evaporation in the area is in excess of 2,600 mm/yr, so the survival of the lake is dependent on riverine inflow from two perennial rivers, the Molo and the Perkerra, and a number of ephemeral channels active only during the rainy seasons. Despite high evaporation rates and having no surface outflow, Lake Baringo remains fresh and the overall salinity of the lake is largely the same as suggested by the earliest analyses conducted in 1929-1930 (salinity of 0.5-0.7 ‰) (Ballot *et al*., 2003). Tarits *et al*. (2006) suggest that this is the result of subsurface groundwater seepage through faulted lavas and permeable sediments. In the past, Lake Baringo is believed to have ranged from a highly alkaline and saline playa-lake during different low-level stages in its history, as marked by the presence of authigenic zeolites formed from $NaCO_3$-rich lake and pore waters (Renaut *et al*., 1999), to an extensive freshwater lake (Kingston *et al*., 2007).

Methodology

Samples from one diatomite (unit #4; 2.606-2.617 Ma) were taken at 10 cm intervals from locality RE26, exposed within the main A-A' type-section of the principal Barsemoi Diatomite sequence (Fig. 1A). The volume of material analysed approximately equates to 30 years per sample (2 cm sample size), based on published sedimentation rates (Deino *et al.*, 2006). In order to remove impurities, diatoms were concentrated using physical and chemical techniques following a modified version of the method outlined by Morley *et al.* (2004). Diatom samples were first soaked in deionised water and freeze-dried in order to aid disaggregation. Organic matter was removed by heating with 30% H_2O_2 at 90°C for 3 hours. Samples were then treated with 5% HCl for 12 hours in order to remove calcium carbonate. After each stage, samples were rinsed with deionised water and centrifuged (1200 rpm for 4 minutes) three times. Diatomite material was then sieved with deionised water at 10 µm and 75 µm in order to optimise the retention of diatom valves and remove both small clay particles and larger silt-sized detrital mineral grains or large diatoms. Using a heavy liquid separation method to isolate materials of different densities, the 10-75 µm fraction was then added to sodium polytungstate (SPT, $3Na_2WO_49WO_3.H_2O$) with a specific gravity of 2.1 sg and continuously centrifuged at 2500 rpm for 20 minutes. The diatom layer was then extracted by pipette and SPT was subsequently removed from the samples using a combination of repeated centrifuge washing with deionised water and a final sieving stage at 10 µm. Following SPT removal, cleaned samples were mixed with deionised water and allowed to settle. Any remaining clay formed a very fine dark band, which was carefully removed using a pipette. The remaining purified diatom samples were dried at 40 °C for 48 hours. Cleaned samples were analysed for $\delta^{18}O_{diatom}$ using stepwise fluorination (Leng and Sloane, 2008) at the NERC Isotope Geosciences Facility in Keyworth.

XRF spectrometry was used to measure the whole-sample geochemistry of the cleaned material. Samples were analysed using PANalytical Axios Advanced XRF spectrometers at the Department of Geology, University of Leicester and the British Geological Survey, Keyworth. Fused glass beads were prepared from approximately 0.1 g of ignited diatom silica powder which had been dried overnight at 105°C to remove moisture.

Powders were mixed with a flux consisting of 80% Li-metaborate and 20% Li-tetraborate at a sample to flux ratio of 1:5 in Pt-Au crucibles which were heated and homogenised at ~1050°C on an oxygen/gas burner system. The resulting melt was cast in a Pt-Au dish to form the fusion beads before cooling. Major element geochemistry was analysed from 32 mm diameter briquettes prepared from 10 g of fine ground powder mixed with ~ 20-25 drops of 7% PVA solution and pressed at 10 tons per square inch.

End-member samples of pure diatom and non-diatom material were also analysed in order to best quantify the isotopic and geochemical signatures of both pure diatom material and potential contaminants (Fig. 2). BFC, is an NIGF within-laboratory pure diatom standard derived from a lacustrine diatomite deposit in California, while #TUFF and #TUFF2, are tephra end-members from within Barsemoi diatomite unit #4. #TUFF was sampled from a 7 cm-thick green-grey ash fall deposit situated approximately 165 cm above the base of unit #4 (Fig. 1C, Fig. 2D,E), while #TUFF2 was taken from a tephra-rich layer at the top of unit #4. Clay particles have been observed to adhere to diatom valves by electrostatic charge (Fig. 2C) and are difficult to extract from within the frustule structure making it difficult to isolate a sample of pure clay material for analysis. Therefore, an additional sample of clay (#CLAY) was identified and analysed using scanning electron microscopy (SEM) and energy dispersive system (EDS) microprobe spot analysis to determine its geochemical composition.

To date, the processes adopted for determining the relative quantity of contamination has been relatively simple as most samples apparently contained a single type of contaminant (e.g. clay or tephra from a single source). Previous work has involved a range of techniques from qualitative methods such as point-counting of silt grains to generate a simple linear mass-balance correction (Morley *et al*., 2005) to the more quantitative geochemical assessment of potential contamination using major and minor trace element geochemistry (XRF) (Lamb *et al*., 2007, Brewer *et al*., 2008, Mackay *et al*., 2011; 2013) or infrared spectroscopy (Swann and Patwardhan, 2011). The XRF techniques have largely lead to a quantification of contamination derived from one or more ‘indicator’ oxides such as Al_2O_3 for clay or silt (Brewer *et al.*, 2008; Mackay *et al.*, 2011). However, sedimentary sections such as the Barsemoi diatomites pose a different challenge due to the occurrence of both clay (Fig. 2A,B) and tephra (Fig. 2D,E). Geochemical similarities between the two components, such as high Fe_2O_3 (Table 1), mean that it was not possible to identify just one elemental oxide that is individually indicative of clay or tephra. Therefore, in order to accurately assess the affect of these different contaminants on $\delta^{18}O_{diatom}$ values, we adopted multivariate statistical analysis to determine what drives variation within our geochemical (XRF) dataset by identifying which oxides can, when considered together, be indicative of clay or tephra contamination. Principal Components Analysis (PCA) was used to explore variation in the chemical composition of the purified samples and thus to establish relationships between different elemental components. PCA was focussed on inter-species correlations and data were centred and standardised in order to calculate a correlation matrix for the data. Statistical analyses were conducted using Canoco ver. 4.5 for Windows and ordination biplots were produced using the associated program CanoDraw (ter Braak and Šmilauer, 2002).

Results and Discussion

Samples covering the whole section (n = 49) were analysed for $\delta^{18}O_{diatom}$ composition. Measured raw values of $\delta^{18}O$ ($\delta^{18}O_{sample}$) vary between +16 and +37‰ and display a marked overall decrease towards the top of unit #4 (Fig. 3). Within-run reproducibility of the diatomite samples averaged 0.34‰ (n=4), respectively, whilst reproducibility of the BFC diatomite standard was 0.22‰ (n=8). The isotopic compositions of diatom and non-diatom end-member components are given in Table 1.

Whole-sample geochemical data for samples from diatomite #4 are shown in figure 3, expressed as weight percentages of major element oxides. Samples from the upper part of the section contain greater proportions of elemental oxides, broadly indicating higher levels of contamination than those towards the base of unit #4. This is in agreement with stratigraphic and sedimentological observations of a gradual transition towards more clayey diatomites near the top of the unit. Diatom samples with the highest proportion of SiO_2 are considered to be the least contaminated and generally occur within the lower half of the sequence. Varying amounts of different elemental oxides are also found to occur naturally within diatom frustules (Brewer *et al*., 2008) and therefore sample concentration values were first normalised to those of the BFC diatomite standard. The geochemical composition of the BFC standard diatomite is taken from Brewer *et al*., (2008).

The marked differences in the geochemistry of the end-member tephra and clay contaminants and the cleanest diatom samples provide a means of determining the level and type of contamination present within samples. Following XRF analysis, samples #4074 (section height = 200 cm) and #4099 (section height = 460 cm) were eliminated from further analysis as it was not possible to obtain sufficiently reliable data (sample #4074: low sample weight; sample #4099: laboratory analysis error). In order to account for any inter-laboratory difference in XRF analyses, we performed dual measurements of the BFC standard: A t-test demonstrates that there is no statistical difference between the two institutions (University of Leicester and British Geological Survey) at the 5% significance level (t = 0.0058; 5% level = 2.1199).

PCA results are summarised in figure 4, which shows the distribution of the cleaned diatomite samples with respect to their concentrations of the various elemental oxides. The results indicated that 85.1% of the variance within the data can be explained by axes 1 and 2 which represent the environmental gradients of the measured elemental oxides (eigenvalues (λ): $\lambda_1 = 0.705$; $\lambda_2 = 0.146$). Figure 4 shows that the oxides cluster in two distinct groups indicating that contamination within the samples arises from two principal sources. This is further enhanced by the geochemical compositions of the contaminant end-member samples and their positions in the ordination biplot. Samples #TUFF and #TUFF2 and tephra-rich samples #4069 and #4134 indicate that levels of tephra contamination can be defined by the relative proportions of CaO, Na_2O and K_2O. The PCA results also suggest that clay contamination within the samples comes from a different source, as indicated by sample #CLAY, and can be quantified by the relative proportions of indicator oxides MgO, Fe_2O_3, Al_2O_3 and TiO_2.

One of the main issues encountered in previous attempts to develop a model to correct for contamination concerns the combination of estimates for both the clay and tephra components and how to accurately account for any potential geochemical overlap. In order to assess the $\delta^{18}O_{sample}$ data for the presence of two contaminants, a three end-member model was developed. The model requires that the relative purity (diatom content) or total contamination (clay and tephra, minus any overlap) proportion of the sample be calculated in order to accurately establish the quantities of the two components. The relative abundance of SiO_2 was scaled, ranging from pure diatomite (~ 93% SiO_2) to tephra (~ 58% SiO_2) and, using this scale, a percentage value was calculated as an indicator of relative purity. From this, total contamination is assumed to represent the remaining proportion of sample material. Proportions of the different elemental oxides analysed were then determined based on the following formula:

$$\%\ contamination\ by\ oxide\ A = \left|\frac{sample_A - pure_A}{contaminant_A}\right| x\ 100$$

Where $sample_A$ is the proportion of oxide A measured within the sample, $pure_A$ is the proportion of oxide A within the cleanest sample and $contaminant_A$ is the value of oxide A within the appropriate end-member (Table 1). The relative proportions of tephra (defined by relative enrichment in CaO, Na_2O and K_2O) and clay (defined by enrichment in Al_2O_3, TiO_2 and MgO) contamination were then ascertained from the average percentages of the relevant oxides (Fig. 5). At this stage, Fe_2O_3 was removed as an indicator oxide for clay contamination as it was found to introduce a bias in the calculation that resulted in consistent over-estimation of clay proportions.

In order to accurately model $\delta^{18}O_{sample}$ data for the effects of contamination, it is necessary to know the $\delta^{18}O$ composition of both tephra and clay. Since, it was not possible to isolate a pure clay sample, the $\delta^{18}O_{clay}$ end-member value was calculated using a regression equation based on the assumption that a linear relationship exists between $\delta^{18}O$ value and the proportion of clay ($r^2 = 0.78$) (Fig. 6A). This generates a $\delta^{18}O_{clay}$ value of +14.4‰. This is close to other published $\delta^{18}O$ values for end-member contaminants ($\delta^{18}O$ of silt = +12.3 ‰, Morley *et al.*, 2005; $\delta^{18}O$ of tephra = +11.6‰, Lamb *et al.*, 2007; $\delta^{18}O$ of tephra = +10.0 ‰, this study).

Using the relative proportions of contaminants and appropriate end-member $\delta^{18}O$ values, raw $\delta^{18}O_{sample}$ data were then corrected for the effect of contamination using the following mass-balance calculation:

$$\delta^{18}O_{modelled} = \delta^{18}O_{sample} - \frac{\left[\left(\frac{\%tephra}{100}\ x\ \delta^{18}O_{tephra}\right) + \left(\frac{\%clay}{100}\ x\ \delta^{18}O_{clay}\right)\right]}{\frac{\%diatom}{100}}$$

Where values of $\delta^{18}O_{tephra}$ and $\delta^{18}O_{clay}$ are given in table 1 and the %tephra and %clay values were determined using the method described above. The resulting $\delta^{18}O_{modelled}$ data are displayed in figure 7. Estimated errors for our modelled isotope corrections are obtained by factoring in an analytical reproducibility of ±0.34‰ for measurements of both $\delta^{18}O_{sample}$ and $\delta^{18}O_{tephra}$. We also associate a conservative error of ±2‰ with our computed value of $\delta^{18}O_{clay}$. In addition to this, we also associate a 15% error with each correction in order to account for the fact that PCA axes 1 and 2 only explain 85.1% of the variation within the geochemical dataset. The application of the three end-member averages model to the $\delta^{18}O_{sample}$ record has the effect of increasing $\delta^{18}O$ values by an average of 2.49‰ (σ = 1.94 ‰; n = 42, values from tephra layers and outliers are excluded). The two samples identified as outliers (#4074 and #4099) were not used in any of the isotope corrections. Additionally, corrected $\delta^{18}O$ values from samples from the two tuff layers within diatomite unit #4 are also not shown.

The removal of the contamination signal from the $\delta^{18}O_{diatom}$ record results in an overall positive shift in $\delta^{18}O_{modelled}$ values. This shift ranges from 0.03‰ (effectively zero) in the least contaminated samples to more than 8‰ towards the top of unit #4 where pure diatomites are replaced by clayey diatomites and total contamination levels approach 40%. A notable feature of the corrected data is that the majority of features of the original $\delta^{18}O_{diatom}$ record are preserved after contamination is accounted for. The resulting modelled isotope curve is less noisy than the original $\delta^{18}O_{sample}$ data (Fig. 7), a feature also common to other diatom $\delta^{18}O$ records that have been corrected for the effects of contamination using major and minor trace element geochemistry (e.g. Brewer *et al.*, 2008; Mackay *et al.*, 2011, 2013). The removal of high-frequency noise from the $\delta^{18}O_{modelled}$ data produces a curve, which exhibits a distinct rhythmical pattern with regular negative excursions of up to 5‰ that occur throughout the sequence. These cycles are driven by variations in the relative balance between precipitation and evaporation within the lake basin that ultimately reflect changes in regional monsoonal circulation, which govern the timing, duration and strength of the rainy seasons in central East Africa. A more detailed palaeoclimatic interpretation of these data is discussed elsewhere (Wilson, 2011; Wilson *et al*., submitted).

Overall, the correction applied to these data has an inherent limitation beyond which sufficient accuracy cannot be guaranteed. We find that the model works well where the total amount of contamination present in samples is assessed to be below 40%. Modelled contamination levels were crosschecked with the PCA axis scores for axes 1 and 2, which correspond approximately to clay and tephra, respectively. PCA axis 1 was found to explain more than 70% of variation in the geochemical dataset and accordingly, there is a strong relationship ($r^2 = 0.92$, $n = 62$) between the PCA axis 1 scores and the calculated proportion of clay which implies that the model has accurately captured clay variation within the samples (Fig. 6B). Correlation between the PCA axis 2 scores and estimated tephra proportion is not as good ($r^2 = 0.39$, $n = 31$ for samples with positive PCA scores), however in samples where tephra is deemed to be a significant component (>7 % content), the model does a better job ($r^2 = 0.66$, $n = 9$).

In addition to the impurities caused by tephra and catchment-derived components such as clays or carbonates, there is also the possibility of secondary isotopic exchange between amorphous diatom silica and sedimentary pore water which could theoretically occur during settling or sedimentation (Schmidt *et al*., 2001). This could potentially limit the applicability of $\delta^{18}O_{diatom}$ as a palaeoclimatic proxy if the diatom frustules are subject to secondary diagenesis. This becomes particularly important in older materials such as the Barsemoi diatomites where both age and subaereal exposure also play a role in the condition of the sediments analysed. While the issue of successive isotopic reactions remains unresolved (Swann *et al*., 2006), it is assumed for the purposes of this study that any secondary isotope exchange only affects the outer hydrous silica layer of the diatom frustule.

The removal of this layer during stepwise fluorination ensures that only the inner, more stable diatom silica is measured for $\delta^{18}O_{diatom}$ and that values of $\delta^{18}O_{diatom}$ can be reliably used for palaeoclimatic reconstruction. It is possible that condensation might continue to alter the hydrous to structural oxygen through time, however the mechanism or extent to which this may occur remains unresolved. Further study is required to fully understand the effects of diagenesis and the extent to which it may limit the application of $\delta^{18}O_{diatom}$ analysis.

Conclusions

We present a novel new approach for assessing the volume and geochemical composition of different types of contamination present within cleaned diatomite material analysed for oxygen isotope composition. It is important to consider and assess potential contamination in order to avoid introducing high-frequency noise to data sets, which can act to mask any climate-driven changes in palaeoclimate records. In some instances, where sedimentary samples may be affected by more than one type of contamination, multivariate statistical analysis of the major and minor trace element geochemistry can be used to identify and differentiate between different contaminants. This coupled approach is used to develop a three end-member mass-balance model to correct $\delta^{18}O_{diatom}$ values and enhance its use as an important palaeoclimate proxy. This technique is applied to $\delta^{18}O_{diatom}$ measurements from one of the Barsemoi Diatomites from the Central Kenyan rift valley which is affected by clay and silt from fluvial inwash from the surrounding catchment as well as air-fall ash deposits from rift-related volcanic activity. Each lake or sedimentary section may pose different contamination challenges, dependent on factors including catchment and regional geology or hydrological setting. Proximity to active volcanic centres, both now and in the past, means that lake sediments can be subject to multiple sources of contamination. Given the impact that any contamination could have on $\delta^{18}O_{diatom}$ data, it is crucial to fully understand both site (hydrological regime and catchment geology) and samples (stratigraphic setting and purity assessment at all stages) prior to isotope analysis. It is therefore vital to understand the nature of both site and samples prior to $\delta^{18}O_{diatom}$ analysis through detailed stratigraphic logging, optical and electron microscopy and geochemical analysis. This approach is thus very important for the evaluation of $\delta^{18}O_{diatom}$ data from more complex sedimentary settings.

ACKNOWLEDGEMENTS

This work was funded by a NERC PhD studentship to KEW (NE/F008635/1) and NERC NIGL Facility grant (IP/1082/1108). KEW was supported by a UCL Graduate School bursary and the Quaternary Research Association Bill Bishop Award. Sampling was conducted under the guidance of Prof. John Kingston (University of Michigan) under Kenyan Ministry of Mines and Geology research permit (OP/13/001/C 1391). The authors wish to thank Nick Marsh (University of Leicester) for help with XRF analysis, Jim Davy (UCL) for help with SEM and EDS analysis, Hilary Sloane (NIGL) for assistance with $\delta^{18}O$ analyses, and Dr George Swann for discussion and advice. We also thank our reviewers for helpful comments and suggestions which improved the manuscript.

FIGURE CAPTIONS

Figure 1

Map showing the location of the Tugen Hills within Kenya (inset, left panel) and the Central Kenyan Rift Valley (left panel) and their relation to the western flank of the main rift (Elgeyo Escarpment). Known outcrops of sediments belonging to the Chemeron Formation (including the Barsemoi Diatomites) are outlined in crosshatch markings. The sampling locality for this study, in a tributary gully to the Barsemoi River is also shown (starred). Modified after Kingston et al. (2007). Photographs depict detail of the Barsemoi exposures: **A** shows the exposed gully section, highlighting diatomite unit #4 across the centre of the image; **B** shows the base of unit #4. The unit has a sharp basal contact with the underlying silt-rich conglomeratic sediments. Photo **C** shows the grey-green tephra layer present within diatomite unit #4, which acts as a marker bed and has been dated to 2.612±0.003 Ma (Deino *et al.*, 2006).

Figure 2

Collection of images taken using scanning electron microscopy (SEM) and light microscopy to assess contamination and purity levels in samples. Images **A** and **B** show fragments of platy clays present in the diatomite samples prior to cleaning, while image **C** illustrates the degree to which secondary clays can form and adhere to diatom frustules. It is extremely difficult to remove these clays by traditional methods thus necessitating the need for further geochemical assessment using x-ray fluorescence (XRF) spectrometry. Images **D** and **E** are optical light microscopy (magnification x1000) pictures of tephra shards from within sample #4069 in diatomite unit #4, coincident with the grey-green air-fall tuff layer. Image **F** shows a sample of purified diatomite material, demonstrating that it is possible to remove a significant portion of non-diatom material via rigorous chemical and physical cleaning techniques.

Figure 3

Raw data for samples of cleaned diatomite material from unit #4 (plotted versus section height) showing changes in oxygen isotope composition and variations in different elemental oxides, measured using XRF analysis.

Figure 4

Ordination biplot showing results of PCA performed on XRF data set in order to explore variation within the geochemical compositions of remnant contamination. End-member samples of tephra (TUFF and TUFF #2), clay and laboratory standard diatomite (BFC) are shown in open triangles. Relative positions of samples identified as outliers, #4074 (laboratory analytical error) and #4099 (low sample weight) are outlined, however these were not included in the analysis. Results show that contamination is strongly influenced by variation along PC axis 1 (70.6% variation explained), representing enrichment in clay (as shown by enrichment in TiO_2, Fe_2O_3, Al_2O_3 and MgO) with a smaller element (14.6% of the variation explained by PC axis 2) controlled by the presence of tephra (enriched in Na_2O, K_2O and CaO).

Figure 5

Ternary plot showing modelled distribution of diatom material with respect to the calculated amounts of two different contamination components. X (tephra) and Z (clay) axes range from 0 to 60% whilst the Y axis (diatom silica) is plotted from 40 to 100%. Values were computed using the three end-member averages model described.

Figure 6

Graph A shows the calculated proportion of remaining clay contamination plotted against the raw $\delta^{18}O$ composition of cleaned samples. Since it was not possible to isolate a pure sample of the clay end-member, the $\delta^{18}O$ value of clay was estimated using the regression equation shown. This generates a $\delta^{18}O_{clay}$ value of 14.4‰. Graph B illustrates the strong correlation between the calculated proportions of clay contamination remaining within samples versus the PCA axis 1 scores generated by PCA.

Figure 7

Calculated amounts of total contamination derived from the abundance of clay and tephra remaining within samples plotted versus stratigraphic height within the section analysed of Barsemoi diatomite unit #4. Ages for this section, as calculated by Deino *et al*. (2006) are given to the left. These are plotted against the raw $\delta^{18}O$ measurements ($\delta^{18}O_{sample}$; open squares). Modelled $\delta^{18}O$ values (black triangles) and associated errors are plotted to the right of the panel and were corrected by the method described in the main text. Also shown are the positions of known tephra layers within the sequence (grey horizontal bands, T symbol).

TABLE CAPTIONS

Table 1

Geochemical and isotopic composition values for end-member components. Elemental totals are expressed as weight percentages. Data for the BFC diatomite is from Brewer *et al*. (2008).

REFERENCES

Ballot A, Pflugmacher S, Wiegand C, Kotut K, Krienitz L. 2003. Cyanobacteria toxins in Lake Baringo. *Limnologica* **33**: 2-9.

Barker PA, Hurrell, ER, Leng MJ, Wolff C, Cocquyt C, Sloane HJ, Verschuren D. 2011. Seasonality in equatorial climate over the past 25 k.y. revealed by oxygen isotope records from Mount Kilimanjaro. *Geology* **39**: 1111-1114.

Bootsma HA. 1993. Spatio-temporal variation of phytoplankton biomass in Lake Malawi, Central Africa. *Verhandlungen Internationale Vereinigung für Limnologie* **25**: 882-886.

Brandriss ME, O'Neil JR, Edlund MB, Stoermer EF. 1998. Oxygen isotope fractionation between diatomaceous silica and water. *Geochimica et Cosmochimica Acta* **62**: 1119-1125.

Brewer TS, Leng MJ, Mackay AW, Lamb AL, Tyler JJ, Marsh NG. 2008. Unravelling contamination signals in biogenic silica oxygen isotope composition: The role of major and trace element geochemistry. *Journal of Quaternary Science* **23**: 321-330.

Chapligin B, Meyer H, Bryan A, Snyder J, Kemnitz H. 2012. Assessment of purification and contamination correction methods for analysing the oxygen isotope composition from biogenic silica. *Chemical Geology* **300-301**; 185-199.

Chapman GR, Lippard SJ, Martyn JE. 1978. The stratigraphy and structure of the Kamasia Range, Kenya Rift Valley. *Journal of the Geological Society of London* **135**: 265-281.

Deino AL, Kingston JD, Glen JM, Edgar RK, Hill A. 2006. Precessional forcing of lacustrine sedimentation in the late Cenozoic Chemeron Basin, Central Kenya Rift, and the calibration of the Gauss/Matuyama boundary. *Earth and Planetary Science Letters* **247**: 41-60.

Juillet-Leclerc A, Labeyrie LD. 1987. Temperature dependence of the oxygen isotopic fractionation between diatom silica and water. *Earth and Planetary Science Letters* **84**: 69-74.

Kingston JD, Deino AL, Edgar RK, Hill A. 2007. Astronomically forced climate change in the Kenyan Rift Valley 2.7-2.55 Ma: implications for the evolution of early hominin ecosystems. *Journal of Human Evolution* **53**: 487-503.

Labeyrie LD. 1974. New approach to surface seawater palaeotemperatures using 18O/16O ratios in silica of diatom frustules. *Nature* **248**: 40-42.

Lamb AL, Leng MJ, Sloane HJ, Telford RJ. 2005. A comparison of the palaeoclimate signals from diatom oxygen isotope ratios and carbonate oxygen isotope ratios from a low latitude crater lake. *Palaeogeography, Palaeoclimatology, Palaeoecology* **223**: 290-302.

Lamb AL, Brewer TS, Leng MJ, Sloane HJ, Lamb HF. 2007. A geochemical method for removing the effect of tephra on lake diatom oxygen isotope records. *Journal of Paleolimnology* **37**: 499-516.

Leng MJ, Barker PA. 2006. A review of the oxygen isotope composition of lacustrine diatom silica for palaeoclimate reconstruction. *Earth Science Reviews* **75**: 5-27.

Leng MJ, Sloane HJ. 2008. Combined oxygen and silicon isotope analysis of biogenic silica. *Journal of Quaternary Science* **23**: 313-319.

Leng MJ, Metcalfe SE, Davies SJ. 2005. Investigating Late Holocene climate variability in Central Mexico using carbon isotope ratios in organic materials and oxygen isotope ratios from diatom silica within lacustrine sediments. *Journal of Paleolimnology* **34**: 413-431.

Mackay AW, Swann GEA, Brewer TS, Leng MJ, Morley DW, Piotrowska N, Rioual P, White D. 2011. A reassessment of late glacial-Holocene diatom oxygen isotope record from Lake Baikal using a geochemical mass-balance approach. *Journal of Quaternary Science* **26**: 627-634.

Mackay AW, Swann GEA, Fagel N, Fietz S, Leng MJ, Morley DW, Rioual P, Tarasov P. 2013. Hydrological instability during the last interglacial in central Asia: a new diatom oxygen isotope record from Lake Baikal. *Quaternary Science Reviews* **66**: 45-54.

Morley CK, Westcott WA, Stone DM, Harper RM, Wigger ST, Karanja FM. 1992. Tectonic evolution of the northern Kenya Rift. *Journal of the Geological Society* **149**: 333-348.

Morley DW, Leng MJ, Mackay AW, Sloane HJ, Rioual P, Battarbee RW. 2004. Cleaning of lake sediment samples for diatom oxygen isotope analysis. *Journal of Palaeolimnology* **31**: 391–401.

Morley DW, Leng MJ, Mackay AW, Sloane HJ. 2005. Late glacial and Holocene atmospheric circulation change in the Lake Baikal region documented by oxygen isotopes from diatom. *Global and Planetary Change* **46**: 221– 233.

Moschen R, Lücke A, Schleser GH. 2005. Sensitivity of biogenic silica oxygen isotopes to changes in surface water temperature and paleoclimatology. *Geophysical Research Letters* **32**: L07708, doi:10.1029/2004GL022167.

Raubitschek S, Lücke A, Schleser GH. 1999. Sedimentation patterns of diatoms in Lake Holzmaar – on the transfer of climate signals to biogenic silica oxygen isotope proxies. *Journal of Paleolimnology* **21**: 437-448.

Renaut RW, Ego J, Tiercelin J-J, Le Turdu C, Owen RB. 1999. Saline, alkaline palaeolakes of the Tugen Hills–Kerio Valley region, Kenya Rift Valley. In *Late Cenozoic Environments* *and Hominid Evolution: A Tribute to Bill Bishop*, Andrews P, Banham P (eds). Geological Society Publishing House: Bath; 41–58.

Rosqvist GC, Leng MJ, Goslar T, Sloane HJ, Bigler C, Cunningham L, Dadal A, Bergman J, Berntsson A, Jonsson C, Wastegård S. 2013. Shifts in precipitation during the last millennium in northern Scandanavia from lacustrine isotope records. *Quaternary Science Reviews* **66**: 22-34.

Round FE, Crawford RM, Mann DG. 1990. *The Diatoms: Biology and Morphology of the Genera*. Cambridge University Press: Cambridge.

Schmidt M, Botz R, Rickert D, Bohrmann G, Hall SR, Mann S. 2001. Oxygen isotopes of marine diatoms and relations to opal-A maturation. *Geochimica et Cosmochimica Acta* **65**: 201-211.

Shemesh A, Charles CD, Fairbanks, RG. 1992 Oxygen isotopes in biogenic silica: Global changes in ocean temperature and isotopic composition. *Science* **256**: 1434-1436.

Swann GEA, Leng MJ. 2009. A review of diatom δ18O in palaeoceanography. *Quaternary Science Reviews* **28**: 384-398.

Swann GEA, Patwardhan SV. 2011. Application of Fourier Transform Infrared Spectroscopy (FTIR) for assessing biogenic silica sample purity in geochemical analyses and palaeoenvironmental research. *Climate of the Past* **7**: 65-74.

Swann GEA, Maslin MA, Leng MJ, Sloane HJ, Haug GH. 2006. Diatom $\delta^{18}O$ evidence for the development of the modern halocline system in the subarctic northwest Pacific at the onset of major Northern Hemisphere glaciation. *Paleoceanography* **21**: doi:10.1029/2005PA001147.

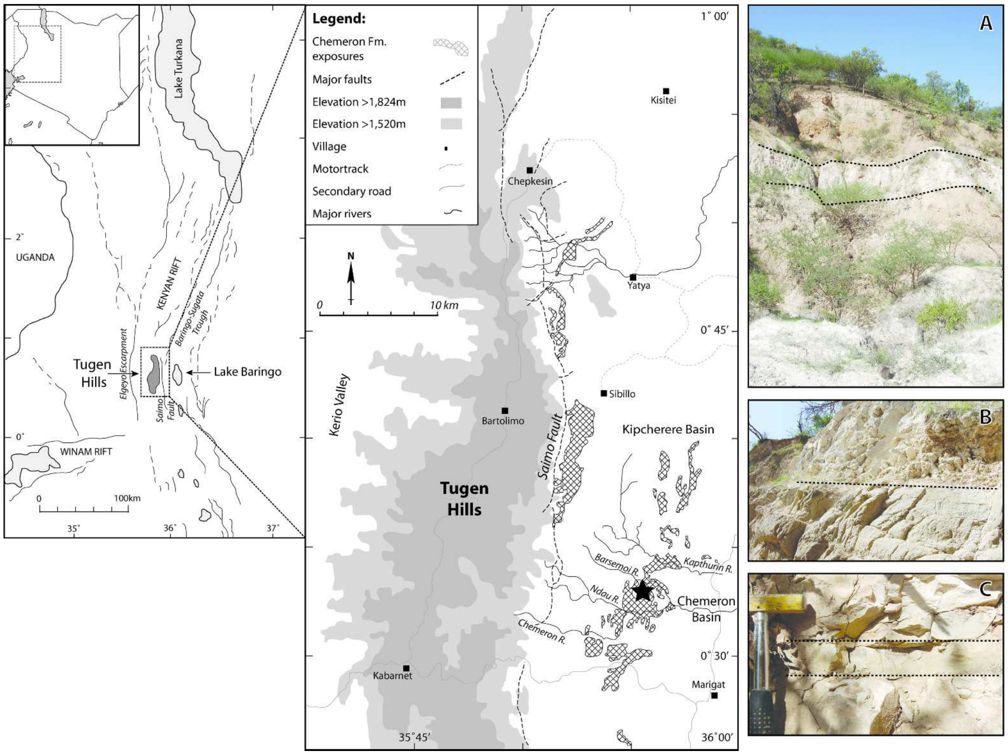
Tarits C, Renaut RW, Tiercelin J-J, Le Hérissé A, Cotton J, Cabon J-Y. 2006. Geochemical evidence of hydrothermal recharge in Lake Baringo, central Kenya Rift Valley. *Hydrological Processes* **20**: 2027-2055.

Ter Braak CJF, Šmilauer P. 2002. *CANOCO Reference Manual and Canodraw for Windows User's Guide: Software for Canonical Community Ordination (Version 4.5)*. Microcomputer Power: Ithaca, USA.

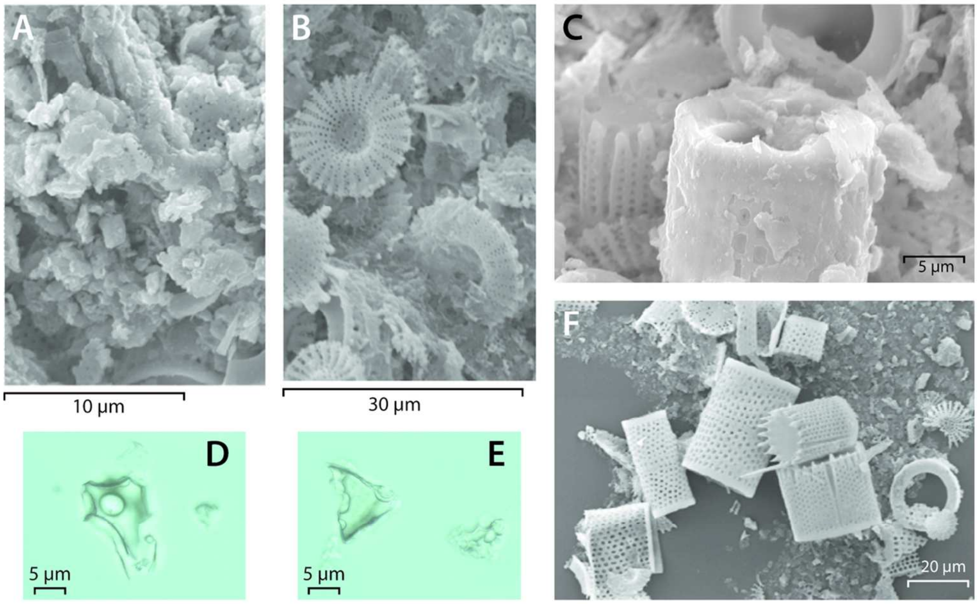
Wetzel RG. 2001. *Limnology*, 3rd edn. Academic, San Diego, USA.

Wilson KE. 2011. *Plio-Pleistocene reconstruction of East African and Arabian Sea palaeoclimates*. PhD thesis. University College London, UK. 301pp.

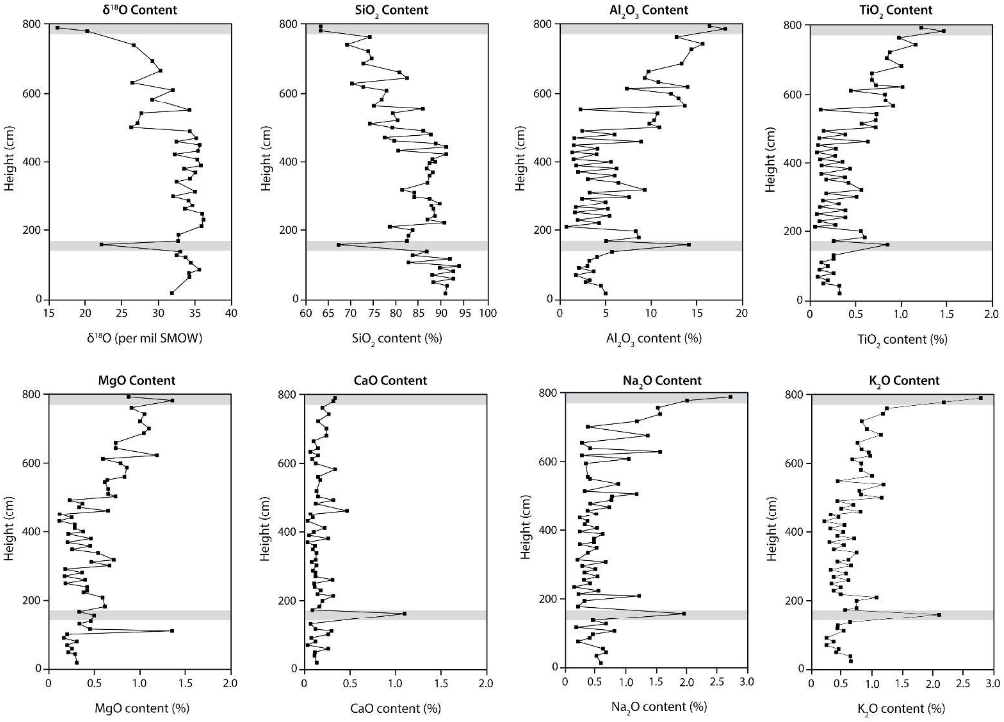
Wilson KE, Leng MJ, Edgar RK, Deino AL, Kingston JD, Maslin MA, Mackay AW. East African lake evidence for Pliocene millennial-scale climate variability. (submitted to *Geology*).



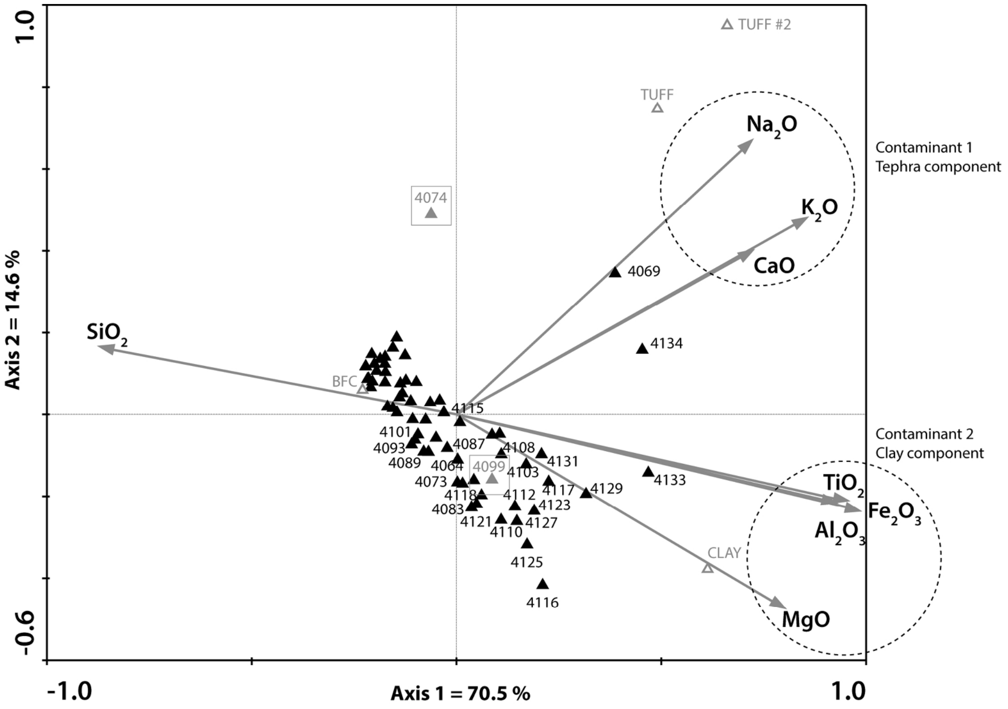
260x194mm (300 x 300 DPI)



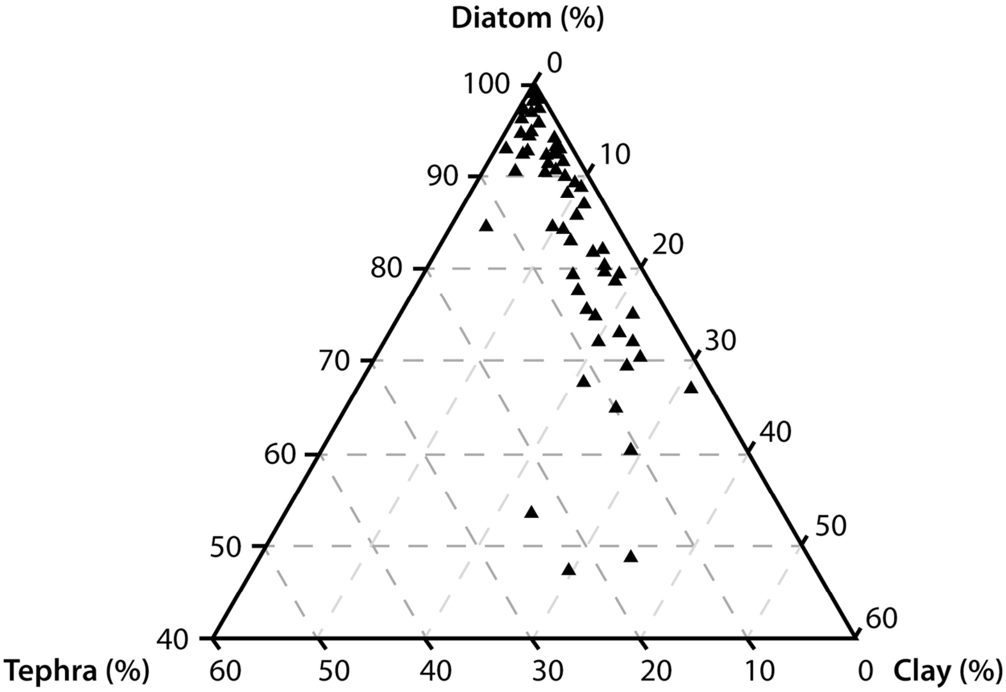
110x70mm (300 x 300 DPI)



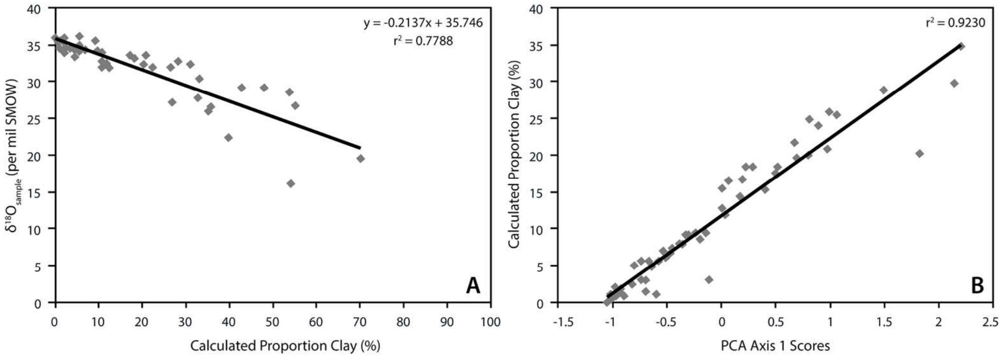
170x122mm (600 x 600 DPI)



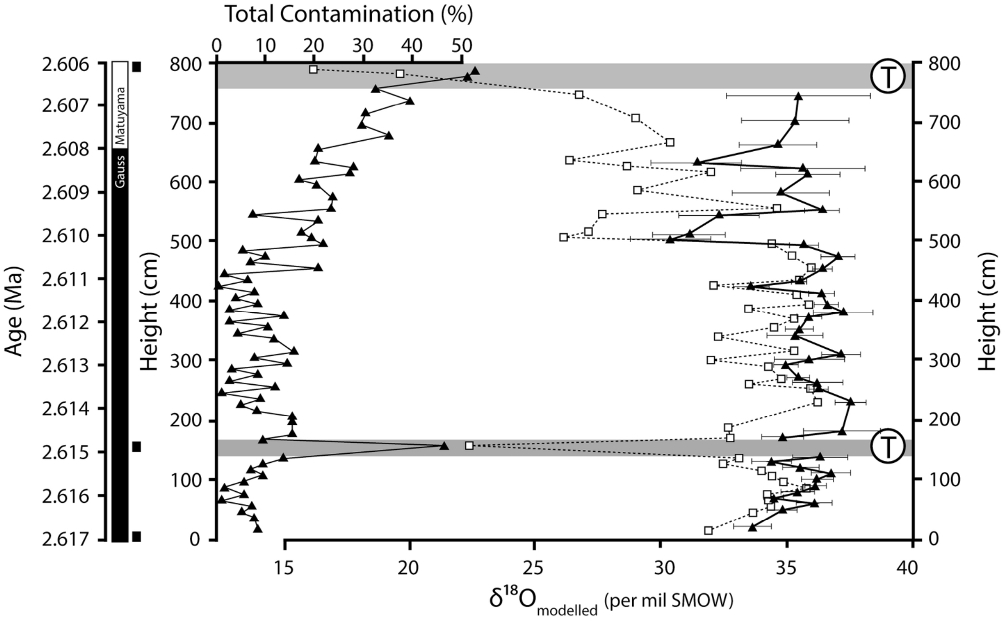
187x131mm (300 x 300 DPI)



76x52mm (600 x 600 DPI)



72x25mm (600 x 600 DPI)



88x54mm (600 x 600 DPI)

Sample	$\delta^{18}O$	SiO_2	Tephra Contamination			Clay Contamination			Fe_2O_3	Total
			CaO	Na_2O	K_2O	Al_2O_3	TiO_2	MgO		
BFC	29.88	91.91	0.33	0.14	0.07	1.38	0.07	0.24	0.39	100.57
TUFF1	10.00	64.63	0.99	4.64	2.70	13.59	0.71	0.72	10.20	98.77
TUFF2	-	58.97	1.03	4.72	4.77	17.16	1.51	0.64	7.16	96.59
CLAY	14.38	58.08	1.36	0.74	0.85	24.62	1.95	1.81	8.28	100.00