Empirical Calibrabration of a Lipid-Based Paleothermometer in the Eastern Cordillera of Colombia

Veronica Anderson
CE 394 K, Fall 2012.

Introduction

In 2004, scientists discovered that the abundances of a number of lipids known as branched GDGT's (glycerol dialkyl glycerol tetraethers) are strongly correlated with the mean annual temperature in soils [*Hopmans et al.*, 2004]. It is believed that these compounds are membrane lipids produced by soil bacteria; as the temperature conditions change in the soil, the relative concentration of each of the nine branched GDGT's varies as the bacteria adapt to different conditions [*Weijers et al.*, 2006]. Because these lipids are very resistant to decomposition and have been found in extremely ancient sediments, these lipids show great potential for use as a proxy for past surface temperatures in geologic studies.

Previous work to calibrate this proxy for temperature reconstructions has relied on empirical calibrations between the measured concentrations of the nine branched GDGT's and mean annual temperature and soil pH from a number of soil samples across the globe [*Peterse et al.*, 2012; *Weijers et al.*, 2007]. Essentially, these studies experimented with a number of multivariate linear regressions between the measured GDGT abundances and these environmental factors, and determined which set of parameters produced the best-fit regression equation. This technique is powerful because it enables them to determine a correlation between these compounds and temperature across a diverse range of environments, soils, and climate conditions; however, because the actual mechanism by which the abundances of these compounds vary is unknown, it has been difficult to determine the best approach to improving the fit of this correlation. In addition, because the reconstructed temperatures are calculated

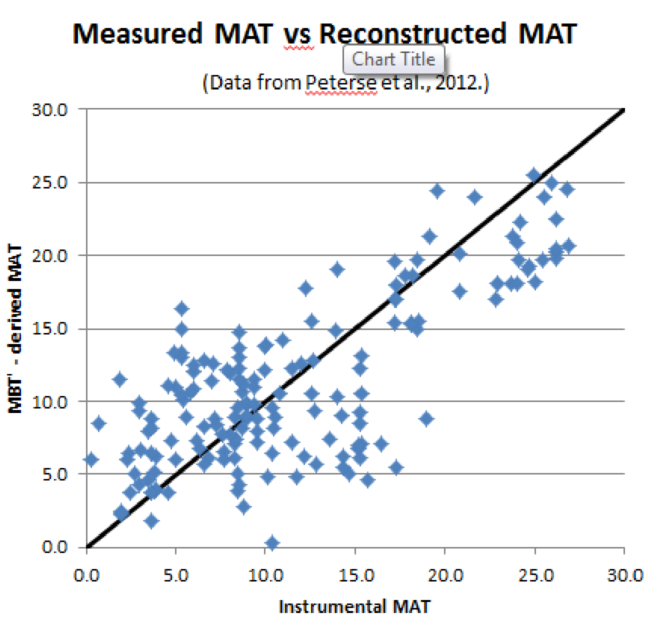

Figure 1: Plot of instrumental mean annual temperature (MAT) vs. the reconstructed value from the most recent calibration. Note the wide scatter about the 1:1 line. This calibration has an $r^2 = 0.59$.

through a purely empirical regression equation, it is unclear how to accurately report errors on these reconstructions. For instance, if one were to use the standard deviations of the residuals from the calibration dataset, this would likely overestimate the error because this is also dependent on the number of points used in the regression.

In my research, I am attempting to determine the magnitude of the effect of a number of sources of error that could be contributing to the scatter in the original regression. The particular source of error that I am attempting to quantify for this project involves the differences between the temperature data used in the calibrations and the actual soil temperature. The empirical calibrations rely on weather station data for the correlations between the measured soil lipids and mean annual temperature; in some cases, the weather station is as far away as a hundred kilometers from the actual soil sample, which would make it unlikely to accurately reflect the temperature of the soil. In addition, several works have noted that soil temperatures are frequently several degrees warmer than the air temperature due to the greater heat capacity of the ground [*Quade et al.*, 2011]. In this my goal was to determine whether the fit improves substantially by using in-situ temperature data over using average temperatures from the nearest weather station, and whether we can improve the fit to station data by using more sophisticated interpolation techniques.

Methods

In order to constrain the actual temperature at each of our soil sampling sites, we buried a set of temperature loggers in 32 sites across an elevation transect of the Eastern Cordillera of Colombia. The mean annual temperature (MAT) at the sites ranged from 10 degrees to 27 degrees, which covers about 2/3 of the range of the original calibration dataset [*Peterse et al.*, 2012]. We also have temperature data from the Colombian national weather service (IDEAM), which I used as the point dataset to test out different interpolation schemes. This field setting represents a "worst-case" scenario for the use of the nearest weather station because of the mountainous terrain – any slight difference between the location of the soil sample and the weather station could result in large discrepancies in climate conditions. In the original global dataset, we do not anticipate that the mismatch will be quite as serious because most sites are not in mountainous terrain.

In order to determine whether any of the interpolation schemes could eliminate the error due to site mismatch, I tried three different interpolation techniques in ArcGIS. For each technique, I took the output raster and used the Sample Raster function to determine the interpolated value at each of my temperature logging sites. I then exported the data as a new shapefile and then matched the contents of the .dbf file with my soil temperature data. Then, I re-ran my regression with the new interpolated temperature data in order to determine whether the scatter in my calibration was significantly reduced.

Nearest Neighbor

In order to characterize the worst-case scenario, I first determined the temperature at each soil sampling site by picking the mean annual temperature measured at the nearest IDEAM weather station. For this, I used the "Near" method in the analysis toolbox, and it added columns to the attribute table containing the FID of the nearest IDEAM station; I then used a table join to extract the mean annual temperature at each of these sites as well.

Linear Interpolation

The simplest interpolation method that I used was a linear interpolation – this is exactly what the authors of the global calibration study did in order to correct sites where they suspected that site mismatch would be a significant problem. In order to perform linear interpolation, I used the "Local Polynomial Interpolation" function in the Geostatistical Analyst toolbox, and selected a polynomial of order 1. I left all of the other parameters as the ArcGIS default values. This produced a reasonable-looking temperature map, with the temperatures appearing to follow the topography fairly closely.

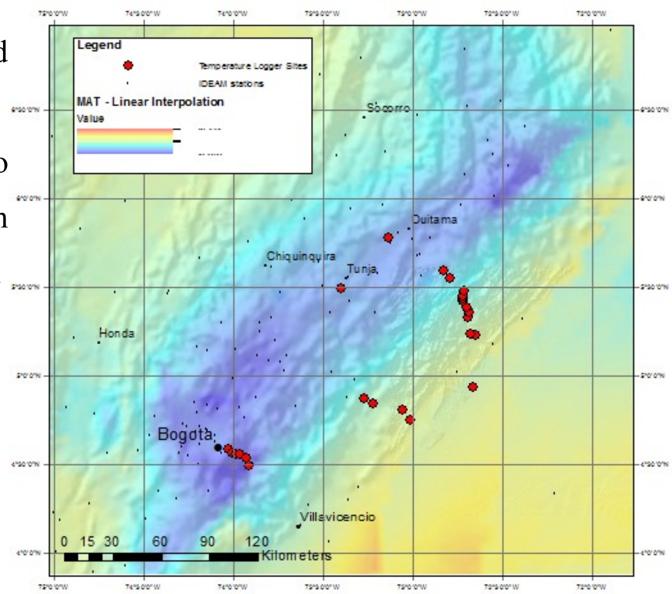

Figure 2: Shaded color map of the linearly interpolated MAT values.

Kriging

As a more sophisticated interpolation technique, I used Kriging (Spatial Analyst Toolbox) with a search radius of 12 and all parameters set to their default values in ArcGIS. As with the Linear Interpolation function, the interpolated field looks quite reasonable, and looks similar to the results obtained by linear interpolation.

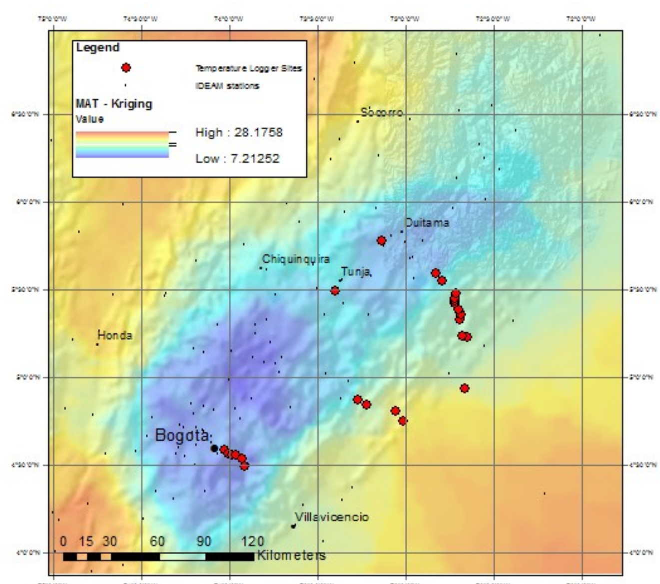

Figure 3: Shaded color map of temperatures interpolated by Kriging.

Spline

Of the three interpolation methods, the spline gave the most unreasonable-looking temperature field. I used the default 12-point model for the spline, and the interpolated surface had some very strange bulls-eye features that had little to do with the elevation of the surface. Not surprisingly, this interpolation technique did not yield good fits to our data.

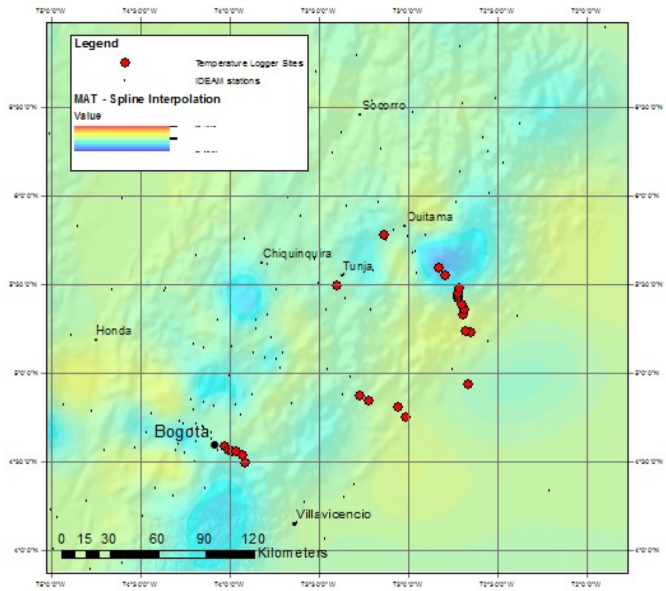

Figure 4: Color-shaded map of mean annual temperatures interpolated by splines.

Results/Discussion

After resampling the mean annual temperature at each of my logger sites, I did a multivariate linear regression between the 9 branched GDGT's and the mean annual temperature. I calculated the correlation coefficient for each model, and compared it to the model fit for the in-situ temperature

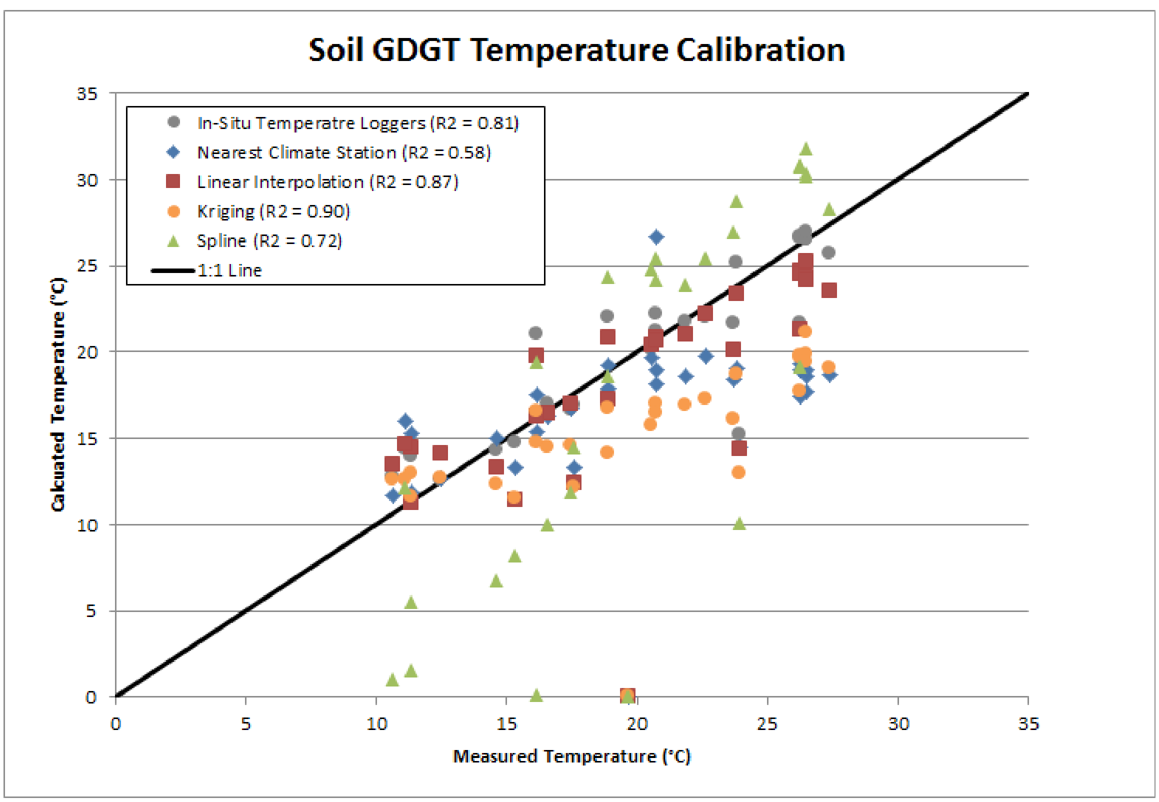

Figure 5: Plot of the instrumental MAT against the reconstructed MAT using each of the different interploation methods.

loggers. I found that linear interpolation actually did the best job of matching the instrumental data, fitting the data with an $r^2 = 0.87$, and minimal structure to the residuals. Interestingly, Kriging produced a fit with a higher correlation coefficient than the in-situ temperature loggers, but it consistently resulted in reconstructed values that under-estimated the true soil temperature. This may be reflecting the fact that soil temperatures are generally warmer than the mean annual air temperature; even if there is a strong linear correlation with air temperature, it will be offset slightly from the true soil temperature. As expected, both the nearest neighbor approach and the spline interpolation resulted in significantly more scatter in their calibration fits, and therefore are not methods that I would recommend for use in global calibrations. However, it should be noted that the errors associated with using the nearest temperature station are likely to be much larger than they would be in the global dataset, due to the significant differences in elevation between our sampling sites and the IDEAM weather stations.

In addition, I also plotted the interpolated values against the actual in-situ data in order to try to determine how closely the interpolated data matches the instrumental data. Interestingly, there are still significant discrepancies between the instrumental and interpolated data, even though the GDGT correlations for linear interpolation and Kriging are similar to what we get for the in-situ temperature loggers. This suggests that the remaining differences between the linearly interpolated data and the in-situ measurements are not contributing significantly to the scatter in the relationship between measured GDGT abundances and mean annual temperature; that scatter appears to be coming from other sources.

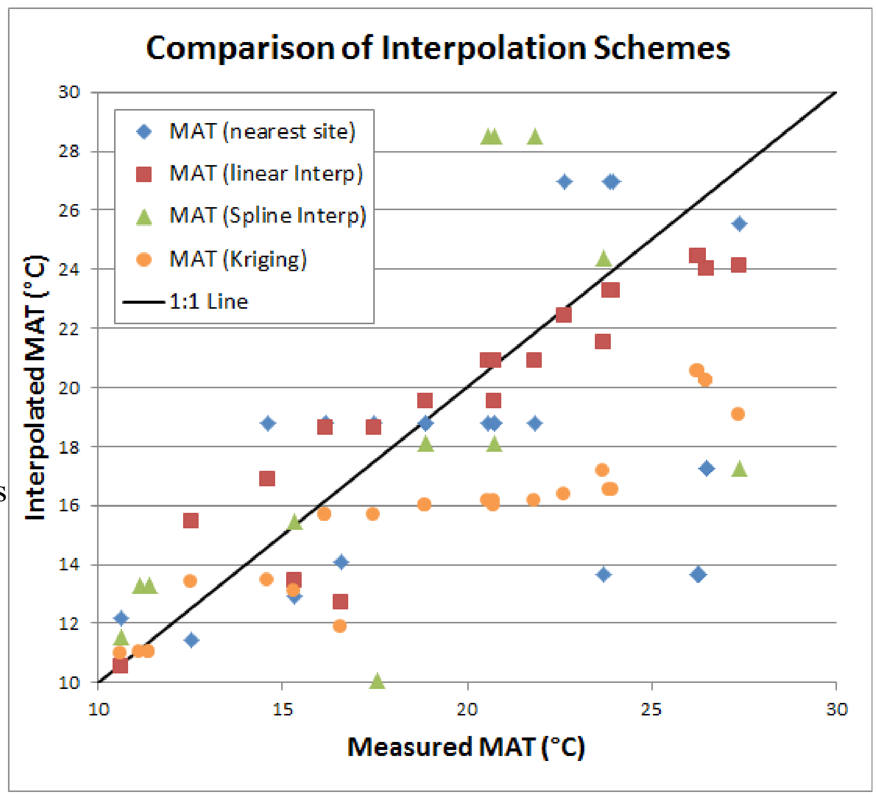

Figure 6: Plot of the interpolated temperatures at each site against the in-situ measurement.

Application to the Global Dataset

One of the aims of this work was to determine if there were any specific techniques that could be used to reduce the error in the calibration of the global dataset. To this end, I downloaded the global calibration database of measured GDGT abundances and station data from the NOAA GHCN (Global Historical Climatology Network) in order to check whether interpolation of the global temperature sites could improve the calibration. Based on my experiences with the Colombia dataset, I decided to try Kriging and linear interpolation, because they had been the most successful with the Colombia data.

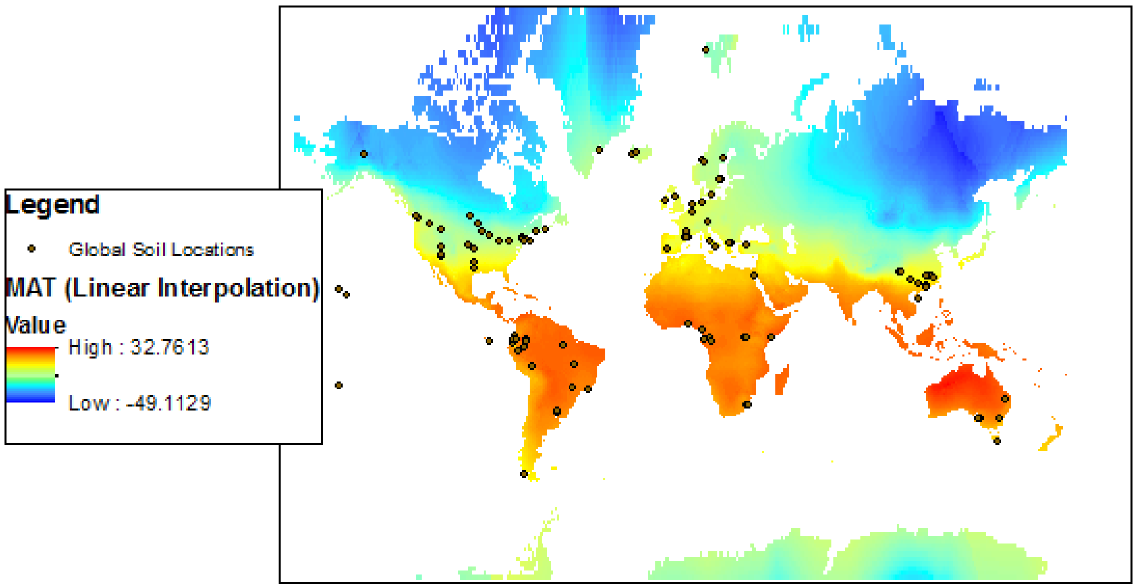

Figure 7: Linearly interpolated temperature values from NOAA GHCN data for the global soil calibration dataset.

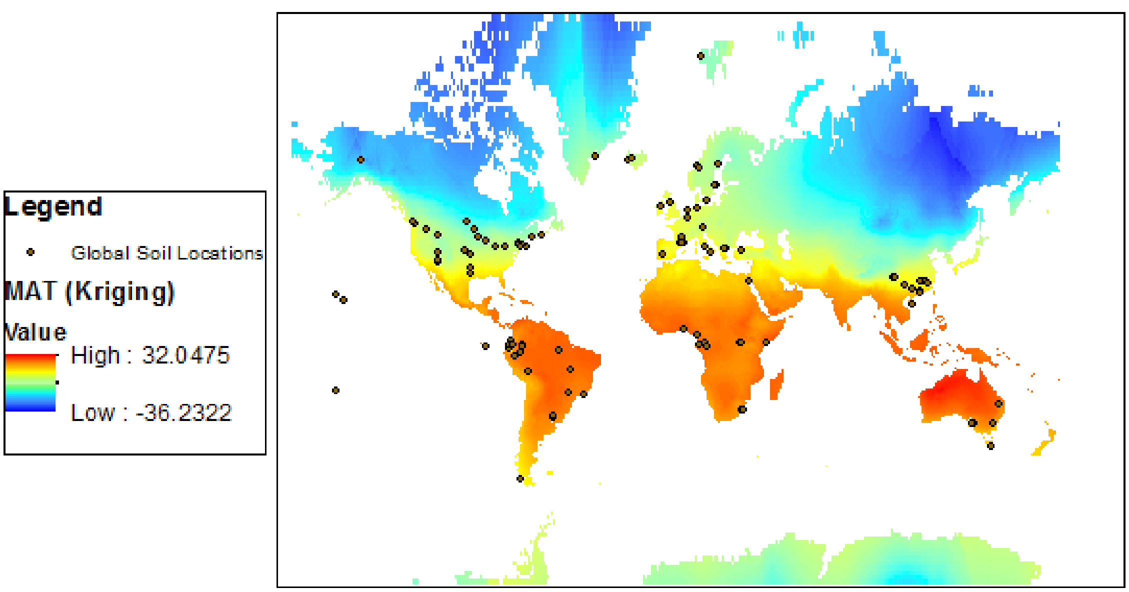

Figure 8: Temperature values interpolated by Kriging from the NOAA GHCN dataset.

As before, the interpolation schemes both performed well – in this case, however, we do not have in-situ temperature data, so I was attempting to see if the calibration could be improved by using one of these interpolation schemes. As it turned out, they did not perform significantly better than the data already being used in the study. Below is a table of the different correlations that I tried and their correlation coefficients.

	R^2 Correlation Coefficient
Temperatures from Peterse et al., 2012	0.622
Linearly interpolated values	0.637
Values interpolated by Kriging	0.632

As you can see from the values in the table, the two interpolation techniques only marginally improved the quality of the fit over the original dataset, suggesting that the temperature data was already a fairly good match.

Conclusions

After trying several different interpolation techniques for comparison with the in-situ temperature loggers in Colombia, it seems that simple linear interpolation is sufficient to produce temperature data that fits the measured GDGT abundances equally well. In addition, this observation is confirmed by my investigation of the global dataset, which shows no significant improvement in the goodness of fit with the two interpolation schemes. This suggests that site mismatch is not actually a significant source of scatter in our calibration; it seems that the scatter may be due to other natural factors such as seasonality biases, and other confounding factors. For future work, I plan to incorporate models of soil temperature as a function of the air temperature and mean annual solar insolation in order to more accurately represent the soil temperature.

References

Hopmans, E. C., J. W. . Weijers, E. Schefuß, L. Herfort, J. S. Sinninghe Damsté, and S. Schouten (2004), A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids, *Earth and Planetary Science Letters*, 224(1-2), 107–116, doi:10.1016/j.epsl.2004.05.012.

Peterse, F., J. van der Meer, S. Schouten, J. W. H. Weijers, N. Fierer, R. B. Jackson, J.-H. Kim, and J. S. Sinninghe Damsté (2012), Revised calibration of the MBT–CBT paleotemperature proxy based on branched tetraether membrane lipids in surface soils, *Geochimica et Cosmochimica Acta*, 96, 215–229, doi:10.1016/j.gca.2012.08.011.

Quade, J., D. O. Breecker, M. Daeron, and J. Eiler (2011), The paleoaltimetry of Tibet: An isotopic perspective, *American Journal of Science*, 311(2), 77–115, doi:10.2475/02.2011.01.

Weijers, J. W. H., S. Schouten, J. C. van den Donker, E. C. Hopmans, and J. S. Sinninghe Damsté (2007), Environmental controls on bacterial tetraether membrane lipid distribution in soils, *Geochimica et Cosmochimica Acta*, 71(3), 703–713, doi:10.1016/j.gca.2006.10.003.

Weijers, J. W. H., S. Schouten, O. C. Spaargaren, and J. S. Sinninghe Damsté (2006), Occurrence and distribution of tetraether membrane lipids in soils: Implications for the use of the TEX86 proxy and the BIT index, *Organic Geochemistry*, 37(12), 1680–1693, doi:10.1016/j.orggeochem.2006.07.018.