This document provides more detail on the forward model described briefly in the main text. We focus on transformations for compositional data and their use, diffusion maps and the forward model.

## 1. Compositional data analysis and coordinate transformations

Composition data are data that lie in a simplex. In other words, each datapoint must contain strictly positive entries which sum to a constant, often one. The dataset of fractional abundances of GDGTs is a compositional dataset. We refer to a single point as a composition.

The majority of multivariate statistics are designed for data that lie in Euclidean space, i.e. unconstrained space. Problems in compositional spaces may arise due to the inherent correlation between compositional parts, or models producing non-compositional predictions (e.g. uncertainty intervals extending beyond 1 for a given compositional part). For that reason, finding a mapping between the simplex and Euclidean space allows standard multivariate techniques to be applied to compositional data.

There are a number of options based on log-ratio transformations. We choose the isometric log-ratio transformation for this work for the reasons laid out by Egozcue et al (2003) [1], the main reasons being that it reflects the true dimensionality of compositional data and provides a one-to-one mapping between the simplex and Euclidean space.

Given a compositional vector $\boldsymbol{x} \in \mathbb{S}^d$ where $\mathbb{S}^d$ is the $d-$dimensional unit simplex, the isometric log-ratio transformation $\text{ilr} : \mathbb{S}^d \to \mathbb{R}^{(d-1)}$ is given by

$$(\text{ilr}(\boldsymbol{x}))^\top = (\text{clr}(x))^\top \Phi$$

where $\Phi$ is a $d \times (d-1)$ matrix, $\text{clr}(\boldsymbol{x})_i = \ln\left(\frac{x_i}{g(\boldsymbol{x})}\right)$ is the centred log-ratio transformation, and $g(\boldsymbol{x})$ is the geometric mean of $\boldsymbol{x}$.

We use the ilr transformation as implemented by the function `pivotCoord` in the R package `robCompositions`. Note that ilr-transformation is unsuitable for data containing zero components. A commonly used strategy is to assume that zero values are not true zeros, but are below some detection limit imposed by the measurement technology. Thus zeros are treated as missing values and a variety of imputation methods can be employed (Martín-Fernàndez et al., 2012). We use the `impCoda` function in the R package `robCompositions` for this purpose.

## 2. Diffusion maps for data visualisation

Diffusion maps (Coifman et al, 2006) are a method for nonlinear dimensionality reduction and visualisation. The building block of a diffusion map is a graph representing the data, in which the vertices are datapoints. These vertices are joined by weighted edges, and a variety of choices exist for assigning weights to these edges. For the implementation we use in the main text, we represent the data as a fully connected graph (i.e., each edge has non-zero weight). The weight assigned to each edge between two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is given by

$$W_{ij} = \exp\left(-\frac{d^2(\boldsymbol{x}_i, \boldsymbol{x}_j)}{2\sigma^2}\right)$$

where $d$ is some distance function and $\sigma$ is a lengthscale parameter to be set later. Essentially this builds a graph in which points which are 'close' subject to $d$ and $\sigma$ have heavily weighted edges and points which are distant have small weights on their edges.

To properly account for the compositional nature of the data, we use a simplicial distance measure rather than the typical Euclidean distance. We choose the Aitchison distance, which is simply the Euclidean distance between ilr-transformed datapoints (Egozcue et al., 2003).

We then used these weights to define a discrete-time, discrete-distance Markov process on the data, with some transition matrix $P$ which depends on $W$ (for full details of the choice of the transition matrix we use in the main text see Haghverdi et al., 2015).

The eigenvectors of this transition matrix are referred to as diffusion components. They represent the dominant modes of variation in the data. The diffusion distance between two points reflects their connectivity in the graph and is related to the probability that one point can be reached from the other in some specified time. The advantage of representing the points by their diffusion components for visualisation purposes is that Euclidean distance in the diffusion space is approximately equivalent to diffusion distance in the original space (Nadler et al., 2006).

## 3. Details of the forward model

The forward model is built on multi-output (or vector-valued) Gaussian Processes. We call this model a forward model in the sense of Haslet et al. (2006), i.e., the basic building block of the model is the assumption that measured compositions arise via some unknown function of temperature plus additional temperature-independent noise. The more traditional regression strategy would be to model the outcome of interest (temperature) as a function of measured predictions (compositions), and indeed this is the approach of the Gaussian Process Regression model described in the main text.

Let $\boldsymbol{x}_i \in \mathbb{S}^d$ denote the $i^{th}$ measured composition, and let $\boldsymbol{x}_i^* \in \mathbb{R}^{(d-1)}$ denote the ilr-transformation of $\boldsymbol{x}_i$. We begin with the very general model

$$\boldsymbol{x}_i^* = \boldsymbol{f}(T_i) + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}(d-1)$ is a function to be specified/determined and $\boldsymbol{\varepsilon}_i$ is a zero-mean random variable independent of temperature.

The function $\boldsymbol{f}$ describes the way in which sea surface temperatures give rise to steady-state GDGT compositions in populations of marine archaeota. It is clear that temperature has some effect, but a well-reasoned mechanistic model has, to our knowledge, not been developed.

In order to capture the model uncertainty associated with the lack of a mechanistic model, we take a Bayesian approach and place a multi-output Gaussian process prior on the function $\boldsymbol{f}$. We also assume that $\boldsymbol{\varepsilon}$ is Gaussian with diagonal covariance matrix $\Sigma$.

In other words, the forward model is as follows

$$\begin{aligned} \boldsymbol{x}^*|T, \boldsymbol{f}, \Sigma &\sim \mathcal{N}(\boldsymbol{f}(T), \Sigma), \\ \boldsymbol{f} &\sim \text{MOGP}(\boldsymbol{K}), \\ \sigma_i &\propto 1,\ i = 1, \ldots, d-1, \\ \Sigma &= \text{diag}(\boldsymbol{\sigma}) \end{aligned}$$

MOGP refers to a zero-mean, multi-output Gaussian process with kernel $\boldsymbol{K}$. There are a number of choices for kernels in multi-output Gaussian process regression models (Alvarez et al., 2012). We choose perhaps the simplest option, the Intrinsic Coregionalisation Model (ICM) with a Matern 3/2 base kernel. The model is implemented in Python 3.6 via the GPy library – the kernel hyperparameters are optimised by maximising the marginal likelihood, and probabilistic predictions are subsequently made using the exact form for the posterior. The code is available as part of the Github repository detailed in the main text.

The most useful outcome of estimating this model is gaining access to the conditional density $p(\boldsymbol{x}^*|T)$, i.e. gaining the ability to predict a distribution over compositions, given a temperature.

Armed with this distribution, a simple application of Bayes' rule allows the model to be inverted and we gain access to $p(T|\boldsymbol{x}^*)$, i.e. a distribution over temperatures given a new composition. In other words, we compute

$$p(T|\boldsymbol{x}^*) = \frac{p(\boldsymbol{x}^*|T)p(T)}{p(\boldsymbol{x}^*)} \tag{1}$$

Here $p(T)$ is a prior distribution over temperatures and reflects our prior beliefs about 'reasonable' sea surface temperatures. For example, we know that sea surface temperatures below $\sim -5°C$ or above $\sim 50°C$ are unreasonable, and our prior should reflect this. If a uniform prior over all temperatures is assumed (i.e. any temperature is possible a priori), the posterior is improper (i.e. does not integrate to 1) due to the large probability mass assigned to all compositions at temperatures far outside the modern temperature range.

Note that the normalising factor $p(\boldsymbol{x}^*) = \int_{\mathbb{R}} p(\boldsymbol{x}^*|T)p(T)\,dT$ is required to ensure that the probability distribution is properly normalised and to allow the computation of quantities of interest such as the predictive mean, $\int_{\mathbb{R}} Tp(T|\boldsymbol{x}^*)\,dT$, and variance, $\int_{\mathbb{R}}(T-\mu)^2 p(T|\boldsymbol{x}^*)\,dT$, of the distribution.

Since the integrals are one dimensional, it is straightforward and computationally cheap to use numerical quadrature to evaluate them.

For the applications in the paper, $p(T)$ is chosen to be a Gaussian distribution, and so the natural choice is Gauss-Hermite quadrature. We use 500 point Gauss-Hermite quadrature. For more details on quadrature in general and numerical methods for integration see Press et al. (2007).

For further details on the implementation of the model see the accompanying code. A schematic of the model is presented in Figure 1.

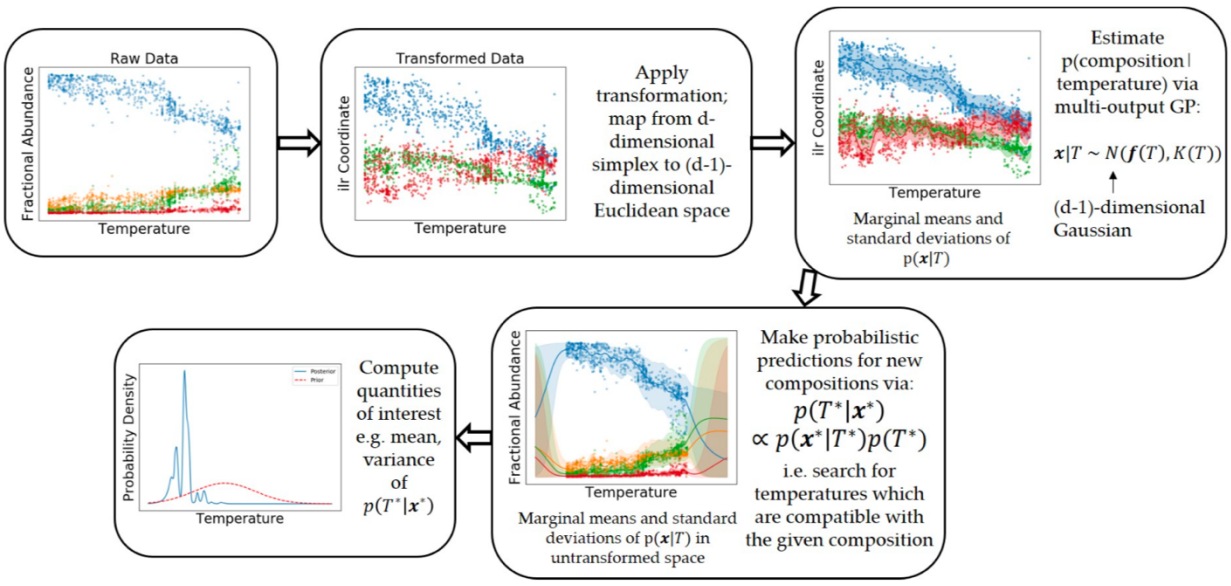


Figure S1: Schematic representation of the forward model

**References:**


Alvarez, M.A., Rosasco, L. & Lawrence, N.D.,: Kernels for vector-valued functions: A review. Foundations and Trends in Machine Learning, 4(3), 195-266. 2012.

Coifman, R.R. & Lafon, S.,: Diffusion maps. Applied and computational harmonic analysis, 21(1), 5-30, 2006.

Egozcue,J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G. & Barcelo-Vidal, C.,: Isometric log ratio transformations for compositional data analysis. Mathematical Geology, 35(3), 279-300, 2003.

Haghverdi, L., Buettner, F. & Theis, F.J.,: Diffusion maps for high- dimensional single-cell analysis of differentiation data. Bioinformatics, 31(18), 2989-2998, 2015.

Haslett, J., Whiley, M., Bhattacharya, S., SalterTownshend, M., Wilson, S.P., Allen, J.R.M., Huntley, B. & Mitchell, F.J.G.,: Bayesian palaeoclimate reconstruction. Journal of the Royal Statistical Society: Series A (Statistics in Society), 169(3), 395- 438, 2006.

Martín-Fernàndez, J.A., Hron, K., Templ, M., Filzmoser, P. & Palarea-Albaladejo, J.,: Model-based replacement of rounded zeros in compositional data: classical and robust approaches. Computational Statistics & Data Analysis, 56(9), 2688- 2704, 2012.

Nadler, B., Lafon, S., Kevrekidis, I. & Coifman, R.R.,: Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators. In Advances in neural information processing systems 955-962, 2006.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P.,: Numerical recipes 3rd edition: The art of scientific computing. Cambridge University Press, 2007.