https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Vegetation reconstruction using plant wax *n*-alkane chain length distribution and $\delta^{13}C$ of multiple chains: A multi-source mixing model using Bayesian framework

Deming Yang [1], Gabriel J. Bowen [1]

[1]Department of Geology and Geophysics, University of Utah, Salt Lake City, UT 84112, United States

Correspondence to: Deming Yang (deming.yang@utah.edu)

Abstract. Plant wax *n*-alkane chain length distribution and $\delta^{13}C$ have been studied in modern ecosystems as proxies to reconstruct vegetation and climate of the past. Studies on modern plants often report both chain-specific *n*-alkane concentrations and $\delta^{13}C$ values. However, studies on geological archives interpret only one proxy, while both carry crucial information on the mixing sources. We propose a multi-source mixing model in a Bayesian framework that evaluates both proxies simultaneously. The model consists of priors that include user-defined source groups and their associated parametric distributions of *n*-alkane concentration and $\delta^{13}C$ with well characterized uncertainties. The mixing process involves newly defined mixing fractions such as fractional leaf mass contribution (FLMC) that can be used in vegetation reconstruction, and fractional source contribution to a specific *n*-alkane homologue (FSC_n). Markov Chain Monte Carlo is used to generate samples from the posterior distribution conditioned on both proxies. We present two case studies with distinct sets of priors. One involves *n*-C_{27}, *n*-C_{29} and *n*-C_{31} alkanes in lake surface sediments of Lake Qinghai, China. The model provides more specific interpretations on the *n*-alkane input from aquatic sources than the conventional P_{aq} proxy. The other involves *n*-C_{29}, *n*-C_{31} and *n*-C_{33} alkanes in lake surface sediments in Cameroon, western Africa. The model produces mixing fractions of forest C_3, savanna C_3, and C_4 plants, offering additional information on the dominant biomes compared to the traditional two-endmember mixing regime. FSC_n can be used to assess the interpretation of associated *n*-alkane δ^2H values, and future versions of the model incorporating lipid H isotope systematics could support integration of this proxy with C isotope and chain length distribution data. Despite the achievements, processes associated with *n*-alkane integration into sedimentary archives have not been incorporated, and the model could be further improved by adding components such as *n*-alkane turnover and transportation. Future studies on modern plants and catchment systems will be critical to develop calibration datasets that advance the strength and utility of the framework.

1 Introduction

Plant wax *n*-alkyl compounds, including *n*-alkanes, *n*-alkanoic acids, *n*-alkanols and *n*-esters are important biomarkers in paleoenvironmental reconstructions. Among them, long-chain *n*-alkanes (27–35 carbons) are the most frequently studied lipid compound class due to their great abundance in higher plants and excellent preservation in sedimentary archives

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

(Diefendorf and Freimuth, 2017; Sachse et al., 2012; Freeman and Pancost, 2014; Liu and An, 2020; e.g., Bush and Mcinerney, 2013). The use of *n*-alkanes as paleoenvironmental proxies relies on our understanding of their biosynthesis and extensive surveys of their distributions in modern plants. *n*-Alkanes of various chain lengths are synthesized via the same acetogenic biochemical pathway in all higher plants, which results in the characteristic of odd-over-even carbon chain length predominance (Chikaraishi et al., 2004; Eglinton and Hamilton, 1967; Hayes, 1993; Kunst and Samuels, 2003; Kolattukudy et al., 1976; e.g., Cheesbrough and Kolattukudy, 1984). The persistence of such chain length predominance in sedimentary archives has been used to inform the state of preservation in plant-derived *n*-alkanes (e.g., Bray and Evans, 1961; Eglinton and Hamilton, 1967; Zech et al., 2013; Buggle et al., 2010; Brittingham et al., 2017).

Despite the shared odd-over-even chain length predominance, the absolute amounts and relative abundances of *n*-alkanes differ between major plant groups, and the distribution of chain lengths in sediments has been used as a chemical fingerprint to reconstruct vegetation composition. In terms of absolute amounts, angiosperms generally produce greater quantities of long-chain *n*-alkanes than gymnosperms (Diefendorf et al., 2015; Diefendorf et al., 2011; Diefendorf and Freimuth, 2017), while terrestrial plants produce greater quantities of long-chain *n*-alkanes than submerged/emergent aquatic macrophytes (Aichner et al., 2010; Liu and Liu, 2016; Ficken et al., 2000; Liu et al., 2015; Mead et al., 2005). In terms of relative abundances, *n*-alkane chain length distributions differ between plant functional types such as trees, shrubs, forbs, grasses and succulents, and between deciduous and evergreen plants (Diefendorf et al., 2015; Diefendorf et al., 2011; Rommerskirchen et al., 2006; Vogts et al., 2009; Carr et al., 2014; Garcin et al., 2014; Magill et al., 2019; Bush and Mcinerney, 2013; Badewien et al., 2015). However, both absolute amounts and relative abundances of *n*-alkanes have been found to be highly variable within groups, as they are influenced by physiological and environmental factors such as photosynthetic pathways, leaf age, temperature, elevation, light intensity and water stress (Macková et al., 2013; Tipple et al., 2013; Koch et al., 2006; Feakins et al., 2016; e.g., Bush and Mcinerney, 2015; Griepentrog et al., 2019; Diefendorf et al., 2011; Liu et al., 2017; Suh and Diefendorf, 2018). The large uncertainties associated with characteristic *n*-alkane chain length distributions have limited their application in quantitative vegetation reconstruction (Bush and Mcinerney, 2013). Several models have been developed to reconstruct vegetation composition from *n*-alkyl lipid chain length distribution (Gao et al., 2011; Jansen et al., 2010; Peaple et al., 2021), but they all require a wide spectrum of chains to be analyzed, which may not be feasible depending on the type of sediment analyzed.

Stable carbon isotopes of chain length specific *n*-alkanes have also been used extensively to indicate changes in vegetation of the past. This is based on the well-established empirical evidence that the distribution of carbon isotope ratios (using the δ notation) of plant *n*-alkanes largely mirrors that of bulk plant tissue, primarily reflecting differences in the photosynthetic pathways among terrestrial plants (e.g., C_3 vs C_4) and the source of carbon in aquatic plants (Diefendorf and Freimuth, 2017; Liu and An, 2020; e.g., Aichner et al., 2010; Mead et al., 2005; Collister et al., 1994). For example, terrestrial C_4 plants produce more ^{13}C-enriched *n*-alkanes than C_3 plants (Rommerskirchen et al., 2006; Vogts et al., 2009; Kristen et al., 2010; e.g., Badewien et al., 2015; Bi et al., 2005). Such a distinction has been used to identify large-scale changes between C_3 and C_4 vegetation based on $\delta^{13}C$ values of *n*-alkanes extracted from sedimentary archives (e.g., Andrae et al., 2018; Uno et al.,

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

2016; Polissar et al., 2019; Schefuß et al., 2003; Niedermeyer et al., 2010; Huang et al., 2000; Freeman and Colarusso, 2001; Vogts et al., 2012; Zhou et al., 2017; Bird et al., 1995; Tipple and Pagani, 2010; Hughen et al., 2004; Schefuß et al., 2011; Castañeda et al., 2009). At a finer scale, carbon isotope compositions of *n*-alkanes from C_3 plants are influenced by both biological and environmental factors, such as taxonomy, plant functional types, water availability, temperature, and altitude (Diefendorf et al., 2010; Liu and An, 2020; Diefendorf et al., 2015; Wu et al., 2017; e.g., Cernusak et al., 2013; Wang et al., 2018). Compared to the uncertainties associated with *n*-alkane chain length distribution, the uncertainties of $\delta^{13}C$ values in C_3 and C_4 plants are relatively constrained and have been incorporated into reconstructions of vegetation cover (Garcin et al., 2014; e.g., Andrae et al., 2018; Polissar et al., 2019; Uno et al., 2016). However, when the reconstructions are based on a linear mixing relationship with a careful selection of one specific *n*-alkane compound (often *n*-C_{29} or *n*-C_{31}), they are associated with the challenge that *n*-alkane production of the selected chain could vary between the end members. If the uncertainties associated with *n*-alkane production are not account for, it is possible that the uncertainty associated with paleo-vegetation reconstructions could be mischaracterized, and that the interpretations could be biased.

Traditional interpretations of *n*-alkane proxies often rely on one line of evidence: either chain length distribution or $\delta^{13}C$ values of one chain. While many studies have reported $\delta^{13}C$ values and relative abundance of multiple *n*-alkane chains with their respective uncertainties, they are often interpreted independently, using the interpretation of one to qualitatively support that of another. A more rigorous approach to interpreting these data could be developed through the use of a proxy system model (Evans et al., 2013), in which the common and unique assumptions underlying the interpretation of both chain-length and isotope data are explicitly represented. Such a model would provide a quantitative framework within which one proxy type can be used to constrain the uncertainties involved in the interpretation of the other, and vice versa. In the case of plant *n*-alkanes, both chain length distribution and $\delta^{13}C$ values of multiple chains carry complementary information related to the mixing of compounds from different vegetation sources. The combination of both proxies, therefore, may help to better constrain uncertainties, reduce biases, and refine our interpretations. This study describes a generic multi-source *n*-alkane mixing model that incorporates both lines of evidence (chain lengths, carbon isotopes) and their associated uncertainties in the interpretation. Using case studies based on published sedimentary *n*-alkane records, the primary goals of this study are: 1) to demonstrate model utility by proposing and evaluating multi-source *n*-alkane mixing regimes; 2) to demonstrate model potential by providing informed interpretations on the mixing regimes with well characterized uncertainties.

2 Methods

2.1 Model structure

The proposed approach is achieved in a Bayesian hierarchical modeling framework (Figure 1), which can leverage information from multiple proxies to provide a robust statistical basis for proxy integration. The hierarchical model is then inverted using Markov Chain Monte Carlo (MCMC) methods (Geman and Geman, 1984) to obtain posterior parameter estimates that are conditioned simultaneously on all proxy data. Similar modeling approaches have been applied to meta-

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

analyses of paleoclimatic/vegetation proxies (e.g., Bowen et al., 2020; Tingley et al., 2012; Li et al., 2010; Garreta et al., 2010), but have not been specifically proposed for proxy interpretation of *n*-alkanes.

2.1.1 Prior distributions

Our understanding of species level *n*-alkane chain length distribution and chain specific $\delta^{13}C$ values is based on empirical evidence from extant plants (Diefendorf and Freimuth, 2017; Sachse et al., 2012; Bush and Mcinerney, 2013). The extensive record of literature and the published empirical data form the basis for prescribing the prior chain-length distributions used in this work.

n-Alkane concentrations (µg/g of dried leaves) are influenced by the taxonomy, growth forms, and growing conditions of the plant (Bush and Mcinerney, 2013; Magill et al., 2019; Vogts et al., 2009; Diefendorf et al., 2011; Rommerskirchen et al., 2006; Han et al., 1968; Ficken et al., 2000; Bush and Mcinerney, 2015; Andrae et al., 2019). Within a source group that is associated with a specific combination of these factors, we expect *n*-alkane concentration to follow a group-specific chain length distribution pattern. The distribution of concentration values for a given chain within a well-defined source group typically follows a log-normal distribution (Garcin et al 2014). Therefore, we assume that for each sample drawn from source group *i* (Figure 1A), *n*-alkane concentrations are derived from a multivariate log-normal distribution:

$$Conc_{1:n,i} \sim multi - ln - N\ (\mu_{1:n,i}, \Omega_{1:n,i})\ , \quad (1)$$

where $\mu_{1:n,i}$ are the means of the natural log-transformed *n*-alkane concentration values, for chains 1 to *n*; $\Omega_{1:n,i}$ is a $n \times n$ variance-covariance matrix of the natural log-transformed *n*-alkane concentration values calculated from empirical data for each source group *i*.

Similarly, *n*-alkane $\delta^{13}C$ values are primarily influenced by taxonomy, photosynthetic pathways, and growing conditions of the plant (Diefendorf et al., 2010; Liu and An, 2020; Diefendorf et al., 2015; Wu et al., 2017; Cernusak et al., 2013; Wang et al., 2018; Rommerskirchen et al., 2006; Vogts et al., 2009; Kristen et al., 2010; e.g., Badewien et al., 2015; Bi et al., 2005). Within a source group *i* that is associated with a specific combination of these factors, we expect *n*-alkane $\delta^{13}C$ to follow a group-specific distribution pattern. Because *n*-alkanes of different chain lengths are synthesized via the same biochemical pathway, $\delta^{13}C_{1:n,i}$ of all chains should be correlated (Chikaraishi et al., 2004; Eglinton and Hamilton, 1967; Hayes, 1993; Cheesbrough and Kolattukudy, 1984; Kunst and Samuels, 2003). Therefore, we assume that for each sample drawn from source group *i* (Figure 1A), $\delta^{13}C_{1:n,i}$ are derived from a multivariate normal distribution:

$$\delta^{13}C_{1:n,i} \sim multi - N\ (\mu^*_{1:n,i}, \Omega^*_{1:n,i})\ , \quad (2)$$

where $\mu^*_{1:n,i}$ are the means of $\delta^{13}C$ for chains 1 to *n*; $\Omega^*_{1:n,i}$ is a $n \times n$ variance-covariance matrix calculated from empirical data for each source group *i*.

Calculation of the variance-covariance matrices does not allow missing values in the empirical dataset. Therefore, data entries with any missing value are removed before calculation of the variance-covariance matrices. The prior distribution parameters allow random samples to be drawn from the prior distributions and used in the Process model calculation (Step 1

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

in Figure 1). By using prior distributions derived entirely from modern plant *n*-alkane data to inform the interpretation of sedimentary alkane data, the version of the model implemented here implicitly assumes no change in the $\delta^{13}C$ value of atmospheric CO_2, and minimal incorporation of reworked and potentially pre-aged *n*-alkanes into sedimentary archives.

2.1.2 Process model

The model consists of a generic mixing process with multiple sources, following the principle of isotope mass balance. First, a fixed number of random draws (*k*) from the prior distributions are used to generate samples of $Conc_{n,i}$ and $\delta^{13}C_{n,i}$ to characterize chain *n* for each source group *i* (similar to Figure 1A). Concentration values are in units of µg of *n*-alkane per gram of dried leaves, and the random draws thus give a distribution of alkane yields per unit leaf mass sampled from the prior. Both concentration and isotope data used to specify the prior distributions typically represent measurements from individual plants, and the process of drawing these *k* random samples is intended to represent the production of an integrated sample from *k* individuals. Although in many cases we would expect more than 50 individuals of a particular source group to contribute to a sedimentary sample, k = 50 was chosen to balance sample representation of prior distributions, process stochasticity, and algorithmic complexity. In general, larger values of *k* are more prescriptive in that the random sample will conform more closely to the prior distribution, and we consider k = 50 to be appropriately conservative in terms of allowing some variation in the samples drawn from the priors.

Second, we calculate the mixture of alkanes from all groups as a function of the fractional leaf mass contribution (*FLMC*) where $FLMC_i$ gives the relative leaf biomass from group i contributing alkanes to the sedimentary sample. Because $\sum_i FLMC_i = 1$; $FLMC_1,\cdots,FLMC_i$ follow a Dirichlet distribution:

$$FLMC_1,\cdots,FLMC_i \sim Dir\ (\alpha_1,\cdots,\alpha_i)\ , \quad (3)$$

Where $\alpha_1,\cdots,\alpha_i$ are the Dirichlet concentration parameters. Using an uninformative prior assuming equal mixture of the sources, the Dirichlet concentration parameters are set to be the same. The values $\alpha_1 = \cdots = \alpha_i = 1$ are chosen assuming a uniform probability of $FLMC_i$ within its range (0, 1). It is important to note that potential differences in leaf mass or lipid turnover are not incorporated into this formulation: the values of $FLMC_i$ may not be directly related to the standing biomass of source *i* if the alkane and/or leaf turnover rates for the groups differ.

Third, the weighted average relative abundance (*RA*) of each *n*-alkane *n* in the mixture can be calculated from the concentrations and mixing fractions:

$$RA_{n,mix} = \frac{\sum_i (FLMC_i \sum Conc_{n,i})}{\sum_i^n (FLMC_i \sum Conc_{n,i})}\ , \quad (4)$$

where in the numerator, $\sum Conc_{n,i}$ is the sum of concentrations from the random draws (*k*) for chain *n* and source group *i*; $\sum_i (FLMC_i \sum Conc_{n,i})$ is the fraction-weighted sum of concentrations for chain *n* of all source groups; in the denominator, $\sum_i^n (FLMC_i \sum Conc_{n,i})$ is the total *n*-alkane concentration for all chains and all source groups. Eq. (4) dictates that $\sum RA_{n,mix} \equiv 1$.

Lastly, for chain *n* and its $\delta^{13}C_{n,mix}$, according to isotope mass balance:

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

$$\delta^{13}C_{n,mix} = \frac{\sum_i [FLMC_i \sum (Conc_{n,i} \times \delta^{13}C_{n,i})]}{\sum_i (FLMC_i \sum Conc_{n,i})}, \quad (5)$$

where in the numerator, $\sum(Conc_{n,i} \times \delta^{13}C_{n,i})$ is the sum of the products of *n*-alkane concentration and δ^{13}C in each round of random draws (k) for source i; $\sum_i[FLMC_i \sum(Conc_{n,i} \times \delta^{13}C_{n,i})]$ is the fraction-weighted sum of the sums for chain n of all source groups; in the denominator, $\sum_i(FLMC_i \sum Conc_{n,i})$ is the same as in the numerator in Eq. (4). Eq. (5) is essentially a concentration weighted isotope mass balance equation.

The process model specifies the numerical relationships between random samples from the prior distributions and the simulated metrics of interest ($RA_{n,mix}$, $\delta^{13}C_{n,mix}$). The simulated metrics are then used in the Data model for further evaluations (Step 2 in Figure 1).

2.1.3 Data model and model inversion

All proxy data are subject to errors that are associated with the proxy observations themselves (Evans et al., 2013). Therefore, $RA_{n,mix}$ and $\delta^{13}C_{n,mix}$ are modeled with their respective measurement error (Figure 1C). Because each of the measured RA values are ratios by nature, they are assumed to be associated with a Cauchy error term while centering around the true value $RA_{n,mix}$:

$$RA_{n,mea} \sim Cauchy\left(RA_{n,mix}, \tau_{n,mea}\right), \quad (6)$$

$$\left(\tau_{n,mea}\right)^2 \sim Gamma\ (shape = 5, rate = 0.05), \quad (7)$$

where $\tau_{n,mea}$ is the measurement error of $RA_{n,mea}$, and the hyper parameters *shape* and *rate* are set to 5 and 0.05, respectively, corresponding to the small error margins of *n*-alkane concentration measurements and the associated RA values, which are often reported to the third decimal digit.

Because the uncertainty of the measured δ^{13}C values typically follows a normal distribution, they are assumed to be associated with a Gaussian error term while centering around the true value $\delta^{13}C_{n,mix}$:

$$\delta^{13}C_{n,mea} \sim N\left(\delta^{13}C_{n,mix}, \sigma_{n,mea}\right), \quad (8)$$

where $\sigma_{n,mea}$ is the measurement error of $\delta^{13}C_{n,mea}$, which is often reported as 1 standard deviation (sd) in the literature.

In the model inversion step (Steps 3 to 5 in Figure 1), MCMC is used to propose samples of all model parameters conditioned on the measured δ^{13}C and RA results.

2.1.4 Model implementation

The model structure described above is coded in the BUGS (Bayesian inference Using Gibbs Sampling) language (Lunn et al., 2012), and implemented in R version 4.0.5 (R Core Team, 2021), using the "rjags" package with the standalone JAGS (Just Another Gibbs Sampler) encoder installed separately (Plummer, 2021). Three chains are run in parallel, and the number of iterations is set at 800,000 to ensure model convergence, with the first 200,000 interactions as burn-ins. Chain thinning is set at one per 240 iterations. Convergence is assessed visually via trace plots (R package "mcmcplots", Curtis, 2018) and

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

with reference to the convergence factor "rhat" (Gelman and Rubin, 1992) and effective sample sizes reported by the "rjags" package. The iteration parameters are chosen to ensure complete convergence with rhat values smaller than 1.01. Average run time is approximately 2 hours.

2.1.5 Model output

Once the model inversion is completed, for each model parameter, its posterior density is summarized via the Kernel density estimation function "density" in the R package "stats" with default settings. For selected parameters, posterior density summaries such as the highest density, median density, and the 89% highest density interval (HDI), are reported using functions in the R package "bayestestR" (Makowski et al., 2019). Out of all the parameters, the posterior densities of two parameters are of special interest in the mixing process. One is $FLMC_i$ (as in Eq. 3, Figure 1E). The other is derived from the posterior samples generated by the MCMC process, which we term as the fractional contribution of source *i* to each specific chain *n* ($FSC_{n,i}$, Figure 1E):

$$FSC_{n,i} = \frac{FLMC_i \sum Conc_{n,i}}{\sum_i (FLMC_i \sum Conc_{n,i})}, \quad (9)$$

where the numerator is the fraction-weighted sum of randomly drawn (*k*) concentrations for chain *n* of source *i*; the denominator is the same as the numerator of Eq. (4). Bivariate density plots and contour lines are used to explore covariation patterns of $FLMC_i$ among the sources.

2.2 Case studies

Here we provide two case studies that demonstrate model characteristics and offer alternative interpretations of previously published *n*-alkane proxy data. We considered that *n*-alkanes extracted from lake surface sediments fulfills most of the assumptions associated with the model structure, and therefore, chose lake surface sediments as the archive of interest. The case studies are not intended to formally test or validate the modeling approach but to illustrate how the model framework can offer an accessible and statistically informative approach for interpretations of *n*-alkanes and similar proxy data.

2.2.1 Evaluating aquatic and terrestrial n-alkane input in sediment samples of Lake Qinghai

2.2.1.1 Background

Lacustrine sediments can incorporate *n*-alkanes that are sourced from both terrestrial and aquatic plants (Aichner et al., 2010; Ficken et al., 2000; Mead et al., 2005). Quantifying contributions from the different sources to specific *n*-alkanes is important because it guides paleoenvironmental interpretation of *n*-alkane proxies. For example, *n*-alkane chain length distribution and stable isotopes ($\delta^{13}C$ and $\delta^{2}H$) have been used to investigate the composition of sedimentary organic matter, and regional hydrological conditions through time (e.g., Aichner et al., 2010; Mead et al., 2005; Huang et al., 2004; Sachse et al., 2004; Schwark et al., 2002; Ficken et al., 1998). Many studies use the proxy $P_{aq} = (C_{23} + C_{25})/ (C_{23} + C_{25} + C_{29} + C_{31})$ (Ficken et al., 2000), as an indicator of terrestrial/aquatic source dominance, to interpret measured isotope ratios of

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

sedimentary *n*-alkanes. It quantifies the relative proportion of mid-chain (*n*-C_{23} and *n*-C_{25}) vs. long chain homologs, based on the evidence that mid-chain alkanes are largely produced by aquatic plants, while long chain alkanes (*n*-C_{29} and *n*-C_{31}) are largely produced by terrestrial plants (Ficken et al., 2000; Gao et al., 2011; Wang and Liu, 2012; Liu and Liu, 2016; e.g., Aichner et al., 2010; Duan and Xu, 2012). However, long chain alkanes (e.g., *n*-C_{27} and *n*-C_{29}) have been found in aquatic macrophytes, while mid-chain alkanes (e.g., *n*- C_{25}) have been found in some terrestrial plants (Duan and Xu, 2012; Wang and Liu, 2012; Gao et al., 2011; Liu et al., 2015; e.g., Aichner et al., 2010; Dion-Kirschner et al., 2020). These findings add some uncertainty to the use of P_{aq} and similar indices as indicators of *n*-alkane contributions from aquatic plants.

$\delta^{13}C$ values of *n*-alkanes have also been used as an indicator of terrestrial or aquatic source dominance (e.g., Aichner et al., 2010; Mead et al., 2005; Liu et al., 2015). This is based on the observation that some aquatic macrophytes that grow in shallow waters display $\delta^{13}C$ values that are distinctive from those of terrestrial C_3 plants and algae due to their ability to utilize dissolved bicarbonate ions as their carbon source (Prins and Elzenga, 1989; Keeley and Sandquist, 1992; e.g., Allen and Spence, 1981). However, the interpretation of $\delta^{13}C$ values for a specific *n*-alkane homologue using a linear two-end mixing relationship will be sensitive to differences in *n*-alkane concentrations between the end member vegetation types, leading to contradictory interpretations from different compounds. The presence of additional (beyond two) potential sources further complicates the interpretation of such data.

To demonstrate how our model can provide quantitative estimates of *n*-alkane mixing fractions of multiple aquatic and terrestrial plant sources, simultaneously conditioned on isotopic and concentration data for all measured chain lengths, we apply the model using published lake surface sediment data from three locations in Lake Qinghai.

2.2.1.2 Data compilation

The region of interest is the Qinghai-Tibetan Plateau in western China, where both freshwater and saline lakes are abundant. The land cover of the region is dominated by alpine meadow, steppe and shrubland, which consist of almost exclusively C_3 plants (Duan and Xu, 2012; Yu et al., 2001; Liu et al., 2015). Most of the lakes in the region are shallow and densely populated with submerged aquatic macrophytes and green algae during the short growing season (Aichner et al., 2010; Liu and Liu, 2016; Liu et al., 2015). Based on these potential sources of *n*-alkanes in the region, we define three source groups:1) terrestrial plants, 2) aquatic macrophytes, and 3) algae.

Per sample *n*-alkane concentrations (µg/g of dried sample) and $\delta^{13}C$ (‰, VPDB) of *n*-C_{27}, *n*-C_{29}, and *n*-C_{31} alkanes are compiled from the literature for terrestrial plants in the region, and aquatic macrophytes and algae in the lakes (Aichner et al., 2010; Liu and Liu, 2016; Liu et al., 2015). The compiled dataset (Supplementary Material) is used to provide parameter estimates for the prior distributions (as explained in section 2.1.1) for each source. Parameters of the multivariate normal distributions are estimated using the R packages “stats” (R Core Team, 2021). To evaluate the assumption that *n*-alkane concentrations follow a log-normal distribution for each chain in each source, the distribution parameters are estimated using the R package “EnvStats” (Millard, 2013). Goodness of fit is visually assessed using quantile-quantile plots (Supplementary Material, Figures S1 and S2). *n*-alkane concentrations and $\delta^{13}C$ values of selected lake surface sediments from Lake Qinghai

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

are compiled from Liu et al. (2015). Other data that characterize the composition and preservation of *n*-alkanes in the geological archives, such as the P_{aq} proxy, are also referenced in the data compilation (Table 1). Because the data are modeled to follow multivariate normal/log-normal distributions (section 2.1.1), the prior distributions should be theoretically assessed in a multivariate space, but such assessment would be difficult to visualize. Instead, for each variable in the multivariate space, the multivariate density is "collapsed" onto a single dimension by integrating the densities in other dimensions (as in Figure 2), using the function "omxMnor" in the R package "OpenMx" (Neale et al., 2016).

2.2.2 Evaluating n-alkane input along a vegetation gradient in lake sediments in western Africa

2.2.2.1 Background

Biome composition and its change over time have profound implications on climatic shifts, and the evolution of the biosphere. The tropical grassland biome has been of particular interest because of its unique $\delta^{13}C$ signature and its important association with climate of the past and mammalian evolution (Polissar et al., 2019; Cerling, 1992; Latorre et al., 1997; Quade and Cerling, 1995; Kaya et al., 2018; Janis et al., 2002; e.g., Bobe and Behrensmeyer, 2004; Strömberg, 2011). Many studies used $\delta^{13}C$ values of long chain *n*-alkanes to provide estimates of C_3/C_4 vegetation cover, based on linear mixing calculations using data from one or more chain lengths (Garcin et al., 2014). In many cases, *n*-C_{31} alkane is used, as its concentration is often similar in C_3 and C_4 plants (e.g., Polissar et al., 2019; Uno et al., 2016). However, this approach does not take advantage of the distinct chain length dominance patterns associated with different biomes. For example, C_4 grasses are associated with much higher production of the *n*-C_{33} alkane than most C_3 plants (Rommerskirchen et al., 2006; Bush and Mcinerney, 2013; Garcin et al., 2014; Krull et al., 2006), while tropical rainforest plants are associated with the very high production of the *n*-C_{29} alkane, (Vogts et al., 2009). Using the information from chain length distribution in addition to $\delta^{13}C$ values has the potential to produce more detailed reconstructions of biome composition beyond the binary C_3/C_4 ratio.

To demonstrate how the model can provide quantitative estimates of *n*-alkane contributions from multiple broadly defined biomes, we apply the model using published lake surface sediment data from three locations in a vegetation gradient in Cameroon, western Africa.

2.2.2.2 Data compilation

The region of interest is Sub-Saharan Africa where rainfall amount and seasonality are the primary determinants of biome types (Sankaran et al., 2005; Aleman et al., 2020). In western Africa, vegetation cover is dominated by rainforest close to the Equator, wooded grassland to the north, and a transitional zone in between (Rommerskirchen et al., 2003; Garcin et al., 2014; Garcin et al., 2012; Huang et al., 2000; Schwab et al., 2015). In particular, tropical forest and savanna woody plant species have been shown to co-occur in this region (Aleman et al., 2020), making it an ideal place to investigate the potential of using *n*-alkane proxies to reconstruct the dominant C_3 biome. Based on the above information, we define three source groups: 1) tropical C_4 plants, 2) savanna C_3 plants, and 3) rainforest C_3 plants. Succulent plants are excluded in the analysis due to their low presence in the region.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Per sample *n*-alkane concentrations (μg/g of dried sample) and $\delta^{13}C$ (‰, VPDB) of *n*-C_{29}, *n*-C_{31}, and *n*-C_{33} alkanes are compiled from the literature for terrestrial plants in western, southwestern, and eastern Africa (Diefendorf and Freimuth, 2017; Ali et al., 2005; Magill et al., 2019; Kristen et al., 2010; Badewien et al., 2015; Vogts et al., 2009; Rommerskirchen et al., 2006). The compiled dataset (Supplementary Material) is used to provide parameter estimates for the prior distributions for each source (as explained in section 2.1.1). Prior parameters are estimated using the same methods as in section 2.2.1.2. Goodness of fit is visually assessed using quantile-quantile plots (Supplementary Material, Figures S2). *n*-Alkane concentrations and $\delta^{13}C$ values of selected lake surface sediments are compiled from Garcin et al. (2014), with estimated fractional C_4 vegetation cover (f_{C4}) referenced from the same publication (Table 2). Following the same methods as in section 2.2.1.2, univariate prior distributions are plotted in Figure 3 as density curves.

2.3 Sensitivity tests

The model results are to some degree sensitive to the prior parameter estimates, which are derived from empirical data but imperfectly known in our case studies. Here we use data associated with case study 2 to explore the influence of prior parameter estimates on model output. To produce a different set of prior parameter estimates, plant samples are selected from western Africa only, as a subset of the sub-Saharan empirical dataset (123 entries out of the 301 entries from the original dataset, Supplementary Material). Prior parameters are estimated from this dataset using the same methods as described in section 2.3.2.2. The resultant prior distributions differ from those of the sub-Saharan dataset primarily in the C_4 source, in that it displays lower *n*-alkane concentrations and less negative $\delta^{13}C$ values (Figure 4). Model sensitivity is assessed via a comparison between the model outputs based on the different priors.

3. Results

3.1 Case study 1

For each of the three data points of the Lake Qinghai case study (Figure 5A – C), the posterior densities of FLMCs of terrestrial plants and aquatic macrophytes vary substantially between the samples (Figure 5D – F), while the distributions of the algal FLMC are almost the same between the samples. FLMC from aquatic macrophytes is the highest in the High P_{aq} sample (QHS13-5S, P_{aq} = 0.34, Table 1), at 0.53 (MAP, summary statistics in Table 3). In comparison, FLMC from aquatic macrophytes is the lowest (MAP close to 0, Table 3) in the Mid-P_{aq} sample (QHS13-7S, P_{aq} = 0.22, Table 1). FLMC from aquatic macrophytes is intermediate in the Low P_{aq} sample (QHS13-9S, P_{aq} = 0.19, Table 1), at 0.09 (MAP, Table 3).
Although the algal FLMC is not well constrained (Figure 5), a strong "trade-off" correlation between algal FLMC and those of the other two sources is apparent in bivariate density plots (Figure 6). The High P_{aq} (QHS13-5S) and Low P_{aq} (QHS13-9S) samples display multiple local modes for FLMCs, represented by multiple bivariate density peaks (as indicated by the contour lines and color gradient, Figure 6).

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

The posterior densities of the algal FSC_n to the sedimentary long chain *n*-alkane pool are consistently low (Figure 7). In comparison, the FSC_{27} of aquatic macrophytes can be substantial in both the High P_{aq} (QHS13-5S) and Low P_{aq} (QHS13-9S) samples, while FSC_{29} of aquatic macrophytes is substantial only in the High P_{aq} sample (QHS13-5S). The terrestrial source shows consistently high FSC_{31}s across all samples.

3.2 Case study 2

For the western African transect case study (Figure 8A – C), the posterior densities for fractional leaf mass contribution of C_4 plants, savanna C_3 plants, and rainforest C_3 plants vary substantially between samples (Figure 8D – F). The C_4 plant contribution is the highest in the High C_4 sample (Rhum, f_{C4} = 0.72, Table 2), at 0.59 (MAP, Table 4). The FLMC of C_4 plants is lowest in the Low C_4 sample (Baro, f_{C4} = 0.05, Table 2), at 0.10 (MAP, Table 4). FLMCs from the savanna C_3 and rainforest C_3 plants are generally less strongly constrained than those of C_4 plants, except for in the Mid-C_4 sample (Asso), which shows a low FLMC from the savanna C_3 plants (MAP close to 0, Figure 8E, Table 4). The MAPs of FLMCs of C_4 plants of the High C_4 (Rhum) and the Low C_4 (Baro) samples fall within the reported mean ± 1sd ranges for f_{C4} at these sites based on satellite imaging (Garcin et al., 2014), (Table 2). The MAP of FLMC ofC_4 plants of the Mid-C_4 sample (Asso) falls outside the reported f_{C4} mean ± 1sd range.

Bivariate density plots show relatively weak "trade off" correlation patterns between the FLMCs of savanna and rainforest C_3 sources in the High C_4 sample (Rhum) and Low C_4 sample (Baro, Figure 9). All three samples display a single mode of most likely vegetation contributions, represented by a single peak in the bivariate density plots (as indicated by the contour lines and color gradient, Figure 9), although a secondary mode with higher rainforest and lower savanna C_3 FLMCs exists at Rhum.

The distributions of FSC_n of all three sources to each specific chain (*n*-C_{29}, *n*-C_{31} and *n*-C_{33}) vary substantially between samples. Within each sample, the rainforest C_3 source consistently shows a higher FSC_{29} value over FSC_{31} and FSC_{33} (Figure 10). In comparison, the C_4 source shows a lower FSC_{29} value over FSC_{31} and FSC_{33} with the exception of the High C_4 sample (Rhum), in which the C_4 FSC_{29} and FSC_{31} are almost identical (Figure 10). Within each sample, the uncertainties of FSC_n vary among the chains, with the High C_4 sample (Rhum) exhibiting the highest overall uncertainty in FSC_{29} and FSC_{31}, and the Low C_4 sample (Baro) exhibiting the highest overall uncertainty in FSC_{33}. The High C_4 sample (Rhum) also shows bimodal distributions in FSC_{29} and FSC_{31} for the rainforest C_3 and savanna C_3 sources. Among the samples, the FSC_n uncertainty of the C_4 source is the lowest.

3.3 Model sensitivity

3.3.1 Sensitivity to parameters of the prior distributions

Using the western Africa prior dataset (Figure 4), with lower *n*-alkane concentration and less negative $\delta^{13}C$ values in the C_4 source, resulted in consistent shifts to higher values of FLMC of the C_4 source relative to the results obtained with the sub-

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Saharan prior dataset (Figure 11). With the published data on fractional C_4 vegetation cover as a reference (Garcin et al., 2014), this set of priors produces FLMC of the C_4 source similar to the reported fractional C_4 vegetation cover (f_{C4}) at the High C_4 site (Rhum), but produces much higher FLMCs than published f_{C4} values at the Mid- and Low C_4 sites (Asso and Baro).

3.3.2 Sensitivity to proxy type

The model shows different sensitivity to chain length distribution and carbon stable isotopes. It is relatively insensitive to chain length distribution: only minor changes in the posterior densities are observed when the likelihood evaluations of RA are removed (left column, Figure 12). By contrast, the model is much more sensitive to the *n*-alkane $\delta^{13}C$ data: the central tendencies of the posterior densities shifted substantially when the likelihood evaluations are removed (right column, Figure 12). Substantial increases in the dispersion of the posterior densities are also observed, indicating more weakly constrained FLMCs (right column, Figure 12).

Discussion

4.1 Interpretations of results

4.1.1 Case study 1

Long chain *n*-alkanes (especially 27–31 carbons) in lake sediments are possibly a mixture of both terrestrial and aquatic sources (e.g., Aichner et al., 2010; Liu and Liu, 2016; Dion-Kirschner et al., 2020; Duan and Xu, 2012; Wang and Liu, 2012). Based on model output, both FLMC and FSC_{27} of aquatic macrophytes are not negligible in the Low P_{aq} sample (QHS13-9S, Figures 5 and 7). By contrast, FLMC and FSC_{27} of aquatic macrophytes are minimal in the Mid-P_{aq} sample (QHS13-7S, Figures 5 and 7). These results suggest that P_{aq} alone may not be a reliable indicator of the relative aquatic macrophyte input into the sedimentary archive. In addition, for the High P_{aq} sample (QHS13-5S), FSC_{29} of aquatic macrophytes is not negligible, while FSC_{31} of the terrestrial source is consistently close to 1 among the samples (Figure 7). This suggests caution when interpretations of terrestrial paleoenvironment are based on the isotope ratios of *n*-C_{29} alkane alone, while previous interpretations based on the *n*-C_{31} alkane are likely more reliable.

The relatively unconstrained FLMC of the algae source (Figure 5) and its tradeoff with that of the terrestrial source (Figure 6) suggest that the possibility of algae being an important biomass source of the lake surface sediment cannot be eliminated, and would impact the interpreted contributions from other sources. This possibility is consistent with the observations that the deep water lake bottom is covered mainly by green algae in Lake Qinghai (Liu et al., 2015). The model approach successfully identified such a possibility, despite the consistently low FSC_{27} of the algal source (Figure 7), most likely due to its limited production of the *n*-C_{27} alkane (Liu et al., 2015). Algae produce greater amounts of short-chain *n*-alkanes than mid-chain and long-chain homologues (Han and Calvin, 1969; Gelpi et al., 1970; Cranwell et al., 1987). Theoretically, the

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

inclusion of *n*-C_{25} or even shorter chains would provide better constraints on both the aquatic macrophyte and the algae sources. However, this option proved impractical due to the many missing values of *n*-C_{25} in the empirical records of terrestrial plants. The other concern would be that mid-chain *n*-alkanes are also produced by microorganisms (Brittingham et al., 2017; Grimalt et al., 1988; Ladygina et al., 2006; Park, 2005), the influence of which is difficult to assess when only odd-chain alkanes are analyzed.

The Tibetan Plateau has been argued to be an ideal region to investigate the input from aquatic macrophytes to the organic matter/*n*-alkane pool in lake sediments, due to the minimal presence of terrestrial C_4 plants in the region (Duan and Xu, 2012; Liu et al., 2015; Aichner et al., 2010; Wang, 2003). While the carbon isotope values of these two *n*-alkane sources largely overlap (Aichner et al., 2010; Liu et al., 2015), they display different chain length dominance patterns: aquatic macrophytes produce relatively high proportions of the *n*-C_{27} alkane, while terrestrial C_4 plants produce relatively high proportions of the *n*-C_{31} alkane (Liu et al., 2015; Badewien et al., 2015; Rommerskirchen et al., 2006; Bush and Mcinerney, 2013; Magill et al., 2019). The consistently low $\delta^{13}C$ values of the *n*-C_{31} alkane among the samples confirm the minimal contribution from terrestrial C_4 plants. In a hypothetical situation where both aquatic macrophytes and terrestrial C_4 plants are present, increased aquatic macrophyte contribution would increase the $\delta^{13}C$ value of the *n*-C_{27} alkane more than the *n*-C_{29} alkane, while increased terrestrial C_4 contribution would increase the $\delta^{13}C$ value of the *n*-C_{31} alkane more than the *n*-C_{29} alkane. The proposed modeling approach is theoretically capable of distinguishing the *n*-alkane input from aquatic macrophytes and terrestrial C_4 plants by leveraging both chain length distribution and $\delta^{13}C$ values of these *n*-alkane chains. While we do not have a case study to clearly demonstrate such a possibility, future studies are encouraged to explore this further.

4.1.2 Case study 2

This case study demonstrates that the proposed framework can leverage chain length distribution and $\delta^{13}C$ of multiple *n*-alkane chains and offer interpretations on the mixing ratios of multiple sources. The Mid-C_4 sample (Asso) shows a high relative abundance of *n*-C_{29} alkane and contrasting $\delta^{13}C$ values among the *n*-alkane chains (Figure 8), which suggests that the most likely mixing regime of the sample is a much higher FLMC of rainforest C_3 plants than savanna C_3 plants. For the high C_4 sample (Rhum) the model also recovered bimodal mixing possibilities between the rainforest and savanna C_3 sources (Figure 10). Such information on the possible mixing patterns of the C_3 biomes cannot be achieved with the traditional two-end member mixing model using *n*-alkane $\delta^{13}C$ values alone.

FLMC as a metric for vegetation reconstruction is not directly comparable with f_{C4} estimated from remote sensing imagery (e.g., MODIS Vegetation Continuous Fields). A carefully selected calibration dataset could support a rigorous comparison of the two metrics. On the other hand, assuming that f_{C4} estimated from satellite imagery is reflecting the true vegetation distribution contributing to the study samples, the differences between FLMC and f_{C4} could inform possible biases associated with FLMC. As presented, the central tendencies of the C_4 FLMC at the Low and Mid-C_4 sites are higher than f_{C4}, while C_4 FLMC at the high C_4 site is lower than f_{C4} (Figure 8). Such a pattern of biases cannot be explained by a higher leaf mass to

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

vegetation cover ratio of the C_3 biomes. One possible explanation is with a spatial pattern of *n*-alkane sourcing and integration. Tropical African lakeshores often host a gradient of mixed vegetation from C_4 grasses/sedges to C_3 trees/shrubs (Vesey-Fitzgerald, 1963; Greenway and Vesey-Fitzgerald, 1969; Howard-Williams and Walker, 1974). If *n*-alkanes in lake sediments are sourced consistently more from the immediate lakeshore region than from further away, the sediments would more likely integrate a mixed vegetation signal than the surrounding regions, which is consistent with the observed differences between FLMC and f_{C4}.

4.1.3 Sensitivity tests

The results of sensitivity test 1 (Figure 11) show that using a different set of prior distributions (Figure 4) can produce somewhat different central tendencies in FLMCs with the same sedimentary *n*-alkane data. This demonstrates that the outputs are model dependent: they provide a basis for interpreting proxy data in the context of a specific model with its associated priors and assumptions. Using the published f_{C4} values as a reference, Prior 2 (the western Africa prior) is associated with overall higher estimated C_4 FLMCs than those with Prior 1 (the sub-Saharan Africa prior, Figure 11). This unidirectional shift in FLMCs is consistent with the expectation of a systematic difference between the two prior distributions. Based on the comparison with satellite-based vegetation cover, this further suggests that Prior 1 is perhaps a better representation of the vegetation sources that produced the *n*-alkane mixtures in the sedimentary samples than Prior 2 (given the caveats of this comparison, discussed above).

The results of sensitivity test 2 (Figure 12) show that $\delta^{13}C$ places stronger constraints on the results than chain length distribution. This demonstrates the importance of stable isotopes as tracers in compound specific mixing models. In contrast to other carbon-isotope based methods, the model proposed here might still provide reasonable resolution of sources when the chain-length distribution of the sources is substantially different.

4.2 Model achievements

The proposed model framework offers a more integrative approach to interpretation of *n*-alkane records that could better leverage these rich data to support more nuanced and specific reconstructions of paleo-vegetation. It has potential advantages over traditional approaches to the interpretation of sedimentary *n*-alkanes.

First, the model provides simultaneous evaluation of chain length distribution and $\delta^{13}C$ of multiple *n*-alkane chains: patterns and uncertainties associated with *n*-alkane production and $\delta^{13}C$ are both incorporated into the posterior distributions. Specifically, it offers a numerical solution (Section 2.1.2) that accounts for the large uncertainty associated with *n*-alkane production among the sources (Figures 2 and 3), which is coupled with the mass balance equation of *n*-alkane $\delta^{13}C$ (Eq. 5). This allows $\delta^{13}C$ to be used to constrain the uncertainty associated with *n*-alkane chain length distributions, which have previously been used in isolation to reconstruct vegetation (Gao et al., 2011; Jansen et al., 2010; Peaple et al., 2021). The model also addresses a common assumption in the interpretation of lipid $\delta^{13}C$ data via linear mixing relationships, in which

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

n-alkane production for a selected chain is assumed to be the same among all sources (Bush and Mcinerney, 2013; Garcin et al., 2014).

Moreover, the case studies demonstrate that the model can be used to explore mixing regimes of multiple sources, which offers alternative interpretations compared to the traditional two-end member mixing regime using $\delta^{13}C$ of one *n*-alkyl lipid. The multi-source framework can be used to assist the interpretations of other environmental proxies. For instance, FSC_n in both case studies (Figures 7 and 10), with well-characterized uncertainties associated with each source, can offer additional information on the interpretation of associated *n*-alkane δ^2H. The apparent fractionation factor of δ^2H between the source water and biosynthesized *n*-alkanes of the plant is influenced by plant growth forms and phylogeny (e.g., Griepentrog et al., 2019; Sessions, 2016; Liu et al., 2016; Sachse et al., 2012). Theoretically, with well-defined mixing sources and their associated δ^2H fractionation factors, the model output can be used to reconstruct δ^2H variation in the source water.

Lastly, the new metric FLMC has the potential to evaluate leaf mass integration patterns in sedimentary archives. The molecular distribution and $\delta^{13}C$ of *n*-alkanes have been used extensively in the assessment of sedimentary organic matter input (e.g., Aichner et al., 2010; Ankit et al., 2017; Kristen et al., 2010; Wiesenberg et al., 2004; Seki et al., 2010; Liu et al., 2020; Hockun et al., 2016; Mead et al., 2005). Because *n*-alkane concentrations (µg/g) involved in the mixing model are ultimately derived from dried leaf mass (Section 2.1.2), the model can potentially make an explicit connection between FLMC (as in Figures 5 and 8) and patterns of organic matter sourcing. Although the model presented here does not yet support these interpretations, addition of components to enable e.g., modeling of H isotope data would be relatively straight forward in future applications.

4.3 Future directions

The proposed mixing model is the bare core of a potentially more comprehensive proxy system model, an approach that has gained traction in recent developments of paleoenvironmental reconstruction (e.g., Bowen et al., 2020; Evans et al., 2013; Tingley et al., 2012; Li et al., 2010; Garreta et al., 2010). A proxy system model is a representation of the complete proxy system that ideally includes four components: the environment, the sensor, the archive, and the observation (Evans et al., 2013). The mixing model here primarily describes the sensor and observation components of the complete proxy system, while the environment and the archive components have not been developed. Future efforts that elaborate on the model structure will provide updated model assumptions, which should be based on systematic investigations of the specific proxy system components. Here are three categories of model improvements to consider.

4.3.1 Improvements on prior distributions

A better characterization of *n*-alkane chain length distribution and $\delta^{13}C$ in modern plants can potentially expand model application, due to the prior distributions' reliance on empirical data. At the same time, substantial gaps in our knowledge of *n*-alkanes still exist, especially from certain underrepresented regions of the world, such as western, central, and southeastern Asia (see global compilations by e.g., Bush and Mcinerney, 2013; Diefendorf and Freimuth, 2017). Future studies should

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

aim at bridging the gaps with systematic surveys of *n*-alkanes in modern plants. The need of a centralized database for plant wax lipids is also worth mentioning, as it will facilitate easy access to published records of both lipid concentrations and stable isotope values. From a different perspective, the development of mechanistic ways to represent the prior covariance patterns in both concentration and $\delta^{13}C$ can reduce model's reliance on complete data records and improve its tolerance with missing values. When the model is applied to *n*-alkanes in sedimentary cores, the $\delta^{13}C$ values of source *n*-alkanes should also be corrected to account for the higher $\delta^{13}C$ values in pre-industrial atmospheric CO_2. These efforts can improve the flexibility of model application and the robusticity of model results.

4.3.2 Better understanding of the processes and the incorporation of additional tracers

Processes such as *n*-alkane turnover and transportation have not been incorporated into the model framework primarily because they are still poorly understood (Diefendorf and Freimuth, 2017; Sachse et al., 2012). For example, factors associated with plant growth, such as primary productivity, leaf lifespan, and leaf turnover patterns (e.g., Burnham, 1989; Greenwood, 1991; Ellis and Johnson, 2013; Tipple et al., 2013; Hauke and Schreiber, 1998; Jetter and SchäFfer, 2001; Suh et al., 2019; Suh and Diefendorf, 2018), can influence the overall *n*-alkane turnover rate of a specific plant source. *n*-Alkane transportation can be another source of systematic biases in the interpretation of *n*-alkane records (e.g., Freimuth et al., 2019; Suh et al., 2019; Yamamoto et al., 2013; Schefuß et al., 2004; Mcfarlin et al., 2019; Nelson et al., 2018). Catchment specific integration effects such as spatial and temporal lipid integration can also affect the interpretation of *n*-alkane records (e.g., Douglas et al., 2014; Freimuth et al., 2021; French et al., 2018; Schefuß et al., 2004; Seki et al., 2010; Vogts et al., 2012; Herrmann et al., 2016; Freimuth et al., 2019; Garcin et al., 2014). Understanding these processes would ideally involve additional tracers or proxies. Future studies that investigate *n*-alkane turnover patterns may consider tracking the total amount of *n*-alkanes produced, and its relationship with the primary productivity of the plants in a year. The incorporation of e.g., pollen, elemental ratios, sedimentary organic matter, and multiple isotope tracers (e.g., Feakins et al., 2018; Freimuth et al., 2019; Seki et al., 2010) as process indicators can help to inform the relative importance of different transportation processes.

4.3.3 Incorporation of environmental factors into the proxy system

Environmental factors, such as temperature and precipitation, are primary determinants of vegetation composition, but they also influence *n*-alkane concentration and $\delta^{13}C$ (sensor values) through plant physiological response. The inclusion of environmental factors into the model framework would naturally involve proxies such as *n*-alkane $\delta^{2}H$, which has been used primarily as an environmental proxy, but interpreted independently (e.g., Liu and An, 2018; Sachse et al., 2012; Sessions, 2016; Liu and Yang, 2008). The multi-level influence from environmental factors to *n*-alkane $\delta^{2}H$ include a static element, such as taxon-specific $\delta^{2}H$ enrichment factors in lipid biosynthesis (although see Newberry et al., 2015; Cormier et al., 2018), and dynamic elements such as precipitation $\delta^{2}H$ and environment-dependent evapotranspiration enrichment (e.g., Cernusak et al., 2016; Douglas et al., 2012; Sachse et al., 2012; Liu and An, 2018; Tipple et al., 2015; Kahmen et al., 2013).

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Incorporating these elements, with the additional interactions between $\delta^{13}C$ and $\delta^{2}H$, into the existing model framework will allow simultaneous evaluation of both isotope tracers in *n*-alkanes. Future studies should consider developing such a model component, which will ultimately permit proxy system interpretation at the environmental level.

5. Conclusion

Traditional interpretations of *n*-alkane proxies often rely on either chain length distribution or $\delta^{13}C$ values of one chain to reconstruct vegetation composition. An alternative approach is to combine the information from both lines of evidence to refine our interpretations. We presented a Bayesian modeling approach that simultaneously evaluates both chain length distribution and stable carbon isotope measurements of sedimentary *n*-alkanes. The model incorporates the uncertainties associated with *n*-alkane concentration and carbon isotope ratios of multiple chains, using a generic mixing process with isotope mass balance. We presented two case studies to illustrate how the model can be applied to the interpretation of sedimentary *n*-alkane records. One involves published long-chain *n*-alkane records from lake surface sediments of Lake Qinghai, China. The other involves published long-chain *n*-alkane records from lake surface sediments along a vegetation gradient in Cameroon, western Africa. Compared to traditional two-end mixing models, our approach can resolve mixing fractions (FLMC) of multiple *n*-alkane sources, providing alternative interpretations to the same *n*-alkane record. The posterior densities of model parameter distributions, such as fractional source contribution to a specific chain (FSC_n), can be used to assess the interpretation of associated proxies such as *n*-alkane $\delta^{2}H$. Despite these achievements, several processes associated with *n*-alkane integration in sedimentary archives are still not accounted for. Future studies on *n*-alkane turnover and transportation will improve our understanding of the biases and constraints associated with *n*-alkane records. The Bayesian model framework could be further improved by adding more mechanistic modeling components such as *n*-alkane integration processes, and additional environmental proxies such as $\delta^{2}H$. The modeling approach represents a continuously evolving framework that can incorporate new understandings and leverage additional proxies in the future.

Code and data availability

All data and code used to conduct the analyses and create figures reported in this paper are archived online (Yang, 2022) and available at https://doi.org/10.5281/zenodo.6236846.

Author contribution

DY conceived, designed, and conducted the analyses with support from GJB. DY prepared the manuscript with contributions from GJB.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This project was sponsored by the National Science Foundation (ABI-1759730).

Acknowledgements

We would like to thank Jamie McFarlin, Kevin Uno, and Brenden Fischer-Femal for their comments and discussions in the model development.

References

Aichner, B., Herzschuh, U., and Wilkes, H.: Influence of aquatic macrophytes on the stable carbon isotopic signatures of sedimentary organic matter in lakes on the Tibetan Plateau, Org Geochem, 41, 706-718, https://doi.org/10.1016/j.orggeochem.2010.02.002, 2010.

Aleman, J. C., Fayolle, A., Favier, C., Staver, A. C., Dexter, K. G., Ryan, C. M., Azihou, A. F., Bauman, D., te Beest, M., Chidumayo, E. N., Comiskey, J. A., Cromsigt, J. P. G. M., Dessard, H., Doucet, J.-L., Finckh, M., Gillet, J.-F., Gourlet-Fleury, S., Hempson, G. P., Holdo, R. M., Kirunda, B., Kouame, F. N., Mahy, G., Gonçalves, F. M. P., McNicol, I., Quintano, P. N., Plumptre, A. J., Pritchard, R. C., Revermann, R., Schmitt, C. B., Swemmer, A. M., Talila, H., Woollen, E., and Swaine, M. D.: Floristic evidence for alternative biome states in tropical Africa, P Natl Acad Sci USA, 117, 28183-28190, 10.1073/pnas.2011515117, 2020.

Ali, H. A. M., Mayes, R. W., Hector, B. L., and Orskov, E. R.: Assessment of *n*-alkanes, long-chain fatty alcohols and long-chain fatty acids as diet composition markers: The concentrations of these compounds in rangeland species from Sudan, Anim Feed Sci Tech, 121, 257-271, https://doi.org/10.1016/j.anifeedsci.2005.02.026, 2005.

Allen, E. D. and Spence, D. H. N.: The differential ability of aquatic plants to utilize the inorganic carbon supply in fresh waters, New Phytol, 87, 269-283, https://doi.org/10.1111/j.1469-8137.1981.tb03198.x, 1981.

Andrae, J. W., McInerney, F. A., Polissar, P. J., Sniderman, J. M. K., Howard, S., Hall, P. A., and Phelps, S. R.: Initial expansion of C_4 vegetation in Australia during the late Pliocene, Geophys Res Lett, 45, 4831-4840, https://doi.org/10.1029/2018GL077833, 2018.

Andrae, J. W., McInerney, F. A., Tibby, J., Henderson, A. C. G., Hall, P. A., Marshall, J. C., McGregor, G. B., Barr, C., and Greenway, M.: Variation in leaf wax *n*-alkane characteristics with climate in the broad-leaved paperbark (*Melaleuca quinquenervia*), Org Geochem, 130, 33-42, https://doi.org/10.1016/j.orggeochem.2019.02.004, 2019.

Ankit, Y., Mishra, P. K., Kumar, P., Jha, D. K., Kumar, V. V., Ambili, V., and Anoop, A.: Molecular distribution and carbon isotope of *n*-alkanes from Ashtamudi Estuary, South India: Assessment of organic matter sources and paleoclimatic implications, Mar Chem, 196, 62-70, https://doi.org/10.1016/j.marchem.2017.08.002, 2017.

Badewien, T., Vogts, A., and Rullkötter, J.: *n*-Alkane distribution and carbon stable isotope composition in leaf waxes of C_3 and C_4 plants from Angola, Org Geochem, 89-90, 71-79, https://doi.org/10.1016/j.orggeochem.2015.09.002, 2015.

Bi, X., Sheng, G., Liu, X., Li, C., and Fu, J.: Molecular and carbon and hydrogen isotopic composition of *n*-alkanes in plant leaf waxes, Org Geochem, 36, 1405-1417, https://doi.org/10.1016/j.orggeochem.2005.06.001, 2005.

Bird, M. I., Summons, R. E., Gagan, M. K., Roksandic, Z., Dowling, L., Head, J., Keith Fifield, L., Cresswell, R. G., and Johnson, D. P.: Terrestrial vegetation change inferred from *n*-alkane $\delta^{13}C$ analysis in the marine environment, Geochim Cosmochim Ac, 59, 2853-2857, https://doi.org/10.1016/0016-7037(95)00160-2, 1995.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Bobe, R. and Behrensmeyer, A. K.: The expansion of grassland ecosystems in Africa in relation to mammalian evolution and the origin of the genus *Homo*, Palaeogeogr Palaeocl, 207, 399-420, http://dx.doi.org/10.1016/j.palaeo.2003.09.033, 2004.
Bowen, G. J., Fischer-Femal, B., Reichart, G. J., Sluijs, A., and Lear, C. H.: Joint inversion of proxy system models to reconstruct paleoenvironmental time series from heterogeneous data, Clim Past, 16, 65-78, 10.5194/cp-16-65-2020, 2020.
Bray, E. E. and Evans, E. D.: Distribution of *n*-paraffins as a clue to recognition of source beds, Geochim Cosmochim Ac, 22, 2-15, https://doi.org/10.1016/0016-7037(61)90069-2, 1961.
Brittingham, A., Hren, M. T., and Hartman, G.: Microbial alteration of the hydrogen and carbon isotopic composition of *n*-alkanes in sediments, Org Geochem, 107, 1-8, https://doi.org/10.1016/j.orggeochem.2017.01.010, 2017.
Buggle, B., Wiesenberg, G. L. B., and Glaser, B.: Is there a possibility to correct fossil *n*-alkane data for postsedimentary alteration effects?, Appl Geochem, 25, 947-957, https://doi.org/10.1016/j.apgeochem.2010.04.003, 2010.
Burnham, R. J.: Relationships between standing vegetation and leaf litter in a paratropical forest: Implications for paleobotany, Rev Palaeobot Palynol, 58, 5-32, https://doi.org/10.1016/0034-6667(89)90054-7, 1989.
Bush, R. T. and McInerney, F. A.: Leaf wax *n*-alkane distributions in and across modern plants: Implications for paleoecology and chemotaxonomy, Geochim Cosmochim Ac, 117, 161-179, https://doi.org/10.1016/j.gca.2013.04.016, 2013.
Bush, R. T. and McInerney, F. A.: Influence of temperature and C_4 abundance on n-alkane chain length distributions across the central USA, Org Geochem, 79, 65-73, https://doi.org/10.1016/j.orggeochem.2014.12.003, 2015.
Carr, A. S., Boom, A., Grimes, H. L., Chase, B. M., Meadows, M. E., and Harris, A.: Leaf wax *n*-alkane distributions in arid zone South African flora: Environmental controls, chemotaxonomy and palaeoecological implications, Org Geochem, 67, 72-84, https://doi.org/10.1016/j.orggeochem.2013.12.004, 2014.
Castañeda, I. S., Mulitza, S., Schefuß, E., Lopes dos Santos, R. A., Sinninghe Damsté, J. S., and Schouten, S.: Wet phases in the Sahara/Sahel region and human migration patterns in North Africa, P Natl Acad Sci USA, 106, 20159-20163, 10.1073/pnas.0905771106, 2009.
Cerling, T. E.: Development of grasslands and savannas in East Africa during the Neogene, Palaeogeogr Palaeocl, 97, 241-247, 1992.
Cernusak, L. A., Ubierna, N., Winter, K., Holtum, J. A. M., Marshall, J. D., and Farquhar, G. D.: Environmental and physiological determinants of carbon isotope discrimination in terrestrial plants, New Phytol, 200, 950-965, https://doi.org/10.1111/nph.12423, 2013.
Cernusak, L. A., Barbour, M. M., Arndt, S. K., Cheesman, A. W., English, N. B., Feild, T. S., Helliker, B. R., Holloway-Phillips, M. M., Holtum, J. A. M., Kahmen, A., McInerney, F. A., Munksgaard, N. C., Simonin, K. A., Song, X., Stuart-Williams, H., West, J. B., and Farquhar, G. D.: Stable isotopes in leaf water of terrestrial plants, Plant Cell Environ, 39, 1087-1102, 10.1111/pce.12703, 2016.
Cheesbrough, T. M. and Kolattukudy, P. E.: Alkane biosynthesis by decarbonylation of aldehydes catalyzed by a particulate preparation from Pisum sativum, P Natl Acad Sci USA, 81, 6613-6617, 10.1073/pnas.81.21.6613, 1984.
Chikaraishi, Y., Naraoka, H., and Poulson, S. R.: Carbon and hydrogen isotopic fractionation during lipid biosynthesis in a higher plant (*Cryptomeria japonica*), Phytochemistry, 65, 323-330, https://doi.org/10.1016/j.phytochem.2003.12.003, 2004.
Collister, J. W., Rieley, G., Stern, B., Eglinton, G., and Fry, B.: Compound-specific $\delta^{13}C$ analyses of leaf lipids from plants with differing carbon dioxide metabolisms, Org Geochem, 21, 619-627, https://doi.org/10.1016/0146-6380(94)90008-6, 1994.
Cormier, M.-A., Werner, R. A., Sauer, P. E., Gröcke, D. R., Leuenberger, M. C., Wieloch, T., Schleucher, J., and Kahmen, A.: ^2H-fractionations during the biosynthesis of carbohydrates and lipids imprint a metabolic signal on the δ^2H values of plant organic compounds, New Phytol, 218, 479-491, https://doi.org/10.1111/nph.15016, 2018.
Cranwell, P. A., Eglinton, G., and Robinson, N.: Lipids of aquatic organisms as potential contributors to lacustrine sediments—II, Org Geochem, 11, 513-527, https://doi.org/10.1016/0146-6380(87)90007-6, 1987.
Curtis, S. M.: mcmcplots: Create Plots from MCMC Output (0.4.3) [code], 2018.
Diefendorf, A. F. and Freimuth, E. J.: Extracting the most from terrestrial plant-derived *n*-alkyl lipids and their carbon isotopes from the sedimentary record: A review, Org Geochem, 103, 1-21, https://doi.org/10.1016/j.orggeochem.2016.10.016, 2017.
Diefendorf, A. F., Leslie, A. B., and Wing, S. L.: Leaf wax composition and carbon isotopes vary among major conifer groups, Geochim Cosmochim Ac, 170, 145-156, https://doi.org/10.1016/j.gca.2015.08.018, 2015.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Diefendorf, A. F., Freeman, K. H., Wing, S. L., and Graham, H. V.: Production of *n*-alkyl lipids in living plants and implications for the geologic past, Geochim Cosmochim Ac, 75, 7472-7485, https://doi.org/10.1016/j.gca.2011.09.028, 2011.

Diefendorf, A. F., Mueller, K. E., Wing, S. L., Koch, P. L., and Freeman, K. H.: Global patterns in leaf ^{13}C discrimination and implications for studies of past and future climate, P Natl Acad Sci USA, 107, 5738-5743, 10.1073/pnas.0910513107, 2010.

Dion-Kirschner, H., McFarlin, J. M., Masterson, A. L., Axford, Y., and Osburn, M. R.: Modern constraints on the sources and climate signals recorded by sedimentary plant waxes in west Greenland, Geochim Cosmochim Ac, 286, 336-354, https://doi.org/10.1016/j.gca.2020.07.027, 2020.

Douglas, P. M. J., Pagani, M., Brenner, M., Hodell, D. A., and Curtis, J. H.: Aridity and vegetation composition are important determinants of leaf-wax δD values in southeastern Mexico and Central America, Geochim Cosmochim Ac, 97, 24-45, https://doi.org/10.1016/j.gca.2012.09.005, 2012.

Douglas, P. M. J., Pagani, M., Eglinton, T. I., Brenner, M., Hodell, D. A., Curtis, J. H., Ma, K. F., and Breckenridge, A.: Pre-aged plant waxes in tropical lake sediments and their influence on the chronology of molecular paleoclimate proxy records, Geochim Cosmochim Ac, 141, 346-364, https://doi.org/10.1016/j.gca.2014.06.030, 2014.

Duan, Y. and Xu, L.: Distributions of *n*-alkanes and their hydrogen isotopic composition in plants from Lake Qinghai (China) and the surrounding area, Appl Geochem, 27, 806-814, https://doi.org/10.1016/j.apgeochem.2011.12.008, 2012.

Eglinton, G. and Hamilton, R. J.: Leaf epicuticular waxes, Science, 156, 1322-1335, doi:10.1126/science.156.3780.1322, 1967.

Ellis, B. and Johnson, K. R.: Comparison of leaf samples from mapped tropical and temperate forests: Implications for interpretations of the diversity of fossil assemblages, Palaios, 28, 163-177, 10.2110/palo.2012.p12-073r, 2013.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Sci Rev, 76, 16-28, https://doi.org/10.1016/j.quascirev.2013.05.024, 2013.

Feakins, S. J., Wu, M. S., Ponton, C., Galy, V., and West, A. J.: Dual isotope evidence for sedimentary integration of plant wax biomarkers across an Andes-Amazon elevation transect, Geochim Cosmochim Ac, 242, 64-81, https://doi.org/10.1016/j.gca.2018.09.007, 2018.

Feakins, S. J., Peters, T., Wu, M. S., Shenkin, A., Salinas, N., Girardin, C. A. J., Bentley, L. P., Blonder, B., Enquist, B. J., Martin, R. E., Asner, G. P., and Malhi, Y.: Production of leaf wax *n*-alkanes across a tropical forest elevation transect, Org Geochem, 100, 89-100, https://doi.org/10.1016/j.orggeochem.2016.07.004, 2016.

Ficken, K. J., Li, B., Swain, D. L., and Eglinton, G.: An *n*-alkane proxy for the sedimentary input of submerged/floating freshwater aquatic macrophytes, Org Geochem, 31, 745-749, https://doi.org/10.1016/S0146-6380(00)00081-4, 2000.

Ficken, K. J., Street-Perrott, F. A., Perrott, R. A., Swain, D. L., Olago, D. O., and Eglinton, G.: Glacial/interglacial variations in carbon cycling revealed by molecular and isotope stratigraphy of Lake Nkunga, Mt. Kenya, East Africa, Org Geochem, 29, 1701-1719, https://doi.org/10.1016/S0146-6380(98)00109-0, 1998.

Freeman, K. H. and Colarusso, L. A.: Molecular and isotopic records of C_4 grassland expansion in the late miocene, Geochim Cosmochim Ac, 65, 1439-1454, https://doi.org/10.1016/S0016-7037(00)00573-1, 2001.

Freeman, K. H. and Pancost, R. D.: 12.15 - Biomarkers for Terrestrial Plants and Climate, in: Treatise on Geochemistry, 2 ed., edited by: Holland, H. D., and Turekian, K. K., Elsevier, Oxford, 395-416, https://doi.org/10.1016/B978-0-08-095975-7.01028-7, 2014.

Freimuth, E. J., Diefendorf, A. F., Lowell, T. V., and Wiles, G. C.: Sedimentary *n*-alkanes and *n*-alkanoic acids in a temperate bog are biased toward woody plants, Org Geochem, 128, 94-107, https://doi.org/10.1016/j.orggeochem.2019.01.006, 2019.

Freimuth, E. J., Diefendorf, A. F., Lowell, T. V., Schartman, A. K., Landis, J. D., Stewart, A. K., and Bates, B. R.: Centennial-scale age offsets of plant wax n-alkanes in Adirondack lake sediments, Geochim Cosmochim Ac, 300, 119-136, https://doi.org/10.1016/j.gca.2021.02.022, 2021.

French, K. L., Hein, C. J., Haghipour, N., Wacker, L., Kudrass, H. R., Eglinton, T. I., and Galy, V.: Millennial soil retention of terrestrial organic matter deposited in the Bengal Fan, Sci Rep-UK, 8, 11997, 10.1038/s41598-018-30091-8, 2018.

Gao, L., Hou, J., Toney, J., MacDonald, D., and Huang, Y.: Mathematical modeling of the aquatic macrophyte inputs of mid-chain *n*-alkyl lipids to lake sediments: Implications for interpreting compound specific hydrogen isotopic records, Geochim Cosmochim Ac, 75, 3781-3791, https://doi.org/10.1016/j.gca.2011.04.008, 2011.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Garcin, Y., Schwab, V. F., Gleixner, G., Kahmen, A., Todou, G., Séné, O., Onana, J.-M., Achoundong, G., and Sachse, D.: Hydrogen isotope ratios of lacustrine sedimentary *n*-alkanes as proxies of tropical African hydrology: Insights from a calibration transect across Cameroon, Geochim Cosmochim Ac, 79, 106-126, https://doi.org/10.1016/j.gca.2011.11.039, 2012.

Garcin, Y., Schefuß, E., Schwab, V. F., Garreta, V., Gleixner, G., Vincens, A., Todou, G., Séné, O., Onana, J.-M., Achoundong, G., and Sachse, D.: Reconstructing C_3 and C_4 vegetation cover using n-alkane carbon isotope ratios in recent lake sediments from Cameroon, Western Central Africa, Geochim Cosmochim Ac, 142, 482-500, https://doi.org/10.1016/j.gca.2014.07.004, 2014.

Garreta, V., Miller, P. A., Guiot, J., Hély, C., Brewer, S., Sykes, M. T., and Litt, T.: A method for climate and vegetation reconstruction through the inversion of a dynamic vegetation model, Clim Dynam, 35, 371-389, 10.1007/s00382-009-0629-1, 2010.

Gelman, A. and Rubin, D. B.: Inference from iterative simulation using multiple sequences, Stat Sci, 7, 457-472, 416, 1992.

Gelpi, E., Schneider, H., Mann, J., and Oró, J.: Hydrocarbons of geochemical significance in microscopic algae, Phytochemistry, 9, 603-612, https://doi.org/10.1016/S0031-9422(00)85700-3, 1970.

Geman, S. and Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE T Pattern Anal, PAMI-6, 721-741, 10.1109/TPAMI.1984.4767596, 1984.

Greenway, P. J. and Vesey-Fitzgerald, D. F.: The Vegetation of Lake Manyara National Park, J Ecol, 57, 127-149, 10.2307/2258212, 1969.

Greenwood, D. R.: The taphonomy of plant macrofossils, in: The Processes of Fossilization, edited by: Donovan, S., Belhaven Press, 141-169, 1991.

Griepentrog, M., De Wispelaere, L., Bauters, M., Bodé, S., Hemp, A., Verschuren, D., and Boeckx, P.: Influence of plant growth form, habitat and season on leaf-wax *n*-alkane hydrogen-isotopic signatures in equatorial East Africa, Geochim Cosmochim Ac, 263, 122-139, https://doi.org/10.1016/j.gca.2019.08.004, 2019.

Grimalt, J. O., Torras, E., and Albaigés, J.: Bacterial reworking of sedimentary lipids during sample storage, Org Geochem, 13, 741-746, https://doi.org/10.1016/0146-6380(88)90096-4, 1988.

Han, J. and Calvin, M.: Hydrocarbon distribution of algae and bacteria, and microbiological activity in sediments, P Natl Acad Sci USA, 64, 436-443, 10.1073/pnas.64.2.436, 1969.

Han, J., McCarthy, E. D., Hoeven, W. V., Calvin, M., and Bradley, W. H.: Organic geochemical studies, ii. A preliminary report on the distribution of aliphatic hydrocarbons in algae, in bacteria, and in a recent lake sediment, P Natl Acad Sci USA, 59, 29-33, 10.1073/pnas.59.1.29, 1968.

Hauke, V. and Schreiber, L.: Ontogenetic and seasonal development of wax composition and cuticular transpiration of ivy (*Hedera helix* L.) sun and shade leaves, Planta, 207, 67-75, 10.1007/s004250050456, 1998.

Hayes, J. M.: Factors controlling ^{13}C contents of sedimentary organic compounds: Principles and evidence, Mar Geol, 113, 111-125, https://doi.org/10.1016/0025-3227(93)90153-M, 1993.

Herrmann, N., Boom, A., Carr, A. S., Chase, B. M., Granger, R., Hahn, A., Zabel, M., and Schefuß, E.: Sources, transport and deposition of terrestrial organic material: A case study from southwestern Africa, Quaternary Sci Rev, 149, 215-229, https://doi.org/10.1016/j.quascirev.2016.07.028, 2016.

Hockun, K., Mollenhauer, G., Ho, S. L., Hefter, J., Ohlendorf, C., Zolitschka, B., Mayr, C., Lücke, A., and Schefuß, E.: Using distributions and stable isotopes of *n*-alkanes to disentangle organic matter contributions to sediments of Laguna Potrok Aike, Argentina, Org Geochem, 102, 110-119, https://doi.org/10.1016/j.orggeochem.2016.10.001, 2016.

Howard-Williams, C. and Walker, B. H.: The vegetation of a tropical African lake: Classification and ordination of the vegetation of Lake Chilwa (Malawi), J Ecol, 62, 831-854, 10.2307/2258958, 1974.

Huang, Y., Shuman, B., Wang, Y., and Webb, T.: Hydrogen isotope ratios of individual lipids in lake sediments as novel tracers of climatic and environmental change: a surface sediment test, J Paleolimnol, 31, 363-375, 10.1023/B:JOPL.0000021855.80535.13, 2004.

Huang, Y., Dupont, L., Sarnthein, M., Hayes, J. M., and Eglinton, G.: Mapping of C*4* plant input from North West Africa into North East Atlantic sediments, Geochim Cosmochim Ac, 64, 3505-3513, https://doi.org/10.1016/S0016-7037(00)00445-2, 2000.

Hughen, K. A., Eglinton, T. I., Xu, L., and Makou, M.: Abrupt tropical vegetation response to rapid climate changes, Science, 304, 1955-1959, doi:10.1126/science.1092995, 2004.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Janis, C. M., Damuth, J., and Theodor, J. M.: The origins and evolution of the North American grassland biome: the story from the hoofed mammals, Palaeogeogr Palaeocl, 177, 183-198, https://doi.org/10.1016/S0031-0182(01)00359-5, 2002.

Jansen, B., van Loon, E. E., Hooghiemstra, H., and Verstraten, J. M.: Improved reconstruction of palaeo-environments through unravelling of preserved vegetation biomarker patterns, Palaeogeogr Palaeocl, 285, 119-130, https://doi.org/10.1016/j.palaeo.2009.10.029, 2010.

Jetter, R. and Schäffer, S.: Chemical composition of the *Prunus laurocerasus* leaf surface. Dynamic changes of the epicuticular wax film during leaf development, Plant Physiol, 126, 1725-1737, 10.1104/pp.126.4.1725, 2001.

Kahmen, A., Schefuß, E., and Sachse, D.: Leaf water deuterium enrichment shapes leaf wax n-alkane δD values of angiosperm plants I: Experimental evidence and mechanistic insights, Geochim Cosmochim Ac, 111, 39-49, https://doi.org/10.1016/j.gca.2012.09.003, 2013.

Kaya, F., Bibi, F., Žliobaitė, I., Eronen, J. T., Hui, T., and Fortelius, M.: The rise and fall of the Old World savannah fauna and the origins of the African savannah biome, Nat Ecol Evol, 2, 241-246, 10.1038/s41559-017-0414-1, 2018.

Keeley, J. E. and Sandquist, D. R.: Carbon: freshwater plants, Plant Cell Environ, 15, 1021-1035, https://doi.org/10.1111/j.1365-3040.1992.tb01653.x, 1992.

Koch, K., Hartmann, K. D., Schreiber, L., Barthlott, W., and Neinhuis, C.: Influences of air humidity during the cultivation of plants on wax chemical composition, morphology and leaf surface wettability, Environ Exp Bot, 56, 1-9, https://doi.org/10.1016/j.envexpbot.2004.09.013, 2006.

Kolattukudy, P. E., Croteau, R., and Buckner, J.: Biochemistry of plant waxes, Chemistry and biochemistry of natural waxes, Elsevier1976.

Kristen, I., Wilkes, H., Vieth, A., Zink, K. G., Plessen, B., Thorpe, J., Partridge, T. C., and Oberhänsli, H.: Biomarker and stable carbon isotope analyses of sedimentary organic matter from Lake Tswaing: evidence for deglacial wetness and early Holocene drought from South Africa, J Paleolimnol, 44, 143-160, 10.1007/s10933-009-9393-9, 2010.

Krull, E., Sachse, D., Mügler, I., Thiele, A., and Gleixner, G.: Compound-specific $\delta^{13}C$ and $\delta^{2}H$ analyses of plant and soil organic matter: A preliminary assessment of the effects of vegetation change on ecosystem hydrology, Soil Biology and Biochemistry, 38, 3211-3221, https://doi.org/10.1016/j.soilbio.2006.04.008, 2006.

Kunst, L. and Samuels, A. L.: Biosynthesis and secretion of plant cuticular wax, Progress in Lipid Research, 42, 51-80, https://doi.org/10.1016/S0163-7827(02)00045-0, 2003.

Ladygina, N., Dedyukhina, E. G., and Vainshtein, M. B.: A review on microbial synthesis of hydrocarbons, Process Biochem, 41, 1001-1014, https://doi.org/10.1016/j.procbio.2005.12.007, 2006.

Latorre, C., Quade, J., and McIntosh, W. C.: The expansion of C_4 grasses and global change in the late Miocene: stable isotope evidence from the Americas, Earth Planet Sc Lett, 146, 83-96, 1997.

Li, B., Nychka, D. W., and Ammann, C. M.: The value of multiproxy reconstruction of past climate, J Am Stat Assoc, 105, 883-895, 10.1198/jasa.2010.ap09379, 2010.

Liu, C., Li, Z., Berhe, A. A., and Hu, B. X.: Chapter Six - The isotopes and biomarker approaches for identifying eroded organic matter sources in sediments: A review, in: Adv Agron, edited by: Sparks, D. L., Academic Press, 257-303, https://doi.org/10.1016/bs.agron.2020.02.005, 2020.

Liu, H. and Liu, W.: *n*-Alkane distributions and concentrations in algae, submerged plants and terrestrial plants from the Qinghai-Tibetan Plateau, Org Geochem, 99, 10-22, https://doi.org/10.1016/j.orggeochem.2016.06.003, 2016.

Liu, J. and An, Z.: A hierarchical framework for disentangling different controls on leaf wax δD *n*-alkane values in terrestrial higher plants, Quaternary Sci Rev, 201, 409-417, https://doi.org/10.1016/j.quascirev.2018.10.026, 2018.

Liu, J. and An, Z.: Leaf wax *n*-alkane carbon isotope values vary among major terrestrial plant groups: Different responses to precipitation amount and temperature, and implications for paleoenvironmental reconstruction, Earth-Sci Rev, 202, 103081, https://doi.org/10.1016/j.earscirev.2020.103081, 2020.

Liu, J., Liu, W., An, Z., and Yang, H.: Different hydrogen isotope fractionations during lipid formation in higher plants: Implications for paleohydrology reconstruction at a global scale, Sci Rep-UK, 6, 19711, 10.1038/srep19711, 2016.

Liu, W. and Yang, H.: Multiple controls for the variability of hydrogen isotopic compositions in higher plant *n*-alkanes from modern ecosystems, Glob Change Biol, 14, 2166-2177, https://doi.org/10.1111/j.1365-2486.2008.01608.x, 2008.

Liu, W., Yang, H., Wang, H., An, Z., Wang, Z., and Leng, Q.: Carbon isotope composition of long chain leaf wax *n*-alkanes in lake sediments: A dual indicator of paleoenvironment in the Qinghai-Tibet Plateau, Org Geochem, 83-84, 190-201, https://doi.org/10.1016/j.orggeochem.2015.03.017, 2015.

Liu, X., Feakins, S. J., Dong, X., Xue, Q., Marek, T., Leskovar, D. I., Neely, C. B., and Ibrahim, A. M. H.: Experimental study of leaf wax *n*-alkane response in winter wheat cultivars to drought conditions, Org Geochem, 113, 210-223, https://doi.org/10.1016/j.orggeochem.2017.07.020, 2017.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D.: The BUGS Book: A Practical Introduction to Bayesian Analysis, CRC Press/Chapman and Hall, Boca Raton, FL2012.

Macková, J., Vašková, M., Macek, P., Hronková, M., Schreiber, L., and Šantrůček, J.: Plant response to drought stress simulated by ABA application: Changes in chemical composition of cuticular waxes, Environ Exp Bot, 86, 70-75, https://doi.org/10.1016/j.envexpbot.2010.06.005, 2013.

Magill, C. R., Eglinton, G., and Eglinton, T. I.: Isotopic variance among plant lipid homologues correlates with biodiversity patterns of their source communities, Plos One, 14, e0212211, 10.1371/journal.pone.0212211, 2019.

Makowski, D., Ben-Shachar, M. S., and Lüdecke, D.: bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework, J Open Source Softw, 4, 1541, https://doi.org/10.21105/joss.01541, 2019.

McFarlin, J. M., Axford, Y., Masterson, A. L., and Osburn, M. R.: Calibration of modern sedimentary δ^2H plant wax-water relationships in Greenland lakes, Quaternary Sci Rev, 225, 105978, https://doi.org/10.1016/j.quascirev.2019.105978, 2019.

Mead, R., Xu, Y., Chong, J., and Jaffé, R.: Sediment and soil organic matter source assessment as revealed by the molecular distribution and carbon isotopic composition of *n*-alkanes, Org Geochem, 36, 363-370, https://doi.org/10.1016/j.orggeochem.2004.10.003, 2005.

Millard, S. P.: EnvStats: an R package for environmental statistics, 2013.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., and Boker, S. M.: OpenMx 2.0: Extended structural equation and statistical modeling, Psychometrika, 81, 535-549, 10.1007/s11336-014-9435-8, 2016.

Nelson, D. B., Ladd, S. N., Schubert, C. J., and Kahmen, A.: Rapid atmospheric transport and large-scale deposition of recently synthesized plant waxes, Geochim Cosmochim Ac, 222, 599-617, https://doi.org/10.1016/j.gca.2017.11.018, 2018.

Newberry, S. L., Kahmen, A., Dennis, P., and Grant, A.: n-Alkane biosynthetic hydrogen isotope fractionation is not constant throughout the growing season in the riparian tree Salix viminalis, Geochim Cosmochim Ac, 165, 75-85, https://doi.org/10.1016/j.gca.2015.05.001, 2015.

Niedermeyer, E. M., Schefuß, E., Sessions, A. L., Mulitza, S., Mollenhauer, G., Schulz, M., and Wefer, G.: Orbital- and millennial-scale changes in the hydrologic cycle and vegetation in the western African Sahel: insights from individual plant wax δD and $\delta^{13}C$, Quaternary Sci Rev, 29, 2996-3005, https://doi.org/10.1016/j.quascirev.2010.06.039, 2010.

Park, M.-O.: New pathway for long-chain *n*-alkane synthesis via 1-alcohol in *Vibrio furnissii* M1, J Bacteriol, 187, 1426-1429, doi:10.1128/JB.187.4.1426-1429.2005, 2005.

Peaple, M. D., Tierney, J. E., McGee, D., Lowenstein, T. K., Bhattacharya, T., and Feakins, S. J.: Identifying plant wax inputs in lake sediments using machine learning, Org Geochem, 156, 104222, https://doi.org/10.1016/j.orggeochem.2021.104222, 2021.

rjags: Bayesian graphical models using MCMC, R package: https://cran.r-project.org/web/packages/rjags/index.html, last access: May 31, 2021.

Polissar, P. J., Rose, C., Uno, K. T., Phelps, S. R., and deMenocal, P.: Synchronous rise of African C_4 ecosystems 10 million years ago in the absence of aridification, Nat Geosci, 12, 657-660, 10.1038/s41561-019-0399-2, 2019.

Prins, H. B. A. and Elzenga, J. T. M.: Bicarbonate utilization: Function and mechanism, Aquat Bot, 34, 59-83, https://doi.org/10.1016/0304-3770(89)90050-8, 1989.

Quade, J. and Cerling, T. E.: Expansion of C_4 grasses in the Late Miocene of Northern Pakistan: evidence from stable isotopes in paleosols, Palaeogeogr Palaeocl, 115, 91-116, 1995.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing [code], 2021.

Rommerskirchen, F., Plader, A., Eglinton, G., Chikaraishi, Y., and Rullkötter, J.: Chemotaxonomic significance of distribution and stable carbon isotopic composition of long-chain alkanes and alkan-1-ols in C_4 grass waxes, Org Geochem, 37, 1303-1332, https://doi.org/10.1016/j.orggeochem.2005.12.013, 2006.

Rommerskirchen, F., Eglinton, G., Dupont, L., Güntner, U., Wenzel, C., and Rullkötter, J.: A north to south transect of Holocene southeast Atlantic continental margin sediments: Relationship between aerosol transport and compound-specific δ13C land plant biomarker and pollen records, Geochem Geophy Geosy, 4, https://doi.org/10.1029/2003GC000541, 2003.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Sachse, D., Radke, J., and Gleixner, G.: Hydrogen isotope ratios of recent lacustrine sedimentary n-alkanes record modern climate variability, Geochim Cosmochim Ac, 68, 4877-4889, https://doi.org/10.1016/j.gca.2004.06.004, 2004.

Sachse, D., Billault, I., Bowen, G. J., Chikaraishi, Y., Dawson, T. E., Feakins, S. J., Freeman, K. H., Magill, C. R., McInerney, F. A., van der Meer, M. T. J., Polissar, P., Robins, R. J., Sachs, J. P., Schmidt, H.-L., Sessions, A. L., White, J. W. C., West, J. B., and Kahmen, A.: Molecular paleohydrology: Interpreting the hydrogen-isotopic composition of lipid biomarkers from photosynthesizing organisms, Annu Rev Earth Pl Sc, 40, 221-249, 10.1146/annurev-earth-042711-105535, 2012.

Sankaran, M., Hanan, N. P., Scholes, R. J., Ratnam, J., Augustine, D. J., Cade, B. S., Gignoux, J., Higgins, S. I., Le Roux, X., Ludwig, F., Ardo, J., Banyikwa, F., Bronn, A., Bucini, G., Caylor, K. K., Coughenour, M. B., Diouf, A., Ekaya, W., Feral, C. J., February, E. C., Frost, P. G. H., Hiernaux, P., Hrabar, H., Metzger, K. L., Prins, H. H. T., Ringrose, S., Sea, W., Tews, J., Worden, J., and Zambatis, N.: Determinants of woody cover in African savannas, Nature, 438, 846-849, 10.1038/nature04070, 2005.

Schefuß, E., Schouten, S., Jansen, J. H. F., and Sinninghe Damsté, J. S.: African vegetation controlled by tropical sea surface temperatures in the mid-Pleistocene period, Nature, 422, 418-421, 10.1038/nature01500, 2003.

Schefuß, E., Versteegh, G. J. M., Jansen, J. H. F., and Sinninghe Damsté, J. S.: Lipid biomarkers as major source and preservation indicators in SE Atlantic surface sediments, Deep Sea Research Part I: Oceanographic Research Papers, 51, 1199-1228, https://doi.org/10.1016/j.dsr.2004.05.002, 2004.

Schefuß, E., Kuhlmann, H., Mollenhauer, G., Prange, M., and Pätzold, J.: Forcing of wet phases in southeast Africa over the past 17,000 years, Nature, 480, 509-512, 10.1038/nature10685, 2011.

Schwab, V. F., Garcin, Y., Sachse, D., Todou, G., Séné, O., Onana, J.-M., Achoundong, G., and Gleixner, G.: Effect of aridity on $\delta^{13}C$ and δD values of C_3 plant- and C_4 graminoid-derived leaf wax lipids from soils along an environmental gradient in Cameroon (Western Central Africa), Org Geochem, 78, 99-109, https://doi.org/10.1016/j.orggeochem.2014.09.007, 2015.

Schwark, L., Zink, K., and Lechterbeck, J.: Reconstruction of postglacial to early Holocene vegetation history in terrestrial Central Europe via cuticular lipid biomarkers and pollen records from lake sediments, Geology, 30, 463-466, 10.1130/0091-7613(2002)030<0463:Ropteh>2.0.Co;2, 2002.

Seki, O., Nakatsuka, T., Shibata, H., and Kawamura, K.: A compound-specific n-alkane $\delta^{13}C$ and δD approach for assessing source and delivery processes of terrestrial organic matter within a forested watershed in northern Japan, Geochim Cosmochim Ac, 74, 599-613, https://doi.org/10.1016/j.gca.2009.10.025, 2010.

Sessions, A. L.: Factors controlling the deuterium contents of sedimentary hydrocarbons, Org Geochem, 96, 43-64, https://doi.org/10.1016/j.orggeochem.2016.02.012, 2016.

Strömberg, C. A. E.: Evolution of grasses and grassland ecosystems, Annu Rev Earth Pl Sc, 39, 517-544, 10.1146/annurev-earth-040809-152402, 2011.

Suh, Y. J. and Diefendorf, A. F.: Seasonal and canopy height variation in *n*-alkanes and their carbon isotopes in a temperate forest, Org Geochem, 116, 23-34, https://doi.org/10.1016/j.orggeochem.2017.10.015, 2018.

Suh, Y. J., Diefendorf, A. F., Bowen, G. J., Cotton, J. M., and Ju, S.-J.: Plant wax integration and transport from the Mississippi River Basin to the Gulf of Mexico inferred from GIS-enabled isoscapes and mixing models, Geochim Cosmochim Ac, 257, 131-149, https://doi.org/10.1016/j.gca.2019.04.022, 2019.

Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, Quaternary Sci Rev, 35, 1-22, https://doi.org/10.1016/j.quascirev.2012.01.012, 2012.

Tipple, B. J. and Pagani, M.: A 35Myr North American leaf-wax compound-specific carbon and hydrogen isotope record: Implications for C_4 grasslands and hydrologic cycle dynamics, Earth Planet Sc Lett, 299, 250-262, https://doi.org/10.1016/j.epsl.2010.09.006, 2010.

Tipple, B. J., Berke, M. A., Doman, C. E., Khachaturyan, S., and Ehleringer, J. R.: Leaf-wax *n*-alkanes record the plant–water environment at leaf flush, P Natl Acad Sci USA, 110, 2659-2664, 10.1073/pnas.1213875110, 2013.

Tipple, B. J., Berke, M. A., Hambach, B., Roden, J. S., and Ehleringer, J. R.: Predicting leaf wax n-alkane 2H/1H ratios: controlled water source and humidity experiments with hydroponically grown trees confirm predictions of Craig–Gordon model, Plant Cell Environ, 38, 1035-1047, https://doi.org/10.1111/pce.12457, 2015.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Uno, K. T., Polissar, P. J., Jackson, K. E., and deMenocal, P. B.: Neogene biomarker record of vegetation change in eastern Africa, P Natl Acad Sci USA, 113, 6355-6363, 10.1073/pnas.1521267113, 2016.

Vesey-Fitzgerald, D. F.: Central African Grasslands, J Ecol, 51, 243-274, 10.2307/2257683, 1963.

Vogts, A., Moossen, H., Rommerskirchen, F., and Rullkötter, J.: Distribution patterns and stable carbon isotopic composition of alkanes and alkan-1-ols from plant waxes of African rain forest and savanna C_3 species, Org Geochem, 40, 1037-1054, https://doi.org/10.1016/j.orggeochem.2009.07.011, 2009.

Vogts, A., Schefuß, E., Badewien, T., and Rullkötter, J.: *n*-Alkane parameters from a deep sea sediment transect off southwest Africa reflect continental vegetation and climate conditions, Org Geochem, 47, 109-119, https://doi.org/10.1016/j.orggeochem.2012.03.011, 2012.

Wang, J., Xu, Y., Zhou, L., Shi, M., Axia, E., Jia, Y., Chen, Z., Li, J., and Wang, G.: Disentangling temperature effects on leaf wax *n*-alkane traits and carbon isotopic composition from phylogeny and precipitation, Org Geochem, 126, 13-22, https://doi.org/10.1016/j.orggeochem.2018.10.008, 2018.

Wang, R. Z.: C_4 plants in the vegetation of Tibet, China: Their natural occurrence and altitude distribution pattern, Photosynthetica, 41, 21-26, 10.1023/A:1025844009120, 2003.

Wang, Z. and Liu, W.: Carbon chain length distribution in *n*-alkyl lipids: A process for evaluating source inputs to Lake Qinghai, Org Geochem, 50, 36-43, https://doi.org/10.1016/j.orggeochem.2012.06.015, 2012.

Wiesenberg, G. L. B., Schwarzbauer, J., Schmidt, M. W. I., and Schwark, L.: Source and turnover of organic matter in agricultural soils derived from *n*-alkane/*n*-carboxylic acid compositions and C-isotope signatures, Org Geochem, 35, 1371-1393, https://doi.org/10.1016/j.orggeochem.2004.03.009, 2004.

Wu, M. S., Feakins, S. J., Martin, R. E., Shenkin, A., Bentley, L. P., Blonder, B., Salinas, N., Asner, G. P., and Malhi, Y.: Altitude effect on leaf wax carbon isotopic composition in humid tropical forests, Geochim Cosmochim Ac, 206, 1-17, https://doi.org/10.1016/j.gca.2017.02.022, 2017.

Yamamoto, S., Kawamura, K., Seki, O., Kariya, T., and Lee, M.: Influence of aerosol source regions and transport pathway on δD of terrestrial biomarkers in atmospheric aerosols from the East China Sea, Geochim Cosmochim Ac, 106, 164-176, https://doi.org/10.1016/j.gca.2012.12.030, 2013.

Yang, D.: SPATIAL-Lab/LipidMM: Arugula (v. 0.9.7), Zenodo [dataset], https://doi.org/10.5281/zenodo.6236846, 2022.

Yu, G., Tang, L., Yang, X., Ke, X., and Harrison, S. P.: Modern pollen samples from alpine vegetation on the Tibetan Plateau, Global Ecol Biogeogr, 10, 503-519, https://doi.org/10.1046/j.1466-822X.2001.00258.x, 2001.

Zech, M., Krause, T., Meszner, S., and Faust, D.: Incorrect when uncorrected: Reconstructing vegetation history using *n*-alkane biomarkers in loess-paleosol sequences – A case study from the Saxonian loess region, Germany, Quatern Int, 296, 108-116, https://doi.org/10.1016/j.quaint.2012.01.023, 2013.

Zhou, B., Bird, M., Zheng, H., Zhang, E., Wurster, C. M., Xie, L., and Taylor, D.: New sedimentary evidence reveals a unique history of C_4 biomass in continental East Asia since the early Miocene, Sci Rep-UK, 7, 170, 10.1038/s41598-017-00285-7, 2017.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Table 1. Lake surface sediment samples used in case study 1, with measured chain specific *n*-alkane concentrations and $\delta^{13}C$ values, originally published by Liu et al. (2015).

Sample	*n*-C_{27} (µg/g)	*n*-C_{29} (µg/g)	*n*-C_{31} (µg/g)	*n*-C_{27} $\delta^{13}C$ ± precision (‰, VPDB)	*n*-C_{29} $\delta^{13}C$ ± precision (‰, VPDB)	*n*-C_{31} $\delta^{13}C$ ± precision (‰, VPDB)	P_{aq}
High P_{aq} (QHS13-5S)	0.57	0.93	1.125	−28.5 ± 0.3	−31.5 ± 0.3	−32.1 ± 0.3	0.34
Mid-P_{aq} (QHS13-7S)	0.42	0.825	0.84	−34.0 ± 0.3	−33.1 ± 0.3	−32.7 ± 0.3	0.22
Low P_{aq} (QHS13-9S)	0.885	1.92	2.22	−32.0 ± 0.3	−32.6 ± 0.3	−32.5 ± 0.3	0.19

Table 2. Lake surface sediment samples used in case study 2, with measured chain specific *n*-alkane concentrations and $\delta^{13}C$ values, originally published by Garcin et al. (2014).

Sample	*n*-C_{29} (µg/g)	*n*-C_{31} (µg/g)	*n*-C_{33} (µg/g)	*n*-C_{29} $\delta^{13}C$ ± 1sd (‰, VPDB)	*n*-C_{31} $\delta^{13}C$ ± 1sd (‰, VPDB)	*n*-C_{33} $\delta^{13}C$ ± 1sd (‰, VPDB)	*f* C_4 ± 1sd (Garcin et al., 2014)
High C_4 (Rhum)	161	195	152	−28.6 ± 0.3	−29.7 ± 0.3	−24.8 ± 0.3	0.72 ± 0.25
Mid-C_4 (Asso)	132	76	55	−32.9 ± 0.2	−31.1 ± 0.8	−23.8 ± 0.8	0.31 ± 0.12
Low C_4 (Baro)	78	97	35	−35.3 ± 0.7	−36.2 ± 0.4	−32.7 ± 0	0.05 ± 0.12

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

Table 3. The Maximum A Posteriori probability estimates (MAP), the medians, and the 89% highest density intervals (HDI) of posterior densities of the mixing fractions as model output using data in published lake surface sediment samples from Lake Qinghai (Figure 4).

Sample		FLMC $_{Terrestrial}$	FLMC $_{Macrophyte}$	FLMC $_{Algae}$
	MAP	0.24	0.53	0.11
High P_{aq} (QHS13-5S)	Median	0.21	0.44	0.32
	89% HDI	[0.05, 0.36]	[0.14, 0.69]	[0.00, 0.72]
	MAP	0.83	0.01	0.09
Mid-P_{aq} (QHS13-7S)	Median	0.69	0.02	0.27
	89% HDI	[0.36, 0.99]	[0.00, 0.07]	[0.00, 0.62]
	MAP	0.61	0.09	0.09
Low P_{aq} (QHS13-9S)	Median	0.54	0.12	0.30
	89% HDI	[0.18, 0.83]	[0.00, 0.24]	[0.00, 0.69]

Table 4. The Maximum A Posteriori probability estimates (MAP), the medians, and the 89% highest density intervals (HDI) of posterior densities of the fractional leaf mass contributions (FLMCs) as model output using data in published lake surface sediment samples from Cameroon (Figure 8).

Sample		FLMC $_{C4\ plants}$	FLMC $_{Savanna\ C3}$	FLMC $_{Rainforest\ C3}$
	MAP	0.59	0.32	0.04
High C_4 sample (Rhum)	Median	0.57	0.27	0.14
	89% HDI	[0.43, 0.72]	[0.00, 0.48]	[0.00, 0.34]
	MAP	0.46	0.03	0.43
Mid-C_4 sample (Asso)	Median	0.47	0.08	0.42
	89% HDI	[0.33, 0.60]	[0.00, 0.26]	[0.25, 0.59]
	MAP	0.10	0.20	0.59
Low C_4 sample (Baro)	Median	0.12	0.31	0.56
	89% HDI	[0.03, 0.21]	[0.00, 0.59]	[0.27, 0.83]

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

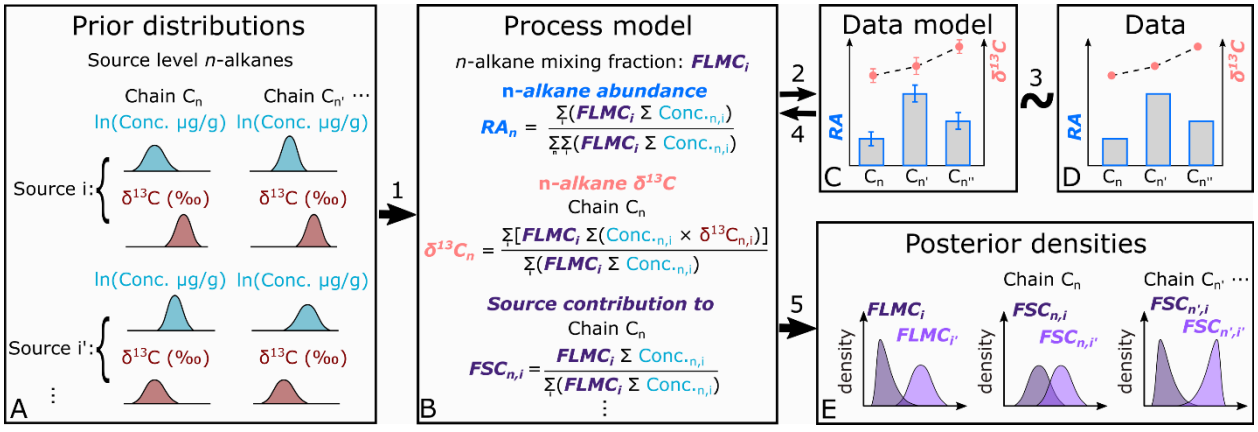

Figure 1. Proposed model structure of the Bayesian hierarchical framework for interpretation of *n*-alkane records; model components are represented by boxes; links between components are numbered and referenced in the Methods section.

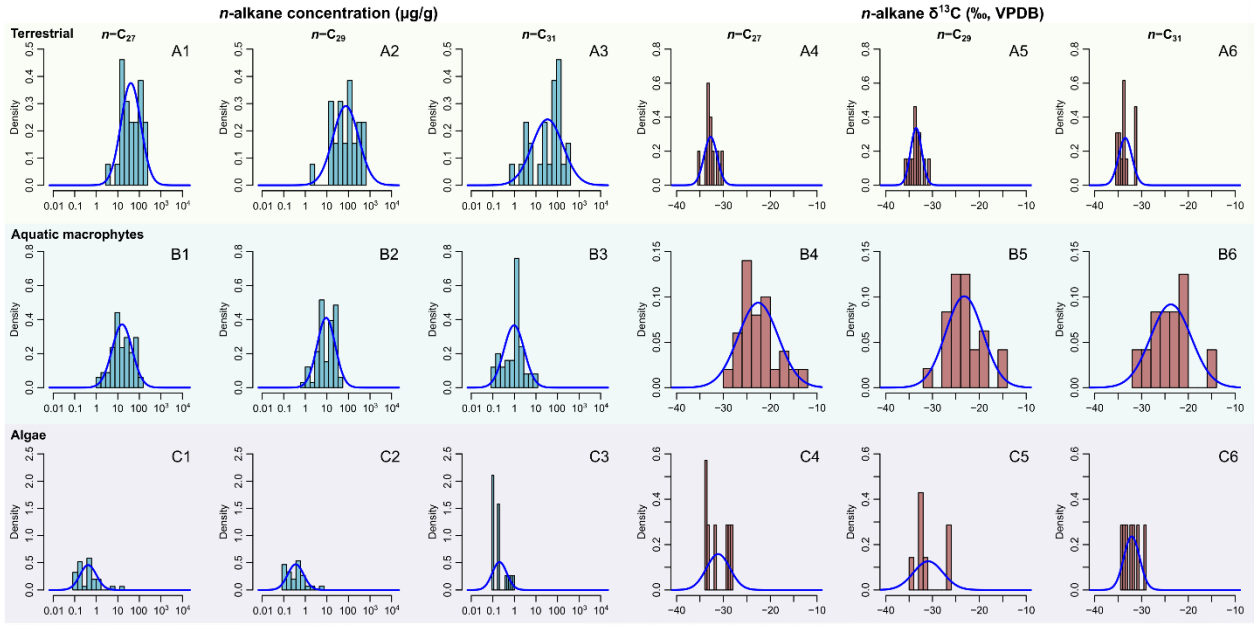

Figure 2. Empirical data of per sample *n*-alkane concentrations (µg/g of dried sample, blue histograms) and $\delta^{13}C$ (‰, VPDB, red histograms) of *n*-C_{27}, *n*-C_{29}, and *n*-C_{31} alkanes and their corresponding prior distributions (blue Gaussian curve overlays) of published plants in the terrestrial, aquatic macrophyte and algae sources; raw data are compiled in Supplementary Material; estimated prior parameters of each source, including the means and variance-covariance matrices, are reported in Tables S1 and S2, in the Supplementary Material.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

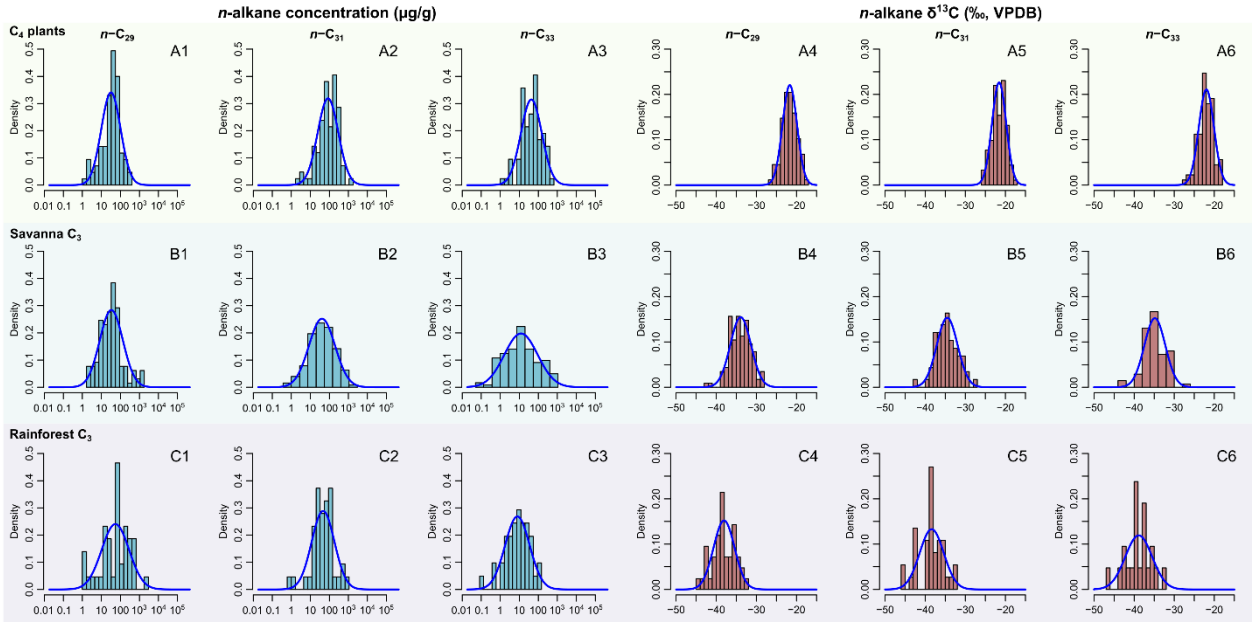

Figure 3. Empirical data of per sample *n*-alkane concentrations (µg/g of dried sample, blue histograms) and $\delta^{13}C$ (‰, VPDB, red histograms) of *n*-C_{29}, *n*-C_{31}, and *n*-C_{33} alkanes and their corresponding prior distributions (blue Gaussian curve overlays) of published plants in the C_4 plants, savanna C_3 and rainforest C_3 sources; raw data are compiled in Supplementary Material; estimated prior parameters of each source, including the means and variance-covariance matrices, are reported in Tables S3 and S4, in the Supplementary Material.

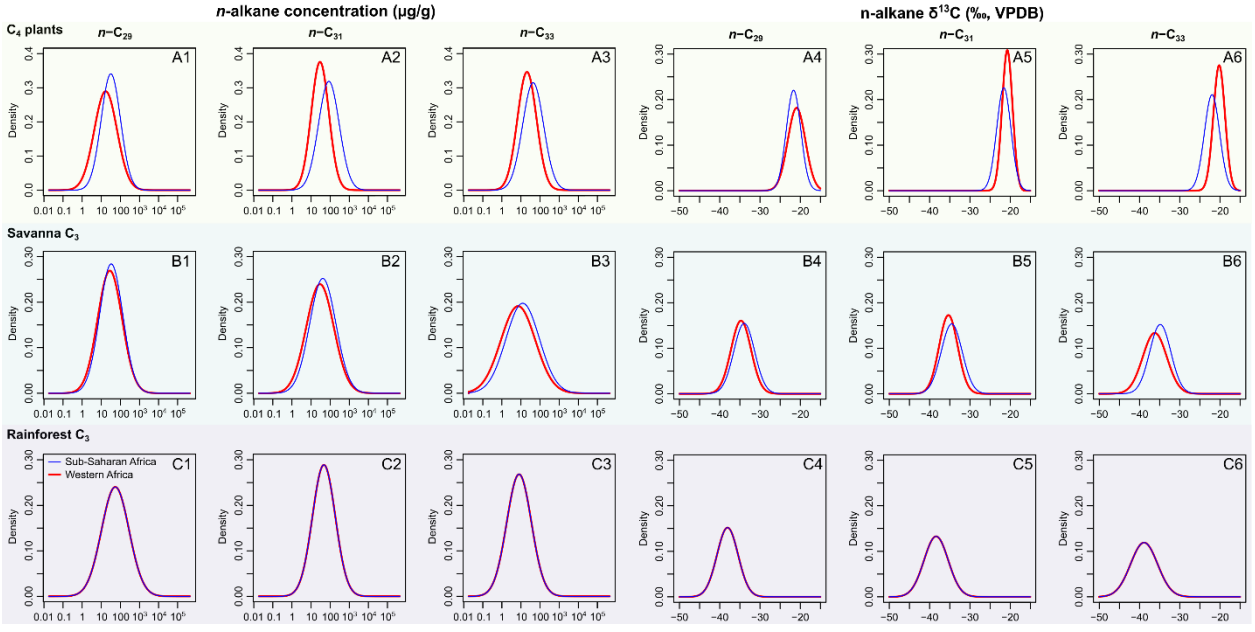

Figure 4. Comparisons of prior distributions based on wester African plant samples (thick red curves) vs. sub-Saharan African plant samples (thin blue curves); left three columns illustrate *n*-alkane concentrations; right three columns illustrate *n*-alkane $\delta^{13}C$ values.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

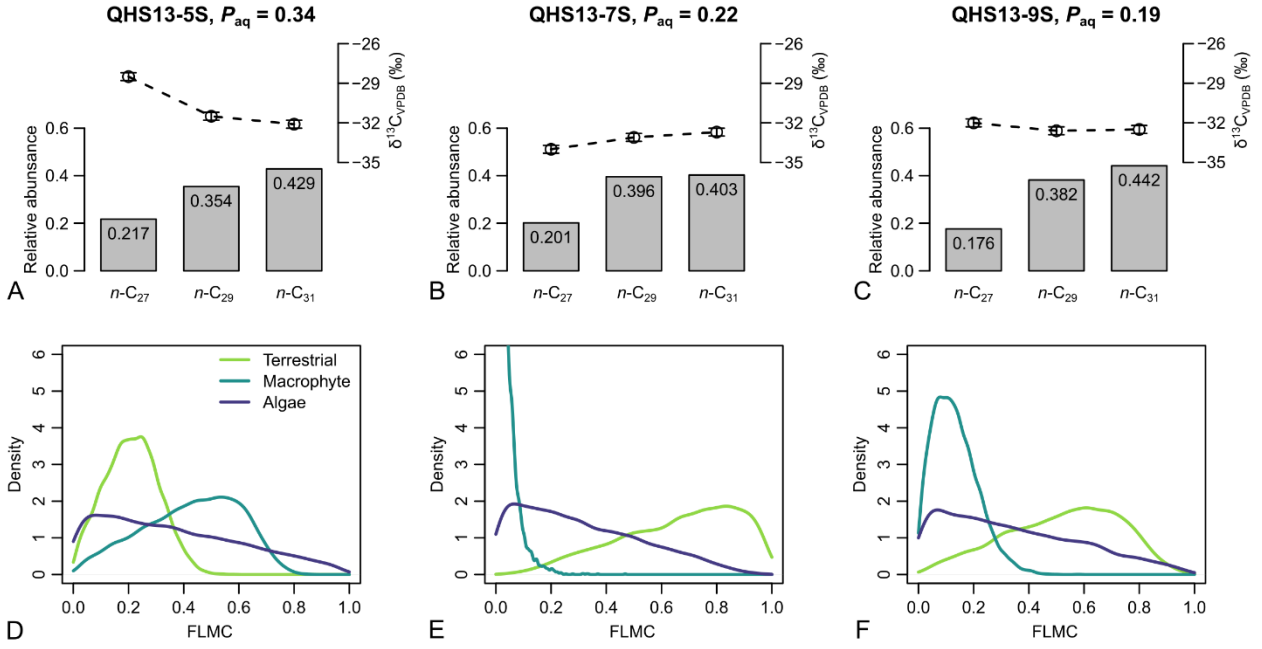

Figure 5. A – C: Sedimentary long-chain *n*-alkanes (*n*-C_{27}, *n*-C_{29} and *n*-C_{31}), their $\delta^{13}C$ (‰, VPDB) and relative abundance values (excluding other chains) in published lake surface sediment samples from Lake Qinghai (Liu et al., 2015); D – F: posterior densities of fractional leaf mass contribution (FLMC) of terrestrial plants, aquatic macrophytes and algae as model output conditioned on the $\delta^{13}C$ and relative abundance values.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

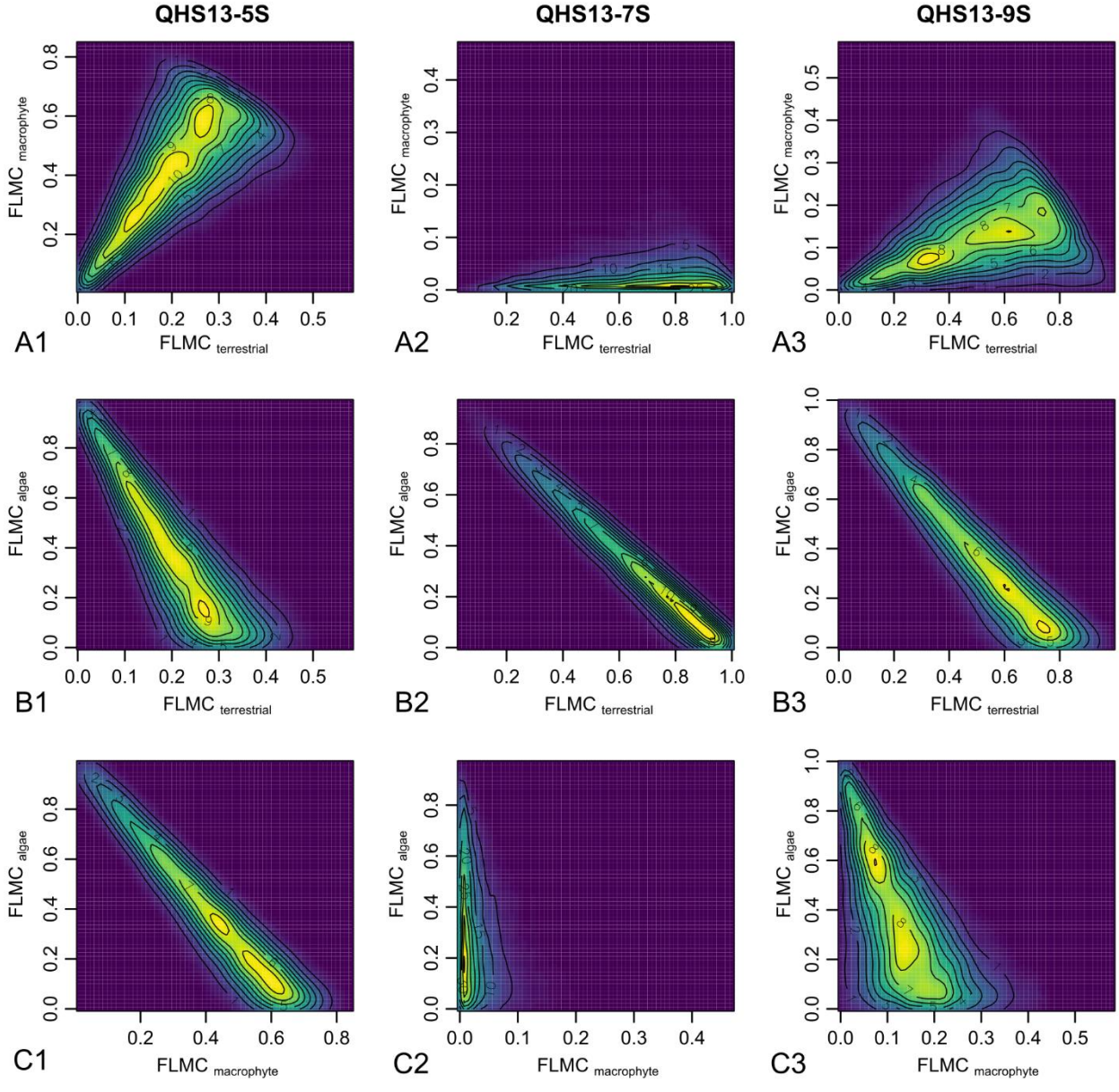

Figure 6. Bivariate density plots of the posterior densities of fractional leaf mass contribution (FLMC) of terrestrial plants, aquatic macrophytes and algae in published lake surface sediment samples from Lake Qinghai (Liu et al., 2015).

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

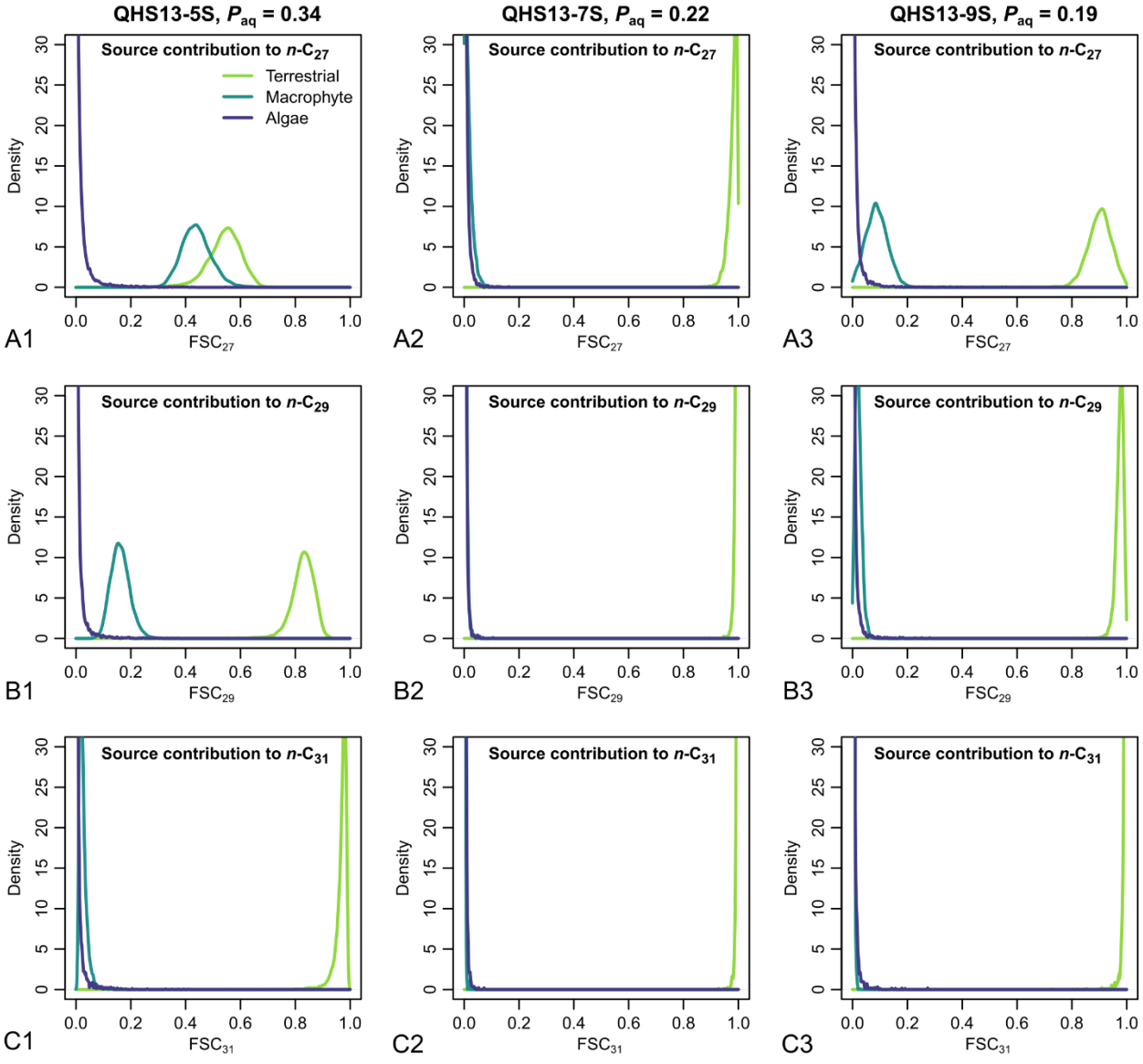

Figure 7. Posterior densities of fractional source contribution (FSC_n) of terrestrial plants, aquatic macrophytes, and algae to each specific alkane chains (n-C_{27}, n-C_{29} and n-C_{31}) in published lake surface sediment samples from Lake Qinghai (Liu et al., 2015).

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

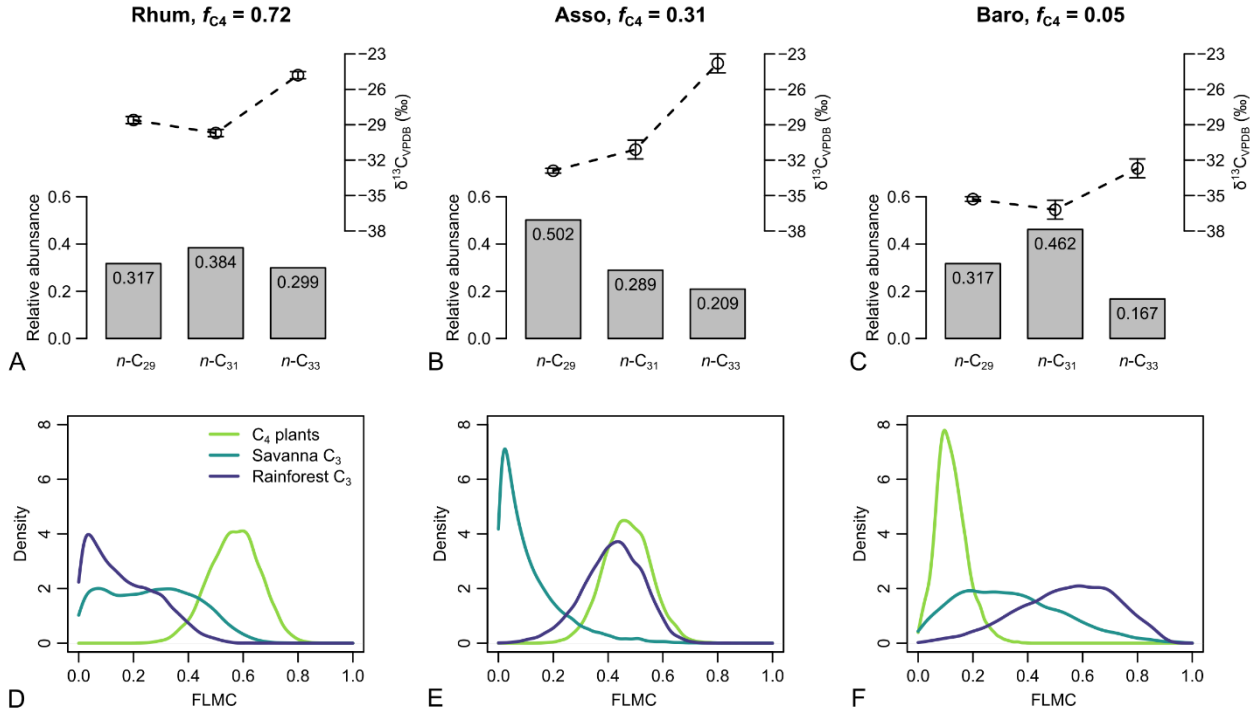

Figure 8. A – C: Sedimentary long-chain *n*-alkanes (*n*-C_{27}, *n*-C_{29} and *n*-C_{31}), their $\delta^{13}C$ (‰, VPDB) and relative abundance values (excluding other chains) in published lake surface sediment samples from Cameroon (Garcin et al., 2014); D – F: posterior densities of fractional leaf mass contribution (FLMC) of tropical C_4 plants, savanna C_3 plants, and rainforest C_3 plants as model output conditioned on the $\delta^{13}C$ and relative abundance values.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

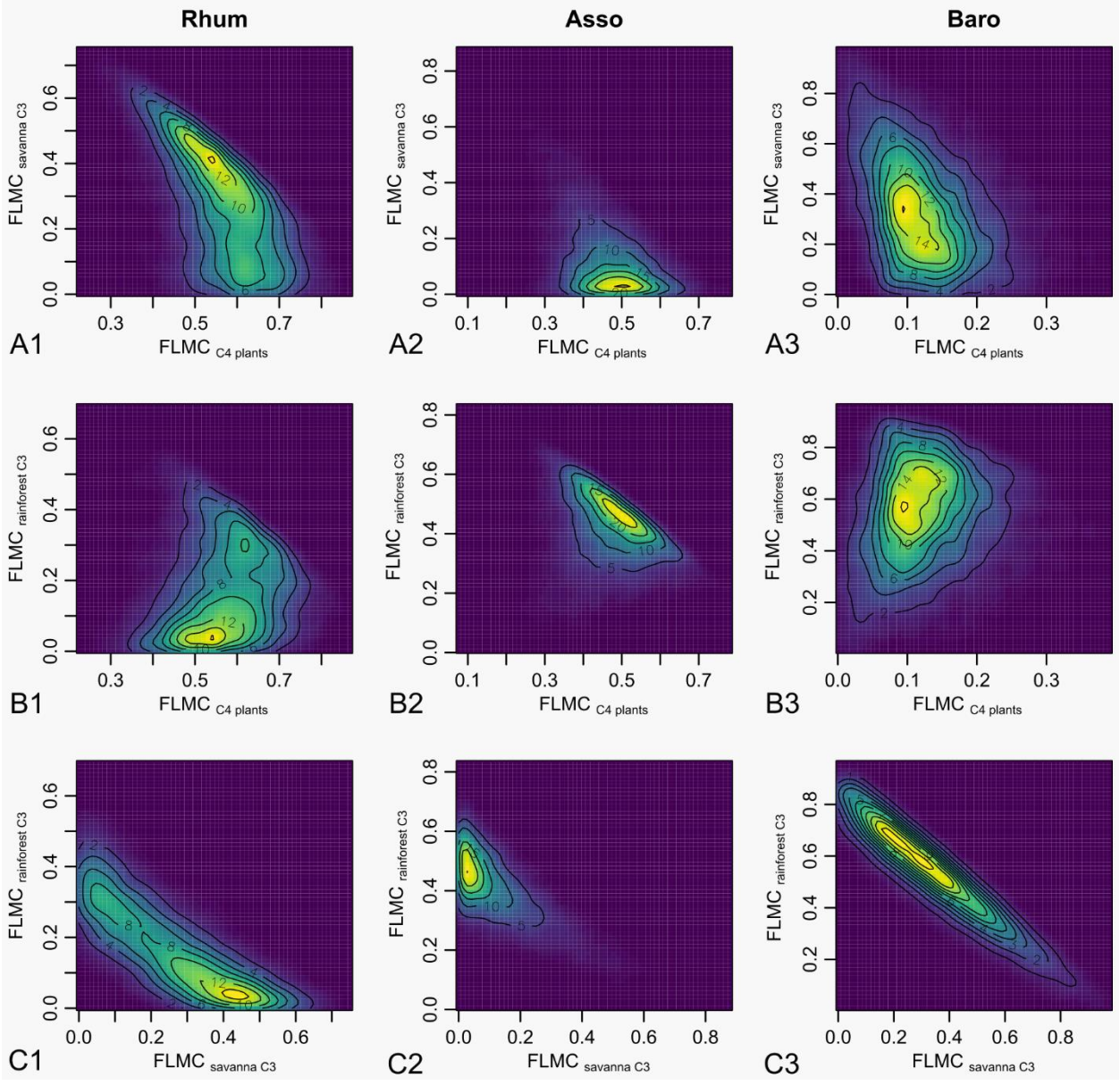

Figure 9. Bivariate density plots of the posterior densities of fractional leaf mass contribution (FLMC) of tropical C_4 plants, savanna C_3 plants, and rainforest C_3 plants in published lake surface sediment samples from Cameroon.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

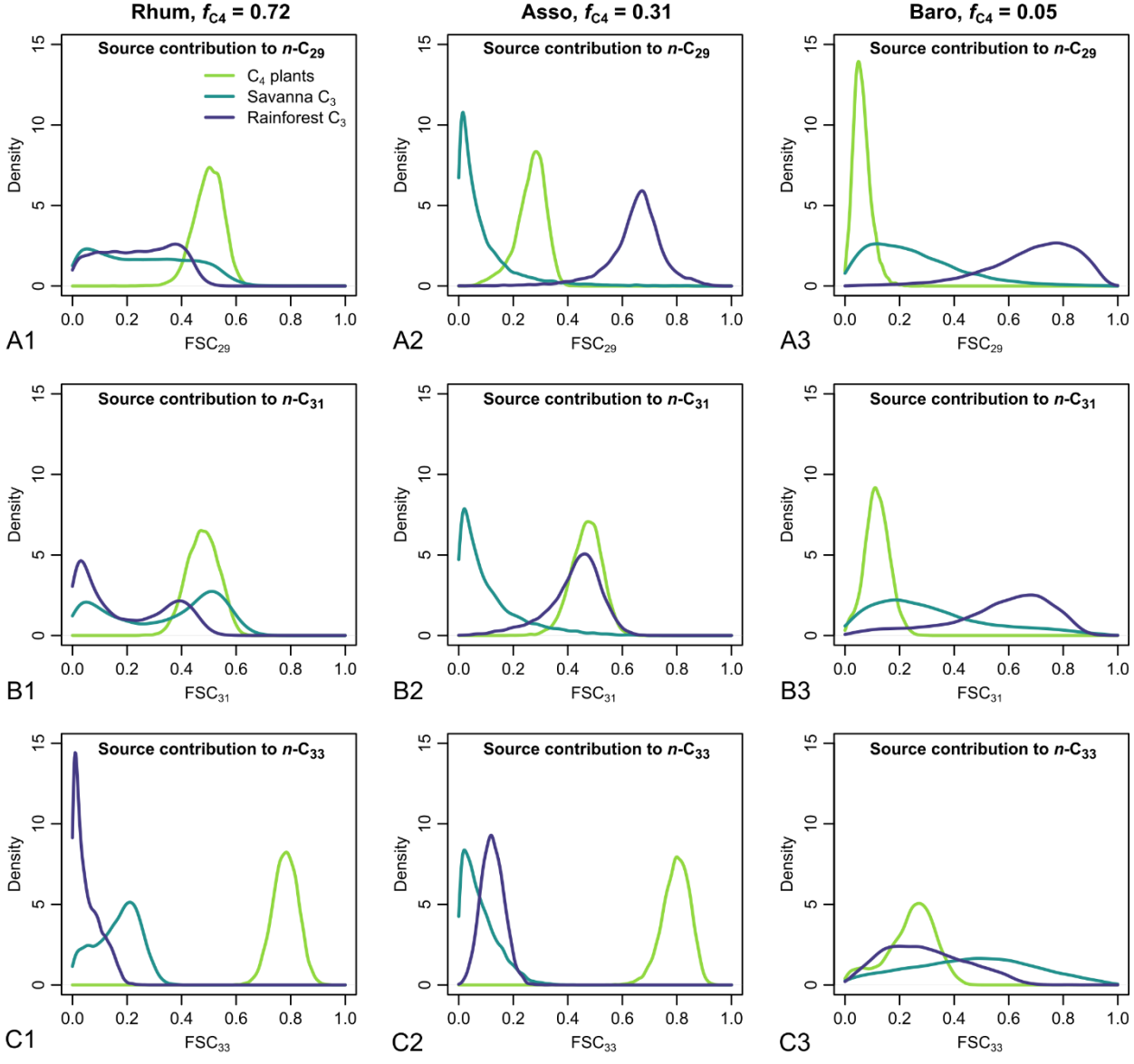

Figure 10. Posterior densities of fractional source contribution (FSC_n) of terrestrial plants, aquatic macrophytes, and algae to each specific alkane chain (n-C_{29}, n-C_{31} and n-C_{33}) in published lake surface sediment samples from Cameroon.

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

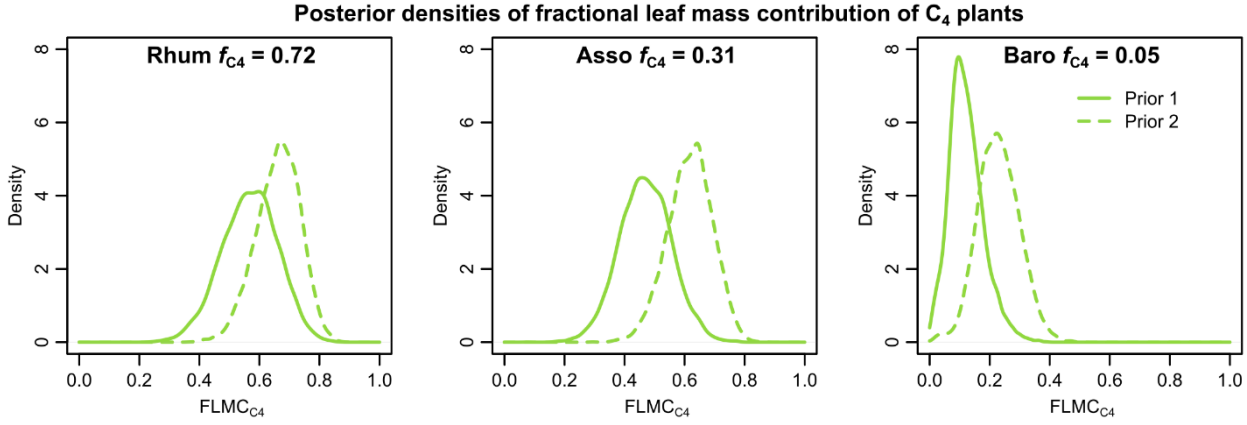

Figure 11. Comparisons of posterior densities of fractional leaf mass contribution of the C_4 source ($FLMC_{C4}$) based on two sets of prior distributions; Prior 1: sub-Saharan Africa dataset; Prior 2 (shaded): western Africa dataset; detailed prior distributions are illustrated in Figure 4; f_{C4} values are from Garcin et al. (2014).

https://doi.org/10.5194/egusphere-2022-23
Preprint. Discussion started: 4 March 2022

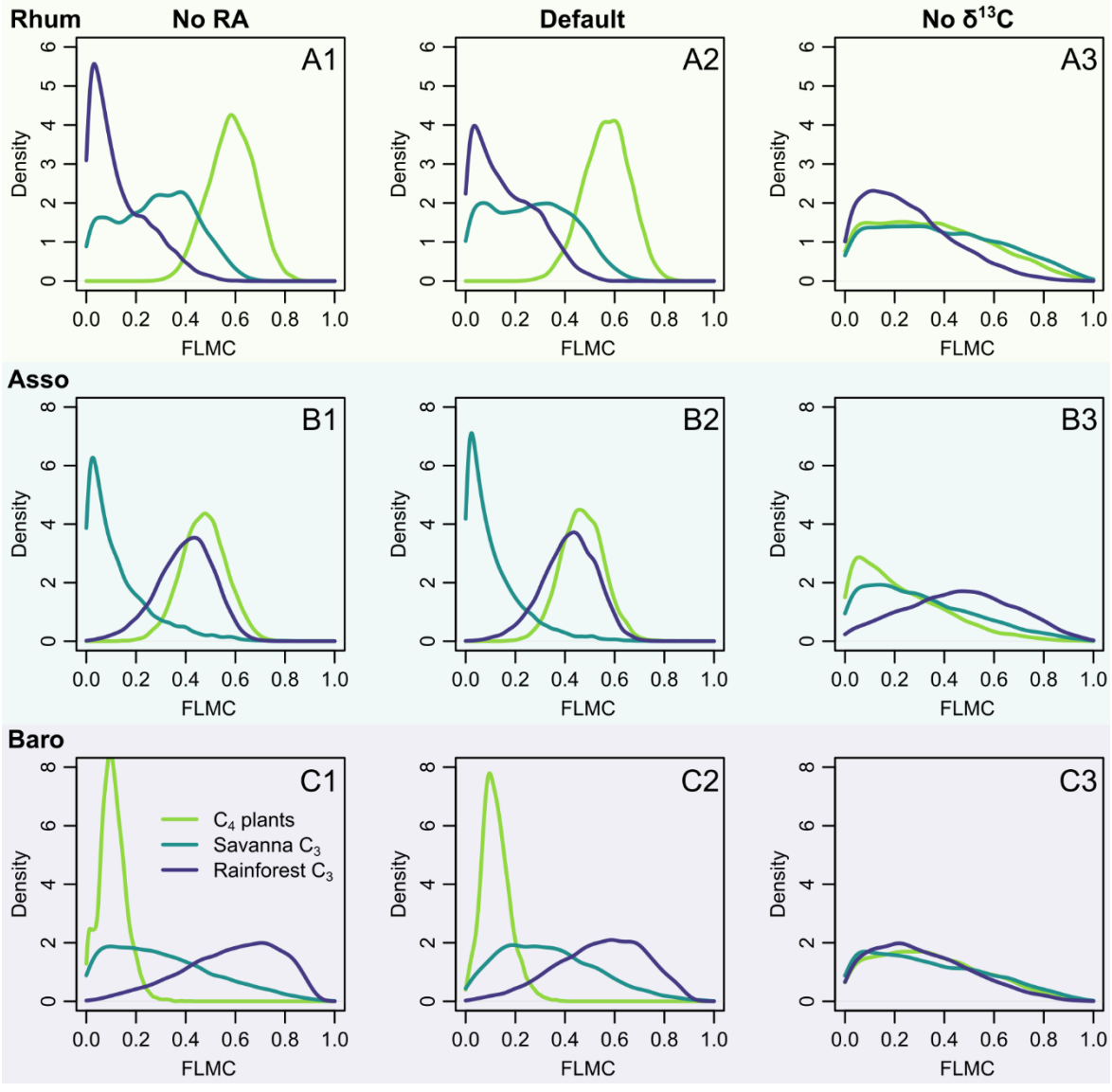

Figure 12. Model sensitivity to proxy type in model inversion (Equations 9 and 10), using case study 2 as an example; the right column (No RA) shows the model output with the likelihood evaluations of RA completely removed from model evaluation; the left column (No $\delta^{13}C$) shows the model output with the likelihood evaluations of $\delta^{13}C$ completely removed from model evaluation.