https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

Biogeosciences Discussions
Open Access
EGU

Development of global temperature and pH calibrations based on bacterial 3-hydroxy fatty acids in soils

Pierre Véquaud[1], Sylvie Derenne[1], Alexandre Thibault[2], Christelle Anquetil[1], Giuliano Bonanomi[3], Sylvie Collin[1], Sergio Contreras[4], Andrew Nottingham[5], Pierre Sabatier[6], Norma Salinas[7], Wesley Phillip Scott[8], Josef P. Werne[8], Arnaud Huguet[1]

[1]Sorbonne Université, CNRS, EPHE, PSL, UMR METIS, Paris, 75005, France
[2]Antea Group, Innovation Hub, 803 boulevard Duhamel du Monceau, Olivet, 45160, France
[3]Dipartimento di Agraria, Università di Napoli Federico II, via Università 100, Portici, NA, 80055, Italy
[4]Laboratorio de Ciencias Ambientales (LACA), Departamento de Química Ambiental, Facultad de Ciencias & Centro de Investigación en Biodiversidad y Ambientes Sustentables (CIBAS), Universidad Católica de la Santísima Concepción, Casilla 297, Concepción, Chile
[5]School of Geosciences, University of Edinburgh, Crew Building, Kings Buildings, Edinburgh EH9 3FF United Kingdom
[6]Univ. Savoie Mont Blanc, CNRS, EDYTEM, Le Bourget du Lac, 73776, France
[7]Instituto de Ciencias de la Naturaleza, Territorio y Energías Renovables, Pontificia Universidad Catolica del Peru, Av. Universitaria 1801, San Miguel, Lima 32, Peru
[8]Department of Geology and Environmental Science, University of Pittsburgh, Pittsburgh, PA 15260, USA

Correspondence to: Arnaud Huguet (arnaud.huguet@sorbonne-universite.fr)

Abstract. 3-hydroxy fatty acids (3-OH FAs) with 10 to 18 C atoms are membrane lipids mainly produced by Gram-negative bacteria. They have been recently proposed as temperature and pH proxies in terrestrial settings. Nevertheless, the existing correlations between pH/temperature and indices derived from 3-OH FA distribution (RIAN, RAN_{15} and RAN_{17}) are based on a small soil dataset (ca. 70 samples) and only applicable regionally. The aim of this study was to investigate the applicability of 3-OH FAs as mean annual air temperature (MAAT) and pH proxies at the global level. This was achieved using an extended soil dataset of 168 topsoils distributed worldwide, covering a wide range of temperatures (5°C to 30°C) and pH (3 to 8). The response of 3-OH FAs to temperature and pH was compared to that of established branched GDGT-based proxies (MBT'$_{5Me}$/CBT). Strong linear relationships between 3-OH FA-derived indices (RAN_{15}, RAN_{17} and RIAN) and MAAT/pH could only be obtained locally, for some of the individual transects. This suggests that these indices cannot be used as paleoproxies at the global scale using simple linear regression models, in contrast with the MBT'$_{5Me}$ and CBT. However, strong global correlations between 3-OH FA relative abundances and MAAT/pH were shown by using other algorithms (multiple linear regression, k-NN and random forest models). The applicability of the k-NN and random forest models for paleotemperature reconstruction was tested and compared with the MAAT record from a Chinese speleothem. The calibration based on the random forest model appeared to be the most robust. It showed

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

Biogeosciences
Discussions
Open Access
EGU

similar trends with previously available records and highlighted known climatic events poorly visible when using local 3-OH FA calibrations. Altogether, these results demonstrate the potential of 3-OH FAs as paleoproxies in terrestrial settings.

Keywords: 3-hydroxy fatty acids; branched GDGTs; soils; global calibration; temperature and pH proxy

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

Biogeosciences Discussions

1. Introduction

Investigating past climate variations is essential to understand and predict future environmental changes, especially in the context of global anthropogenic change. Direct records of environmental parameters are available for the last decades, the so-called "instrumental" period. Beyond this period, proxies can be used to obtain indirect information on environmental parameters. A major challenge is to develop reliable proxies which can be applied to continental environments in addition to marine ones. Indeed, available proxies have been mainly developed and used in marine settings, as the composition and mechanism of formation of marine sedimentary cores is less complex than in continental settings, which are highly heterogeneous. Several environmental proxies based on organic (e.g. the alkenone unsaturation index ($U^{k'}_{37}$; Brassell et al., 1986) and inorganic (Mg/Ca ratio and $^{18}O/^{16}O$ ratio of foraminifera; Emiliani, 1955; Erez and Luz, 1983) fossil remains were notably developed for the reconstruction of sea surface temperatures.

Some of the existing proxies are based on membrane lipids synthesized by certain microorganisms (Eglinton and Eglinton, 2008). These microorganisms are able to adjust the composition of their membrane lipids in response to the prevailing environmental conditions in order to maintain an appropriate fluidity and to ensure the optimal state of the cellular membrane (Singer and Nicolson, 1972; Sinensky, 1974; Hazel and Williams, 1990; Denich et al., 2003). The structure of glycerol dialkyl glycerol tetraethers (GDGTs), which are membrane lipids biosynthesized by archaea and some bacteria, is especially known to be related to environmental conditions. Archaeal GDGTs are constituted of isoprenoid alkyl chains ether-linked to glycerol, whereas bacterial GDGTs are characterized by branched alkyl chains instead of isoprenoid ones. The latter compounds are ubiquitous in terrestrial (De Jonge et al., 2014; Naafs et al., 2017; Peterse et al., 2012; Weijers et al., 2007) and aquatic environments (Loomis et al., 2012; Peterse et al., 2015; Weber et al., 2015). These branched GDGTs (brGDGTs) are produced by still unidentified bacteria, although some of them may belong to the phylum *Acidobacteria* (Sinninghe Damsté et al., 2011, 2014, 2018). The analysis of brGDGTs in a large number of soils distributed worldwide showed that the relative distribution of these compounds is mainly related to mean annual air temperature (MAAT) and soil pH (Weijers et al., 2007; Peterse et al., 2012; De Jonge et al., 2014). Despite improvements in brGDGT analytical methods and development of refined calibration models (De Jonge et al., 2014; Dearing Crampton-Flood et al., 2020), the Root Mean Square Error (RMSE) associated with MAAT reconstruction using the global brGDGT calibrations in soils remains high (>4°C). Thus,

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


development of new molecular proxies, independent of and complementary to brGDGTs, is essential to improve the reliability of temperature reconstructions in such settings.

Recent studies have unveiled the potential of another family of lipids − 3-hydroxy fatty acids (3-OH FAs) − for temperature and pH reconstructions in terrestrial (Wang et al., 2016, 2018; Huguet et al., 2019) and marine (Yang et al., 2020) settings. 3-OH FAs with 10 to 18 carbon atoms are specifically produced by Gram-negative bacteria and are bound to the lipopolysaccharide (LPS) by ester or amide bonds (Wollenweber and Rietschel, 1990; Wollenweber et al., 1982). Three types of 3-OH FAs can be distinguished, with either *normal* chains or branched chains, *iso* or *anteiso*.

The analysis of 3-OH FAs in soils showed that the RAN_{15} and RAN_{17} indices, defined as the ratio of C_{15} or C_{17} *anteiso* 3-OH FA to *normal* C_{15} or C_{17} 3-OH FA, were negatively correlated with MAAT along the three mountains investigated so far: Mts. Shennongjia (China; Wang et al., 2016), Rungwe and Majella (Tanzania and Italy, respectively; Huguet et al., 2019). This suggests that Gram-negative bacteria producing these fatty acids similarly respond to colder temperatures with an increase in *anteiso*-C_{15}/C_{17} vs. *n*-C_{15}/C_{17} 3-OH FAs, in order to maintain a proper fluidity and optimal state of the bacterial membrane, the so-called homeoviscous adaptation mechanism (Hazel and Eugene Williams, 1990; Sinensky, 1974). Nevertheless, the relationships between RAN_{15} and MAAT along the three mountain transects showed the same slopes but different intercepts (Wang et al., 2016; Huguet et al., 2019), suggesting that regional calibrations may be more adapted to apply RAN_{15} as a temperature proxy in soils. In contrast, a significant calibration between RAN_{17} and MAAT could be established using combined data from the three mountain regions (Wang et al., 2016; Huguet et al., 2019).

Another index, the RIAN, defined as the cologarithm of the sum of *anteiso* and *iso* 3-OH FAs divided by the sum of *normal* homologues, was shown to be strongly negatively correlated with soil pH along the three aforementioned mountains (Wang et al., 2016; Huguet et al., 2020), reflecting a general relative increase in normal homologues compared to branched (*iso* and *anteiso*) ones with increasing pH. This mechanism was suggested to reduce the permeability and fluidity of the membrane for the cell to cope with lower pH (Beales, 2004; Denich et al., 2003; Russell et al., 1995).

3-OH FA indices were recently applied for the first time to the reconstruction of the temperature and hydrological changes over the last 10,000 years in a speleothem from China (Wang et al., 2018), showing the potential of 3-OH FAs as independent tools for environmental reconstruction in terrestrial settings. A very recent study based on marine sediments from the

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


North Pacific Ocean suggested that the distribution of 3-OH FAs could also be used to reconstruct sea surface temperature (Yang et al., 2020).

Even though these results are promising, the linear regressions between pH/MAAT and 3-OH FA indices in terrestrial environments are still based on a rather small dataset (ca. 70 soil samples; Wang et al., 2016; Huguet et al., 2019). The aim of this study was to investigate the applicability of 3-OH FAs as MAAT and pH proxies at the global level using an extended soil dataset and a more developed statistical approach. 3-OH FA distribution from 54 soils was collected along globally distributed elevational transects (Tibet, Italy, Peruvian Andes and Chile) and was combined with data previously published by Wang et al. (2016; Mt Shennongjia, China), Huguet et al. (2019; Mt. Rungwe, Tanzania and Mt. Majella, Italy) and Véquaud et al. (under revision; Mts. Lautaret-Bauges, France), leading to a total of 168 samples. Even though reconstruction using the global brGDGT calibrations present large uncertainties, there are widely used as MAAT/pH proxies. They can be considered as a reference proxy and were analyzed concomitantly to 3-OH FAs in the dataset. In addition to linear regressions, non-parametric, machine learning models were used to improve the global relationships between 3-OH FA distribution and MAAT/pH. Finally, these new models were tested and compared by applying them to a speleothem archive (Wang et al., 2018) representing to date the only available MAAT record derived from 3-OH FA proxies in continental setting.

2. Material and methods

2.1. Soil dataset

2.1.1. Study sites

The dataset of the present study is comprised of the globally distributed surface soils previously analyzed for brGDGTs and 3-OH FAs and collected along 4 altitudinal transects: Mts. Shennongjia (China; Yang et al., 2015; Wang et al., 2016), Rungwe (Tanzania ; Coffinet et al., 2017; Huguet et al., 2019), Majella (Italy; Huguet et al., 2019) and Lautaret-Bauges (France; Véquaud et al., under revision). This set was extended with surficial soils (0-10 cm) from 4 additional elevational transects described below, located in Italy, Tibet, Peru and Chile (Table 1).

Soil samples were collected from 13 sites along Mount Pollino in the Calabria region (Italy) between 0 and 2,200 m above sea level (a.s.l.) (Table 1). Mt. Pollino is located in the

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


155 calcareous Apennine range and is 2,248 m a.s.l. It is framed to the northwest by the Sierra de
156 Prete (2,181 m high) and to the south by the Pollino Abyss. The alpine to subalpine area (above
157 2,100 m a.s.l.) is characterized by the presence of Mediterranean grasslands (*Festuca bosniaca*,
158 *Carex kitaibeliana*) and the presence of sinkholes (Scalercio et al., 2014; Todaro et al., 2007).
159 The mountainous vegetation (over 1,200 m a.s.l.) is dominated by *Fagus sylvatica* forests and,
160 at the treeline, by scattered *Pinus leucodermis* (Bonanomi et al., 2020). The soil is poorly
161 developed and dominated by calcareous soils. Between 0 to 1,200 m a.s.l (Scalercio et al., 2014
162 and reference therein), Mt. Pollino is characterized by the presence of *Q. ilex* forests or shrubs.
163 Climate along this mountain is humid Mediterranean, with high summer temperatures and an
164 irregular distribution of rainfall throughout the year with pronounced summer drought (39.5%
165 in winter, 23.7% in spring, 29.2% in autumn, 7.6% in summer; average annual precipitation:
166 1,570 mm; see Todaro et al., 2007). MAAT is comprised between 7 °C (2,200 m a.s.l) and 18
167 °C (0 m a.s.l; Scalercio et al., 2014). MAAT along Mt. Pollino was estimated using a linear
168 regression between two MAAT (16°C at 400 m a.s.l and 10°C at 1,600 m a.s.l.) from the
169 meteorological data (Castrovillari station) recorded by Scalercio et al. (2014). The pH of the
170 soils analyzed in the present study ranges between 4.5 and 6.8 (Table 1).
171 Soil samples were collected from 17 sites along along Mount Shegyla between 3,106
172 and 4,474 m a.s.l. (southeastern Tibet, China), as previously described by Wang et al. (2015).
173 Different climatic zonations are observed along this high-altitude site (2,700 to 4,500 m a.s.l):
174 (i) a mountainous temperate zone between 2,700 and 3,400 m, (ii) a subalpine cold temperate
175 zone between 3,400 and 4,300 m and (iii) a cold alpine zone above 4,300 m. Plant species, such
176 as brown oak (*Q. semecarpifolia*) or common fir (*Abies alba*) are abundant within the
177 mountainous and subalpine levels. In the cold subalpine zone, the forest fir (*Abies georgei var.*
178 *smithii)* is endemic to western China. In the cold alpine zone, coniferous species (*Sabina*
179 *saltuaria*) as well as species typical of mountainous regions such as *Rhododendron* are
180 observed. MAAT was estimated using a linear regression between 7 measured MAAT from the
181 data recorded by Wang et al. (2015). The average MAAT along the transect is 4.6°C, with a
182 minimum of 1.1 °C at ca. 4,500 m a.s.l. and a maximum of 8.9 °C at ca. 3,100 m a.s.l. (Table
183 1). Soil pH ranges between 4.6 and 6.4 (Table 1).
184 Soils were sampled from 14 sites in the Peruvian Andes along the Kosñipata transect,
185 located in south-eastern Peru, in the upper part of the Madre de Dios/Madeira watershed, east
186 of the Andes Cordillera (Nottingham et al., 2015). This transect (190 m to 3,700 m a.s.l) is well-
187 documented and is the object of numerous ecological studies (Malhi et al., 2010; Nottingham
188 et al., 2015). There is a shift in vegetation zonation with increasing elevation, from tropical

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


lowland forest to montane cloud forest and high-elevation 'Puna' grassland. The tree line lies between 3,200 and 3,600 m a.s.l. For the 14 sites sampled in this study, the lower 13 sites are forest and the highest site is grassland. The 14 sites are part of a network of 1 ha forest plots (Nottingham et al., 2015); for each 1 ha plot, 0-10 cm surface soil was sampled from 5 systematically distributed locations within each 1 ha plot. Mean annual precipitation does not vary significantly with altitude (mean =2448 $mm.y^{-1}$, SD = 503 $mm.y^{-1}$; Rapp and Silman, 2012; Nottingham et al., 2015). MAAT is comprised between 26.4 °C at 194 m altitude and 6.5°C at 3644 m altitude (Table 1). The pH is characteristic of acidic soils (3.4 - 4.7; Table 1). Further information on these sites and soils is available in Nottingham et al. (2015).

Soil samples were collected from 10 sites between 690 m and 1,385 m a.s.l. from the lake shore (20 to 50 m offshore) of 10 Andean lakes located in Chile (38–39°S) within the temperate forest (Table 1). High-frequency measurements (every hour) of *in situ* soil temperature over a period of one year are available for the different sampling sites. MAAT is comprised between 5.75°C and 9.2°C. Soil pH ranges between 4.4 and 6.8 (Table 1).

2.1.2. pH measurement

Following sampling, soils were immediately transported to the laboratory and stored at -20 °C. Soil samples from the Peruvian Andes, Mt. Pollino and Mt. Shegyla were then freeze-dried, ground and sieved at 2 mm. The pH of the freeze-dried samples was measured in ultrapure water with a 1:2.5 soil water ratio. Typically, 10 ml of ultrapure water were added to 4 g of dry soil. The soil solution was stirred for 30 min, before decantation for 1 hand pH measurement (Carter et al., 2007).

2.2. Lipid analyses

BrGDGTs and 3-OH FAs were analyzed in all samples from the Peruvian Andes, Chilean Andes, Mt. Pollino and Mt. Shegyla.

2.2.1. 3-OH FA analysis

Sample preparation for 3-OH analysis was identical to that reported by Huguet et al. (2019) and Véquaud et al. (under revision). Soil samples were subjected to acid hydrolysis (3 M HCl) and extracted with organic solvents. This organic fraction was then rotary-evaporated, methylated in a 1M HCl-MeOH solution at 80 °C for 1 h and separated into three fractions over an activated silica column: (i) 30 ml of heptane/EtOAc (98: 2), (ii) 30 ml of EtOAc and (iii) 30

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


ml of MeOH. 3-OH FAs contained in the second fraction were derivatized at 70°C for 30 min with a solution of *N,O*- bis(trimethylsilyl)trifluoroacetamide (BSTFA) – Trimethylchlorosilane (TMCS) 99:1 (Grace Davison Discovery Science, USA) before gas chromatography-mass spectrometry (GC-MS) analysis.

3-OH FAs were analyzed with an Agilent 6890N GC-5973N using a Restek RXI-5 Sil MS silica column (60 m × 0.25 mm, i.d. 0.25 µm film thickness), as previously described (Huguet et al., 2019). 3-OH FAs were quantified by integrating the appropriate peak on the ion chromatogram and comparing the area with an internal standard (3-hydroxytetradecanoic acid, 2,2,3,4,4-d5; Sigma-Aldrich, France). The internal standard (0.5 mg/ml) was added just before injection as a proportion of 3 µl of standard to 100 µl of sample, as detailed by Huguet et al. (2019). The different 3-OH FAs were identified based on their retention time, after extraction of the characteristic *m/z* 175 fragment (*m/z* 178 for the deuterated internal standard; cf. Huguet et al., 2019).

The RIAN index was calculated as follows (Wang et al., 2016 ; Eq. 3) in the range C_{10}-C_{18} :

$$RIAN = -\log[(I + A)/ N] \quad (3)$$

where I, A, N represent the sum of all *iso*, *anteiso* and *normal* 3-OH FAs, respectively.

RAN_{15} and RAN_{17} indices are defined as follows (Wang et al., 2016; Eq. 4 and 5):

$$RAN_{15} = [\textit{anteiso}\ C_{15}] / [\textit{normal}\ C_{15}] \quad (4)$$

$$RAN_{17} = [\textit{anteiso}\ C_{17}] / [\textit{normal}\ C_{17}] \quad (5)$$

Analytical errors associated with the calculation of RIAN, RAN_{15} and RAN_{17} indices are respectively 0.006, 0.3 and 0.2 based on the analysis of one sample injected nine times during the analysis and five samples injected in triplicates.

2.2.2. brGDGT analysis

Sample preparation for brGDGT analysis was similar to that reported by Coffinet et al. (2014). Briefly, ca. 5-10 g of soil was extracted using an accelerated solvent extractor (ASE 100, Dionex-ThermoScientific, USA) with a dichloromethane (DCM) / methanol (MeOH) mixture (9: 1) for 3×5 min at 100 °C and a pressure of 100 bars in 34 ml cells. The total lipid extract was rotary evaporated and separated into two fractions of increasing polarity on a column of activated alumina: (i) 30 ml of heptane: DCM (9: 1, v:v) ; (ii) 30 ml of DCM: MeOH (1: 1, v:v). GDGTs are contained in the second fraction, which was rotary evaporated. An

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


aliquot (300 µL) was re-dissolved in heptane and centrifuged using an Eppendorf MiniSpin centrifuge (Eppendorf AG, Hamberg, Germany) at 7000 rpm for 1 min.

GDGTs were then analyzed by high pressure liquid chromatography coupled with mass spectrometry with an atmospheric pressure chemical ionisation source (HPLC-APCI-MS) using a Shimadzu LCMS 2020. GDGT analysis was performed using two Hypersil Gold silica columns in tandem (150 mm × 2.1 mm, 1.9 µm; Thermo Finnigan, USA) thermally controlled at 40 °C, as described by Huguet et al. (2019). This methodology enables the separation of 5- and 6-methyl brGDGTs. Semi-quantification of brGDGTs was performed by comparing the integrated signal of the respective compound with the signal of a C_{46} synthesized internal standard (Huguet et al., 2006) assuming their response factors to be identical.

The MBT'$_{5Me}$ index, reflecting the average number of methyl groups in 5-methyl isomers of GDGTs and considered as related to MAAT, was calculated according to De Jonge et al. (2014; Eq. 1):

$$\text{MBT'}_{5\text{Me}} = \frac{[Ia+Ib+Ic]}{[Ia+Ib+Ic]+[IIa+IIb+IIc]+[IIIa]} \quad (1)$$

The CBT index, reflecting the average number of cyclopentyl rings in GDGTs and considered as related to pH, was calculated as follows (Peterse et al., 2012; Eq. 2):

$$CBT = -\log\left(\frac{[Ib]+[IIb+II'b]}{[Ia]+[IIa+II'a]}\right) \quad (2)$$

The Roman numerals correspond to the different GDGT structures presented in De Jonge et al. (2014). The 6-methyl brGDGTs are denoted by an apostrophe after the Roman numerals for their corresponding 5-methyl isomers. Analytical errors associated with the calculation of MBT'$_{5Me}$ and CBT indices are 0.015 and 0.02, respectively, based on the analysis of three samples in triplicate among the 44 soil samples.

2.3. Statistical analysis

In order to investigate the correlations between environmental variables (pH, MAAT) and the relative abundances of bacterial lipids (brGDGTs and 3-OH FAs) or the indices based on these compounds, pairwise correlation matrices were performed in addition to single or multiple linear regressions. As the dataset is not normally distributed, Spearman correlation was used with a confidence level of 5%.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


Principal component analyses (PCA) were performed on the different soil samples to identify the relationships between MAAT/pH and 3-OH FA/brGDGT distribution or the indices derived from these compounds.

Other models than ordinary least squares or single/multiple regression were used in this study. The first model is the k-nearest neighbur (k-NN) algorithm, which is a supervised learning method (e.g. Gangopadhyay et al., 2009). A training database composed of N "input-output" pairs is initially constituted to estimate the output associated with a new input x. The method of the k-neighbors takes into account the k training samples whose input is the closest to the new input x, according to a distance to be defined. This method is non-parametric and is used for classification and regression. In k-NN regression, the result is the value for this object, which is the average of the values of the k nearest neighbors.

The second model is the random forest algorithm, which is also a supervised learning method used, among other things, for regressions (e.g. Ho, 1995; Denisko and Hoffman, 2018). This model works by constructing a multitude of decision trees at training time and producing the mean prediction of the individual trees. Decision tree learning is one of the predictive modeling approaches used to move from observations to conclusions about the target value of an item. Decision trees where variables are continuous values are called regression trees.

The training phase required for the random forests, k-NN and multiple linear regression was performed on 75% of the sample set with an iteration of ten cross-validations per model. Data selection was performed randomly on the dataset but with a stratification modality according to the MAAT or the pH to limit the impact of extreme values. Then, the robustness and precision of the different models were tested on the remaining 25 % of samples, considered as an independent dataset. Simple and Multiple linear regressions, PCA, k-NN and random forest models were performed with R software, version 3.6.1 (R Core Team, 2014) using the packages - tidymodels (version 0.1.0)- kknn (version 1.3.1), ranger (version 0.11.2). A web application is available online (https://athibault.shinyapps.io/paleotools) for the reconstruction of 3-OH FA-derived MAAT using the machine learning models proposed in the present study.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


318 **3. Results**

319 **3.1. Distribution of bacterial lipids**

320 *3.1.1. 3-OH FAs*

321 3-OH FAs were identified in the whole dataset, representing eight elevation transects
322 and 168 samples (Supplementary table 1; Yang et al., 2015; Wang et al., 2016; Coffinet et al.,
323 2017; Huguet et al., 2019; Véquaud et al., under revision). Their chain lengths range between
324 8 and 26 C atoms, indicating that these compounds have various origins (bacteria, plants, and
325 fungi; Zelles, 1999; Wang et al., 2016 and reference therein). The homologues of 3-OH FAs
326 with 10 to 18 C atoms are considered to be produced exclusively by Gram-negative bacteria
327 (Szponar et al., 2003; Wollenweber and Rietschel, 1990) and will be the only ones considered
328 in the following. Compounds with an even carbon number and *normal* chains were the most
329 abundant 3-OH FAs in all samples (mean 67.9 % of the total 3-OH FAs, Standard Deviation
330 (SD) 6.8%), with a predominance of the *n*-C_{14} homologue (21.9%, SD 3.23%; Fig. 1). *Iso* (mean
331 22.9%, SD 5.01%) and *anteiso* (mean 6.33 %, SD 1.79%) isomers were also present. It must be
332 noted that *anteiso* isomers were only detected for odd carbon-numbered 3-OH FAs (Coffinet et
333 al., 2017; Huguet et al., 2019; Wang et al., 2016; Yang et al., 2015).

334 The distribution of 3-OH FAs in the soils of the different altitudinal transects did not
335 show a large variability (Fig. 1). Thus, there was no major difference in the relative abundances
336 of most of the 3-OH FAs (*i*-C_{11}, *a*-C_{11}, *n*-C_{11}, *i*-C_{12}, *a*-C_{13}, *n*-C_{13}, *i*-C_{14}, *n*-C_{15}, *i*-C_{16}, *a*-C_{17} and
337 *n*-C_{17}) between the 8 study sites, even though slight differences could be observed for some
338 compounds as detailed below. For example, the Peruvian samples were characterized by higher
339 average proportions of *n*-C_{18} 3-OH FA and lower contribution of the *n*-C_{10} and *n*-C_{12}
340 homologues than those from the other transects. Soils from Mt. Shegyla were characterized by
341 lower average proportions of *n*-C_{14} 3-OH FAs and higher abundances of *i*-C_{17} compounds
342 compared to the other transects (Fig. 1).
343
344 *3.1.2. brGDGTs*

345 The relative abundances of brGDGTs were compared between the same transects as
346 for 3-OH FAs, representing a total of 168 samples. The 5- and 6-methyl isomers were separated
347 in most of the samples (Fig. 2) , except in older datasets, i.e. soils from Mts. Rungwe (Coffinet
348 et al., 2014, 2017) and Shennongjia (Yang et al., 2015: Sup. tables 2, 3 and Sup. Fig. 1).

349 The brGDGT distribution was dominated by acyclic compounds (Ia, IIa, IIa', IIIa,
350 IIIa') which represent on average ca. 84% of total brGDGTs (SD = 14.7%; Fig. 2). The

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


351 tetramethylated (Ia-c; mean 37.4%, SD) of 21.4%) and the pentamethylated (IIa-c; 45.2%, SD
352 13.5%) brGDGTs were predominant over the hexamethylated ones (IIIa-c; Fig. 2). The 5-
353 methyl isomers were on average present in a higher proportion (mean 71.9%, SD 23.4%) than
354 the 6-methyl compounds (Fig. 2).
355 High variability of the brGDGT distribution was observed among the different
356 transects. The relative abundance of brGDGT Ia was much higher in the Peruvian soils (mean
357 83%, SD 12.6%) than in the other transects (mean between 17.3% and 61.7%; Fig. 2). The 5-
358 methyl isomers were more abundant than the 6-methyl isomers for all sites except for Mt.
359 Pollino (mean 5-methyl = 44%, SD=11.7%) and Mt. Majella (mean 5-methyl = 33.7 %, SD =
360 5.5%; Fig. 2).
361

362 **3.2. 3-OH FA and brGDGT-derived indices**

363 *3.2.1. 3-OH FA*
364 The RIAN index roughly varied by an order of magnitude among the eight elevation
365 transects (Table 1). The RIAN index ranged from 0.37 to 0.67 for the Peruvian Andes, 0.23 to
366 0.56 for Mt. Shegyla, 0.15 to 0.34 for Mt. Pollino, 0.21 to 0.53 for the Chilean Andes, 0.26 to
367 0.80 for Mt. Rungwe (Huguet et al., 2019), 0.16 to 0.46 for Mt. Majella (Huguet et al., 2019),
368 0.20 to 0.69 for Mt. Shennongjia (Wang et al., 2016) and 0.13 to 0.56 for the French Alps
369 (Véquaud et al., under revision).
370 The RAN_{15} varied greatly among the different sites (Table 1). It was in the same range
371 along Mts. Rungwe (1.04-5.73) and Majella (0.68-6.43; Huguet et al., 2019). In contrast, its
372 upper limit was higher for Mts. Shennongjia (0.67-10.77; Wang et al., 2016), Shegyla (4.07-
373 12.17), Pollino (2.41-10.26), the Peruvian Andes (2.45-13.77) and the French Alps (1.44-
374 12.26). The range of variation in RAN_{15} was narrower for the Chilean Andes (3.82-6.40).
375 The RAN_{17} values were similar among the different altitudinal transects (Table 1),
376 ranging from 1.72 to 3.90 along Mt. Shegyla, 0.68 to 6.43 along Mt. Majella (Huguet et al.,
377 2019), 1.19 to 4.54 along Mt. Pollino, 1.91 to 4.25 for the Chilean Andes and 1.12 to 3.57 along
378 Mt. Shennongjia (Wang et al., 2016). The range of RAN_{17} values was narrower for Mt. Rungwe
379 (0.33-1.62; Huguet et al., 2019) and the Peruvian Andes (0.61-2.39) and wider for the French
380 Alps (0.89-6.42; Véquaud et al., under revision) compared to the other sites.
381

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


382 *3.2.2. brGDGT*

383 The range of variation in the MBT'$_{5Me}$ index was homogeneous along most transects
384 (0.32-0.63; Table 1), except the Peruvian Andes, with higher values (0.58-0.98; Table 1).
385 Regarding the CBT, it showed similar ranges along Peruvian Andes (0.97-1.81) and Mt.
386 Shegyla (0.76-1.87; Table 1) on the one hand and along Mts. Shennongjia (0.14-1.43; Yang et
387 al., 2015), Rungwe (0.16-1.01; Coffinet et al., 2017) and Pollino (0.28-1.09) on the other hand.
388 The CBT values varied within a narrow range along Mt.Majella (0.21-0.74; Huguet et al., 2019)
389 and within a wide range along the French Alps (0.25-2.23; Véquaud et al., under revision).
390

391 **3.3. Principal component analysis and clustering of 3-OH FA and brGDGT**
392 **distribution**

393 Principal component analyses were performed to refine the comparison of bacterial
394 lipid distribution (3-OH FAs and brGDGTs) among the different elevational transects.
395

396 *3.3.1. 3-OH FA*

397 The first two axes of the 3-OH FA PCA explained 39.1% of the total inertia in the
398 dataset (Fig. 3a). Dimension 1 (23.9%) opposed samples from Mt. Pollino in the right quadrant
399 to Peruvian soils and samples from Mt. Shennongjia. Dimension 2 (15.2%) especially separated
400 individuals from Chile and Mt. Rungwe. The Wilks' test showed that the location of the
401 sampling sites was the best variable discriminating the distribution of the individuals in the
402 PCA.

403 Principal component analysis performed on the temperature (RAN_{15}, RAN_{17}) and pH
404 (RIAN) indices derived from 3-OH FAs showed that most of the inertia was carried by the first
405 two axes of the PCA (Axis 1 = 56.09%; Axis 2 = 35.29%; Supp. Fig. 2). The first axis was
406 highly correlated with the RAN_{15} (r = 0.87) and RAN_{17} (r = 0.93) as well as with MAAT (r=-
407 0.67), while Axis 2 showed strong correlations with the RIAN (r = 0.96) and pH (r = -0.61).
408 The PCA allowed visualizing relationships at the scale of the whole dataset, between MAAT
409 and RAN_{15} and RAN_{17} (r= -0.61; r = -0.64 respectively) and between pH and RIAN (r = -0.53).
410
411

412 *3.3.2. brGDGT*

413 The first two axes of the brGDGT PCA explained 65.3% of the total inertia in the
414 dataset (Fig. 3b). Dimension 1 (35.60%) strongly discriminated soils from Mt. Rungwe and, to

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


a lesser extent, Mt. Majella, in the right quadrant from those from Mt. Shegyla and Peruvian Andes in the left quadrant. The former two sites were also discriminated negatively along dimension 2 (29.7%). As for the 3-OH FAs, Wilks' test showed that the location of the sampling sites was the best variable discriminating the distribution of the brGDGTs in the PCA.

4. Discussion

4.1. 3-OH FA and brGDGT-derived proxies

Previous studies conducted on soils from individual altitudinal transects revealed (1) local linear relationships between MAAT/pH and 3-OH FA indices and (2) the potential for combined calibrations using simple linear regressions (Wang et al., 2016; Huguet et al., 2019; Véquaud et al., under revision). In the present study, the existence of linear relationships between 3-OH FA-derived indices and environmental variables was further investigated using an extended soil dataset and the corresponding results were compared with those derived from brGDGT proxies.

4.1.1. Relationships between pH and bacterial lipid-derived proxies

The relationship between RIAN and pH was investigated along each of the altitudinal transects (Fig. 4a; Sup. Table 4). No significant linear relationship was obtained for the Peruvian Andes, Mts. Rungwe, Pollino and Majella (Huguet et al., 2019) and weak to moderate correlations were observed along Mts. Shegyla and Lautaret-Bauges (R^2 = 0.29-0.46; Sup. Table 4). In contrast, strong regressions between RIAN and pH were observed along Mt. Shennongjia (R^2 = 0.71) and in Chilean Andes (R^2 = 0.66). A weak linear relationship between RIAN and pH (R^2=0.27; RMSE = 0.99; $p = 8.63 \times 10^{-14}$) was also obtained when considering the 168 samples for the eight elevation transects altogether. Therefore, our results confirm the general influence of pH on the relative abundance of 3-OH FAs (Huguet et al., 2019) but suggest that strong linear correlations between RIAN and pH can only be obtained (i) at a local level and (ii) only for some of the sites.

As previously suggested (Huguet et al., 2019), the absence or weakness of linear correlations between RIAN and pH may be at least partly due to the small range of variation of pH (<2 units) along some mountains, such as Mts. Rungwe, Majella, and the Peruvian Andes (Fig. 4a; Table 1, Huguet et al., 2019). Transects for the Peruvian Andes, Mts. Majella and Rungwe were also characterized by the absence of relationships between pH and the brGDGT-derived CBT index, supporting the hypothesis that low pH ranges limit the potential of

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


obtaining linear relationships between indices based on bacterial lipids and pH. Nevertheless, the existence of a narrow pH range was not the only limiting factor in obtaining a strong linear regression between RIAN and pH. Indeed, MAAT rather than soil pH was the dominant driver of soil bacterial diversity and community composition for the Peruvian transect (using 16S rRNA sequencing (Nottingham et al., 2018); and using phospholipid fatty acids (Whitaker et al., 2014), consistent with the weak correlation between soil pH and bacterial lipids. The weakness of the RIAN-pH relationship may also be partly due to the heterogeneity of soils encountered along a given altitudinal transect, representing specific microenvironments and to the large diversity of bacterial communities in soils from different elevations (Siles and Margesin, 2016).The 3-OH FA distribution strongly varied from a Gram-negative bacterial species to another (Bhat and Carlson, 1992), hence a potentially large variability of RIAN values among soils from a given transect. Altogether, these results suggest that linear models are not the most suitable for establishing a global calibration between RIAN and pH in soils.

Concerning GDGTs, moderate to strong relationships between brGDGT-derived CBT index and pH were observed along 4 of the 7 altitudinal transects investigated (Fig. 4b; Sup. Table 4). All the individual linear relationships between CBT and pH, where present, had similar slopes and ordinates and share (for most of the samples) the same 95% prediction intervals (p-value <0.5). This resulted in a strong linear relationship between CBT index and pH values for the whole dataset ($R^2 = 0.65$; RMSE = 0.73; $n = 158$), which is slightly weaker than the global calibration ($R^2 = 0.70$; $n = 170$) proposed by (Peterse et al., 2012). Altogether, these results confirm that, in contrast with the RIAN, the CBT index can be applied at a global scale using a simple linear regression model.

The discrepancy in relationships between temperature and brGDGTs and 3-OH FAs might partly be due to differences in the relative abundance of these lipids among bacterial communities. The brGDGTs are produced by a more restricted and less diverse number of bacterial species than 3-OH FAs, which are arguably biosynthesized by a large diversity of Gram-negative bacteria species (e.g. Wakeham et al., 2003, Zelles et al., 1995; Zelles, 1999). So far, only bacteria from the *Acidobacteria* phylum were identified as putative brGDGT producers in soils (Damste et al., 2018). The hypothetical lower diversity of brGDGT producers, in contrast with 3-OH FAs might explain the more homogenous response and lower scatter of the relationships between pH and CBT index. Moreover, the CBT index is a ratio based on a restricted number of compounds, representing the direct dependence of the degree of cyclisation of bacterial GDGTs on pH. Conversely, the RIAN index is calculated from the relative abundances of all the individual 3-OH FAs between C_{10} and C_{18} (Wang et al., 2016). It

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


cannot be ruled out that some of the compounds used to calculate the RIAN index are preferentially synthesized, as part of the homeoviscous mechanism, in response to environmental variables other than pH. This calls for a better understanding of the ecology of 3-OH FA-producing bacteria and their adaptation mechanisms.

4.1.2 Relationships between MAAT and bacterial lipid-derived proxies

RAN_{15} was previously shown to be correlated with MAAT along Mts. Rungwe, Majella and Shennongjia (Wang et al., 2016; Huguet et al., 2019). Moderate to strong linear correlations (R^2 0.49-0.79) between RAN_{15} and MAAT were similarly observed along most of the individual transects investigated (Fig. 5a; Sup. Table 4), except along the Chilean and Lautaret-Bauges transects. The individual correlations do not share the same 95% prediction intervals and even when some of them present similar slopes, the regression lines display significantly different intercepts (p-value > 0.05) (Fig. 5a). This supports the hypothesis of a site-dependent effect of the linear RAN_{15}-MAAT relationship previously made by Huguet et al. (2019).

Similarly to RAN_{15}, RAN_{17} was moderately to strongly correlated (R^2 0.27-0.81) with MAAT along 5 out of 8 individual transects (Fig. 5b; Sup. Table 4). The small range of variation in MAAT along the Chilean transect (6.0-9.2 °C) (Table 1), associated with that of the RAN_{15} /RAN_{17}, could explain the lack of a linear relationship between the MAAT and these indices. As for the French Alps (Mts Lautaret-Bauges), the influence of local environmental parameters (pH and to a lesser extent soil moisture and grain size, related to vegetation and soil types, or thermal regimes associated with the snow cover) on 3-OH FA distribution was shown to be predominant over that of MAAT (Véquaud et al., under revision). In contrast with RAN_{15}, the linear regressions between RAN_{17} and MAAT along Mts. Shegyla, Shennongjia, Rungwe and the Peruvian Andes transects share prediction intervals at 95% and have similar slope and intercept values (p-value <0.05; Fig. 5b; Sup. Table 4), suggesting that RAN_{17} could be a more effective global proxy for MAAT reconstructions than RAN_{15}.

In order to test the hypothesis that RAN_{17}, rather than RAN_{15}, is a more effective global proxy for MAAT, the global calibrations between RAN_{15}/RAN_{17} and MAAT based on the entire soil dataset (n = 168) were compared. The two linear regressions had similar moderate determination coefficients (R^2 = 0.37 and 0.41 for RAN_{15} and RAN_{17}, respectively) and similar high RMSE (RMSE = 5.46°C and 5.28°C for RAN_{15} and RAN_{17}, respectively; Sup. Table 4). For all transects (except for the Mt Majella RAN_{17}/MAAT relationship), the individual local regressions between RAN_{15}/RAN_{17} and MAAT outperformed the proposed global linear

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


516 calibrations in terms of determination coefficients (0.49-0.81) and RMSE (1.98-3.57 °C; Sup.
517 Table), suggesting that local rather than global linear transfer functions based on RAN_{15} or
518 RAN_{17} may be more appropriate for paleotemperature reconstructions in soils.
519 The difficulties in establishing global linear RAN_{15}/RAN_{17}-MAAT calibrations may
520 partly be due to the fact that microbial diversity, especially for 3-OH FA-producing Gram-
521 negative bacteria (Margesin et al., 2009; Siles and Margesin, 2016), can vary greatly from one
522 soil to another, resulting in variation of the RAN_{15}/RAN_{17} indices, as also assumed for the
523 RIAN. The strong regional dependence of the 3-OH FA distribution may thus explain the weak
524 correlation between 3-OH FA-derived indices (RAN_{15}, RAN_{17} and RIAN) and environmental
525 variables (MAAT/pH) at a global level. This regional dependency was further supported by the
526 PCA of the relative abundance of 3-OH FAs across the global dataset, which showed that the
527 individuals were grouped based on the sampling location (Fig. 3a).
528 In addition to 3-OH FAs, the relationships between brGDGT distribution and MAAT
529 were investigated along the six transects for which the 5- and 6-methyl brGDGT isomers were
530 separated (Mts Shegyla, Pollino Majella, Lautaret-Bauges, Peruvian Andes and Chilean
531 Andes). These individual transects showed moderate to strong relationships between MAAT
532 and MBT'_{5Me} (R^2 0.34-0.75; Fig. 6 and Sup. Table), with similar slopes and ordinates (except
533 for the Peruvian Andes) and shared 95% prediction intervals for most of the samples. A distinct
534 relationship between MBT'_{5Me} and MAAT was observed along the Peruvian Andes and Mt
535 Majella transects (Fig. 6a), as also observed for the RIAN and RAN_{15} indices (Figs 4a and 5a).
536 The singularity of the Peruvian soils is also visible on the PCA performed on the brGDGT
537 distribution (Fig 3b), where the samples from this region are pooled separately from the rest of
538 the dataset. This specific trend is difficult to explain, even though the Peruvian Andes are
539 subjected to warmer climatic conditions (Table 1) than the other temperate transects, which
540 may in turn affect the nature of the microbial communities encountered in the soils and the
541 bacteria lipid distribution (Hofmann et al., 2016; Siles and Margesin, 2016).
542 A moderate linear relationship between MAAT and MBT'_{5Me} (R^2 = 0.56, RMSE = 3.65
543 °C, n = 104; Sup. Table) was observed after combining the data for the five aforementioned
544 altitudinal transects. This global relationship was more robust and accurate than those obtained
545 between the RAN_{15}/RAN_{17} and MAAT (Sup. Table 4). This confirms that the MBT'_{5Me} index
546 can be applied at a global scale using a simple linear regression model as previously shown (De
547 Jonge et al., 2014; Naafs et al., 2017), in contrast with the RAN_{15} and RAN_{17} proxies, for which
548 only strong local calibrations with MAAT were found.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


549 As a similar conclusion was obtained for the RIAN-pH proxy, it appears necessary to use
550 more complex models to develop global calibrations between 3-OH FA-derived proxies and
551 MAAT/pH. This novel method allows taking into account the complexity and specificity of
552 each environmental site.
553
554 **4.2. Development of new models for the reconstruction of MAAT and pH from 3-**
555 **OH FA**

556 Most calibrations between lipid distribution and environmental variables are based on:
557 (i) a simple linear regression model (e.g. for brGDGTs: De Jonge et al., 2014; Wang et al.,
558 2016; Naafs et al., 2017) – most often the ordinary least square regression – as it is simple and
559 easy to implement and understand; or (ii) multiple regression models (e.g. Weijers et al., 2007;
560 Peterse et al., 2012), describing the relationships between an explained variable and several
561 explanatory variables. The latter can reveal the presence of linear relationships among several
562 known variables but cannot take in account non-linear influences, which can occur in complex
563 environmental settings. As previously discussed, robust global calibrations between 3-OH FA-
564 derived indices (RIAN, RAN_{15} and RAN_{17}) and MAAT/pH could not be established using a
565 simple linear regression model.
566 Therefore, other models (non-parametric, and involving machine learning) were tested
567 to potentially establish stronger statistical relationships between 3-OH FA distributions and
568 pH/MAAT at the global level. In the present study, three different machine learning models
569 were compared: the multiple linear regression model, the random forest model (e.g. Ho, 1995;
570 Denisko and Hoffman, 2018) and the k-NN model (k-nearest neighbours; e.g. Gangopadhyay
571 et al., 2009). The multiple linear regression model was chosen as a complement to the simple
572 linear regression model, as it allows determination of linear relationships between MAAT/pH
573 and the individual relative abundances of 3-OH FAs, instead of indices derived from the latter.
574 The three models, based on a supervised machine learning approach (i.e. the expected response
575 to the model is known), were applied to the total soil dataset (n=168).
576 The multiple linear regression model yielded a strong relationship between 3-OH FA
577 relative abundances and MAAT (R^2= 0.70; Fig. 7a; Eq.6):
578 $MAAT\ (°C) = -17.28 \times [nC_{10}] + 274.88 \times [iC_{11}] + 1570.7 \times [aC_{11}] - 441.78 \times [nC_{11}] - 17.68$
579 $\times [nC_{12}] + 136.19 \times [iC_{13}] + 266.88 \times [aC_{13}] + 36.67 \times [nC_{13}] - 179.19 \times [iC_{14}] + 2.71 \times$
580 $[nC_{14}] + 50.74 \times [iC_{15}] - 236.81 \times [aC_{15}] + 101.98 \times [nC_{15}] - 44.74 \times [iC_{16}] + 57.22 \times$
581 $[nC_{16}] + 48.93 \times [iC_{17}] - 167.99 \times [aC_{17}] + 42.16 \times [nC_{17}] + 42.16 \times [nC_{18}] - 7.66$

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


582 *(n* = 168; R^2 = 0.70; RMSE = 3.69 °C) (6)

583 This model greatly improves the accuracy of MAAT prediction and the associated error
584 (RMSE = 3.69°C) in comparison with the global linear relationship between RAN_{15}/RAN_{17} and
585 MAAT (R^2=0.37 and 0.41; RMSE =5.46 °C and 2.28 °C for RAN_{15} and RAN_{17}, respectively;
586 Sup. Tables 2).

587 Both the random forest and k-NN models (Figs. 7b, c) reliably predicted MAAT, with
588 even higher determination coefficients (R^2 = 0.74 and 0.78, respectively) and RMSE (3.22 and
589 2.73 °C, respectively) than simple or multiple linear regressions. These models have the
590 advantage of accounting for the environmental complexity and the multiplicity of parameters
591 that can influence bacterial membrane lipid distribution and demonstrate the use of 3-OH FAs
592 as effective temperature proxies in terrestrial environments.

593 Regarding pH, a robust global linear relationship could not be obtained with the RIAN.
594 In contrast, the multiple regression model provided a moderate correlation with pH (R^2 = 0.38;
595 Fig. 8a; Eq. 7):

596 $pH = -2.79 \times [nC_{10}] - 40.69 \times [iC_{11}] - 112.007 \times [aC_{11}] - 31.072 \times [nC_{11}] - 6.404 \times [nC_{12}] +$
597 $57.095 \times [iC_{13}] - 4.598 \times [aC_{13}] + 6.784 \times [nC_{13}] + 20.253 \times [iC_{14}] - 3.710 \times [nC_{14}] - 13.179$
598 $\times [iC_{15}] - 21.691 \times [aC_{15}] - 11.448 \times [nC_{15}] - 50.177 \times [iC_{16}] - 2.668 \times [nC_{16}] - 1.871 \times [iC_{17}]$
599 $- 12.571 \times [aC_{17}] + 6.345 \times [nC_{17}] - 4.634 \times [nC_{18}] - 8.262$

600 *(n* = 168; R^2 = 0.38; RMSE = 0.89) (7)

601 The k-NN model (R^2 = 0.48 and RMSE = 0.79; Fig. 8b) did not clearly improve the
602 estimation of pH compared to the multiple linear regression (Fig. 8a). The k-NN model is based
603 on the estimation of the mean distances between the different samples. Its constraints lie in the
604 fact that, by definition, if a range of values is more frequent than the others, then it will be
605 statistically predominant among the k closest neighbours. In the present case, this regression
606 method would not be the most suitable to establish a global calibration between lipid
607 abundances and pH, due to the large predominance of soils with pH comprised between 5 and
608 7 (Table 1).

609 Last, the random forest model appeared to be the most robust (R^2 = 0.65) and accurate
610 (RMSE = 0.67) for pH reconstruction based on 3-OH FA abundance (Fig. 8c) and present a
611 model as robust as global CBT index. In summary, the three tested models provided global
612 calibrations with pH, contrary to the simple linear regression (Fig. 4a). They might be further
613 improved by increasing the number of samples analyzed and the representativeness of the
614 different pH values within the dataset.

615

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


616 **4.3. Paleoclimate application of the new 3-OH FA/MAAT models**
617
618 The random forest and k-NN models were similar in terms of robustness and precision
619 for MAAT reconstruction based on 3-OH FA abundance (Figs. 7b, c). The applicability of these
620 global calibrations for paleotemperature reconstructions was thus tested and compared with the
621 MAAT record from a Chinese speleothem (HS4 stalagmite) covering the last 9,000 years BP
622 and based on the RAN_{15} index (Fig. 9a; Wang et al., 2018). To the best of our knowledge, this
623 archive represents the only published application of 3-OH FAs as a paleotemperature proxy in
624 terrestrial settings.
625
626 *4.3.1 Comparison of the random forest and k-NN global MAAT calibrations*
627 MAAT estimates derived from the k-NN calibration ranged between 6.8 and 14.3 °C
628 over the last 9,000 years (Fig. 9b). Abrupt shifts in MAAT of more than 5 °C were observed
629 between 2,000 and 4,000 yrs BP. These large variations in MAAT over such short periods of
630 time are higher than the RMSE of the calibration and appear excessive, as the expected
631 amplitude of MAAT during the Holocene is expected to be up to ca. 2-3 °C (Liu et al., 2014).
632 The bias in MAAT estimates may be because the k-NN method is better suited for uniformly
633 distributed datasets, which is not the case here, as the individual transects heterogeneously
634 cover a wide range of temperatures. The application of a global calibration at the local scale –
635 that of the HS4 stalagmite – using the k-NN method and based on the similarities among
636 samples, does not appear appropriate. Such a calibration might be improved by extending the
637 dataset with samples more equally distributed across a wider range of global climatic gradients.
638 The random forest model yielded MAAT estimates between 11.1 and 17.3°C, i.e. a
639 smaller range than the k-NN algorithm (Fig. 9c). The amplitude of the shifts observed between
640 2,000 and 4,000 yrs BP was ca. 3°C, which is climatically more consistent than the variations
641 obtained with the k-NN method. Furthermore, the application of the global random forest
642 calibration provided similar temperature trends as those derived from the local RAN_{15}
643 calibration by Wang et al. (2018). These results suggest that the random forest calibration is
644 more reliable than the k-NN one. This can be explained by the intrinsic definition of the random
645 forest algorithm, which averages the results of several independent models (so-called decision
646 trees), thus reducing the variance and thus the forecast error on the final model.
647
648

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


649 *4.3.2. Comparison of the random forest and RAN_{15} calibrations for MAAT*
650 *reconstruction*

651 We compared the temperature records derived from the global random forest
652 calibration and local MAAT/RAN_{15} transfer function (Wang et al., 2016) (Fig. 9). Application
653 of the local RAN_{15} calibration to the HS4 stalagmite yielded an average MAAT of ca. 18.4 °C
654 over the most recent part of the record (last 800 yrs; Fig. 9), consistent with the MAAT of 18
655 °C recorded *in situ* by a temperature logger (Hu et al., 2008;Wang et al., 2018). In contrast,
656 absolute MAAT estimates derived from the random forest model were on average 14.7 °C over
657 the last 800 yrs and were generally lower than those obtained from the local RAN_{15} calibration
658 over the whole record. Altogether, these results suggest that the random forest model tends to
659 underestimate absolute MAAT, in contrast with the RAN_{15} calibration proposed by Wang et al.
660 (2016). This discrepancy may be due the fact that the calibration proposed in the present study
661 is based on a global dataset, with samples subject to a large variety of environmental and
662 climatic conditions, whereas the RAN_{15}-MAAT transfer function by Wang et al. (2016) was
663 constructed using soil samples from a regional altitudinal transect, located at only 120 km
664 distance from the stalagmite site (Wang et al., 2018).

665 Even though the local calibration by Wang et al. (2016) provides more accurate
666 absolute MAAT values than the present global random forest model, both calibrations roughly
667 generate similar qualitative MAAT trends over time. A regular slight decrease in temperature
668 of ca. 1 °C was observed between 9,000 and ca. 1,000 yrs BP based on the local RAN_{15}
669 calibration (Fig. 9a; Wang et al., 2018). This general decreasing trend was also visible when
670 using the random forest model, but mainly between 9,000 and 4,000 yrs BP in agreement with
671 the general trend recorded by the $\partial^{18}O$ record (mixture of temperature and hydrological signals,
672 Wang et al., 2018) of the HS4 stalagmite (Fig 9 c,d; Hu et al., 2008). In addition, both the global
673 random forest, local RAN_{15} calibrations and the $\partial^{18}O$ record allowed the identification of
674 several climatic events in the Northern hemisphere, in agreement with the reconstructed total
675 solar irradiance (TSI, Steinhilber et al., 2009, Fig.9e). Thus, both models highlighted, with
676 slightly different amplitudes, the Medieval Warm Period (800-1000 years BP) and Little Ice
677 Age (LIA; 200-500 years BP) periods (Ljungqvist, 2010; Mann et al., 2008; Wang et al., 2018).
678 The LIA event is particularly well represented by the global random forest calibration, in line
679 with the decrease in the TSI (Fig. 9 c,e) associated with a relative increase in the $\partial^{18}O$ of HS4
680 carbonates (dry/cool event, Wang et al., 2018). Before the MWP, the global random forest
681 calibration shows large oscillations, which can be assumed to be representative of TSI
682 variations between 500 and 1,300 yrs BP. Similarly, an important cooling event, well correlated

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


683 with a significant decrease in the TSI (Fig. 9 a, c, e), was recorded by the two calibrations at
684 1300 yr BP.
685 The global random forest calibration also highlighted two cooling events, poorly
686 represented by the local RAN_{15} calibration: one between 2,800 and 3,000 yrs BP and another
687 ca. 4,200 yrs BP ago (Bond et al., 2001; Mayewski et al., 2004). The event at 4,200 yrs BP is
688 consistent with the $\partial^{18}O$ and solar irradiance records and is referenced in the literature as the
689 "4.2 kiloyear event" (deMenocal, 2001). This intense drought event was suggested to have had
690 a major impact on different civilizations (collapses, migrations; (Bini et al., 2019; Gibbons,
691 1993; Li et al., 2018; Staubwasser et al., 2003). Thus, in some parts of China, the production of
692 rice fields sharply decreased during this period, leading to a decrease in population (Gao et al.,
693 2007).
694 The global random forest calibration additionally shows a cooling period between 4,000
695 yrs and 3,000 yrs BP, with a cooling (-1°C/800 years) between 4,000 yrs and 3,200 yrs BP,
696 followed by an abrupt cooling between 3,200 years BP and 3,000 yrs BP (-2°C/200 years). This
697 cooling period, represented more accurately by the random forest model than by the local
698 RAN_{15} calibration, is consistent with the trends derived from $\partial^{18}O$ and solar irradiance records.
699 It culminates with a cold episode at 3000 yrs BP, also known as Late Bronze Age Collapse
700 (Kaniewski et al., 2013). Indeed, this cold episode, combined with droughts, may have led to a
701 decrease in agricultural production in China, contributing to the degradation of trade routes and
702 ultimately to the collapse of Bronze Age civilizations (Knapp and Manning, 2016; Weiss,
703 1982). Last, the global random forest calibration also highlights two additional cold events,
704 between 5,600 and 5,900 yrs BP, as well as around 7,100 yrs BP, corresponding to solar
705 irradiance minima (Bond et al., 2001; Mayewski et al., 2004).
706 The first application of the random forest calibration to a natural archive shows the
707 potential of 3-OH FAs as paleotemperature proxies at a global scale, as known and documented
708 climatic events were recorded, with a similar RMSE (2.6 -2.7 °C) as that of the local calibration
709 by Wang et al. (2016). This RMSE is also much lower than the one related to the latest global
710 MAAT-brGDGT calibrations (> 4 °C; De Jonge et al., 2014; Naafs et al., 2017; Dearing
711 Crampton-Flood et al., 2020), even though the latter are based on a larger number of soil
712 samples than the global 3-OH FA model proposed in the present study. In summary, we
713 demonstrate that 3-OH FAs are promising and effective temperature proxies for terrestrial
714 settings, complementary to, and independent of, the brGDGTs. We expect that analyses of 3-
715 OH FAs in a larger number of globally distributed soils will further improve the accuracy and
716 robustness of the global random forest calibration for paleotemperature reconstruction.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


5. Conclusions

3-OH FAs have been recently proposed as environmental proxies in terrestrial settings, based on local studies. This study investigated for the first time the applicability of these compounds as MAAT and pH proxies at the global scale using an extended soil dataset across a series of globally distributed elevation transects (n = 168). Strong linear relationships between 3-OH FA-derived indices (RAN_{15}, RAN_{17} and RIAN) and MAAT/pH could only be obtained locally, for some individual transects, suggesting that these indices cannot be used as paleoproxies at the global scale through this kind of model. Other algorithms (multiple linear regression, k-NN and random forest models) were tested and, in contrast with simple linear regressions, provided strong global correlations between MAAT/pH and 3-OH FA relative abundances. The applicability of the k-NN and random forest models for paleotemperature reconstruction was tested and compared with the MAAT record from the unique available record: a Chinese speleothem. The calibration based on the random forest model appeared to be the most robust and showed similar trends to previous reconstructions and known Holocene climate variations. Furthermore, the global random forest model highlighted documented climatic events poorly represented by the local RAN_{15} calibration. This new global model is promising for paleotemperature reconstructions in terrestrial settings and could be further improved by analyzing 3-OH FAs in a larger number of globally distributed soils. This study demonstrates the major potential of 3-OH FAs as MAAT/pH proxies in terrestrial environments through the different models presented and their application for paleoreconstruction.

Data availability. All data are available in the Supplementary tables.

Author contributions. P.V. performed the lipid and statistical analyses and wrote a first draft of the paper., A.H. and S.D. supervised the work of P.V. and corrected the first draft, P.V. and A.T. developed the different models, G.B., A.N., W.P.S., N.S., J.P.W. and S.C. provided samples and/or associated data, and all the co-authors reviewed and commented on the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. We thank Sorbonne Université for a PhD scholarship to P.V. and the Labex MATISSE (Sorbonne Université) for financial support. The EC2CO program (CNRS/INSU – BIOHEFECT/MICROBIEN) is thanked for funding of the SHAPE project. A.H. and S.C. are grateful for funding of the ECOS SUD/ ECOS ANID #C19U01 project. We are grateful to

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

751 Jérôme Poulenard for discussions on soil characteristics, and for comments on the manuscript.
752 We thank Dr. Juntao Wang and Prof. Jinzheng He for having provided soils from Mt. Shegyla.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

References

Beales, N.: Adaptation of Microorganisms to Cold Temperatures, Weak Acid Preservatives, Low pH, and Osmotic Stress: A Review, Comprehensive Reviews in Food Science and Food Safety, 3(1), 1–20, doi:10.1111/j.1541-4337.2004.tb00057.x, 2004.

Bhat, U. R. and Carlson, R. W.: A new method for the analysis of amide-linked hydroxy fatty acids in lipid-As from gram-negative bacteria, Glycobiology, 2(6), 535–539, doi:10.1093/glycob/2.6.535, 1992.

Bini, M., Zanchetta, G., Persoiu, A., Cartier, R., Catala, A., Cacho, I., Dean, J. R., Di Rita, F., Drysdale, R. N., Finné, M., Isola, I., Jalali, B., Lirer, F., Magri, D., Masi, A., Marks, L., Mercuri, A. M., Peyron, O., Sadori, L., Sicre, M.-A., Welc, F., Zielhofer, C. and Brisset, E.: The 4.2 ka BP Event in the Mediterranean region : an overview, Climate of the Past, 15(2), 555–577, 2019.

Bonanomi, G., Zotti, M., Mogavero, V., Cesarano, G., Saulino, L., Rita, A., Tesei, G., Allegrezza, M., Saracino, A. and Allevato, E.: Climatic and anthropogenic factors explain the variability of Fagus sylvatica treeline elevation in fifteen mountain groups across the Apennines, Forest Ecosystems, 7(1), 5, doi:10.1186/s40663-020-0217-8, 2020.

Bond, G., Kromer, B., Beer, J., Muscheler, R., Evans, M. N., Showers, W., Hoffmann, S., Lotti-Bond, R., Hajdas, I. and Bonani, G.: Persistent Solar Influence on North Atlantic Climate During the Holocene, Science, 294(5549), 2130–2136, doi:10.1126/science.1065680, 2001.

Brassell, S. C., Eglinton, G., Marlowe, I. T., Pflaumann, U. and Sarnthein, M.: Molecular stratigraphy: a new tool for climatic assessment, Nature, 320(6058), 129–133, doi:10.1038/320129a0, 1986.

Carter, M. R., Gregorich, E. G. and Gregorich, E. G.: Soil Sampling and Methods of Analysis, CRC Press., 2007.

Coffinet, S., Huguet, A., Williamson, D., Fosse, C. and Derenne, S.: Potential of GDGTs as a temperature proxy along an altitudinal transect at Mount Rungwe (Tanzania), Organic Geochemistry, 68, 82–89, doi:10.1016/j.orggeochem.2014.01.004, 2014.

Coffinet, S., Huguet, A., Pedentchouk, N., Bergonzini, L., Omuombo, C., Williamson, D., Anquetil, C., Jones, M., Majule, A., Wagner, T. and Derenne, S.: Evaluation of branched GDGTs and leaf wax n-alkane δ2H as (paleo) environmental proxies in East Africa, Geochimica et Cosmochimica Acta, 198, 182–193, doi:10.1016/j.gca.2016.11.020, 2017.

Damsté, J. S. S., Rijpstra, W. I. C., Hopmans, E. C., Weijers, J. W. H., Foesel, B. U., Overmann, J. and Dedysh, S. N.: 13,16-Dimethyl Octacosanedioic Acid (iso-Diabolic Acid), a Common Membrane-Spanning Lipid of Acidobacteria Subdivisions 1 and 3, Appl. Environ. Microbiol., 77(12), 4147–4154, doi:10.1128/AEM.00466-11, 2011.

Damsté, J. S. S., Rijpstra, W. I. C., Hopmans, E. C., Foesel, B. U., Wüst, P. K., Overmann, J., Tank, M., Bryant, D. A., Dunfield, P. F., Houghton, K. and Stott, M. B.: Ether- and Ester-Bound iso-Diabolic Acid and Other Lipids in Members of Acidobacteria Subdivision 4, Appl. Environ. Microbiol., 80(17), 5207–5218, doi:10.1128/AEM.01066-14, 2014.

Damste, J. S. S., Rijpstra, W. I. C., Foesel, B. U., Huber, K. J., Overmann, J., Nakagawa, S., Kim, J. J., Dunfield, P. F., Dedysh, S. N. and Villanueva, L.: An overview of the occurrence of

© Author(s) 2020. CC BY 4.0 License.


793 ether- and ester-linked iso-diabolic acid membrane lipids in microbial cultures of the
794 Acidobacteria: Implications for brGDGT paleoproxies for temperature and pH, Org. Geochem.,
795 124, 63–76, doi:10.1016/j.orggeochem.2018.07.006, 2018.

796 De Jonge, C., Hopmans, E. C., Zell, C. I., Kim, J.-H., Schouten, S. and Sinninghe Damsté, J.
797 S.: Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in
798 soils: Implications for palaeoclimate reconstruction, Geochimica et Cosmochimica Acta, 141,
799 97–112, doi:10.1016/j.gca.2014.06.013, 2014.

800 Dearing Crampton-Flood, E., Tierney, J. E., Peterse, F., Kirkels, F. M. S. A. and Sinninghe
801 Damsté, J. S.: BayMBT: A Bayesian calibration model for branched glycerol dialkyl glycerol
802 tetraethers in soils and peats, Geochimica et Cosmochimica Acta, 268, 142–159,
803 doi:10.1016/j.gca.2019.09.043, 2020.

804 deMenocal, P. B.: Cultural Responses to Climate Change During the Late Holocene, Science,
805 292(5517), 667–673, doi:10.1126/science.1059287, 2001.

806 Denich, T. J., Beaudette, L. A., Lee, H. and Trevors, J. T.: Effect of selected environmental and
807 physico-chemical factors on bacterial cytoplasmic membranes, Journal of Microbiological
808 Methods, 52(2), 149–182, doi:10.1016/S0167-7012(02)00155-0, 2003.

809 Denisko, D. and Hoffman, M. M.: Classification and interaction in random forests, Proc Natl
810 Acad Sci USA, 115(8), 1690–1692, doi:10.1073/pnas.1800256115, 2018.

811 Eglinton, T. I. and Eglinton, G.: Molecular proxies for paleoclimatology, Earth and Planetary
812 Science Letters, 275(1), 1–16, doi:10.1016/j.epsl.2008.07.012, 2008.

813 Emiliani, C.: Pleistocene Temperatures, The Journal of Geology, 63(6), 538–578,
814 doi:10.1086/626295, 1955.

815 Erez, J. and Luz, B.: Experimental paleotemperature equation for planktonic foraminifera,
816 Geochimica et Cosmochimica Acta, 47(6), 1025–1031, doi:10.1016/0016-7037(83)90232-6,
817 1983.

818 Gangopadhyay, S., Harding, B. L., Rajagopalan, B., Lukas, J. J. and Fulp, T. J.: A
819 nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles,
820 Water Resources Research, 45(6), doi:10.1029/2008WR007201, 2009.

821 Gao, H., Zhu, C. and Xu, W.: Environmental change and cultural response around 4200 cal. yr
822 BP in the Yishu River Basin, Shandong, J GEOGR SCI, 17(3), 285–292, doi:10.1007/s11442-
823 007-0285-5, 2007.

824 Gibbons, A.: How the Akkadian Empire Was Hung Out to Dry, Science, 261, 985,
825 doi:10.1126/science.261.5124.985, 1993.

826 Hazel, J. R. and Eugene Williams, E.: The role of alterations in membrane lipid composition in
827 enabling physiological adaptation of organisms to their physical environment, Progress in Lipid
828 Research, 29(3), 167–227, doi:10.1016/0163-7827(90)90002-3, 1990.

829 Hofmann, K., Lamprecht, A., Pauli, H. and Illmer, P.: Distribution of Prokaryotic Abundance
830 and Microbial Nutrient Cycling Across a High-Alpine Altitudinal Gradient in the Austrian

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


831 Central Alps is Affected by Vegetation, Temperature, and Soil Nutrients, Microb Ecol, 72(3),
832 704–716, doi:10.1007/s00248-016-0803-z, 2016.

833 Hu, C., Henderson, G. M., Huang, J., Xie, S., Sun, Y. and Johnson, K. R.: Quantification of
834 Holocene Asian monsoon rainfall from spatially separated cave records, Earth and Planetary
835 Science Letters, 266(3), 221–232, doi:10.1016/j.epsl.2007.10.015, 2008.

836 Huguet, A., Coffinet, S., Roussel, A., Gayraud, F., Anquetil, C., Bergonzini, L., Bonanomi, G.,
837 Williamson, D., Majule, A. and Derenne, S.: Evaluation of 3-hydroxy fatty acids as a pH and
838 temperature proxy in soils from temperate and tropical altitudinal gradients, Organic
839 Geochemistry, 129, 1–13, doi:10.1016/j.orggeochem.2019.01.002, 2019.

840 Huguet, C., Hopmans, E. C., Febo-Ayala, W., Thompson, D. H., Sinninghe Damsté, J. S. and
841 Schouten, S.: An improved method to determine the absolute abundance of glycerol
842 dibiphytanyl glycerol tetraether lipids, Organic Geochemistry, 37(9), 1036–1041,
843 doi:10.1016/j.orggeochem.2006.05.008, 2006.

844 Kaniewski, D., Campo, E. V., Guiot, J., Burel, S. L., Otto, T. and Baeteman, C.: Environmental
845 Roots of the Late Bronze Age Crisis, PLOS ONE, 8(8), e71004,
846 doi:10.1371/journal.pone.0071004, 2013.

847 Knapp, A. B. and Manning, S. W.: Crisis in Context: The End of the Late Bronze Age in the
848 Eastern Mediterranean, American Journal of Archaeology, 120(1), 99–149,
849 doi:10.3764/aja.120.1.0099, 2016.

850 Li, C.-H., Li, Y.-X., Zheng, Y.-F., Yu, S.-Y., Tang, L.-Y., Li, B.-B. and Cui, Q.-Y.: A high-
851 resolution pollen record from East China reveals large climate variability near the
852 Northgrippian-Meghalayan boundary (around 4200 years ago) exerted societal influence,
853 Palaeogeography, Palaeoclimatology, Palaeoecology, 512, 156–165,
854 doi:10.1016/j.palaeo.2018.07.031, 2018.

855 Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., Smith, R. S.,
856 Lohmann, G., Zheng, W. and Elison Timm, O.: The Holocene temperature conundrum, Proc
857 Natl Acad Sci U S A, 111(34), E3501–E3505, doi:10.1073/pnas.1407229111, 2014.

858 Ljungqvist, F. C.: A new reconstruction of temperature variability in the extra-tropical northern
859 hemisphere during the last two millennia, Geografiska Annaler: Series A, Physical Geography,
860 92(3), 339–351, doi:10.1111/j.1468-0459.2010.00399.x, 2010.

861 Loomis, S. E., Russell, J. M., Ladd, B., Street-Perrott, F. A. and Sinninghe Damsté, J. S.:
862 Calibration and application of the branched GDGT temperature proxy on East African lake
863 sediments, Earth and Planetary Science Letters, 357–358, 277–288,
864 doi:10.1016/j.epsl.2012.09.031, 2012.

865 Malhi, Y., Silman, M., Salinas, N., Bush, M., Meir, P. and Saatchi, S.: Introduction: Elevation
866 gradients in the tropics: laboratories for ecosystem ecology and global change research, Global
867 Change Biology, 16(12), 3171–3175, doi:10.1111/j.1365-2486.2010.02323.x, 2010.

868 Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S. and Ni,
869 F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over
870 the past two millennia, Proceedings of the National Academy of Sciences, 105(36), 13252–
871 13257, doi:10.1073/pnas.0805721105, 2008.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

872 Margesin, R., Jud, M., Tscherko, D. and Schinner, F.: Microbial communities and activities in
873 alpine and subalpine soils: Communities and activities in alpine and subalpine soils, FEMS
874 Microbiology Ecology, 67(2), 208–218, doi:10.1111/j.1574-6941.2008.00620.x, 2009.

875 Mayewski, P. A., Rohling, E. E., Curt Stager, J., Karlén, W., Maasch, K. A., David Meeker, L.,
876 Meyerson, E. A., Gasse, F., van Kreveld, S., Holmgren, K., Lee-Thorp, J., Rosqvist, G., Rack,
877 F., Staubwasser, M., Schneider, R. R. and Steig, E. J.: Holocene climate variability, Quaternary
878 Research, 62(3), 243–255, doi:10.1016/j.yqres.2004.07.001, 2004.

879 Naafs, B. D. A., Gallego-Sala, A. V., Inglis, G. N. and Pancost, R. D.: Refining the global
880 branched glycerol dialkyl glycerol tetraether (brGDGT) soil temperature calibration, Organic
881 Geochemistry, 106, 48–56, doi:10.1016/j.orggeochem.2017.01.009, 2017.

882 Nottingham, A. T., Whitaker, J., Turner, B. L., Salinas, N., Zimmermann, M., Malhi, Y. and
883 Meir, P.: Climate Warming and Soil Carbon in Tropical Forests: Insights from an Elevation
884 Gradient in the Peruvian Andes, BioScience, 65(9), 906–921, doi:10.1093/biosci/biv109, 2015.

885 Nottingham, A. T., Fierer, N., Turner, B. L., Whitaker, J., Ostle, N. J., McNamara, N. P.,
886 Bardgett, R. D., Leff, J. W., Salinas, N., Silman, M. R., Kruuk, L. E. B. and Meir, P.: Microbes
887 follow Humboldt: temperature drives plant and soil microbial diversity patterns from the
888 Amazon to the Andes, Ecology, 99(11), 2455–2466, doi:https://doi.org/10.1002/ecy.2482,
889 2018.

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

894 Peterse, F., Moy, C. M. and Eglinton, T. I.: A laboratory experiment on the behaviour of soil-
895 derived core and intact polar GDGTs in aquatic environments, Biogeosciences, 12(4), 933–943,
896 doi:10.5194/bg-12-933-2015, 2015.

897 R Core Team, R: A language and environment for statistical computing. R Foundation for
898 Statistical Computing, Vienna, Austria, 2014.

899 Russell, N. J., Evans, R. I., ter Steeg, P. F., Hellemons, J., Verheul, A. and Abee, T.: Membranes
900 as a target for stress adaptation, International Journal of Food Microbiology, 28(2), 255–261,
901 doi:10.1016/0168-1605(95)00061-5, 1995.

902 Scalercio, S., Bonacci, T., Mazzei, A., Pizzolotto, R. and Brandmayr, P.: Better up, worse down:
903 bidirectional consequences of three decades of climate change on a relict population of Erebia
904 cassioides, J Insect Conserv, 18(4), 643–650, doi:10.1007/s10841-014-9669-x, 2014.

905 Siles, J. A. and Margesin, R.: Abundance and Diversity of Bacterial, Archaeal, and Fungal
906 Communities Along an Altitudinal Gradient in Alpine Forest Soils: What Are the Driving
907 Factors?, Microb Ecol, 72(1), 207–220, doi:10.1007/s00248-016-0748-2, 2016.

908 Sinensky, M.: Homeoviscous Adaptation—A Homeostatic Process that Regulates the Viscosity
909 of Membrane Lipids in Escherichia coli, PNAS, 71(2), 522–525, doi:10.1073/pnas.71.2.522,
910 1974.

© Author(s) 2020. CC BY 4.0 License.


911 Singer, S. J. and Nicolson, G. L.: The Fluid Mosaic Model of the Structure of Cell Membranes,
912 Science, 175(4023), 720–731, doi:10.1126/science.175.4023.720, 1972.

913 Staubwasser, M., Sirocko, F., Grootes, P. M. and Segl, M.: Climate change at the 4.2 ka BP
914 termination of the Indus valley civilization and Holocene south Asian monsoon variability,
915 Geophysical Research Letters, 30, 1425, doi:10.1029/2002GL016822, 2003.

916 Steinhilber, F., Beer, J. and Fröhlich, C.: Total solar irradiance during the Holocene,
917 Geophysical Research Letters, 36(19), doi:https://doi.org/10.1029/2009GL040142, 2009.

918 Szponar, B., Kraśnik, L., Hryniewiecki, T., Gamian, A. and Larsson, L.: Distribution of 3-
919 Hydroxy Fatty Acids in Tissues after Intraperitoneal Injection of Endotoxin, Clin Chem, 49(7),
920 1149–1153, doi:10.1373/49.7.1149, 2003.

921 Todaro, L., Andreu-Hayles, L., D'Alessandro, C., Gutiérrez, E., Cherubini, P. and Saracino, A.:
922 Response of Pinus leucodermis to climate and anthropogenic activity in the National Park of
923 Pollino (Basilicata, Southern Italy), Biological Conservation, 137, 507–519,
924 doi:10.1016/j.biocon.2007.03.010, 2007.

925 Véquaud P. , Derenne S. , Anquetil C., Collin S. , Poulenard J., Sabatier P., Huguet A. :
926 Influence of environmental parameters on the distribution of bacterial lipids in soils from the
927 French Alps: Implications for paleo-reconstructions , Organic Geochemistry, under revision.

928 Wakeham, S. G., Pease, T. K. and Benner, R.: Hydroxy fatty acids in marine dissolved organic
929 matter as indicators of bacterial membrane material, Organic Geochemistry, 34(6), 857–868,
930 doi:10.1016/S0146-6380(02)00189-4, 2003.

931 Wang, C., Bendle, J., Yang, Y., Yang, H., Sun, H., Huang, J. and Xie, S.: Impacts of pH and
932 temperature on soil bacterial 3-hydroxy fatty acids: Development of novel terrestrial proxies,
933 Organic Geochemistry, 94, 21–31, doi:10.1016/j.orggeochem.2016.01.010, 2016.

934 Wang, C., Bendle, J. A., Zhang, H., Yang, Y., Liu, D., Huang, J., Cui, J. and Xie, S.: Holocene
935 temperature and hydrological changes reconstructed by bacterial 3-hydroxy fatty acids in a
936 stalagmite from central China, Quaternary Science Reviews, 192, 97–105,
937 doi:10.1016/j.quascirev.2018.05.030, 2018.

938 Wang, J.-T., Cao, P., Hu, H.-W., Li, J., Han, L.-L., Zhang, L.-M., Zheng, Y.-M. and He, J.-Z.:
939 Altitudinal Distribution Patterns of Soil Bacterial and Archaeal Communities Along Mt.
940 Shegyla on the Tibetan Plateau, Microb Ecol, 69(1), 135–145, doi:10.1007/s00248-014-0465-
941 7, 2015.

942 Weber, Y., De Jonge, C., Rijpstra, W. I. C., Hopmans, E. C., Stadnitskaia, A., Schubert, C. J.,
943 Lehmann, M. F., Sinninghe Damsté, J. S. and Niemann, H.: Identification and carbon isotope
944 composition of a novel branched GDGT isomer in lake sediments: Evidence for lacustrine
945 branched GDGT production, Geochimica et Cosmochimica Acta, 154, 118–129,
946 doi:10.1016/j.gca.2015.01.032, 2015.

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

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.


950 Weiss, B.: The decline of Late Bronze Age civilization as a possible response to climatic
951 change, Climatic Change, 4(2), 173–198, doi:10.1007/BF00140587, 1982.

952 Whitaker, J., Ostle, N., Nottingham, A. T., Ccahuana, A., Salinas, N., Bardgett, R. D., Meir, P.
953 and McNamara, N. P.: Microbial community composition explains soil respiration responses to
954 changing carbon inputs along an Andes-to-Amazon elevation gradient, Journal of Ecology,
955 102(4), 1058–1071, doi:https://doi.org/10.1111/1365-2745.12247, 2014.

956 Wollenweber, H. W. and Rietschel, E. T.: Analysis of lipopolysaccharide (lipid A) fatty acids.,
957 Journal of Microbiological Methods, 11(3,4), 195–211, 1990.

958 Wollenweber, H.-W., Broady, K. W., Luderitz, O. and Rietschel, E. T.: The Chemical Structure
959 of Lipid A, European Journal of Biochemistry, 124(1), 191–198, doi:10.1111/j.1432-
960 1033.1982.tb05924.x, 1982.

961 Yang, H., Lü, X., Ding, W., Lei, Y., Dang, X. and Xie, S.: The 6-methyl branched tetraethers
962 significantly affect the performance of the methylation index (MBT′) in soils from an altitudinal
963 transect at Mount Shennongjia, Organic Geochemistry, 82, 42–53,
964 doi:10.1016/j.orggeochem.2015.02.003, 2015.

965 Zelles, L.: Fatty acid patterns of phospholipids and lipopolysaccharides in the characterisation
966 of microbial communities in soil: a review, Biol Fertil Soils, 29(2), 111–129,
967 doi:10.1007/s003740050533, 1999.

968

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

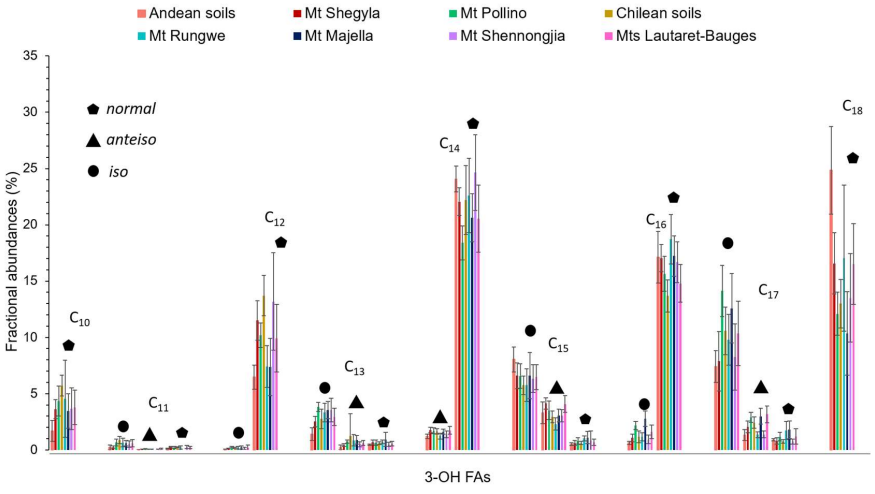

Figure 1. Average distribution of 3-OH FAs along the 8 altitudinal transects investigated in this study. Data from Mts. Majella and Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from Wang et al. (2016). Data from Mts. Lautaret-Galibier were taken from Véquaud et al. (under revision).

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

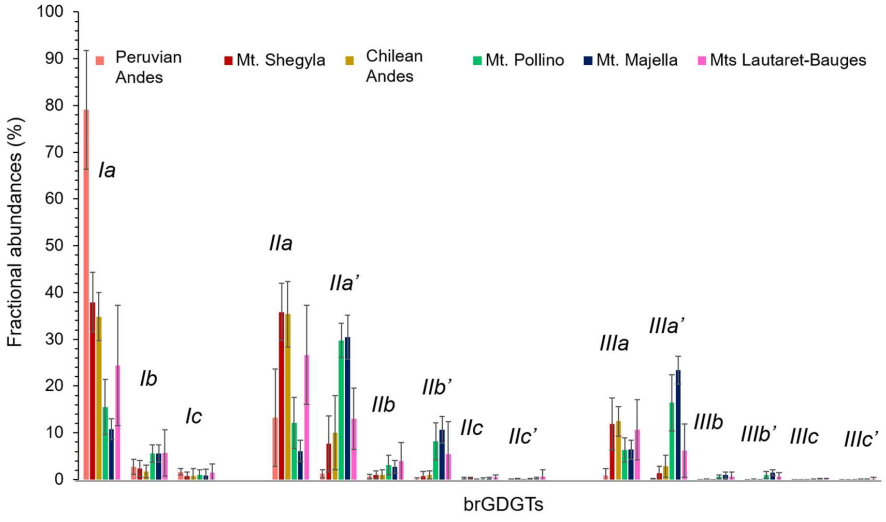

Figure 2. Average distribution of 5- and 6-methyl brGDGTs, along Mts. Shegyla, Pollino Majella, Lautaret-Bauges, Peruvian Andes and Chilean Andes. Data from Mt. Majella were taken from Huguet et al. (2019). Data from Mts. Lautaret-Galibier were taken from Véquaud et al. (under revision).

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

Biogeosciences
Discussions
Open Access
EGU

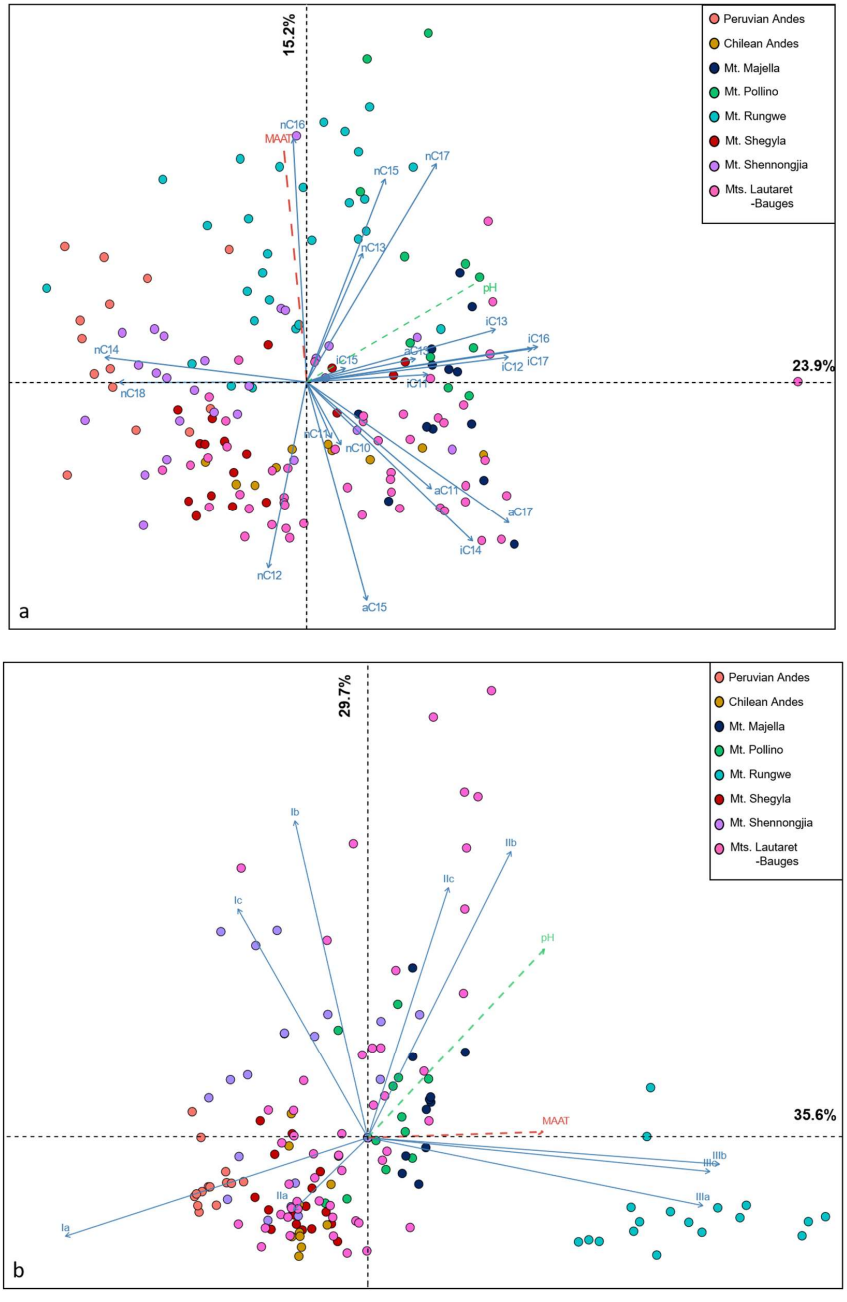

Figure 3. PCA biplot of (a) individual 3-OH FA relative abundances and (b) individual brGDGT relative abundances in soil samples from the 8 altitudinal transects.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

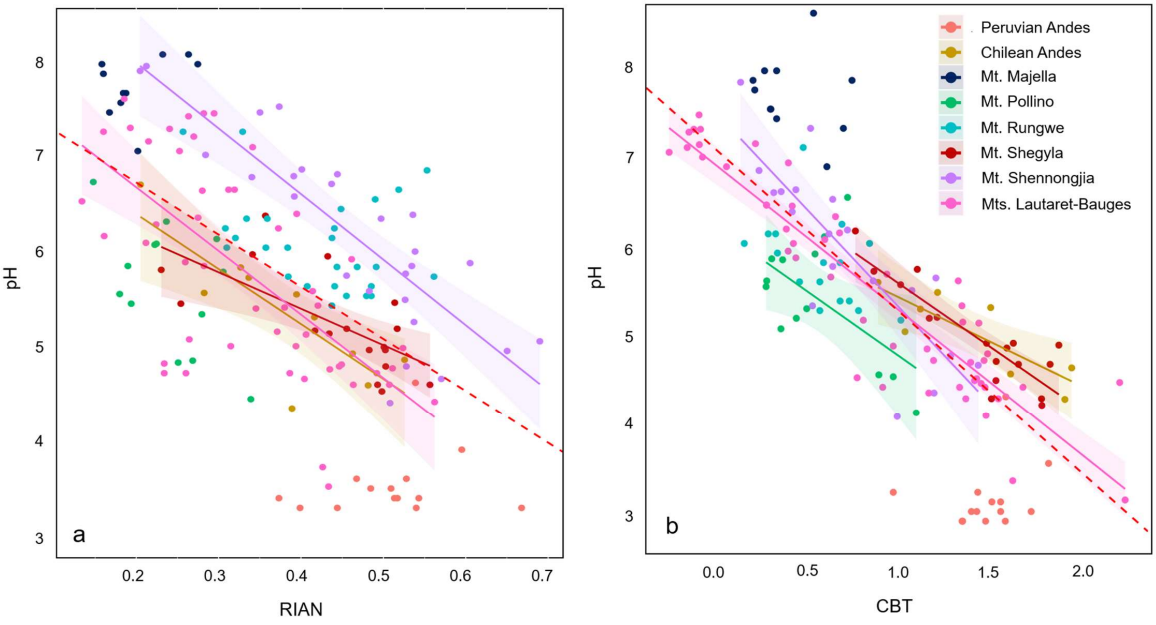

Figure 4. Linear regressions between (a) pH and RIAN and (b) pH and CBT along the 8 altitudinal transects investigated. Dotted lines represent the 95% prediction interval for each regression and colored areas represent the 95% confidence interval for each regression. Data for Mts. Majella and Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from Wang et al. (2016). Data from Mts. Lautaret-Galibier were taken from Véquaud et al. (under revision). Only significant regressions ($p < 0.05$) are shown.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

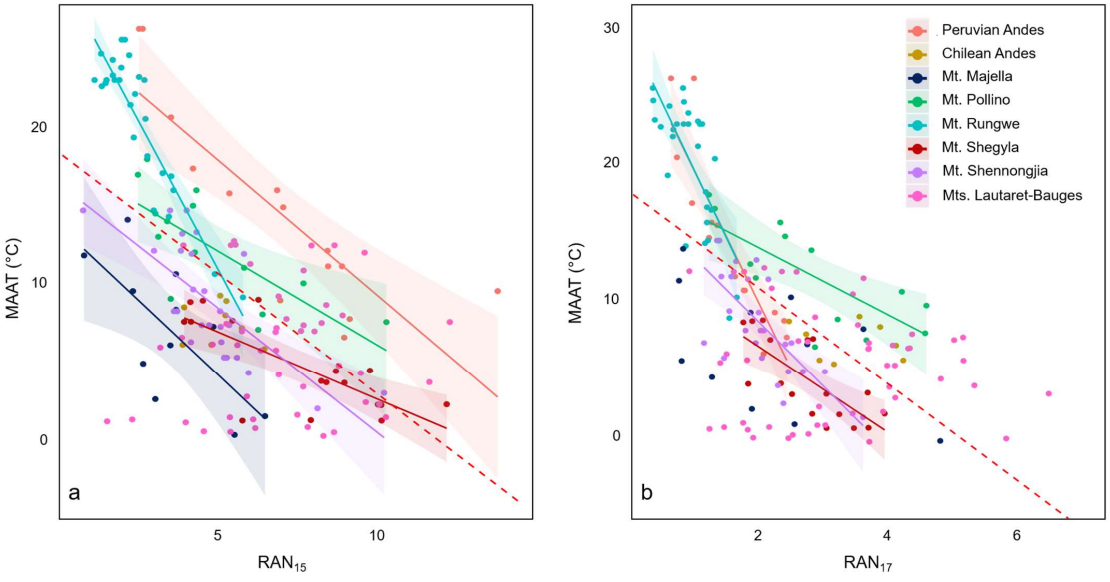

Figure 5. Linear regressions between (a) MAAT and RAN_{15} and (b) MAAT and RAN_{17} along the 8 elevational transects investigated. Dotted lines represent the 95% prediction interval for each regression and colored areas represent the 95% confidence interval for each regression. Data from Mts. Majella and Rungwe were taken from Huguet et al. (2019). Data from Mt. Shennongjia were taken from Wang et al. (2016). Data from Mts. Lautaret-Galibier were taken from Véquaud et al. (under revision). Only significant regressions ($p < 0.05$) are shown.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

Biogeosciences
Discussions
Open Access
EGU

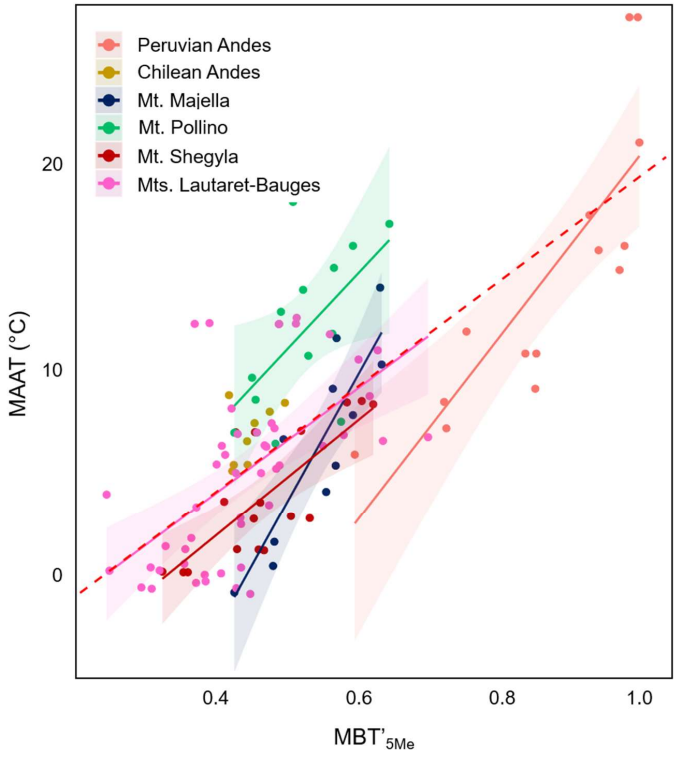

Figure 6. Linear regressions between (a) MAAT and MBT'$_{5Me}$ along 6 of the 8 elevational transects investigated. Dotted lines represent the 95% prediction interval for each regression and colored areas represent the 95% confidence interval for each regression. Data from Mt. Majella were taken from Huguet et al. (2019). Data from Mts. Lautaret-Galibier were taken from Véquaud et al. (under revision). The global soil calibration by De Jonge et al. (2014) was applied to all these transects. Only significant regressions ($p < 0.05$) are shown.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

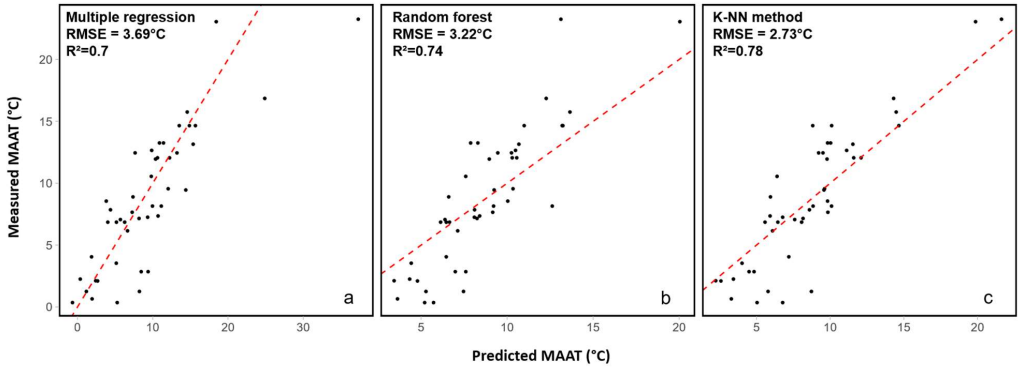

Figure 7. Results of the three different models tested to reconstruct the MAAT from 3-OH FA distribution: (a) Multiple regression, (b) random forest, (c) k-NN method.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

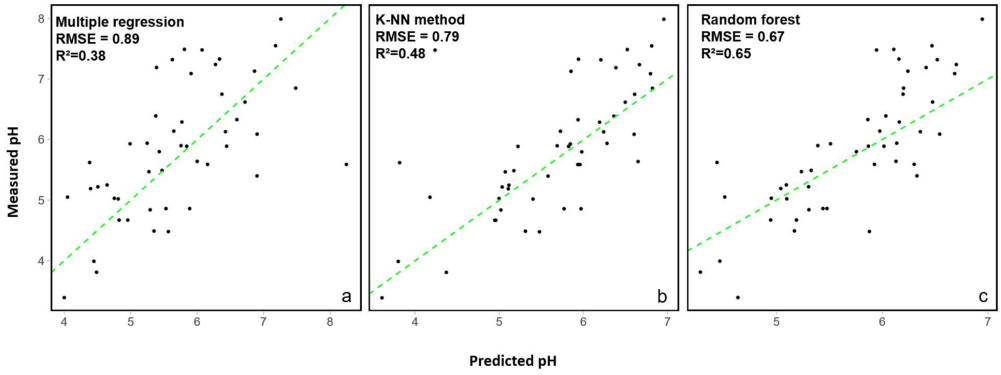

Figure 8. Results of the three different models tested to reconstruct the pH from 3-OH FA distribution: (a) Multiple regression, (b) k-NN method, (c) random forest.

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

Biogeosciences
Discussions
Open Access
EGU

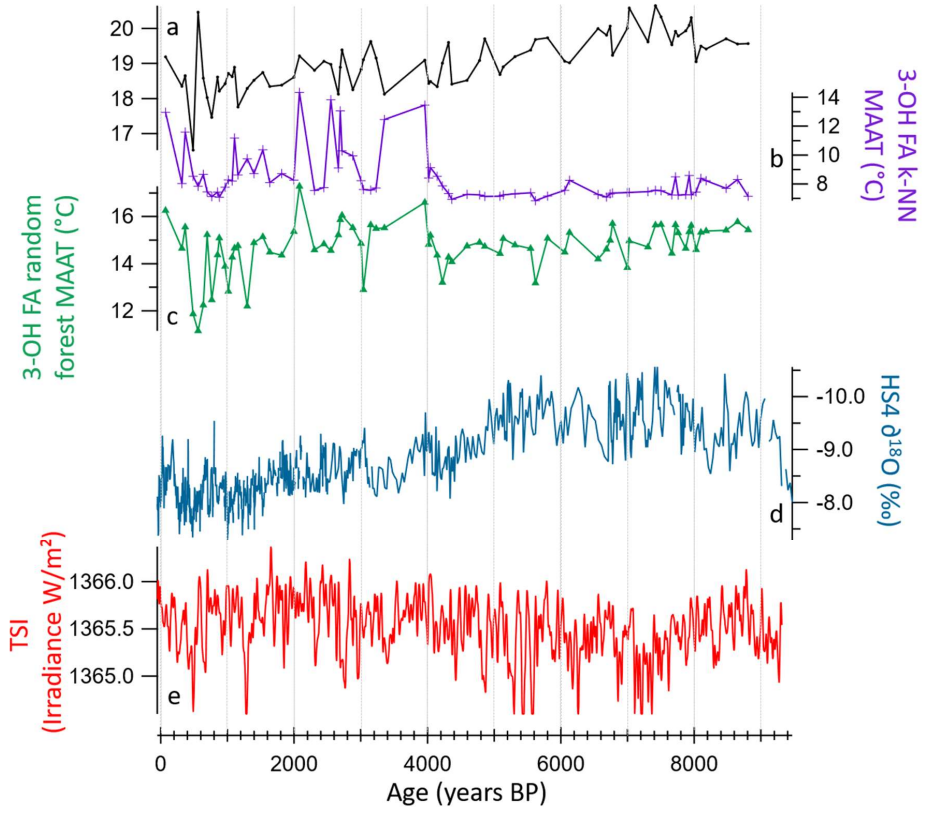

Figure 9. Comparison of the 3-OH FA model-MAAT record with other time-series and proxy records for the HS4 speleothem (Wang et al., 2018). (a) RAN_{15}-MAAT record reconstructed using a local Chinese calibration (Wang et al., 2016; Wang et al., 2018). (b) 3-OH FA k-NN model-MAAT (c) 3-OH FA random forest model-MAAT. (d) The $CaCO_3$ oxygen isotope record (Hu et al., 2008b). (e) Total solar irradiance (TSI; W/m²) during the Holocene (past 9300 years) based on a composite described in Steinhilber et al. (2009).

https://doi.org/10.5194/bg-2020-428
Preprint. Discussion started: 24 November 2020
© Author(s) 2020. CC BY 4.0 License.

Biogeosciences
Discussions
Open Access
EGU

ID	Location	Altitude (m)	MAAT(°C)	pH	RAN_{15}	RAN_{17}	RIAN	MBT'_{5Me}	CBT
1	Peruvian Andes	194	26.4	3.7	2.45	0.96	0.47	0.96	0.97
2	Peruvian Andes	210	26.4	4	2.56	0.61	0.60	0.97	1.81
3	Peruvian Andes	1063	20.7	4.7	3.46	0.70	0.54	0.98	1.58
4	Peruvian Andes	1500	17.4	3.5	4.15	0.93	0.51	0.91	1.39
5	Peruvian Andes	1750	15.8	3.6	5.30	1.32	0.51	0.92	1.50
6	Peruvian Andes	1850	16	3.5	6.81	1.23	0.54	0.96	1.72
7	Peruvian Andes	2020	14.9	3.4	7.00	1.19	0.54	0.95	1.58
8	Peruvian Andes	2520	12.1	3.7	8.40	1.59	0.53	0.74	1.43
9	Peruvian Andes	2720	11.1	3.6	8.42	1.73	0.48	0.83	1.55
10	Peruvian Andes	3020	9.5	3.4	13.78	2.21	0.44	0.83	1.34
11	Peruvian Andes	3200	8.9	3.5	6.91	2.35	0.37	0.71	1.42
12	Peruvian Andes	3025	11.1	3.5	8.86	1.74	0.52	0.82	1.55
13	Peruvian Andes	3400	7.7	3.4	9.10	2.39	0.40	0.71	1.35
14	Peruvian Andes	3644	6.5	3.4	8.93	2.03	0.67	0.58	1.47
15	Mt. Shegyla, Tibet	3106	8.9	5.53	6.22	2.02	0.51	0.59	1.21
16	Mt. Shegyla, Tibet	3117	8.9	6.43	4.47	1.86	0.36	0.57	0.76
17	Mt. Shegyla, Tibet	3132	8.8	6.01	4.07	1.72	0.43	0.61	0.86
18	Mt. Shegyla, Tibet	3344	7.6	6.03	5.40	2.80	0.34	0.51	1.10
19	Mt. Shegyla, Tibet	3355	7.5	5.87	4.09	2.71	0.23	0.44	1.01
20	Mt. Shegyla, Tibet	3356	7.5	5.52	3.87	2.14	0.25	0.42	1.16
21	Mt. Shegyla, Tibet	4030	3.7	5.21	8.21	3.64	0.43	0.49	1.59
22	Mt. Shegyla, Tibet	4046	3.6	4.68	8.37	3.00	0.49	0.52	1.50
23	Mt. Shegyla, Tibet	4050	3.6	4.61	8.94	2.47	0.50	0.44	1.78
24	Mt. Shegyla, Tibet	3912	4.3	5.04	9.74	2.30	0.48	0.40	1.82
25	Mt. Shegyla, Tibet	3918	4.3	4.68	8.67	1.80	0.56	0.45	1.77
26	Mt. Shegyla, Tibet	4298	2.1	5.04	10.00	2.78	0.50	0.45	1.65
27	Mt. Shegyla, Tibet	4295	2.2	4.87	12.17	3.90	0.50	0.42	1.53
28	Mt. Shegyla, Tibet	4304	2.1	5.26	10.10	3.20	0.46	0.46	1.62
29	Mt. Shegyla, Tibet	4479	1.1	5.26	10.11	3.42	0.52	0.35	1.48
30	Mt. Shegyla, Tibet	4479	1.1	5.07	5.71	3.00	0.50	0.35	1.53
31	Mt. Shegyla, Tibet	4474	1.1	5.24	7.88	3.65	0.42	0.32	1.87
32	Mt. Pollino, Italy	0	18	6.78	2.71	1.19	0.15	0.50	0.72
33	Mt. Pollino, Italy	200	17	6.19	2.41	1.28	0.30	0.63	0.54
34	Mt. Pollino, Italy	400	16	6.13	4.26	2.29	0.22	0.58	0.37
35	Mt. Pollino, Italy	600	15	6.14	4.15	2.36	0.22	0.55	0.31
36	Mt. Pollino, Italy	800	14	4.53	3.34	2.77	0.34	0.51	1.09
37	Mt. Pollino, Italy	1000	13	5.41	3.06	1.83	0.28	0.48	0.36
38	Mt. Pollino, Italy	1200	12	6.37	4.21	1.91	0.24	0.55	0.60
39	Mt. Pollino, Italy	1400	11	5.62	5.77	4.16	0.18	0.52	0.50
40	Mt. Pollino, Italy	1600	10	4.93	7.64	4.54	0.27	0.44	0.89
41	Mt. Pollino, Italy	1800	9	4.91	3.45	3.17	0.25	0.45	0.97
42	Mt. Pollino, Italy	2000	8	5.52	6.35	4.52	0.19	0.56	0.44
43	Mt. Pollino, Italy	2100	7.5	5.91	10.26	3.62	0.19	0.42	0.28
44	Mt. Pollino, Italy	2200	7	5.85	6.21	2.82	0.31	0.47	0.28
45	Chilean Andes	690	9.2	5.38	5.01	3.51	0.42	0.41	1.03
46	Chilean Andes	870	8.9	5.62	5.21	2.43	0.39	0.49	1.12
47	Chilean Andes	891	7.9	4.94	5.18	2.69	0.53	0.44	1.61
48	Chilean Andes	915	NA	6.75	4.67	4.25	0.21	NA	NA
49	Chilean Andes	980	8.5	5.63	3.87	3.83	0.28	0.46	1.50
50	Chilean Andes	985	5.8	4.67	6.41	3.12	0.48	0.41	1.90
51	Chilean Andes	1125	6.0	5.00	3.83	4.18	0.46	0.42	1.94
52	Chilean Andes	1151	6.0	5.89	4.74	2.89	0.33	0.43	0.89
53	Chilean Andes	1196	7.1	5.79	5.70	4.07	0.34	0.43	1.21
54	Chilean Andes	1385	NA	4.43	4.85	1.91	0.39	0.41	NA

Table 1. List of the soil samples collected along Mts. Shegyla, Pollino, Peruvian Andes and Chilean Andes, with corresponding altitude (m), MAAT (°C), pH and 3-OH FA/brGDGT-derived indices.