Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-428-AC2, 2021

Interactive comment on "Development of global temperature and pH calibrations based on bacterial 3-hydroxy fatty acids in soils" *by* Pierre Véquaud et al.

Pierre Véquaud et al.

pierre.vequaud@upmc.fr

Received and published: 18 February 2021

We thank the anonymous reviewer for these comments. A detailed list of changes and arguments answering to the different comments is provided below.

Reviewers' comments:

Reviewer #2: Review

-I have finished the review of the manuscript "development of global temperature and p calibrations based on bacterial 3-hydroxy fatty acids in soils", submitted to biogeosciences by Pierre Vequaud and co-authors. In general, the authors present a good

contribution to the developing field of 3-OH fatty acids, by compiling several published and novel altitudinal gradients. The performance of the local and compiled linear regressions are compared with that of brGDGT lipids, for those sites where 5- and 6-methyl compounds were separated (and where the MBT'5ME could thus be calculated). This is an interesting dataset, and a necessary next step in the development of 3-OH fatty acids as temperature proxies.

Author comment: We would like to thank the reviewer for his positive comments and acknowledging the work made to evaluate 3-OH FAs as temperature and pH proxies on a global scale.

- However, the performance of the three additional models (multiple linear regression, random forest and k-nearest neighbor), that move away from the previously established RIAN, RAN15 and RAN17 ratios, are not evaluated thoroughly. There is no explanation of what the models represent, and why they allow a better correlation between the3-OH fatty acids and MAAT or pH, compared to the simple linear regression. There is for instance no analysis of residuals, and no indication of maximum reconstructed temperatures. I would recommend the authors to look at a recent calibration paper (fi Dearing-Crampton Flood, 2020), to see what the state-of-the-art in lipid proxy calibration is. As it stands, the authors will have a difficult time convincing reader that the more complex models are the preferred choice when doing 3-OH FA-based climate reconstructions.

Author comment: As suggested by the reviewer, the detailed characteristics of the models proposed in this work will be added. Thus, a study of the residuals of these models will be proposed, and in addition to the R^2 and RMSE values, a table of the variance values of the residuals, model bias and upper/lower limits of pH and MAAT estimates will be presented (taking as an example Table 2 in the paper by Dearing-Crampton Flood, 2020). The advantage of the three models of the present manuscript over simple linear regression models is to take into account the whole suite of bacterial lipids (in our case C10-C18 3-OH FAs) to estimate MAAT and pH values, which allows overcoming the limitations related to the use of a single index. In addition, the Random Forest and k-NN models are non-parametric and non-linear models, allowing to take into account the potential non-linear influences of MAAT/pH on the fractional abundances of 3-OH FAs. The functioning of the different models will be clarified in the revised manuscript, with more didactic explanations. Quantitative evaluations of the influence of each 3-OH FA on the MAAT/pH calibrations will also be proposed, thus providing mechanistic details on the models. These explanations should highlight the advantages of these new models and especially their ease of use through the web application provided in this study.

-When developing a MAAT-calibration, the authors should also mention the following weakness: there is currently only 3 samples with MAAT > 15 âËĞD ËĞC in the dataset (maximum temperature: +- 20 âËĞDCËĞ). Do these samples influence the MAAT calibration disproportionally?

Author comment: There are > 30 samples collected in locations with MAAT > 15 °C, especially from Mt. Rungwe (Tanzania) and Peruvian Andes. Nevertheless, as explained in L. 305 "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". Thus, the developed models include 25 % of the dataset, considered as independent from the training dataset. The test dataset, randomly chosen, indeed contains only 3 samples with MAAT > 15 °C. Nevertheless, the cross validations and stratification modalities allow to reduce the influence of extreme values. To go further, the models proposed in this study were tested again without the extreme temperature values. This did not change the proposed MAAT estimates. To take into account the reviewers' comment, a discussion will be added in the revised version of this manuscript on the potential influence of "extreme" MAAT values on our models.

Minor comments:

L 91. Include 'bacterial' before 'lipids'. L 129. 'more developed statistical approach', perhaps rephrase as 'further development of the statistical approaches'?

Author comment: This sentence will be modified as suggested.

L 148. For easy comparison: include the altitudinal, pH and MAAT rangesnof the previously published transects. L 200-201. Soil sensors have been used, but the calibration is done with mean annual air temperature (MAAT) instead. As the soil sensor data is not used or discussed further, I'd remove their mention here, and use the MAAT for this site as well. L 322 (and further in the manuscript). As far as I can see, the supp. tables referred to are not present. Please include these tables in a revised version.

Author comment: pH and MAAT ranges will be added for all the previously published transects. In the supplementary tables we will provide the new data obtained from Mt Pollino, Shegyla, Peruvian and Chilean Andes. For the other previously published transects, the reader will be directed towards the references of the corresponding papers. This point will be clarified in the revised version of this manuscript. We will remove the mention of the soil sensors. We will check that the supp. Tables are included in the revised version.

L 541. Perhaps the authors can refer here to De Jonge et al., 2019, who argue that temperature can modulate the soil bacterial composition, and the dependency between MBT'5ME and soil temperature?

Author comment: This reference will be added to the manuscript.

L551. Here the authors should explain better how 'more complex models' allow to take the complexity of each site into account. At this point, in the manuscript, we have observed that the linear dependency between the 3-OH fatty acids is better on a local scale, than on a global scale. What mechanism do the authors propose for this, and how can a more complex model correct for it?

Author comment: It should be highlighted that the relative abundances of the 3-OH

FAs were linearly correlated with MAAT/pH at a local scale, but only along some of the transects. The absence of linear relationships at a local scale for part of the sites, and more generally at a global scale, may at least “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)”, as specified in the manuscript (L. 435). The models proposed in this study for MAAT/pH reconstruction are based on the whole suite of 3-OH FAs (C10-C18) instead of indices. They allow better capturing the complexity of the microoenvironments found along each transect and the variability of the lipid distribution, in contrast with linear models. In addition, two of the three models presented in this paper (k-NN and random forest) are both non-parametric and non-linear and present the advantage of taking into account the potential non-linear relationships between the relative abundances of 3-OH FAs and environmental parameters (here MAAT / pH).

L567. Have the authors done a selection of the fatty acids that are necessary for the model? For instance, forward selection or reverse selection? Does the model not suffer from overfitting? Is there any fatty acids that are generally present in low abundance (and can thus be absent in geological archives) that are important for the regression? If yes, is it prudent to include these low abundance compounds in the model as well?

Author comment: In order to prevent any overfitting of the models, several steps were followed and will be better explained in the next version of the manuscript: (1) Cross-validation of the training dataset to find the optimal combination of 3-OH FA relative abundances and maximize the model performance. (2) No pre-processing (i.e.no preselection) of the individual 3-OH FAs. All the 3-OH FA homologues, whatever the abundance, were used in the different models to keep the maximum variability and take into account the specificity and complexity of each altitudinal transect. (3) Testing of the model on a randomly selected dataset (i.e. test dataset) that was not used to build the model. In order to clarify the importance of the different 3-OH FAs in the

models, figures presenting the variance of the different homologues in each model will be proposed in the next version of the manuscript.

L 587. The random forest model and k-NN model need to be explained much better before the results are presented. What are they based on and how do they compare variability in the lipid distribution with the MAAT? L 593. Same comments as at L 567.

Author comment: As explained above, each model will be better explained, and the results of these different new models will be directly and statistically compared to those based on previously defined indices (RAN15, RAN17 and RIAN).

L 651. Can the authors comment on the probable source of the 3-OH fatty acids in the speleothem? Can we assume that all lipids are derived from the soil, or is there a (variable) proportion produced in the cave environment as well?

Author comment: Wang et al. (2018), who analyzed 3-OH FAs in this speleothem, suggested that these compounds are largely derived from the overlying soils based on geochemical and microbiological analyses. Indeed, as specified by these authors, “the broad similarity of 3-OH-FA distributions in the overlying soils and stalagmites, supported by the site-specific analyses of bacterial diversity and transport pathways, supports a major contribution of 3-OH-FAs from Gram-negative bacteria dwelling in the overlying soils to the stalagmite samples”. As also stated by Wang et al. (2018), this does not totally exclude that a proportion of the lipids may be derived from the cave ecosystem. This point will be mentioned in the revised manuscript.

Refs: Dearing Crampton-Flood E., Tierney J. E., Peterse F., Kirkels F. M. S. A. and Sinninghe Damsté J. S. (2020) BayMBT: A Bayesian calibration model for branched glycerol dialkyl glycerol tetraethers in soils and peats. Geochimica et Cosmochimica Acta 268, 142–159. De Jonge C., Radujkovi′c D., Sigurdsson B. D., Weedon J. T., Janssens I. and Peterse F. (2019) Lipid biomarker temperature proxy responds to abrupt shift in the bacterial community composition in geothermally heated soils. Organic Geochemistry, S0146638019301275.

References

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

Siles, J.A., Margesin, R., 2016. Abundance and Diversity of Bacterial, Archaeal, and Fungal Communities Along an Altitudinal Gradient in Alpine Forest Soils: What Are the Driving Factors? Microbial Ecology 72, 207–220.

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

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-428, 2020.