***Response to Interactive comment on* "OPTiMAL: A new machine learning approach for GDGT-based palaeothermometry" *by* Yvette L. Eley et al.**

Formatting notes:

Reviewers / Other comments are in bold italic

Our responses are in plain type

Proposed changes in manuscript or quotes from existing text are in italic

Response to Reviews:

Referee Yige Zhang: Received and published: 12 September 2019

I think TEX86 is wickedly charming. We all have heard about the misdeed of TEX86, but also realized that often times TEX86 is the only thing that could help us to learn about the ocean temperatures, especially in the greenhouse climates.

This proxy has probably already passed its pessimism stage, and currently in the realism phase. Eley et al is a timely contribution to help us better understand this important proxy. They applied the cutting-edge machine-learning tools to improve the SST estimates using GDGTs, with the concept of identifying nearest neighbors in the global core-top dataset. This is innovative, but I do have some concerns detailed below.

We thank Yige Zhang for his comments and agree that GDGT proxies are heading into a 'realism phase', and hope that our submission provides new analytical approaches for improving the robustness of GDGT-based paleothermometry.

I suggest moderate revision of the MS before it can be accepted by CP.

Although this was not explicitly explained in the MS, I assume the authors used the percentage of each individual GDGT, when [GDGT-0]+[GDGT-1]+[GDGT-2]+[GDGT3]+[Cren]+[Cren']=100%. If this is the case, these 6 variables are not independent from each other. Instead, they are often dominated by the variations of [GDGT-0] and [Cren], the two major GDGTs. To show this, I did a simple calculation of the T&T15 global core-top dataset, which yielded an average [GDGT-0]+[Cren] of 88% of all GDGTs. This means that the variability of [GDGT-1], [GDGT-2], [GDGT-3] or [Cren'] that you see might be largely explained by the changes of [GDGT-0] or [Cren]. This is one of the reasons that TEX86 only considers the minor GDGTs, and uses a ratio. Ratios are good, as demonstrated by numerous cases in geochemistry.

With that being said, I agree with the authors that by using a subset of GDGTs like the TEX86, we are losing some information. We also discussed this in the Rind Index paper of Zhang et al., 2016. We realized that Ring Index values are dominated by [GDGT-0] and [Cren], as illustrated in Fig. 2. So the real difference between OPTIMAL and TEX86 (and any TEX86 calibrations, BAYSPAR, Kim or Liu) is not 6 dimensions vs. 1.

If the authors would like to, they can try something like use 2 subgroups of GDGTs - major and minor ones, with the total of each equals 100%; or some other forms of 6 ratios, normalizing GDGTs to one in the major, and one in the minor category. There's still going to be some dependency between the 6 variables, but they are closer to the 6 dimensions than the original treatment.

If they decide not to pursue these alternatives, I'd like to see that at least they acknowledge the interdependency of GDGT% data.

We will amended the text in the final submission to make it explicit that we are using the percentage of each individual GDGT. Regarding the issue of dependency, our analysis suggests there are only five free parameters. Machine learning tools should be able to pick up on this correlation and effectively ignore one of the parameters (or one parameter combination). For example, we do find that the GP emulator has a very broad kernel in at least one dimension, signaling this. In principle, we could have considered only five of six parameters. The smaller scale of some of the parameters is automatically accounted for by the trained kernel size in GP regression, or by normalising to the appropriate dynamical range in our initial investigation.

Another issue is the extrapolating from the modern calibration data set. Nobody likes extrapolations. But there are these mesocosm studies from NIOZ that demonstrated the response of GDGTs to ~40ºC temperatures. The archaea found in hot spring continues to increase their ring numbers until ~100ºC.

We refer back to the extensive responses to Tierney, concerning the nature of the constraints on the form of the temperature dependence of GDGT assemblages - that although there is evidence that ring number increases with temperature, the form of this relationship is characterised by empirical model fitting where we have modern environmental sampling (Kim et al. 2010; Tierney & Tingley et al. 2014; Zhang et al., 2016; OPTiMAL), but that extrapolation of these models beyond the calibration data is problematic, in the absence of other quantified constraints on the behaviour of this system.

I agree that we might not know the absolute temperature above ~30ºC very well, but I wouldn't call them "inappropriate use" that "impacts the confidence".

This is a matter of judgement. If the lack of understanding of the nature of GDGT temperature dependency above 30ºC results in paleo-SST estimates that we do "*not know ...very well*", that is grounds for a reduced confidence. In this context, as argued above, using such high temperature estimates to generate global mean surface temperature and climate sensitivity estimates (Zhu et al. 2019), in our judgment, is not appropriate without substantial caveats. We recognise however that the word 'inappropriate' may imply a subtle subtext of something being done in bad faith, and we have therefore amended this in the revised manuscript, replacing it with the words "provides absolute temperature reconstructions and uncertainty estimates that are relatively unconstrainted and therefore speculative".

In fact, the beauty of TEX86 is that it works in greenhouse climates and tropics when Uk'37 maxes out, carbonates are diagenetically C2 altered and seawater Mg/Ca is difficult to constrain.

As noted above: we do not wish to stop people using GDGT-based thermometry for the reconstruction of greenhouse climate states. We simply wish to urge caution in the application of this method when the fossil GDGT assemblages, from any time period, are strongly non-analogous to the modern calibration data on which this proxy rests. As we demonstrate this appears to be an increasing problem with increasing sample age, but it does not preclude the use of this proxy in greenhouse climate states where the fossil data are well-constrained by the modern calibration.

There is clearly an issue with the applicability of GDGT-based paleothermometry to high temperature conditions (>30ºC), but it is appropriate to end with a clear statement of our position: we would be truly delighted to secure a sound basis for the form of GDGT temperature dependency above 30ºC. As we, and all comments / reviewers note, applying GDGT paleothemometry at high temperatures would be extremely valuable for paleoclimate studies. Wishing for something, however, does not make it so. Instead, as a community we need to think and work creatively to address this problem, through culture or other manipulation studies, or through robust proxy-proxy inter-comparisons from ancient samples. Rather than the promotion of any single fix-all solution as *the answer,* we believe that advances will come by paying attention to the difficult questions. Amidst the complexities of nature, we should beware hubris.

Additional references (not found in our original submission) referred to in our response – these will be incorporated into our final revised submission:

Bale, N. et al. (2019) Applied and Environmental Microbiology 85(20) e01332-19
Cadillo-Qiuroz, H. et al. (2012) PLOS Biology, https://doi.org/10.1371/journal.pbio.1001265
Dunkley Jones, T. et al. (2013) Earth-Science Reviews, 125, 123-145
Elling, F. et al. (2017) Environmental Microbiology 19(7), 2681–2700
Hollis, C. et at. (2019) Geosci. Model Dev., 12, 3149–3206
Liu, X-L. et al. (2014) Marine Chemistry, 116, 1-8.
Lunt, D. J. et al. (2012) Clim. Past Discuss., 8, 1229- 787 1273.
Qin., W. et al. (2015) PNAS 112 (35) 10979-10984
Schouten, S. et al. (2007) Organic Geochemistry 38, 1537-1546
Wuchter, C. et al. (2004) Paleoceanography 19, PA4028, doi:10.1029/2004PA001041, 2004
Zhang, Y. G. et al. (2016) Paleoceanography, 31, 220-232
Zhu, J. et al. (2019) Science Advances 5 (9), eaax1874

.