

Reply to Reviewer 1 [in square brackets]
4 June 2019

1. General comments.

The paper by Hollis et al. provides a well written summary of the proxy methods used to derive past climate states and greenhouse gas levels from the geological record – specifically for the latest Paleocene (LP), Paleocene-Eocene thermal maximum (PETM) and early Eocene climatic optimum (EECO). The focus on this geological interval is appropriate - as stated by the authors – as at around 56 Ma the world experienced a geologically short (220 ky) and strong warming, the PETM, that was followed by a series of additional hyperthermals (e.g., ETM-2) and the longer duration EECO from 53 to 50 Ma. The PETM in particular offers a possible analog for current and forecast climate warming due to its short duration and the range of pCO₂ reconstructed for the PETM spanning current to projected greenhouse gas levels under different projections.

The authors provide a comprehensive yet succinct review of both paleontological proxies as well as geochemical proxies of temperature of waters and of the air, past CO₂ levels, as well as a range of paleoclimate parameters including annual and seasonal precipitation based primarily on terrestrially derived proxies. Further, their review highlights the challenges – strengths and weaknesses – posed by such a wide range of methods and approaches, ranging from concerns over precision versus accuracy, alternative methods of analysis and interpretation, taphonomy, limited geographic coverage, and the issues around compiling such disparate data into a database that can be used by the paleoclimate modelling community (model-model and model-data comparisons). As such, the article largely succeeds, offering a set of guidelines – as the article states – for the selection, compilation and analysis of such proxy data.

The article closes with a proposal for an “atlas” of climate conditions arrayed on an agreed paleogeography for the three selected time intervals to constrain and provide insight into the mechanisms controlling past hyperthermals via ‘database ver. 0.1’. The article draws on the expertise of a large set of authors – the DeepMIP team – whose research foci span these topics. The manuscript therefore represents a substantial contribution to modelling science and is appropriate for the journal. I have few concerns, and these mainly reflect topics dealt with in less depth than I would consider necessary, or restricting discussion to particular research teams, as detailed in the next section. In my specific comments I have focused on my area of expertise; paleobotanical reconstruction of past pCO₂ and climate, so “5 Terrestrial proxies for air temperature.”

[Many thanks for the positive feedback and constructive review.]

2. Specific comments

Section 5 in its title only states ‘Terrestrial proxies for air temperature’, when this section also considers precipitation (e.g., p. 31 line 24, in section 5.2 or in section 5.3 at line 15 on p. 32), and should be re-titled as ‘Terrestrial proxies for climate’.

[We prefer to keep the section 5 title as it is because this paper is focussed on temperature proxies. Although many of the approaches discussed can be used to reconstruct precipitation,

we wish to keep this manuscript focussed on temperature and CO₂. Precipitation warrants full treatment in a separate paper.]

Further, as leaf physiognomic methods as well as NLR approaches have been applied to reconstructing latest Paleocene to EECO precipitation (e.g., Greenwood et al. 2010; Pross et al. 2012; Eldrett et al. 2014; West et al. 2015; Suan et al. 2017; Hyland et al. 2018), the leaf-based and NLR proxies for precipitation should be included, e.g., CLAMP estimates growing season precipitation as well as that of the wettest 3 months and the driest 3 months, but these are not differentiated on p. 31 line 24, nor discussed at any point in section 5.2.

[We agree that these suggestions are all important and we will action them in a subsequent paper with a greater focus on precipitation proxies.]

5.2 Leaf morphology-based approaches. This section is very focused on the univariate leaf margin analysis (LMA) and the multivariate CLAMP, and ignores other methods such as digital leaf physiognomy (DiLP; Peppe et al. 2011), and also (as noted above and below) barely mentions the use of leaf physiognomy for reconstructing precipitation (annual and seasonal values) in PETM and EECO studies (e.g., Leaf Area Analysis [LAA] and DiLP). Ignoring DiLP is unfortunate, but perhaps forgivable as DiLP has been applied to very few Paleogene floras to date. However, LAA (Wilf et al. 1998; Peppe et al. 2011) has been applied to a wider set of floras, including LP to EECO floras (e.g., Greenwood et al. 2010; Sunderlin et al. 2011; Smith et al. 2012; West et al. 2015; Lowe et al. 2018) and – as noted – yields data on precipitation that is critical to the DeepMIP project.

[We agree that this section is focussed on LMA and CLAMP. We mention some other leaf morphology-based methods such as DiLP and the Nearest Neighbour Approach in Section 5.2.1, but do not provide details. This is primarily due to the relatively limited application of those other leaf morphology-based temperature proxies to date, and also to the limited scope for detailed discussion in this current manuscript. Precipitation proxies such as LAA were not discussed in any detail due to scope and space. The paper focus is on temperature for all proxies but the reviewer has brought an important point to our attention, that we do not make that sufficiently clear in Section 5, even though it is clearly stated in Section 1 Introduction. We certainly don't want to minimise the importance of precipitation and the hydrological regime in general, we believe this topic warrants another whole paper. This is the reason that LAA and the precipitation aspects of CLAMP were not discussed fully here.]

The discussion of weaknesses and strengths in this section is also out of date as it misses current and recent literature on LMA and CLAMP.

[Happy to be updated here, we certainly don't want to miss any key references. The most recent CLAMP methodology reference seems to be Yang et al. (2015), which we have included, but there are of course more recent articles that apply LMA/CLAMP, are these the ones being referred to as missing? We have now also included some of the additional references Reviewer 1 mentioned in his comments.]

Citing Carpenter et al. (2012) at line 22, p. 30 is idiosyncratic as an example as this paper proposed a method that has been never used in another study, whereas LAA (Wilf et al. 1998) is cited, but this univariate method's application to estimate Paleogene annual precipitation is not mentioned, despite LAA being used in multiple studies of the LP to EECO interval.

[Carpenter et al. (2012) proxies are for temperature and are included in the Paleogene temperature database with this paper. It therefore seemed appropriate to include the reference in the main text as an example, even though it is certainly a fair point about the limited application of the proxies involved. A key LAA reference is cited, also as an example, but since it is a precipitation proxy, we did not discuss it further in this paper.]

5.3 Nearest Living Relative Analysis. I disagree that the 'most widely used method' is the Coexistence Approach (CA), particularly when applied to the LP to EECO interval of concern.

[We have revised this wording, noting that even if CA were the most widely used method, this does not imply that CA is the best method for the early Paleogene]

By focusing on a method that has been widely criticized – and further is not the most widely used NLR-proxy used in existing studies of the PETM and EECO – the authors create problems rather than solve them. My recommendation (as argued below) is that CA should not be the focus of recommendations. Grimm and Denk (2012) and Grimm and Potts (2016) posed useful concerns about the analytical approach that underpins the 'coexistence approach' as employed by the PALAEOFLORA group. Principally CA's reliance on single taxa to define upper and lower limits.

Grimm directs deep time climate researchers towards statistical methods such as CRACLE (Harbert and Nixon 2015) and an earlier iteration of my bioclimatic approach (Greenwood et al. 2005; see also Ballantyne et al. 2010) where we took a more objective nice succinct summary of these concerns and the more appropriate probability based approaches is given by Hyland et al. (2018).

In essence the concerns are that:

1) genera that are low species richness today (e.g., Ginkgo, taxodioid Cupressaceae such as Glyptostrobus, Metasequoia and Sequoia, etc) may be restricted today in their climate range for reasons other than climate, and may have occupied wider (or just different) climatic spaces in the Paleogene when they were clearly far more geographically widespread and so potentially more ecologically varied than today – this argues against using single taxa such as Ginkgo or Metasequoia (both of which are commonly present in LP, PETM and EECO macrofloras) to define hard limits;

2) that climatic tolerance may have evolved in some lineages – e.g., palms are often used in Paleogene hyperthermal studies to constrain winter temperatures, however pre-Eocene palms likely didn't include the most cold-tolerant subtribes present today as they were a late Eocene or even Miocene radiation (Reichgelt et al. 2018 and references therein);

3) issues with identification of the NLRs – plant taxa may show the same leaf or pollen morphology in Eocene fossils as today, but come from plants that were phylogenetically sister taxa – i.e. not the same genus – with morphologically different flowers or other plant organs to the NLR, so the climate tolerance of the fossil's NLR may be a nonmatch – e.g., Platanaceae and Betulaceae leaves and pollen present in many North American floras may show 'Platanus' or 'Carpinus' leaf or pollen morphology, but are from plants with extinct taxon flowers and other organs (Macginitia and Palaeocarpinus); methods like CA are inherently subjective because 'outliers' – e.g., those relictual taxa noted in point 1, or that 'sit outside' the climate range of the majority of taxa – are arbitrarily excluded from the CA, or arbitrarily used to set a limit for the range of possible estimates;

[We agree with these concerns and note that point 3 is an issue with all NLR approaches, not specific to CA. We have revised the text substantially in accordance with these comments and removed the focus on CA.]

4) in the past there were concerns about where NLR climate range data were coming from – a lack of standardization of data sources and data quality control.

Due to all of the above concerns, a number of authors have adopted Grimm's recommendations or independently came to a similar solution (e.g., Ballantyne et al. 2010). For example, the mathematical approach developed in R by my colleague Tammo Reichgelt – a method with comparable assumptions to CRACLE – where we use probability density functions and Monte Carlo runs to select the highest probability climate estimate based on the whole suite of fossil taxa's NLRs, including taxa at different taxonomic ranks (Family, tribe or genus) to allow for differing degrees of confidence of matching fossils to an extant NLR.

[Use of probability density analysis solves several issues with NLR, and we now emphasise the advantages of this approach in our revision.]

Furthermore, Greenwood et al. (2017), Hyland et al. (2018) and Reichgelt et al. (2018) and others derive the NLR climate range data from the international online portal GBIF.org which warehouses distribution data from most of the world's university, museum and government herbaria and equivalent, with clear policies on data quality and ownership, so verifiable data. The GBIF records are not without bias (geographical gaps or low data density for reasons of local and national politics or economics – rich countries lots of good data, poor countries not so much), but they do constitute the best available, i.e. best practice. Climate data is queried using either WorldClim (Hijmans et al. 2005; Fick and Hijmans 2017) or one of its comparable global climate surfaces (interpolated met station data and digital elevation model coupled with some atmospheric physics to fine tune effects of slope, aspect, continentality etc).

I would recommend the authors consider advocating as a standard that NLR methods use GBIF for the source of their extant taxon distributions, with a mathematical climate surface such as WorldClim (Hijmans et al. 2005; Fick and Hijmans 2017) as the analytical procedure to derive the climate range data.

[The quality of the NLR database is of vital importance for the accuracy of the climate estimates, and many of the existing paleodatabases have a strong regional bias and unclear

policy in regard to ownership and data quality control. We are grateful that the reviewer raised this issue, and fully agree that GBIF and WorldClim are currently the best options available. We have amended the text accordingly.]

Finally, a suggestion. In Ballantyne et al. (2010), Greenwood et al. (2017) and Lowe et al. (2018) we employ what we call a consensus or an ensemble approach, where multiple terrestrial climate proxies – including both geochemical and paleobotanical methods in some instances – are combined and assessed using probability density analysis. We think this is the solution to the problem of competing proxies; query the data to see where they overlap and are most consistent. I would invite the authors to consider recommending such an approach.

[Yes, multiproxy approaches where possible should be advocated. We touched on this in the first sentence of section 5.2.4 but with insufficient direction. We have added a statement here and in 5.3.4 directing to the examples that Reviewer 1 gives above.]

3. Technical corrections.

p. 4, line 5, typo: Paleogene, not Paleogone

p. 4, line 15, omission: cite Greenwood and Wing (1995) as the LAT global N & S hemisphere latitudinal compilation predecessor of Huber and Caballero (2011).

p. 4, Line 32, typo? I would think ‘end-member’ rather than without the hyphen.

p. 5, line 24: cite also Eldrett et al. (2014) as well as Suan et al. (2017) as these authors document from terrestrial plant palynomorphs this same point for the PETM at high northern latitudes.

p. 22, lines 9-34: I would like to see mention here the concerns expressed by Eberle et al. (2010) on a possible seasonal bias in Arctic TEX86 reconstructions.

[All corrections accepted]

4. References cited. [all added]

Ballantyne, A.P., Greenwood, D.R., Sinninghe Damsté, J.S., Csank, A.Z., Eberle, J.J., and Rybczynski, N. 2010. Significantly warmer Arctic surface temperatures during the Pliocene indicated by multiple independent proxies. *Geology*, 38 (7): 603–606, doi:10.1130/G30815.1

Eberle, J.J., Fricke, H.C., Humphrey, J.D., Hackett, L., Newbrey, M.G. and Hutchison, J.H. 2010. Seasonal variability in Arctic temperatures during early Eocene time. *Earth and Planetary Science Letters*, 296(3-4): 481-486.

Eldrett, J. S., Greenwood, D. R., Polling, M., Brinkhuis, H., and Sluijs, A. 2014. A seasonality trigger for carbon injection at the Paleocene–Eocene Thermal Maximum. *Clim. Past*, 10: 759–769, <https://doi.org/10.5194/cp-10-759-2014>.

Fick, S.E. and Hijmans, R.J., 2017. WorldClim 2: new 1°AR km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12): 4302-4315.

GBIF (GlobalBiodiversityInformationFacility). Open Access Biodiversity Data, available at: <http://gbif.org>

Greenwood, D.R. and Wing, S.L., 1995. Eocene continental climates and latitudinal temperature gradients. *Geology*, 23(11): 1044-1048.

Greenwood, D.R., Basinger, J.F., Smith, R.Y. 2010. How wet was the Arctic Eocene rainforest? Estimates of precipitation from Paleogene Arctic macrofloras. *Geology*, 38(1): 15–18, doi: 10.1130/G30218.1

Greenwood, D.R., Keefe, R.L., Reichgelt, T., and Webb, J.A. 2017. Eocene paleobotanical altimetry of Victoria's Eastern Uplands. *Australian Journal of Earth Sciences*, 64(5): 625-637.

Grimm, G. W. and Denk, T. 2012. Reliability and resolution of the coexistence approach- A revalidation using modern day data. *Rev. Palaeobot. Palyno.*, 172; 33–47.

Grimm, G. W. and Potts, A. J. 2016. Fallacies and fantasies: the theoretical underpinnings of the Coexistence Approach for palaeoclimate reconstruction. *Clim. Past*, 12: 611–622, <https://doi.org/10.5194/cp-12-611-2016>

Harbert, R. and Nixon, K. 2015. Climate reconstruction analysis using coexistence likelihood estimation (CRACLE): A method for the estimation of climate using vegetation. *Am. J. Bot.*, 102: 1277– 1289.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas, *Int. J. Climatol.*, 25: 1965–1978.

Hyland, E.G., Huntington, K.W., Sheldon, N.D. and Reichgelt, T., 2018. Temperature seasonality in the North American continental interior during the Early Eocene Climatic Optimum. *Clim. Past*, 14(10): 1391-1404.

Lowe, A.J., Greenwood, D.R., West, C.K., Galloway, J.M., Reichgelt, T., and Sudermann, M. 2018. Plant community ecology and climate on an upland volcanic landscape during the Early Eocene Climatic Optimum: McAbee Fossil Beds, British Columbia, Canada. *Palaeogeography, Palaeoclimatology, Palaeoecology* 511: 433-448, doi:10.1016/j.palaeo.2018.09.010

Peppe, D.J., Royer, D.L., Cariglino, B., Oliver, S.Y., Newman, S., Leight, E., Enikolopov, G., Fernandez~AR~ Burgos, M., Herrera, F., Adams, J.M. and Correa, E., 2011. Sensitivity of leaf size and shape to climate: global patterns and paleoclimatic applications. *New Phytologist*, 190(3): 724-739.

Pross, J., Contreras, L., Bijl, P.K., Greenwood, D.R., Bohaty, S.M., Schouten, S., Bendle, J.A., Röhl, U., Tauxe, L., Raine, J.I., Huck, C.E., van de Flierdt, T., Jamieson, S.S.R., Stickley, C.E., van de Schootbrugge, B., Escutia, C., Brinkhuis, H., and IODP Expedition 318 Scientists. 2012. Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. *Nature*, 488: 73–77, doi: 10.1038/nature11300

Reichgelt, T., West, C.K. and Greenwood, D.R. 2018. The relation between global palm distribution and climate. *Scientific Reports* 8(1): 4721, doi: 10.1038/s41598-018-23147-2

Smith, R.Y., Basinger, J.F., and Greenwood, D.R. 2012. Early Eocene plant diversity and dynamics in the Falkland flora, Okanagan Highlands, British Columbia, Canada. *Palaeobiodiversity and Palaeoenvironments*, 92(3): 309–328, doi: 10.1007/s12549-011-0061-5

Suan, G., Popescu, S.M., Suc, J.P., Schnyder, J., Fauquette, S., Baudin, F., Yoon, D., Piepjohn, K., Sobolev, N.N. and Labrousse, L. 2017. Subtropical climate conditions and mangrove growth in Arctic Siberia during the early Eocene. *Geology*, 45(6): 539-542.

Sunderlin, D., Loope, G., Parker, N.E. and Williams, C.J., 2011. Paleoclimatic and paleoecological implications of a Paleocene–Eocene fossil leaf assemblage, Chickaloon Formation, Alaska. *Palaios*, 26(6): 335-345.

Wilf, P., Wing, S.L., Greenwood, D.R. and Greenwood, C.L., 1998. Using fossil leaves as paleoprecipitation indicators: an Eocene example. *Geology*, 26(3): 203-206.