

## Reply to reviewer 1

Jonkers and co-authors give an update on the PALMOD data base of marine paleotracers for the last 130 k. This is a necessary and significant update for the PALMOD data base. The previous version was limited to  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$  from benthic foraminifera, Jonkers et al. now include planktic foraminifera stable isotopes, and other proxies such as Mg/Ca and carbonate and biogenic silica content. They also include temperature reconstructions for each of the sites. The authors claim to have merged age models with proxy data, assigning an age value to each downcore sample. This is a useful update, which saves users from the necessity to interpolate the age models to the data depth scales.

We would like to thank the reviewer for their positive evaluation of version 2 of the database.

I am worried by the presentation of the data. The authors choose two formats: R and LiPD files. However, the way they have structured the data makes it very hard to look at it at a glance. For R (.RDS) files R needs to be used. I am an advanced python programmer, and I wasn't able to quickly access the data. Both in R and LiPD the data sets of each coring sites were saved without column names, and I find no easy way to see a depth, age, proxy list on screen. I thought LiPD would be easier, since I know that inside a LiPD folder the data are saved as .csv, however, the problem is the same: There is no reference at each file to know what I am seeing. Of course, it could be that I am not knowledgeable enough to open these files. But I am myself a data person, so if I had a problem accessing the data base, it is reasonable to think that many other users will have issues too.

Thank you for this important feedback. From experience with the previous version of the database we learned that our prime user base works with R, but we acknowledge that this limits the use of the data with different softwares. For that reason, we also provide a LiPD version of the database that can be read using a wide range of softwares (including Python).

Palaeo(climate) data are inherently complex given the high demands on metadata and contextual information and the need for age-depth models (which are both uncertain and likely to change in the future). We fully acknowledge that this complexity may render reading the data somewhat difficult or non-intuitive. For this reason, we paid special attention to clearly describe the data structure and we provided links to online resources to facilitate reading and using the data. In the revised version we will update the links following the comment by Julien Emilie-Geay and add additional detail in the text.

Perhaps the issues with reading the data arose because of conversion from the specific R format to a Python readable format or failure to make use of the python utilities to access the LiPD files. Both formats are structured similarly and the data are provided separate from the metadata (analogous to a netcdf file where the variable information is provided separately from the data and separate commands need to be used to access the metadata and the data). Importantly, the R-formatted data do contain column names (see screenshot below).

```

Console Terminal Background Jobs
R 4.5.1 ~ /UniNextCloud/PalMod/R_database/database_scripts/PALMOD/ ↗
> site <- "GeoB3359_3"
> dat <- readRDS(paste0("/Users/lukas/UniNextcloud/PalMod/PalMod_data_QC/PALMOD_v2_RDS_with_age/", site, "_age.RDS"))
> head(dat$dat)
  meanAge_kaBP Age_kaBP_2.5percentile medianAge_kaBP Age_kaBP_97.5percentile depth_merged Age_ka_BP CaCO3_percent TOC_percent bSiO2_percent
1          NA              NA              NA              NA              NA          0.03      7.156          1.40          1.967          4.8
2          NA              NA              NA              NA              NA          0.08      7.400          1.50          1.892          NA
3          NA              NA              NA              NA              NA          0.13      7.644          1.27          1.895          5.1
4          NA              NA              NA              NA              NA          0.18      7.889          1.84          1.809          NA
5          NA              NA              NA              NA              NA          0.23      8.133          1.06          1.853          5.4
6          NA              NA              NA              NA              NA          0.28      8.378          1.25          1.678          NA
  SST_12_degC UK37_prime recal_UKSST_2.5 recal_UKSST_50 recal_UKSST_97.5
1      17.5      0.6215      14.08580      16.77813      19.57781
2      17.4      0.6182      13.96192      16.71961      19.47067
3      17.5      0.6215      14.01745      16.77948      19.56596
4      17.6      0.6248      14.14173      16.89350      19.62375
5      17.6      0.6248      14.17981      16.90338      19.63713
6      17.7      0.6281      14.30275      17.00807      19.76461
>

```

I recommend the authors to re-include netcdf files of each coring sites as part of the data base (they were included in the first version of the PALMOD database). These files are more universal than LiPD and R files, and readable by different software types. In addition, I recommend the authors to produce a more human-readable version of the data base. These could just be the csv files inside the LiPD directories, if we have in each file explicit information of what each column is. Otherwise, I am sad to say that this important data product will be useful for just a very few R and LiPD experts.

We appreciate the reviewers' concern about the complexity of the data, but unfortunately inclusions of rich metadata, chronological information, documentation of the age-depth modelling approach and posterior age-depth models and seawater temperature estimates render some complexity unavoidable. From our experience with version 1, we are confident that, with the documentation in the article and the online resources, the data synthesis is useful for a wide audience.

In the original manuscript we did not mention that the palaeodata and chronological information can be downloaded as human readable csv files on the lipdverse website. We will make this clear in the revision and we hope that this addresses the concerns of the reviewer.

However, as

The netcdf format is not specifically intended to accommodate the complexity of palaeodata and conversion to this format leads to loss of information (also mentioned in Jonkers et al. (2020)). In particular, the metadata won't be as easily queryable and there is no standard way to include the seawater temperature ensembles. The reason why netcdf files were included in version 1 was to ensure readability with the custom made software PDV (Langner and Mulitza 2019), but PDV is no longer actively maintained and as of yet no longer works on all platforms. Thus, given the availability of the data in the universally readable json (lipd) format as well as existing online resources to interact with the data, we see no reason to offer a netcdf version of the synthesis too.