

Author Comment to 'Review' in RC2

General Comments

1. Improvable Data References:

References for the source data should be added. As part of the terms of use for CMIP5 data, apart from the existing acknowledgement, data collections should be referenced in the article's body and cited in the reference list. For CMIP5 references could be found on the IPCC DDC web page for the AR5 Reference Data Archive: http://www.ipcc-data.org/sim/gcm_monthly/AR5/Reference-Archive.html

Thank you for making us aware of more specific references that we could use to acknowledge the modeling groups and their specific experiments. We will extend Table 2 to include the experiment and model specific references.

2. Improvable provenance information:

CMIP5 datasets used should be specified using the full DRS (Data Reference Syntax) including their versions and tracking_ids. All additionally needed datasets like 'psl' and 'ts' (p.3 l. 30) should be specified. An additional table is suggested.

We will add the the original 'version_history' and 'tracking_id' global attributes of each of the models for both the lgm and piControl runs to the respective variable in our dataset. The CMIP variables 'psl' and 'ts' were used for the re-referencing to 10m height. We will specify their names in the manuscript, such that it is clear which CMIP variables were used for the re-referencing.

3. Question: Why did the authors not use the 3D CMIP5 datasets?

The authors state that only the selected four models provided all the required variables. Could the authors explain why they did not use the 3D fields of the model output but based their study on the interpolated (post-processed) surface variables? The 3D fields were provided by more modeling centers. As the authors apply a vertical interpolation to 10 m height for some variables, the direct model output seems to be better suited as source data. Moreover, the number of models could be increased, on which the data is based. These 3D variables are e.g. 'ta' instead of 'tas', 'hus' ('huss') or 'so' ('sos'). Especially the sea surface salinity anomaly, which is currently based on only two CMIP5 datasets (p.5 ll. 24/25), will become more reliable.

The atmospheric 3D fields of CMIP/PMIP are provided on pressure levels. The geopotential height corresponding to the atmospheric pressure level closest to the earth's surface (the 1000 hPa pressure level) is generally ~150-200m above the ocean (CMIP variable 'zg'). The use of the 3D fields on the level of the air-sea interface would thus mean an extrapolation over many tens of meters height for all of the atmospheric variables in our dataset. We expect that the native 2 and 10 meter fields calculated by each model (internally consistent) will provide a better approximation of the near-surface state. The re-referencing of temperature and specific humidity to 10 m done by us is a minor correction compared

to this. The choice of surface fields indeed forces us to reduce the number of models we can base our data on. We noticed however that we can add the GISS-E2-R model (which was left out before because we selected r1i1p1 ensembles only) to our dataset. We will add the GISS-E2-R output to our dataset, which will make all atmospheric variables be based on five models instead of the previous four.

For sea surface salinity, the use of the ocean 3D field would indeed be an improvement. It would make it possible to use four of the five atmospheric models (namely CNRM-CM5, GISS-E2-R, MIROC-ESM and MRI-CGCM3). We decided in addition to calculate the monthly climatological sea surface salinity based on the 'mon' 3D salinity (so) output for IPSL-CM5A-LR, as we have not found the 'monClim' IPSL-CM5A-LR data for piControl variable 'so'. In this way, all variables in a new version of the dataset are based on the same 5 models.

4. Reuse of the data:

As the paper on the 'CORE forcing fields' is cited as reference for the usefulness of the chosen spatial-temporal resolution of the provided datasets for common ocean-only model runs, it should be made accessible (e.g. on Zenodo) if possible. Alternatively, have the authors used the datasets for the forcing of a second ocean-only model run to show the reusability of the dataset, yet?

The Large and Yeager (2004) report is publicly available at the NCAR/UCAR "OpenSky"-repository: (<http://opensky.ucar.edu/islandora/object/technotes:434>). We will add this URL to the reference. Experiments with different versions of the CORE forcing data have been done by many modelling groups before – see for example Griffies et al. (2009). The CORE forcing is also used for the Ocean Model Intercomparison Project (OMIP) within CMIP6 (Griffies et al., 2015).

5. Further reuse of the data:

The authors state that data users could adjust the data using the spread of the CMIP5 model results (p. 6 ll. 29-31). Then the authors need to provide this information in their data.

We understand this can be a valuable additional variable for the potential user, and will add it to the dataset for all variables.

Specific comments

6. Please delete 'CMIP-type' as additional characterization of complex fully coupled models, as it is unclear what that means and it does not add information.

We will remove this wording.

7. Data files do not contain any history of the applied commands. cdo writes information on the applied commands into the global attribute 'history'. This provides useful information about dataset creation. Why is that not in the file?

The authors wanted to provide clean files to the user alongside the detailed procedure description in the manuscript. As the reviewer points out, the exact procedure is indeed saved in the history (by both CDO and NCO) and could be useful for the potential user. We will keep the full history of all files as well as their appended file history as a global attribute in an updated version of the dataset.

8. Data files could include more information not only on the above-mentioned history but also on the methodology. I suggest, the authors add the data doi as a reference to the global attributes, which leads the data user to the doi page with further information.

We will add a global attribute 'references' with the ESSDD manuscript DOI and the NIRD Research Data Archive dataset DOI. As the NIRD Research Data Archive does not allow reservation of DOI's, we can not know before publishing the dataset online what the DOI of the dataset will be. However, we are able to refer to the previous version and state that the user should check for newer versions of the dataset. Similarly, we can not know the final DOI of an ESSD article, so we can only refer to the ESSDD article.

9. Why was the unit of precipitation_flux changed from the NetCDF/CF recommended and within CMIP5/PMIP3 used kg m⁻² s⁻¹ to mm/day? The unit should not be changed if not required.

The 1979-2000 GXGXS Precipitation Climatology employed for the CORE forcing is in mm/d (Large and Yeager, 2004), and the authors wish to present a dataset that can be easily used in combination with a models' original CORE forcing. We will clarify this in Sect. 3.4 and will provide the conversion factor (86400) used such that the units can be set back to kg m⁻² s⁻¹ without data loss. The conversion makes the units of our variable pr deviate from the CF-1.6 convention, which we will note in the manuscript.

10. The provided datasets same as the CMIP5 datasets should comply with the NetCDF/CF conventions. This seems to be the case, though I did not check it. Then the version of the convention should be specified in the global attributes as described at: <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html>.

We will use one of the available online CF compliance checkers to make sure that the dataset follows NetCDF/CF conventions, and will add the version of the convention as a global attribute in the updated dataset.

11. The authors should add a sentence on the relation of PMIP3 and CMIP5 (PMIP4 and CMIP6 resp.) for readers less familiar with these large intercomparison projects.

We will clarify the terms PMIP and CMIP in the introduction of the manuscript.

12. Is there a reason why the current version 1.9.7 of the cdo package was not used but the old version 1.7.0? Moreover, on the cdo's page 1.7.0 cannot be downloaded (<https://code.mpimet.mpg.de/projects/cdo/files>). The authors should consider using the current or a more recent version of the cdos.

The somewhat older 1.7.0 version of CDO gives to our knowledge no different results than later versions for the functions we applied to make our dataset. We however see that the use of the most up-to-date version of CDO is desirable, and we are able to use CDO version 1.9.3 on our systems to remake the dataset. We did a test, and see no differences in the result when using 1.9.3 as compared to version 1.7.0.

Technical Corrections

13. CMIP stands for *Coupled* Model Intercomparison Project.

Thank you for noting this mistake – we will correct it in the manuscript.

14. 'lgm' and 'piControl' are the CMIP5 experiment acronyms. It is confusing and unnecessary to introduce the additional acronyms 'LGM' and 'PI' for them.

We chose to use LGM and PI for readability, but as this causes confusion we will remove these acronyms from the manuscript.

References

Griffies, S. M., A. Biastoch, C. Böning, F. Bryan, G. Danabasoglu, E. P. Chassignet, M. H. England, R. Gerdes, H. Haak, R. W. Hallberg, W. Hazeleger, J. Jungclaus, W. G. Large, G. Madec, A. Pirani, B. L. Samuels, M. Scheinert, A. Sen Gupta, C. A. Severijns, H. L. Simmons, A. M. Treguier, M. Winton, S. Yeager and J. Yin, 2009: Coordinated Ocean-ice Reference Experiments (COREs). *Ocean Modelling*, 26, 1–46.

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., Chassignet, E. P., Curchitser, E., Deshayes, J., Drange, H., Fox-Kemper, B., Gleckler, P. J., Gregory, J. M., Haak, H., Hallberg, R. W., Heimbach, P., Hewitt, H. T., Holland, D. M., Ilyina, T., Jungclaus, J. H., Komuro, Y., Krasting, J. P., Large, W. G., Marsland, S. J., Masina, S., McDougall, T. J., Nurser, A. J. G., Orr, J. C., Pirani, A., Qiao, F., Stouffer, R. J., Taylor, K. E., Treguier, A. M., Tsujino, H., Uotila, P., Valdivieso, M., Wang, Q., Winton, M., and Yeager, S. G.: OMIP contribution to CMIP6: experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project, *Geosci. Model Dev.*, 9, 3231-3296, <https://doi.org/10.5194/gmd-9-3231-2016>, 2016.