

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2022-218-RC1>, 2023
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-218

Anonymous Referee #1

Referee comment on "Using the COAsT Python package to develop a standardised validation workflow for ocean physics models" by David Byrne et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-218-RC1>, 2023

This manuscript appears to describe a python package that can be used to compare models with observations and quantify the differences in a reproducible way. It gives a very nice example of this comparison using two versions of the NEMO model. There are some fantastic moments in this manuscript, for example the section about vertical interpolation options on line 281 is wonderful. One of my struggles in reviewing this paper has been that it is not clear if it is a documentation paper for the COAsT python package. I do not wish to review the whole of the COAsT package because its goals are a bit nebulous. The scope of this paper is narrower and makes more sense to me, but I still feel that the manuscript could more clearly state that the COAsT python package contains other tools that are not useful for comparing models with observations in this way.

The programming decisions here make sense in the operational context in which this paper was written: a context in which many versions of the same coastal model are being run, with the goal of improving its realism, and in which the same operations are being performed many times. The list in the introduction to the paper reads like a list of generic values that are important for all scientific code. I hope that it can be better tuned to make it clear why particular choices were taken. The software framework described here puts the data in a very specific format: this specific format is an advantage in this context, and the goal of this work is not to write general code for comparing any model with any observations.

I recommend that the authors make significant changes in response to my comments. Some updates to the documentation of the COAsT package may also be appropriate.

Major comments:

1. On first read, it was not clear to me that using classes like the "Gridded" class had real benefits. It seemed to me that this data could simply be stored as an xarray dataset, and the relevant dimension names could be input into any plotting or calculation functions. I eventually realized that if you were performing similar operations multiple times, putting all of this information into an object where the details are abstracted away from the user probably reduces errors. But I didn't understand that until I had gone away and thought about it a lot. Please rewrite the beginning of the paper to emphasize this and any other advantages of classes that I may have missed.

Perhaps this is the same point, but I was confused by the sentence "By providing a middle layer into the workflow, it is much simpler to apply the analysis technique to multiple data sources, to share it with others and to expand upon it in the future." I do not understand what "providing a middle layer to the workflow" means, and I would like to understand more about why classes were chosen for this task.

2. I can't find any examples of this python package being used on gridded datasets that are not based on NEMO. It is fine if this package (and hence framework) is actually mainly designed for NEMO data, but then the paper should clearly state this. If this package will be applied to other gridded datasets, I'd like to see a discussion of how different kinds of data (netcdf, zarr, binary) could be read by the package, e.g. via xarray. Lines 30-37 say some really important things about the need for lazy loading, but it's no good having lazy loading if I have to rewrite all my data in a different format in order to even load it into the package.

In addition, I am not sure that this package makes full use of lazy loading. If the data is in netcdf format with no use of kerchunk indices, then you must load the whole netcdf file in order to access the data. The dask tutorial on the COAST website is a bit lacking here. Certainly computation can be delayed and some parallelization should be possible because the objects are based on xarray, but again it's not clear why building these new classes is helpful, because the user has to use xarray/dask in order to parallelize anyway. Why not just use xarray objects directly?

3. It seems to me that COAST is a bit of an "everything but the kitchen sink" package at the moment. Having an expandable code base is nice, but xarray already exists and some more clarity on the goals of COAST would certainly help people to understand what is going on here.

4. If this manuscript is meant to document the python package (and the first half of the manuscript suggests that it is), then I'd like to see a significant discussion of testing. Part of having an expandable code base is having well-designed tests. I see the package has some testing set up. Good code coverage is also necessary for the testing, so that untested code isn't constantly being added.

5. I like the Matched Harmonic Analysis section, but I'd like to see a bit more context at the beginning. What is the overall goal of the comparison?

6. I was not able to understand the description of CRPS provided between line 290 and 299. Please provide more detail on what F_x and y represent.

7. The code actually used to make the figures presented here does not appear to be available anywhere (potentially it is located somewhere in the package, but its location is not given). For a paper that talks about reproducibility, I think that the plotting code should at least be provided. Ideally the datasets used to generate the figures would also be made available, but I understand that they might be too large.

Minor points:

1. I'd like to see some citations for technical concepts like lazy loading, chunking etc. I know that traditional references for these concepts may not be available, but I think non-expert readers would benefit from some references.

2. I would also like to see more citations for concepts introduced between line 145 and line 165. e.g. "the estimation of tides is a vital step for the validation of sea surface height in our regional models" Please reference an example. "Non-tidal residual signals can be generated by many processes but in coastal regions the modest (sic) significant are generated by atmospheric processes". How do we know this?

3. Figures 1 and 3 have colorbars with white in the middle. I would recommend choosing a

different colorbar so that we can see all the observations.

Typos:

1. Line175: "quick and easy" should be "quickly and easily"