Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-257-RC1, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

Geoscientific Model Development Discussions

# *Interactive comment on* "Parallel I/O in FMS and MOM5" *by* Rui Yang et al.

**Nikolay V. Koldunov (Referee)**

koldunovn@gmail.com

Received and published: 13 November 2019

The paper describes the implementation and tuning of the parallel I/O in MOM.

This is a very good and very timely study, as the parallel I/O continues to be one of the major problems in the current Earth System Science codes. This type of description is usually ether never put together as a text or in a best-case scenario just "collect dust" as a technical report. The authors did a great job of describing in detail their technical development as a paper, and I wish there are more such descriptions in the future. The sharing of this information is very important so that the progress in the field is faster.

I am in general happy with the paper, while having a couple of suggestions that authors free to agree or disagree with. Paper can be published after minor revision.

General points



To make the paper even more useful it would be nice to discuss several additional details.

Short description of how hard it was to implement parallel I/O using each of the libraries (maybe person/month estimate?), what is the user experience with each of the libraries (are they easy to install and support?).

Mentioning another parallel I/O solutions, that become popular in Earth Science (e.g. XIOS http://www.ifremer.fr/docmars/html/doc.coupling.xios.html ) or even something outside of the ocean modelling world (e.g. https://csmd.ornl.gov/adios ) would help the unexperienced reader to be more aware of available software solutions.

Maybe you can speculate about the applicability of your results to unstructured mesh ocean models, that usually store their results in netCDF as long 1D vectors?

Your data-intensive benchmark, although it serves the purpose well, is not very realistic. I think your results will shine even more if you can show how beneficial parallel I/O is in realistic simulations. In Koldunov et al., 2019 we showed that in our case for relatively small setup (about 600 000 surface points) running on 1152 cores the price of the serial I/O in "operational" simulation is only about 5%. For the user that typically has tasks of this size it is not a very large price to pay, and maybe investments in the parallel I/O are not necessary. It would be great if you can run, say, a year of model simulations with typical I/O workload (e.g. in our case its monthly means) on different number of cores with serial and parallel I/O and estimate the amount of time (in %) the I/O takes from the total run time.

Minor point

For figures 3 to 7 please add PE/node as a second x-axis (e.g. on the top). This will make it easier to interpret.

Nikolay Koldunov

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-257, 2019.