

YAC 1.2.0: New aspects for coupling software in Earth system modelling.
M. Hanke, R. Redler, T. Holfeld, and M. Yastremsky.

This version of the paper is significantly improved. The paper's focus is much improved

and many problems have been addressed. However, the authors did not adequately address the major shortcoming noted by both reviewers in the original paper, lack of performance data. In addition, there are a few new issues.

- Section 4 has several spelling and grammatical errors, that section should be reviewed carefully

 - measurement -> measurement

 - "The performance of the exchange step itself was in the tested setup good for all node counts." needs to be written more clearly

 - "The measurements suggest that the time for the exchange step is mainly influenced by external factors like network utilisation of other applications running at the same time." is something I think is incorrect.

- I think the new section discussing the remapping cost is probably not correct and needs to be updated. That's page 15, lines 12-16. The authors state performance was random and fell over 2 orders of magnitude. I believe the performance should vary with nodes/cores and timing numbers should be reasonably reproducible.

 - My sense, based on the statement in the paper, is that the timing may need to be redone, with appropriate barriers and/or a larger sample size. You cannot measure one exchange in these kinds of tests, the time required is too small and is often shorter than the resolution of the timers.

 - The point is that over the course of a long coupled run, coupling is done thousands or millions of times and it's that aggregate time that's important.

 - At high resolution or high pe counts, the coupling time can be a bottleneck.

 - To get an accurate representation of the time, the ping pong test should couple data back and forth for at least a few wall-seconds in total, maybe 1000s or 100s of thousands of times. I would like to see a scaling curve, similar to Figure 3 for the ping pong test for a couple of resolutions that accurately reflect the communication/remapping cost.

In addition, I think the time for remapping and the time for coupling could be separated in the presentation of the data. It would be interesting to know how much of the ping pong time is spent remapping (including halo update and application of mapping weights) and how much is spent in the concurrent transfer of data. There are cases when remapping is not needed.

- There is a set of timing data called "file" on figure 3 but very little information is provided in the text how this is implemented (is it parallel I/O or "read and scatter" or something else) and there is no accompanying discussion. The results are interesting and I feel I am left hanging. I think some discussion needs to be added to discuss that set of data. This is interesting data.

- Both reviewers felt the paper fell short on performance data and the paper has not substantially improved in that area. As indicated above, some scaling curves for remapping and data transfer would make the paper significantly better. I also think additional data should be provided on the weights generation, for additional resolutions and more mapping options. The introduction also talked about how high resolution was the reason fast weights generation was important, but there are no high resolution results in the paper. The authors

suggested moderate resolution is not interesting and no "significant challenge", but then the authors only present results at moderate resolution. I feel this inconsistency needs to be addressed.

- Several of the authors' responses to the reviewers comments did not seem particularly well thought out. Without focusing on that too much, I think there are a few important points that don't necessarily reflect on the current paper but might on YAC.

First, a round robin or similar decomposition can be an excellent decomposition in cases where the cost per gridcell varies in time and/or space, like in radiation calculations (diurnal), sea ice coverage (seasonal), or land cover models (cost can vary diurnally and seasonally by location). This decomposition probably deserves more consideration than "We do not mind not supporting round robin like decompositions"

Second, the requirement to provide connectivity information as part of the weights generation is not state-of-the-art. The author's response, "ICON can do this", "models with diffusion and advection operations will have the information readily available", and "it was too hard in OASIS4" are poor excuses. If the tools in YAC are being designed to work generally, then other models will not be ICON, diffusion and advection operators may implicitly carry the connectivity information but accessing that information to pass to YAC might not be straight-forward. In addition, there is no guarantee that the same halo is required for a diffusion or advection operation and for the YAC weights generation. Finally, the fact that OASIS4 was not particularly successful in computing connectivity does not mean it's not possible. Other weights generation tools are able to compute connectivity on the fly.

In summary, I believe the paper needs to be further revised to

- review section 4 for grammar and spelling.
- show accurate timing data for the ping pong test and if possible, to separate the remapping and coupling times into distinct terms.
- add some discussion of the "file" data on figure 3.
- extend the weights generation performance data further to include several additional mapping options including especially nearest neighbor.
- provide timing results for the ping pong, file reading, and weights generation on at least one high resolution configuration as this is a major driver of the current implementation.