Comment on gmd-2022-214

Peter Düben (Referee)

Referee comment on "Improving scalability of Earth system models through coarse-grained component concurrency – a case study with the ICON v2.6.5 modelling system" by Leonidas Linardakis et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-214-RC1, 2022

This is a good paper addressing the important topic of concurrency that has so far not been explored enough in Earth system modelling. The paper is well written but could be more concise. The theoretical analysis is backed up with convincing performance measurements that show the usefulness of concurrency. The paper will be ready for publication in GMD if the following minor comments are addressed.

More significant points:

- Page 9 seems to be a broken in the published PDF. I opened the document with 3 different PDF viewers, but I never got to see the entire page.
- Overall, the document is longer than it would need to be. E.g. the model descriptions could be shortened as documentations to the models are provided as references. This is a subjective opinion by me, so feel free to disagree, but maybe you can consider making the document shorter.
- Section 3.2: This is a very interesting discussion. However, I am not sure how much of this is new and how much of this is common knowledge in computer science. If it is new, it should be made more prominent (e.g. in abstract and intro). If it is not new, maybe provide references more prominently.
- Figure 5: I am not a big fan of the figure. The left side is supposed to show a serial workflow but is showing two boxes in "parallel". Why not add a time dimension to the figure and streamline the left side into a single line of tasks? Maybe you can also use the nomenclature of Section 3.2 (W,N…)?
- I am missing a bit more of a discussion what happens when the components are not strictly independent. E.g. when running the radiation scheme concurrently to the rest of the atmosphere model. The independence is quite a significant limitation for couple Earth system model components.
- Paragraph starting with L369: These seem to be quite significant limitations of the approach. Could you also provide timing results with GM parametrisation, IO etc. to give the reader an impression how significant the limitations are?

- Can you say anything about the energy cost when using concurrency (or not)? No big problem if not.
- Maybe I missed the information, but do you state somewhere how long the individual experiments were and how many you have run for each performance measure to reduce measurement errors?
- I am not quite sure whether you present results from the old and the new machine at DKRZ since it is useful (2 generations of machines), or just since the machine has changed in the meanwhile. Both would be OK but maybe you can justify somewhere.

Minor points:

L6-7: "The novel…" This sentence is a bit odd and "function parallel technique" is unclear at this stage.

L41: "…cannot efficiently scale" That is quite a statement. Can you provide a reference?

Figure 1: "sea-land mask is in color" ?

L113: Better say "local grid-spacings of 600m"

L158: Maybe it is just me, but I am not sure what a "trophic level" is.

L176: "and may only need… components" Is unclear and should be revised.

L190: "in of"

You could cite work of ESiWACE from a couple of years ago (but no worries if you disagree): https://zenodo.org/record/1453858#.Y0KPyi8Rp9c

L319: I do not understand this sentence and it should be rephrased.

L349: Is there a performance hit when guaranteeing bit-reproducible results? I assume this only holds if the concurrent parts are running on the same hardware?

L377: "No such effort…" I think this is an understatement. Maybe say that those parameters have not been optimised?

L410: si -> is

L474: "does not apply". Well, I guess this is always only an approximation. -> "is a bad approximation due to the unusual super-linear…"

L513: remove "though"