Comment on gmd-2022-171
Malcolm Roberts (Referee)

Referee comment on "ICON-Sapphire: simulating the components of the Earth System and their interactions at kilometer and subkilometer scales" by Cathy Hohenegger et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-171-RC1, 2022

Referee review of Hohenegger et al: “ICON-Sapphire: simulating the components of the Earth System and their interactions at kilometer and subkilometer scales”

General comments

In this article the authors describe a new model ICON-Sapphire that has been run at a variety of very high global resolutions of 10km and finer for coupled and ESM simulations, in order to explicitly represent processes that have previously been parameterised. Model simulations are short, one or multiple years for some of the reference simulations. Some initial evaluation of the model simulations is shown as well and information on model performance is given.

The article is well aligned with GMD in terms of model development and description, and very useful to inform the community of such impressive progress in terms of global model resolution advancements and future prospects for their use in climate simulation.

My main concern is the model evaluation part of the manuscript. I understand that it is difficult, with only one year of coupled simulation (which is not initialised and hence difficult to compare to any particular observational data year), to produce evaluation of the model. However, I have significant concerns about the evaluation shown. In particular with regards to the ocean, it is terribly unclear how the one year of simulation differs from its previous spin-up, and hence what the impact of coupling ~5km atmosphere and ocean
models together is in that one year. Perhaps by showing similar diagnostics at the start and end of the one year simulation, or some differences over that one year period, it would be clearer what the simulation achieves.

I suspect the above problems are enhanced by the imbalance in the top of atmosphere radiation. I understand that this is a “tuning” aspect that has not yet been addressed, but a more explicit showing of this bias (and presumably the inherent surface temperature and other consequent biases) would give a clearer view of the current capability of the model rather than what the model might be able to do in future.

Several of the comments below refer to phrases used such as resolving convection/convective storms. I will not insist, but my colleagues more expert in convection processes than me would point out convection happens all the way to the sub-metre scale, so perhaps you could make this point early on in the manuscript and then use whatever phrase you choose.

Detailed comments

L27: “grid spacing of 10 km would at least permit...” – I think I understand the sense of this, but maybe it could be more explicit. I think you are saying 10 km is enough to use explicit representation of convection (i.e. switch off the parameterisation of convection), just saying permit is a bit imprecise on this point.

L57: Again possibly semantic, but “resolved explicitly” perhaps could be “represented explicitly”.

L83: again, resolving convective storms vs representing them.

Fig. 3: I confess I struggle to follow the logic of the timestepping illustrated with so many lines and symbols.

L225: If I understand, the rainfall over land is not put back into the ocean currently?
Given that you talk about the importance of water and energy cycles, is this not a big problem (perhaps it would be once the model is run for longer)?

L 245: This text suggests (but does not explicitly say) that $Z^*$ is used – perhaps it could be clearer if that is the case?

L260: I guess more details may be in Korn et al, but perhaps a little more detail would be useful on the sea-ice. Is EVP not marginal at these scales in terms of its assumptions? Perhaps this could be mentioned, or else justified, given similar discussion of the atmosphere setup. Also a single category zero-layer thermodynamics model is relatively simple these days and could be noted?

L290: Much emphasis at the start of the paper is the importance of moisture and energy. But it is unclear how well the coupler conserves fluxes of heat and moisture across the interface. You say no correction is applied for global conservation, but do you know how well the model conserves globally and can this be stated?

L305: Does the land surface not need some spin-up too, given the different resolutions of the initial data and the model?

L323: Perhaps this bit could go into an appendix, rather interrupting the flow about the model.
L336: Maybe these numbers could go in a table, they might then be easier to digest?

L357: Similarly there are lots of numbers in this paragraph too, making it hard to absorb.

Table 2: Just to check, for Δz: L, is the vertical spacing over 5 levels giving 5700m on the last level?

L395: Ocean spin-up. I’m not sure I understand this, and I don’t understand the metric of the biases at L402 (global mean?). The biases in a forced ocean model are generally rather constrained by the forcing (certainly at the surface). A figure would help to illustrate what the biases look like spatially, since global mean biases can hide compensating errors.

In addition, can you clarify that the spun-up ocean state uses all prognostics variables – temperature, salinity, currents, sea level height etc. It is unclear, later on, when you compare the simulation with observation, how much is included in the initial state.

Before 4.1.1: perhaps you could say something in this section about the difficulties of assessing models on such short simulations, given issues with spin-up, coupled model shock, lack of TOA balance, lack of directly comparable observation year etc. I think this echoes through the next sections, but is being slightly ignored and I’m sure this is an issue that future groups will grapple with too. As mentioned in the overall comments, a more convincing assessment of the one year of coupled model data (that is demonstrably different from the ocean spin-up, for example) would be very welcome.

To a suspicious eye, it also feels a little that metrics have been chosen where they look
good, and others mentioned but any poor representation is blamed on lack of TOA balance or similar. Could you perhaps better motivate the metrics you show to convince readers otherwise?

L427: but also that you have not spun-up the land surface, it is just taken from initial conditions? Presumably one can expect the land surface to take considerable time to spin-up, e.g. soil moisture etc.

L428: it is a shame that spatial plots are not shown here but referenced to another paper that is not available yet.

L429/430: Can you include a reference for mid-latitude storms dominating meridional transport, convection for vertical transport?

Fig 5: Based on 1 year (1 season) of data. Is this robust – e.g. different runs, or longer runs (I know these have not been done)? Any climate analysis would consider 1 year to be far too little data to base assessment on, also given the TOA. Are you saying that the TOA makes no difference to the large scale circulation? How much is the simulation able to change from the initial conditions in such a short time?

Fig. 6: My looking at the “all points” figure would also suggest a double ITCZ through the year in the model, with ITCZ not fully propagating across the equator.
L460: As noted before, and perhaps it is a small point, but convection happens at all scales. You are resolving the part above your grid spacing, but I think many would argue you are not resolving convection, you are explicitly representing it. As you suggest above with shallow convection and clouds, you are still missing some important processes, which you are choosing not to parameterise.

L465: You state here that the hydrological budget is closed in ICON-Sapphire, but you have not shown that previously. Can you say more about this earlier in the manuscript? Indeed it is not closed in the usual sense, as you do not have the rivers returning fresh water to the ocean, as I understand it.

L466: The low soil moisture, is this the initialisation or the simulation?

L468 & Fig. 7: The spin-up of the ocean is presumably forced by observed precipitation, and (somehow, undefined) I assume the ocean salinity is constrained during spin-up? If so then this figure is a given (is it not, if not then please say why), because in one year the model will not change these large-scale patterns and hence you are mostly showing the spun-up state.

Figure 8: I don’t understand what the zonal mean of the correlation (which is not clearly defined in the text) is meant to show. Why not show spatial maps (as in Wu et al) to demonstrate where the correlations are positive/negative and hence suggest mechanisms.

Figure 9: Can you add the uncertainties for the observations into the table. At the moment it implies much better known values (in 2020 or any other year) that is actually the case, and add to the caption that the observed values are for (historic) periods, not 2020.
L484: Again I have some trouble with this. If the ocean was spun-up with observed forcing, then of course after 1 year it will still look like that state. Can you say anything about what changes in the ocean over the 1 year that would suggest otherwise?

L495: “..which may affect the Indian summer monsoon” – do you have a reference for this statement?

L497: I think I might be able to see one TC path in the North Atlantic in JJA, but they are not obviously visible to me otherwise unless you label them.

Figure 11: Please label the solid and dashed lines in panels e&g. I’m also not quite sure that the d&f panels are meant to convey, without any reference to observations/reanalysis to compare against.

Figure 13: I assume that in (b) the region shown is blown up in size, if so please can you add that to the caption.

Figure 15: I’m struggling to take much from this figure. The precipitation is very noisy in the contours and it is difficult to see the interaction between wind and precipitation as suggested.
L547: Do you mean minimalistic parameterised physics?

L572: Could you plot the 1.25 km driving model data as well. As with the figures above, it is difficult to see that you are demonstrating this model can do something different from other (e.g. lower resolution) models when you are only presenting one figure and no process validation.

L623: As before I’m not convinced that Fig. 19 says very much about the performance of the 10km model. You note that the bias is inherited from the spinup, so I think it would be useful to show the spinup field as well, and/or the difference to the 10km model. You do not show how much this field is reset each seasonal cycle, hence it is unclear what the 10km model is contributing.

**Technical corrections**

L103: ...climate processes are represented physically – perhaps explicitly rather than physically?

L366: Finland

L383: "already now" – perhaps just one or the other, they both mean the same thing.

L447: observations
L588: simulation