This paper describes an huge amount of cutting-edge work. It is well written and I have only minor comments. Basically, I think it is acceptable as-is. That said, I was a bit disappointed with the paper.

Major Comments:

1. Coupled km-scale simulations unlock a lot of interesting questions, but the results shown here seemed more of the "we made pretty pictures" variety. Having written overview papers for new model releases myself, I can commiserate – these papers are a lot of work to write and hard to make interesting to read. That said, I had a couple of questions going into the paper which might be useful to reflect on:
    a. **What features in the coupled system are improved by storm-resolving scales?** Some features can't help but become realistic as they become explicitly resolved. Orographic precipitation is an obvious example. Getting these things right should fix classic problems in coupled models like dynamic vegetation die-off in the Amazon due to precipitation biases or ocean circulation biases due to incorrect bathymetry. Identifying classic biases which you expect your GSRM to get right, then checking whether this happens would be interesting.
    b. **What features in the coupled system are NOT improved by storm-resolving scales**? Do you have a sense for what the canonical problems of GSRMs will be? What did you struggle to get right in your simulations?
    c. **What do we get out of global coupled k-scale models that was missing from prescribed-SST runs or regional simulations?** In this context, it would be nice to see prescribed-SST companion simulations and/or regional simulations.

2. I'm also disappointed by the discussion (particularly section 4.1.1) about:
    a. **Model tuning**: 4 W/m2 is a huge imbalance. I'm confused why you didn't insist on tuning the model better before running these simulations. How can you hope to do multi-decadal simulations with such a large radiative imbalance? I don't expect you to redo the simulations, but acknowledging the problem and explaining why you didn't want to or couldn't tune the model would be interesting. Ignoring the issue leaves the reader feeling like they missed something.
    b. **Model drift**: Fig 4 shows that TOA energy is adjusting rapidly over the course of this simulation. It would be nice to see a similar graphic for global-average surface temperature. Lack of discussion of what this drift means was conspicuously absent from the paper. Do you really think you could do a long simulation with this configuration? Do you have plans for fixing the drift?

3. I think a lot of the ocean analysis is naïve because it doesn't acknowledge that it takes the ocean a long time to drift away from its initial condition. Thus a lot of the analysis is probably more reflective of having an initial condition which looks like observations rather than that your ocean dynamics are working correctly. You can get a sense of initial condition versus equilibrated model bias by comparing the output from a coarse-resolution coupled run at initialization, after 1 year, and after 500 yrs. I bet most of your fields of interest look a lot more like the initial condition than the 500 yr value. I'm not sure this means you should

throw out your ocean analysis, but I do think you need to clearly articulate the potential source of good skill.

4. I'd like to hear more about conservation properties of the model. If I understand correctly, you've designed your schemes to have decent conservation properties and don't have a mass or energy fixer. I'd really like to see a plot of the unexplained global-average water and energy leak over time. This seems to me like it could be a huge problem for your mult-decadal simulation aspirations.

Grammar, spelling, and details:

1. ~L130: it seemed odd this paragraph doesn't include citations for readers to find out more about each of these component models, but then I realized that you go into a lot more detail about each component in following sections. If convenient, citations to the overview papers on each component model would be useful here. If not, please note that details about these models are given in section 2.
2. L135: You say here that the atmosphere can only be run in uniform global or regional modes, but Fig 2 and elsewhere in the text talks about nesting, which seems like it uses several different resolutions in one run. The text also seems to imply that nesting is only available in regional simulations, which seems odd. Why can't you do nested regions inside a global run? Also, I think of telescoping as identical to nesting: you divide each tile of a coarse outer grid into finer but uniform grid cells and run a regional version of the model on this patch of fine-resolution cells. This fits with the idea of a telescope extending in a few discrete segments rather than continuously deforming to extend and retract. I think the way your ocean model works is that resolution is allowed to vary smoothly throughout the domain.
3. Section 2.1: It would be good to mention in this section that aerosols are prescribed. You say this on line 303 in the I/O section, but readers will expect to hear about aerosol treatment in the atmos description.
4. Fig 3: Wow, this plot is cool – it has so much info. I'd prefer if dynamics and transport dots had their own line in the legend since closed and open dots obviously mean something uniquely different. Also, maybe add titles for the left and right columns of the legend since left is atm and right is ocean?
5. Is land seen as just another atm process? It seemed odd that some components coupled via YAC but land doesn't. A sentence explaining why would be useful.
6. I thought Bjorn said at Pan-GASS that you have another turbulence option (Deardorff?) which was unintentionally acting as a shallow convection scheme. Is that worth mentioning here in the same spirit as you mention 2 moment microphysics and RTE-RRTMGP but don't use it?
7. Does ICON include horizontal turbulent mixing, or just vertical? It seems like horizontal mixing will be important at the hectometer scales you run at.
8. L195: citation for Richtmyer and Morton numerical scheme?
9. L246 – unclear what "latter" refers to.
10. L278: What do "processes" refer to here? I think you mean the land model, atmosphere model, etc. I tend to think of these as "component models" with "processes" being

particular physics schemes within a component… but that might be idiosyncratic of me. The concept of "neighboring processes" seems odd since processes have no spatial relationship to each other. I think you mean the process called before or after in sequential time splitting?

11. L281: Doesn't the atmosphere just compute wind stress and provides that to whatever land model wants it? The way it's written, it sounds like the atmosphere provides a different wind stress to sea ice versus ocean.
12. L286: I think you should delete "and" between "wind" and "vectors"
13. L298: It would be handy to point out that 30" is equal to ~900m at the equator.
14. L325: "single-precision 32-bit float arrays are now kept in memory". I think you mean that "output is now stored in single instead of double precision, reducing memory requirements by a factor of 2". "kept in memory" sounds like you mean the data is kept in cache instead of slower-access disk and "single-precision 3d bit" is redundant.
15. L175 – I'm confused how you obtain good performance on GPUs if you use PSrad for all calculations in this paper and PSrad only runs on CPUs – I would have thought having some processes on CPU and others on GPU would result in excessive communication overhead and slow runs. Is radiation running in parallel with other atm processes? Or is it just that you call radiation so infrequently (every 15 min!) that it doesn't matter? I suspect you are forced to run radiation so infrequently precisely because it is on CPU.
16. L339: I think 40 TB/month is for *5 km dx*. It would be useful to point this out and to also say how much storage space you're using for the 2.5 km grid (which I expect is 4x more = 120TB/month!).
17. L369: "half productive" is bad grammar.
18. Adding another panel to Fig 4 showing the annual cycle of global-average surface T would be useful.
19. L447: observations misspelled.
20. L452: I'm unclear how negative TOA radiative imbalance would lead to enhanced radiative cooling. First, what's your sign convention? Does negative radiative imbalance imply that the planet is losing heat? If so, I would think enhanced radiative cooling would *cause* the radiative imbalance. But TOA radiative imbalance could also be caused by an excessive planetary albedo.
21. Fig 11: caption skips panels e and g.
22. Fig 12: panel a and e seem redundant.
23. L554: "my" should be "may"
24. Fig 15: I thought sea breeze was a weak example of 2.5 km resolution since it is also captured pretty well at 25 km resolution. Also, this graphic would be a lot better using wind vectors rather than colors for just zonal wind.
25. L586: "latter" rather than "later"?