We thank the first reviewer for his thorough reading of our manuscript and insightful comments that helped clarify our manuscript and strengthen our validation. For simplicity, we rewrote the reviewer's comment below (in black) and respond to them point-by-point (in blue).

———————————————————————————————————

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.

Thanks.

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.

We understand the reviewer main concern about the model evaluation part related to the ocean very well where, with a one-year simulation, the slow ocean dynamics can a priori not drift too far away from its initial conditions, an issue that we didn't explicitly discuss in the submitted version. The reason we think it is still important to validate the ocean state, even for one year, is that one could already investigate interesting scientific questions just based on one year of simulation, and we thus have to make sure that there is no obvious bias, even in the ocean. Despite the long experience of running ICON at low resolution in a coupled mode (Jungclaus et al., 2021), that the ocean works and couples correctly to the atmosphere is not given. In fact, several bugs were actually discovered during the development phase, bugs related to the momentum coupling between ocean and atmosphere. We thus added in this revised version both the note of caution related to the slow ocean dynamics and the motivation for still validating the ocean at the beginning of section 4.1 on lines 423-431.

Second, based on the suggestions of both reviewers (see below for our detailed answers to the specific comments), we analysed how much of the biases we saw in G_AO_5km are just inherited from the spin-up, or are developing through the simulation. We conducted this analysis for salinity (Fig. 7) and for the barotropic streamfunction and water transport (Fig. 9), see our response to the corresponding reviewer comments below.

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.

The reason we originally kept this part rather short is that there is another paper by Mauritsen et al. 2022 (now published) that specifically deals with this aspect of tuning. But from the comments of both reviewers we now feel that we should say more on this aspect. First, we added the seasonal cycle of surface temperature in Figure 4. Second, we explicitly explained our current strategy to get rid of this imbalance. The TOA imbalance is mostly related to a preponderance of low clouds. These arise when using the Smagorinsky scheme, which has a mixing cutoff. Our formulation sets the eddy diffusivities to zero if the Richardson number is greater than the eddy Prandtl number. This cutoff unrealistically inhibits mixing, both because of well known limitations of the Smagorinsky scheme in simulating the transition to turbulence (Porté-Agel et al., 2000), and because of a failure to incorporate the effect of moist processes. As a result, over cold and moist surfaces, insufficient ventilation of the boundary layer occurs, causing moisture to build up and resulting in excessive low clouds. In ongoing experiments, we have explored adding a small amount of background mixing at interfaces between saturated and unsaturated layers where the equivalent potential temperature decreases upward, mimicking the effects of buoyancy reversal (Mellado, 2017). Low clouds respond sensitively to this background mixing, what provides a convenient control on their amount and on their influence on the top of the atmosphere energy budget. Ongoing work is exploring theoretical justifications for the choice of the background mixing, but it may also be set empirically, as a way to provide a better representation of the statistics of low clouds. We added these considerations on lines 450-461. Finally we partly rewrote the second paragraph of the conclusions to better summarize the current capability of ICON-Sapphire (see lines 715-726). Many investigations do not require long simulations, e.g. investigating processes that control the seasonal migration of the rainbelts, or interactions between ocean-atmosphere and ice as shown in Fig. 14, or effects of small-scale ocean features on large-scale properties of the atmosphere. But it is clear that the current imbalance prevents decadal simulation as the model cools too much. So although the simulations may not be adequate to answer every question yet, we find it important to make the community aware of those simulations as they already allow to investigate aspects of the climate system in a novel light.

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.

Yes we agree. There are different terminologies in use in the literature and different opinions on it. We will use “represent convection explicitly” when talking about convection specifically as, we agree, we don’t resolve the full spectrum of convection, and “resolve storms” as, in terms of storms, we do resolve them. Resolving storms also corresponds to the new terminology in use for atmosphere-only kilometer-scale models which are called (by a part of the community) “storm-resolving”.

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.

Yes we meant that 10 km is enough to use an explicit representation of convection. We rephrased accordingly, see lines 28-29.

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

Changed accordingly (line 58).

L83: again, *resolving* convective storms vs representing them.

For storms (see above), we will keep resolve but we will remove explicitly as it sounds a bit redundant (lines 83-84).

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

We are in sympathy with the reviewer comment but we haven’t found a better representation. Given the technical flavor of GMD, we find it important to show such a figure. Also the figure conveys very well the complex and maybe questionable logic of the timestepping in ICON.

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)?

We agree that this would be a big problem for longer simulations. The reason we turned the discharge off was that all the discharge of one river was happening in one ocean grid cell, which is an incorrect assumption with a grid spacing of 5 km. Moreover, the river reservoirs need to be properly initialized, otherwise too much fresh water ends up in the ocean in the first few months of the simulation. These two aspects have been since then corrected and new ICON-Sapphire simulations can use discharge. We adapted the text to clarify this, see lines 236-238.

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

No z is used, we added this information on lines 256-257.

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?

We agree with the reviewer's intuition that with kilometre and subkilometre scales we approach regimes where some of the standard assumptions of sea-ice modelling may become questionable (see e.g. Rigeisen at al, https://doi.org/10.5194/tc-13-1167-2019). The reason for choosing the EVP rheology in this work was its superior computational efficiency compared to the VP rheology (and this was the primary motivation for developing EVP at all). The question when VP or EVP starts to become invalid is hard to answer. Recent work (Koldunov et al, https://doi.org/10.1029/2018MS001485, Spreen et al https://doi.org/10.5194/tc-11-1553-2017, Wang et al https://doi.org/10.1002/2016GL068696) suggests that the EVP rheology provides at high-resolution of 4.5 km a good compromise between physical realism and computational efficiency. To take the reviewers comment into account, we have changed the corresponding part of the text, see lines 270-273.

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?

Most configurations use the 1-nearest-neighbor exchange of data, as the grid geometry is identical between atmosphere and ocean over the area of exchange. Thus, conservation of quantities by the interpolation is not an issue here.

Conservative remapping is only applied in G_AO_tel. We did not investigate global conservation properties for this particular model configuration. From century long simulations of the coarse resolution model version described in Jungclaus et al (2021), we know that e.g. the water in this model configuration is closed (which includes conservative remapping of evaporation and precipitation). To clarify this, we rephrased the sentence “No correction is applied for global conservation” to “The coupler and our coupling strategy are designed to conserve fluxes”, see lines 301-302.

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

We didn’t spin up soil moisture as it is not clear how to best do this. We cannot run the model over many years to wait until soil moisture equilibrates and starts from this state, as this is computationally too expansive. The other possibility would be to run the land model offline, either forced by observations or by coarse-resolution model output. But in this case, the land surface model would equilibrate to the observation climatology or coarse-resolution climatology, which does not have to be the same as the one obtained if forced by the 5-km atmosphere simulation. We checked the evolution of soil moisture (see Fig. R1 below). Despite the lack of spin-up, four out of the five soil layers are already equilibrated by September in the northern mid-latitudes, and three out of the five layers in the tropics. We added this information in the text on lines 318-321.

L323: Perhaps this bit could go into an appendix, rather interrupting the flow about the model

We agree that this part might interrupt a bit the flow. We shortened it (see lines 340-341), avoiding the description of the CDO operators, a description which is now available in the new release of the CDO documentation (https://doi.org/10.5281/zenodo.7112925).

L336: Maybe these numbers could go in a table, they might then be easier to digest?

We agree and did a table, now Table 2 and correspondingly adapted the text on line344-345.

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

We simplified this paragraph by only talking about numbers in PFlops (rather than PFlops and nodes) and removed unnecessary repetition of the grid spacing, see lines 368-380.

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

5700 m is the thickness of the last layer, 0.065m the thickness of the first layer. This is different in the ocean, where the first layer is thicker than the second layer. We clarified this in the caption of Table 2 (now Table 3).

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.

We rewrote the beginning of this part to better explain how we spin up the ocean, see lines 406-409. Concerning the bias, we agree and rephrased the corresponding lines (see lines 416-420). As shown in Fig. R2 below, the SST is consistently too cold except along 60S, north of 50N and along the western coast of North America. The global mean SST bias is -0.53 K. Since the SST bias does not show too much spatial structure, and since we already have many figures, we didn't include a figure but indicated in the text where the bias is positive/negative. We also do not show a figure for the salinity bias (see Fig. R3b) as the bias tends to be small due to the use of salinity relaxation, except in the Arctic. This is now mentioned on lines 419-420. See also our reply to the comment related to Fig. 7.

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.

We confirm that the spin-up ocean simulation employs all prognostic variables, an information that we added on line 426.

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.

We agree. We added before the beginning of 4.1.1 (lines 421-431) the difficulty of comparing to observations due to the lack of directly comparable observation year, which motivated our use of year-to-year variability in the observations besides the climatological mean in our previous analysis. We also now mention both the difficulty of validating the ocean given the shortness of the integration period as well as the motivation for still doing so. We don't mention the lack of TOA balance yet, as this is a result of the validation. And as described below in response to the specific comments concerning Fig. 7 and Fig. 9, we added a more convincing assessment of the one year of coupled model data versus its spin-up.

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?

We disagree with the reviewer's comment. We looked at obvious basic features of the climate system, namely TOA radiation balance, temperature, large-scale circulation of the atmosphere, precipitation, soil moisture, salinity, ocean circulation and ocean-atmosphere interactions. We believe that our comparison was fair, but what may have led to this impression is that, in the more summarizing sentences of the manuscript, we generally only stated that ICON-Sapphire reproduces basic features of the climate system. We now added in the abstract (see lines 9-10) as well as at the beginning of section 4 (see lines 393-394) that ICON-Sapphire can reproduce basic features of the climate system even though some aspects would require further improvements. We also removed the last summarizing paragraph of section 4.1.1 which only summarized the positive aspects, and partly rewrote the second paragraph of the conclusions to better highlight what we can capture well and what remain bottle necks (see lines 715-726).

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.

The soil moisture indeed needs time to spin up, but surprisingly, this seems to happen quite fast, see our reply to above comment related to L.305 and Fig. R1 below.

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

We apologize. We didn't want to include the plots as we already have many plots in the paper and those plots are indeed in another paper. That paper has now been published (https://a.tellusjournals.se/articles/10.16993/tellusa.54/). Note also that, in response to the comments about the TOA imbalance and its effects, we now added a new panel in Fig. 4 showing the seasonal cycle of temperature.

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

We rephrased the sentence to "In the atmosphere, storms (baroclinic eddies) dominate the energy transport in the extratropics, whereas atmospheric convection dominates the vertical energy transport in the tropics.", see lines 475-476.

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?

We indeed only have one year for the simulation, but for the observations, we included 10 individual years to get an estimate of internal variability. We agree that, although the first year falls into the internal variability, we cannot exclude that the second year might not, but a priori, we would not expect this to be the case. The atmosphere has a very short memory, weather systems loose predictability in about two weeks, so if there was something really wrong, we should have seen it already in the first year. Also weather forecasts, whose main aim is to predict the large-scale circulation of the atmosphere, are conducted with models that have TOA imbalances. Maybe if we would run the model for many years and maybe if the TOA imbalance would change the temperature gradient drastically, this could have an effect on the large-scale circulation, but we don't want to go into too many conjectures here, also because we will fix the TOA imbalance before running decadal simulations.

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.

Yes, the issue is that G_AO_5km keeps two clear bands of rain over the western Pacific at all times. As mentioned in the text, the equator stands out as very dry (in terms of precipitation) and this leads to two clear bands of precipitation with the southern band being too zonal, reaching too far east and remaining south of the equator even during boreal summer. We added this information in the text when discussing Fig. 6 on lines 490-492.

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.

We agree and changed the text to "explicitly representing" as noted above (see line 506).

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.

We meant the hydrological budget over land, which we checked and was closed. We added "over land" to clarify (line 512).

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

As shown from Fig. R1 below, it is not from the initialisation but from the simulation as soil moisture decreases with time. We added this information (line 513).

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.

Yes, as usual in uncoupled ocean simulations there is sea surface salinity relaxation active. The salinity is relaxed towards observed 10m salinity of the PHC3 climatology with a time constant of 3 months. We added this information on lines 413-414.

We partly disagree with the reviewer that we are mostly showing the spin-up state. As discussed previously in the paper, salinity biases arise at the mouth of big rivers, because we neglect river discharge, and, salinity biases in the tropics, are consistent with precipitation biases. Hence, being related to the absence of river discharge and to the simulated precipitation pattern, the pattern of the salinity bias in the tropics is distinct from the one at the end of the spin-up period. This is confirmed by Fig. R3 below where we show the salinity bias in the last full month of the spin-up (December 2019) and the same month in G_AO_5km (December 2020). Except for the salinity bias in the Arctic, the pattern looks distinct. To clarify this point, we updated the discussion of Fig. 7 in the text, now mentioning which biases are inherited from the spin-up and which not, see lines 517 and 521-523..

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.

For each grid point, we computed the correlation between (a) SST and latent heat flux and (b) SST and precipitation and then took the zonal average of that value. We clarified this in the text on lines 524-525. The reasons for showing zonal means rather than spatial maps are twofolds. First, this allows us to indicate the internal variability in the observations. As mentioned by the reviewer in one of his previous comments, validating one year of simulation is difficult as it is not clear which year the model is representing. Second, we were struck by Fig. 3 of Wu et al., where the correlation between SST and precipitation was positive everywhere in the climate model whereas there were clear meridional differences in observations. To check this, a cross-section of the zonally averaged correlation coefficient is sufficient. We added this motivation for looking at the zonally averaged correlation in the text, see lines 525-528. Understanding differences in the spatial pattern of the correlation would be topic for an own study, as shown by the paper of Wu et al. dedicated to that question.

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.

We added the uncertainties for the observations into the table, as given in Griffies et al. (2016) and in Donohue et al. (2016) for the Drake passage. We also added to the caption that the observed values are for historic periods.

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?

Following the reviewer's suggestion, we compared the transport values obtained in one year in G_AO_5km to the transport values from the spin-up simulation, see the updated Table in Fig. 9. Compared to the mean transport obtained in the spin-up, the transport, with the exception of the Bering Strait, is weaker, with values out of one standard deviation (computed to quantify internal variability) for all the passages but the Indonesian Throughflow and the Mozambique Channel. Hence the coupling leads to systematic differences. Except for the Florida Bahamas Strait, the weaker transport of G_AO_5km is in better agreement with observations. The weaker transport is consistent with weaker wind stress in G_AO_5km compared to the spinup simulation, also expressed in a weaker barotropic streamfunction. We added these considerations on lines 540-544.

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

We added a reference to Seo (2017), see line 550.

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.

We didn't label TCs as this would imply that we indeed track them. We thus reformulated the sentence to "Tropical cyclone tracks are also visible, for instance off the coast of central America in the eastern Pacific in JJA . " to highlight one particular example (see lines 552-553).

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.

We added labels. Panels d&f have only illustrative purpose. Since this kind of simulations are new, and structurally very different from low-resolution models, which we now recall when discussing Fig. 11 on lines 555-556, we find it interesting to illustrate how the ITCZ and the atmosphere over the Gulf Stream look like in such simulations. Also observations from vertical velocity (in d&f panels) would have to be taken from field experiments, but we don't have a field experiment coinciding with the simulated period. This is different in panels a-c where we could make use of the EUREC^{4}A field experiment which happened right at the start of the simulation, where one can expect the simulation to still reproduce the observed weather.

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.

We added this 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.

We removed this figure as we agree that it is not the best representation and, as also pointed by the second reviewer, there is no surprise that the model can reproduce land-sea breezes. We just kept this information in the text, but also shortened that paragraph, see lines 600-601.

L547: Do you mean minimalistic parameterised physics?

Yes, but since we removed that paragraph, comment doesn't apply anymore.

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.

We are not sure that showing the 1.25 km driving model will add much information. With a grid spacing of 1.25 km, the driving model is also able to reproduce the observed features. A model with a much lower resolution would not be able to do so, but we don't have such a simulation. Also the goal of this section is to show that we can not only run ICON-Sapphire globally but also on limited areas, which is of interest for many applications. But we on purpose don't do an extensive validation as there is nothing very novel in being able to reproduce patterns of mesoscale organization with high enough resolution on limited domains and we wanted to concentrate in the paper on the more novel aspects. We adapted the beginning of section 4.2 (lines 609-610) to better convey this message and not raise false expectations. Having said this, note that we do validate the simulated cloud cover on lines 629-633.

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.

The most important contribution of the 10km model is that it captures the mesoscale eddies in contrast to the coarser resolution 40km model. The imprint of these eddies in the 10km model are seen, for example, in the phytoplankton field in the North Atlantic in July (Fig. R4), whereas the fields simulated by the 40km model are much smoother. Previous studies have shown that the impact of mesoscale eddies on local carbon export production can be as large as 50% (Harrison et al., 2018). We agree that in terms of the large-scale pattern, differences between the 10km and the 40km simulations are small, but we can already see differences in the seasonal cycle (Fig. R5) where the performance of the 10-km simulation is slightly better than the one of the 40-km.

Taking these considerations into account, we added the following paragraph at the end of section 4.4 (lines 689-696): "The large-scale pattern displayed by G_OC_10km in Fig. 18a is reminiscent of the large-scale pattern displayed by the 40-km spin-up simulation. The two fields correlate with a correlation coefficient of 0.89 for the yearly mean. However, the mesoscale structures displayed by Fig. 18b are clearly absent in the spin-up simulation and we can already see differences during the blooming season, with G_OC_10km simulating lower chlorophyll concentration than the spin-up simulation, in better agreement with observations. In the northern hemisphere, peak values are 1.5 mg m^{-3} in the 40-km spin-up simulation, 1.25 mg m^{-3} in G_OC_10km and 0.88 mg m^{-3} in observations. Whether these differences are due to the representation of mesoscale eddies is too early to tell, but is a further argument for being able to run ICON-Sapphire as an ESM on longer time scales."

Technical corrections

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

Changed accordingly (line 105).

L366: Finland

Corrected (line 377).

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

Corrected (line 396), also other instances in the text.

L447: observations

Corrected (line 494).

L588: simulation

Corrected (line 641).

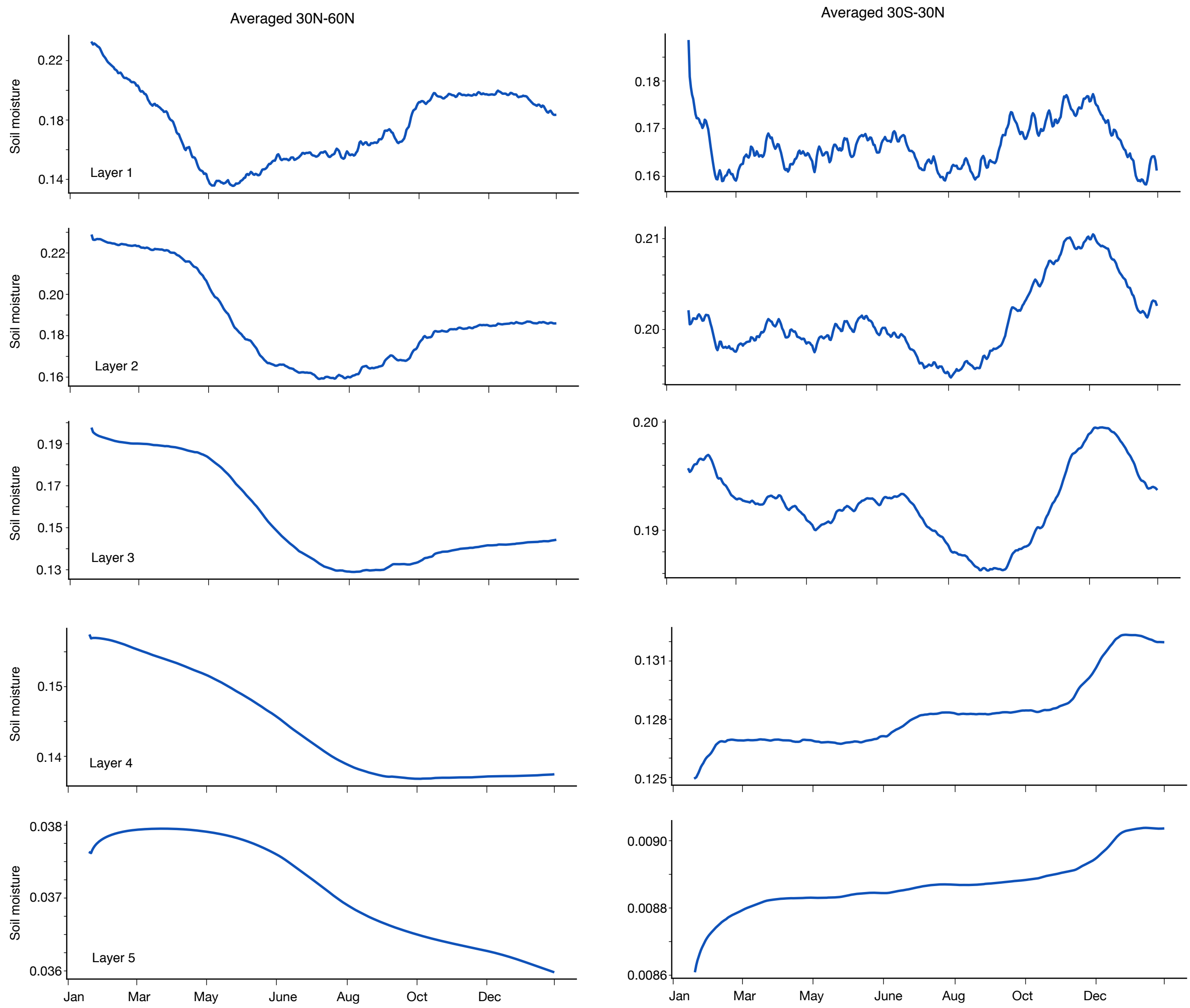

Figure R1: Time evolution of soil moisture in each soil layer and averaged over 30N-60N (left) as well as 30S-30N (right).

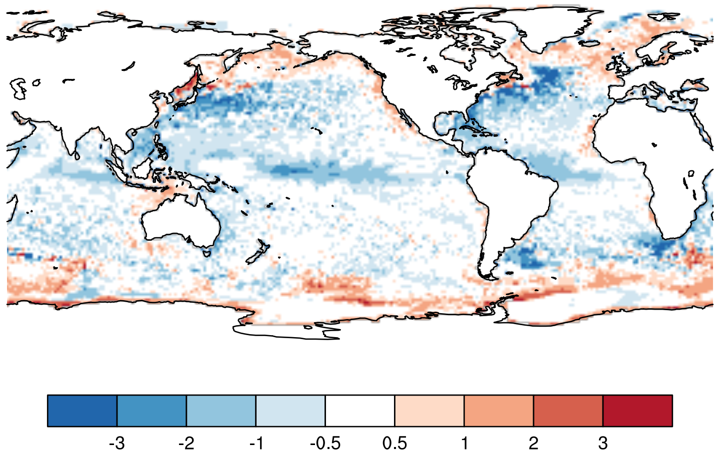

Figure R2: SST bias (K) computed as the monthly mean difference between the spin-up simulation and the Ocean Reanalysis System 5 for the last full month of the spin-up period (December 2019).

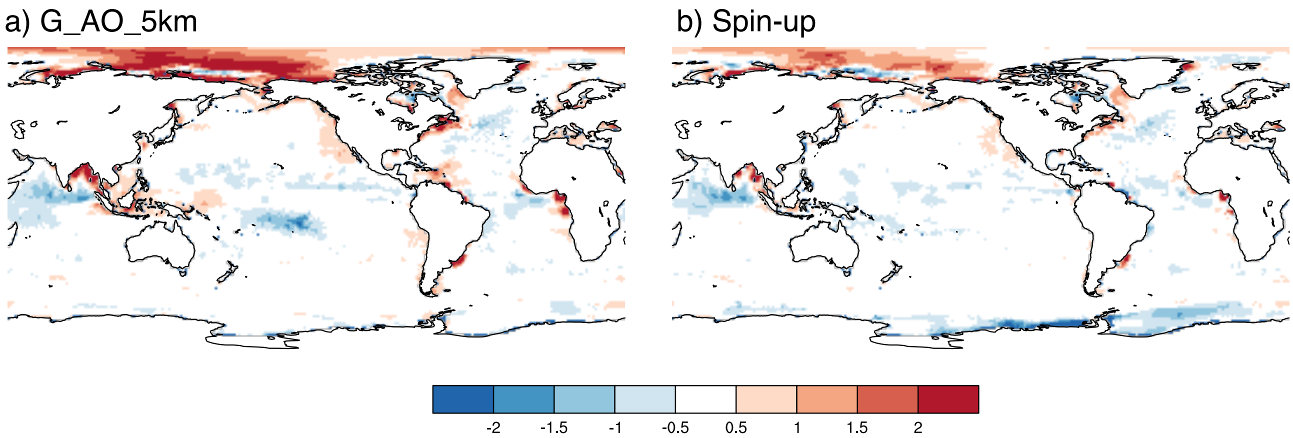

Fig. R3: Monthly mean (December) salinity bias (g kg^{-1}) in G_AO_5km and in the spin-up ocean simulation. Observations from PHC climatology.

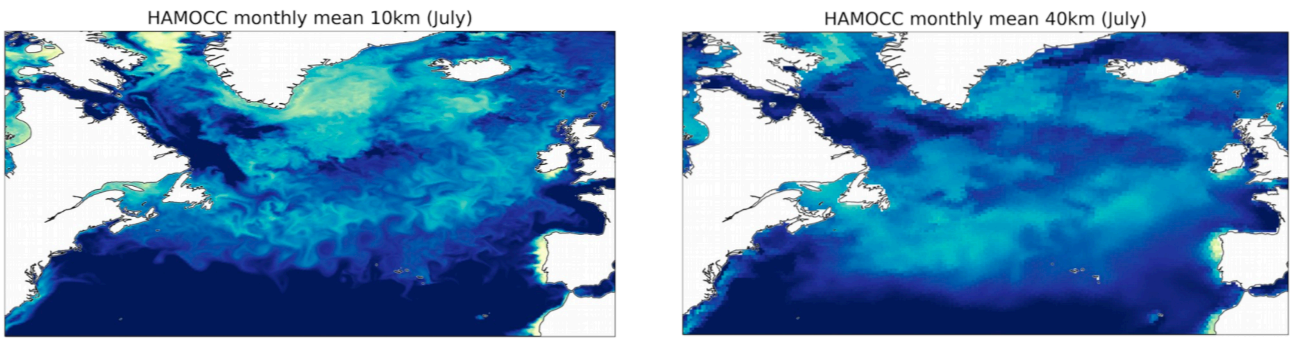

Fig. R4: Chlorophyll-a concentration in the 10km and 40km simulation over North Atlantic in July.

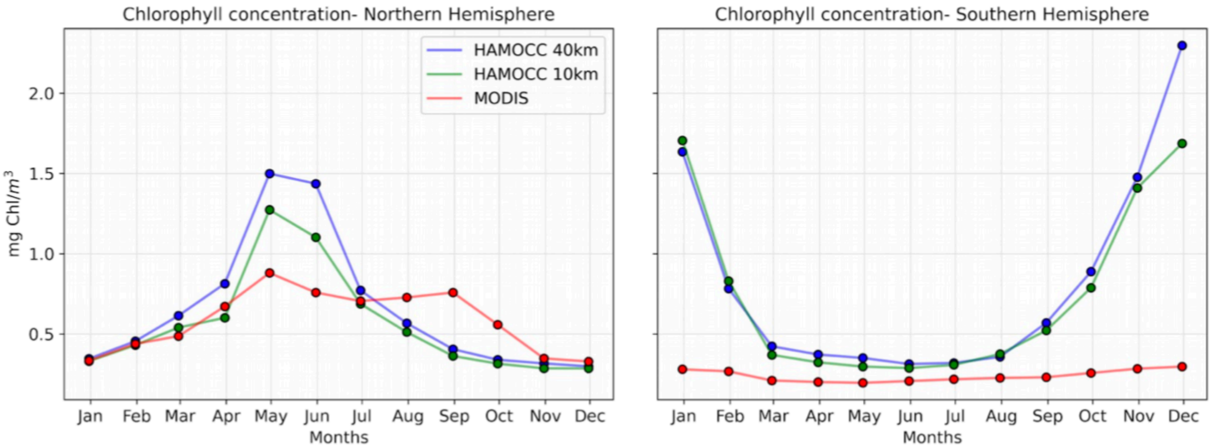

Fig R5. Seasonal cycle of chlorophyll-a concentration in the Northern (left) and Southern (right) hemisphere simulated by the 40km and 10km model, compared to observations.