Review of Schüller et al., Quantifying Coupling Errors in Atmosphere-Ocean-Sea Ice Models: A Study of Iterative and Non-Iterative Approaches in the EC-Earth AOSCM

Schüller et al. investigate the intrinsic errors associated with a standard method of coupling atmosphere and ocean models, namely explicit parallel coupling. They do this, if my understanding is correct, by comparing a series of short single column model experiments in the Arctic and Pacific to a control run produced by a novel approach: iterating the experiment a large number of times and using coupling variables from the previous iteration to force the ensuing iteration. They also compare to serial coupling approaches in which atmosphere is run before ocean (and vice versa).

The reduction of coupling errors is a very important problem in climate and weather modelling. While Schuller et al.'s control run methodology may be not be practical for systematic use (see discussion below), their experiments produce a number of intriguing new results which may inform model development. Of particular note is the sensitivity of the coupling to the sea ice albedo parameterisation, and to a lesser extent the cloud parameterisations. This study should therefore be published. I have two major questions/concerns, which are each set out below as a discussion followed by recommendations, followed by a few more minor suggestions.

1. **The experiment design: SWR and its relationship to standard coupled model approaches**

If I have understood correctly, the authors produce (for example) a 2-day SCM run in the MOSAiC region from 0000Z 14-04-2020 to 0000Z 16-04-2020 by the following means. The coupled model is initialised from ERA5 and CMEMS and run forward for 2 days, using additional ERA5 atmospheric forcing for context where necessary, with all coupling variables saved at every coupling instance. Once the experiment ends, it is repeated from the same initial conditions, but at each coupling instance *T*, the saved variables from the coupling instance *T+1* in the previous iteration are used as input to each model, while coupling output is saved for use in the following iteration.

(Aside: I hope I am right here about the 'jump forward' to *T+1*; without it, I could not see how the model evolution would actually change from the first iteration to the second – and for the updating process to work, the first two iterations' evolution must differ. Specifically, at iteration 2, the initial coupling exchange would otherwise just read the initial coupling variables from the first iteration, which would be identical by design; the first model timestep would then evolve in exactly the same way, and the second coupling exchange output variables would also be identical between iterations, etc.)

The authors' experiment design corresponds, as they state, to Schwarz waveform relaxation, in which an equation is repeatedly numerically integrated over a time domain and multiple space domains, with each spatial domain repeatedly updating boundary conditions. It is described as a contrast to standard explicit parallel coupling, in which each spatial domain calculates the latest boundary conditions for use by its neighbours in the ensuing coupling period.

While the SWR method is useful for the present idealised study, I think it is technically and scientifically very different to standard explicit coupling, and attempting to use it for longer experiments might be problematic, because of chaotic variations in run evolution. For iteration N, the coupling variables from iteration N-1 are used as input. This will alter the run evolution, but it will also, presumably, impose some control on how the run evolution varies, as the atmosphere bottom boundary condition is tied closely to the previous iteration for the entire duration of the run. For the numbers of iterations considered here by the authors, this would presumably impose strict controls on how much the run evolution can vary relative to the first iteration (the standard, explicit parallel solution). For a run of sufficient length (e.g. a year) this seems unrealistic, and might indeed mean that much of the effects of improved coupling cannot be realised. An alternative outcome, too, is that the atmosphere evolution might vary chaotically from iteration to iteration such that eventually it is completely unrelated to its own lower boundary condition, the coupling variables from the previous iteration – which might cause a model error.

A more comparable approach may be the implicit solvers found throughout the atmosphere, ocean and sea ice submodels, which iterate not over the whole model period, but simply over every timestep. I am not asking the authors to carry this out, but it seems that an interesting experiment would be to carry out the SWR method coupling period by coupling period: run the parallel models forward for a coupling period and repeat, using the outputs from iteration *N* as inputs for iteration *N+1*. After a fixed number of iterations, continue to the next coupling period. I can see why such an experiment might not be useful for the present study: after the first coupling period, the model evolution would begin to vary, such that errors between standard coupling and control would no longer be easy to interpret. However, this approach seems to be scientifically more appropriate for wider use (even if still computationally very expensive).

This would maybe not matter, if it were not for the authors' L335: 'Such a large number of iterations is infeasible for long or many runs.' I think the problems of using the authors' method for long runs are more fundamental than this, and the method is better understood as a means of producing a control for these short idealised experiments, rather than as an alternative means of coupling.

Recommendations:

- The experiment design needs to be set out more clearly from the start, notably the fact that the iteration is performed over the whole model run (or otherwise, if I have completely misunderstood this).
- For the reasons above I think it needs to be made clear that the control run methodology is not a practical way of running a coupled climate model for longer than a few days – or if the authors disagree, justification should be given. This is not a criticism of the study; it is a highly appropriate methodology for the purpose on which the authors actually use it.
- I would like the authors to clarify in the paper how the evolution of the second iteration will differ from the first, if the coupling output from the first iteration drives coupling input for the second – is it because of a time offset?

2. The sea ice albedo behaviour and the authors' solution

The behaviour of the SWR method in the presence of warm air intrusions over sea ice is fascinating, and provides evidence that the coupling problem is qualitatively different where sea ice, or specifically melting sea ice, is present. Figure 10 in particular is quite shocking: the point that in the old (LIM) sea ice model, flipping the albedo from its cold to its melting value results in surface temperature swings of 30 degrees. I would be interested to know if the authors have any idea as to a) the mechanisms, in particular negative feedbacks, behind this behaviour; b) why it is so greatly attenuated in SI^3.

The authors' albedo parameterisation test provides convincing evidence that the albedo parameterisation is responsible for most of the observed instability (presumably not all, given that the experiment still fails to converge in 3 out of 84 cases?).

On the surface, there's an interesting problem here: in the physical world, the transition in albedo between cold and melting snow/ice is presumably pretty sharp. By introducing a smooth variation, the algorithm performance might be improved despite making the model less representative of reality. In practice, the values the authors choose for epsilon are sufficiently small that this effect is probably negligible – and explainable by the fact that the model aggregates temperature over large areas that would in the real world make the transition much smoother (and in the real world the atmosphere and sea ice can react instantly to one another).

Nevetheless, it is interesting that the sigmoid parameterisation is able to resolve the oscillatory behaviour so effectively despite being so close to the original, abrupt parameterisation (due to the small values of epsilon). The ranges of Ti over which the oscillations occur in Figure 10 are sufficiently large that they could not be directly affected by the authors' sigmoid parameterisation; is this because there is a large number of variables that are 'out of balance' by the end of the iterations, driving the uncontrolled negative feedback? Presumably at some crucial point in the iterations, small oscillations must grow in the tiny region of the state space that is affected by the authors' sigmoid parameterisation?

Recommendations: it would be really interesting if the authors could give some ideas as to

- the mechanisms behind this oscillation
- why it is improved in SI^3
- how the sigmoid albedo parameterisations resolve it, despite being near-identical to the abrupt parameterisations in all but a very small part of the state space

In addition, I wonder if their statement in L457-8 that the albedo discontinuity is responsible for the non-convergence needs qualifying, given that 3 experiments still fail to converge – the albedo is obviously the dominant factor, but perhaps there are other parameterisations that are still causing problems?

3. Some additional minor comments

- Section 3 in general: it might aid understanding to have a table of variables that are exchanged between the atmosphere, and the ice and ocean
- L251 and 261: perhaps a word of explanation here to describe the physical interpretation of 'right-hand sides' would be appropriate
- Figure 4: It is interesting that relative error stays constant for around the first 7 iterations. Could the authors comment in the paper on why this is?

- L403: recommend to reword ‘it is possible to study warm air intrusions in detail’