Comments for:
Quantifying Coupling Errors in Atmosphere-Ocean-Sea Ice Models: A Study of Iterative and Non-Iterative Approaches in the EC-Earth AOSCM

by V. Schüller *et al.*
https://doi.org/10.5194/egusphere-2025-1342

Referee : charles.pelletier@ecmwf.int

General

As its title suggests, the work presented in this manuscript focuses on quantifying systematic coupling errors committed when coupling ocean and atmosphere models due to the inevitable "lag" in the perceived exchange fields across their shared interface.

The authors first provide their motivations, expanding on why classical coupling methods can be qualified as "non-converged". Then, they proceed to introduce the Schwarz Waveform Relaxation (SWR) method as a framework for investigating the convergence (or lack thereof) of non-overlapping coupling problems, which the ocean – atmosphere one falls into. The authors then apply the SWR formalism to the coupled ocean—atmosphere single-column model (SCM) version of EC-Earth, a state-of-the-art Earth System Model (ESM) participating in CMIP exercises. They focus on two realistic test cases, including one involving sea-ice presence, which, to my knowledge, had not been investigated (with this lens) in the existing literature. Further investigations lead to pinpointing two specific parametrizations (the atmospheric convection scheme, and the sea-ice albedo) as potential causes of numerical instability in the coupling, which could only be identified through the SWR lenses. The authors then come to the more general conclusion that ensuring the "regular" (in the mathematical sense) character of ESM physical cores are most likely a necessary condition for ensuring the well-posedness of the underlying coupled ocean – atmosphere problem they are attempting at solving.

The manuscript is of very high quality, didactive, and contains engaging information for ESM modellers. In my opinion, it builds up on more theoretical applied mathematics literature and makes a considerable step towards applications to state-of-the-art climate models. Despite being set in a relatively idealised setting (which is perfectly understandable, considering the novel approach the authors are pursuing), the authors hint towards clear recommendations for ESM models. In that sense, this manuscript is an excellent and exciting submission for GMD.

That said, I do think that the manuscript could be made clearer, especially to the GMD readership. There also are key unaddressed methodological limitations, which I think could be at least presented and briefly discussed. I therefore am recommending this paper for publication under major revision, provided the authors address the comments listed below. “Major” and “minor” does not mean “important” and “unimportant” comments – I think all should be accounted for, but I expect the “minor” ones to be more easily dealt with.

Major comments

1. Introducing SWR and its relationship to well posedness

L. 20 – 23: I think this important part should be expanded to suit the lenses of a typical ESM (or components thereof) modeller. Before talking about SWR algorithms, it would be nice to more explicit about what current coupling methods do (or do not do), algorithmics-wise, in terms that are more intuitive to Earth system modellers.

I would say that for practical (efficiency) reasons, both models must be coupled at a frequency lower than either model’s time step. and that this introduces errors (lack of feedback within one coupling window) and/or latency/staggering of the surface BC from one model to the other. And that this is unavoidable.

To be more concrete, give one example, e.g. for the parallel EC-Earth algorithm (around L. 230 – 235). At this point, I think introducing any sort of mathematical notations, or notions like “coupling window”, would not really help but just saying that openIFS computes the fluxes using SSTs which are delayed compared to the model’s clock, and that NEMO gets fluxes that are also late (and on top of that, these fluxes were computed from “belated” SSTs).

L. 24 – 25: I’d be slightly more explicit and say that SWR algorithms can provide “the” reference solution (if the problem is well-posed, it is “the” solution, not “a”) that classical (first iteration only) methods can be compared against, thus giving an estimate of the committed error. The manuscript could also be more expansive about what SWR methods bring with respect to error quantification: without doing them, the actual proper solution to the coupled problem is not known (we don’t even know whether there is one!), usually what comes out of the first iteration is just accepted as is.
There’s a lot of literature of the impact of the coupling frequency on coupled model performance (e.g., Lebeaupin Brossier, 2009; Scoccimarro, 2017; Li, 2020), which I think is an illustration of what the manuscript is covering there, but maybe more familiar / graspable to the typical GMD reader. This parallel might be worth being done there.

L. 28 – 31: Drop the “typically”, they’re always developed independently. I have two scientific comments/questions regarding these few lines.

  a. We don’t know whether in their state-of-the-art implementations, air-sea coupling problems are well- or ill-posed, but given that they include many parametrizations that have often (always?) been developed with concerns pertaining to physical realism rather than well-posedness (and that’s fine and understandable), it probably is a safe bet to consider them ill-posed until the contrary has been proven, or at least observed (like the manuscript does to some extent).

b. The manuscript says that if SWR converges, then the coupled is well-posed. Is the reciprocal true? If the answer is not known, then it's also worth being explicitly written there.

L. 257 – 262: this is an interesting paragraph. What does "obey regularity" mean? My understanding is that:
a) if the RHS is not regular enough, then the coupled problem is ill-posed.
b) a coupled problem can be ill-posed even if the RHS is regular
The manuscript is stating a). Is b) true also?
I'm not sure I understand the logics of L. 257, with the "or".
I would more pragmatical and say that since all we have is a), then we should at least strive to have regular RHS, and then elaborate about physical cores there.
But then, if b) is true (which I'm not sure of), then having regularity does not provide a guarantee of well-posedness either. It would be a necessary condition, not a sufficient one.

2. <u>(Non-)linear free surface, and "NEMO3.6 vs NEMO4.0"</u>

L. 120: In **our** 3.6 NEMO version, the volume of the oceanic column is constant… This (important) difference is known to as "nonlinear free surface" (or vertical varying layers) to the community. I think explicitly saying it might make sense. And that option was already available is NEMO3.6, so "our" NEMO3.6 would clear up potential confusions ("our" NEMO4 as well, because it's also possible to run without this option in NEMO4). I would stick to linear free surface vs nonlinear free surface (or equivalently, non-VVL vs VVL). That is what the distinction is about, the NEMO3.6 vs NEMO4 one is artificial and coming from specific physical choices.
L. 175 – 179: IMO, that part is a good example of the fundamental difference between a non-VVL and a VVL model (NEMO3.6 vs NEMO4, in your case). There's no salt being exchanged between the IFS and NEMO in either model versions. But a linear free surface does need a **virtual** salt flux, because the ocean volume is kept constant, and the prognostic variable for salinity is actual salinity (salt per unit volume). So, salinity will decrease/increases via removal/injection of water mass. On the other hand, a VVL model uses salt per unit **area** as an effective prognostic variable, and the variations in ocean volume takes care of concentrating/diluting the salt upon water mass removal/injection.
That's not the point of the manuscript, so don't be expansive on this. But I do recommend sticking to sharper terminology (linear free surface and/or (non-)VVL) whenever discussing implications on your study. The boundary condition for salinity is one.
Same applies to the heat boundary condition, for which the internal enthalpy of mass exchanges (E-P terms) only apply with the VVL. The manuscript happens to neglect them anyway, which is a reasonable assumption.

3. <u>One unaddressed key limitation of the SWR wrapper (from my understanding)</u>

L. 282 – 286: I understand the point about doing the SWR with the wrapper, etc. But there is a serious potential drawback that is never addressed: I have a strong gut feeling that this approach, while indeed more flexible, is a lot greedier than the (comparatively intrusive) method of Marti. Marti's method will iterate over each coupling window separately and then proceed to the next window when convergence is reached. Whereas the SWR wrapper re-runs the full simulation. It's not a showstopper for your paper, especially considering it uses a 1D model for which cost does not come into question. But I still think that this point is worth being mentioned and at least mentioned. The SWR wrapper is indeed less intrusive, and that's a strength, but it's also probably much less efficient, which is a drawback.

4. <u>Comparing coupling methods based upon "lowest error"</u>

L. 476 – 478: I am not convinced that "coupling method that produces the lowest error" is a relevant evaluation metric. Especially considering ocean-first and atmosphere-first look like they're running a real close race. An ESM user is probably interested in a method that limits cases where the error is large. IMO, it doesn't matter to them whether the error is small, or very small, which is what the current criterium will discriminate. The user probably wants to avoid large errors, where "large" is a tolerance they are willing to accept.
This ranking criterium (lowest error) is then used as ground base for delivering one of the key conclusions of the manuscript; that ocean-first methods might be better in the presence of sea ice. But I'm not convinced that the comparison method is relevant in the first place. I think a safer conclusion (and the manuscript does say it already), is that the presence of sea ice shuffles the cards again. I would need more evidence, with other criteria, to be convinced with ocean-first as a "best candidate".
Incidentally, these conclusions arose from running one test case, in one SCM configuration. As of now, it is also not clear whether they would be robust in other configurations.

Minor comments

L. 40: replace "e.g." with "in our case"

L.54: for clarity, I would split this sentence in two. "...is robust in ice-free condition. This allows us to have a reference state to compare classical one-iteration coupling to, for which already after two days, temperature coupling errors can reach up to several degrees in the atmospheric boundary layer."

L. 57: Say NEMO 4.0.1 instead of "newer model versions".

L. 88: the first term of (2) is turbulence (and it's local), but the second is not, it's a nonlocal convective contribution, which I would refrain from describing as "turbulent". To me, (2) contains turbulent and convective contributions to vertical transport on subgrid scales.

L. 97: might be worth specifying that in the SCM, the Coriolis parameter is constant.

L. 103: not a huge fan of the word "local" in this context. To me, gradients are local properties of the function, because it only involves the vicinity of the point it's evaluated at (but numerical schemes assessing them are nonlocal) … I know what's meant, what I'm not sure what the relevant term should be there. Maybe just say that these terms do not involve gradients.

L. 106: "incompressible" on top of Boussinesq and hydrostatic is relevant.

L. 113: the four **large-scale liquid ocean** prognostic variables are… large-scale (or "resolved") because when used, at least TKE is also prognostic. "Liquid ocean" because the sea-ice model also has a bunch of prognostic variables.

L. 118: might be worth explicitly saying that the unlike the solar heat flux, the nonsolar heat flux does not penetrate. Which is the reason why both are treated separately.

L. 124: I wouldn't say LIM3 and SI3 are "equivalent", especially considering some of your results presented further down suggest SI3 has much better convergence properties than LIM3 (although I'm not convinced yet that it's due to the sea-ice model, it could also be the NEMO model update, and/or the VVL). "Similar" is a better fit, and it doesn't dampen your message.

L. 132: Having the bottom sea-ice temperature set to the local seawater freezing point is a (Dirichlet) boundary condition for the sea ice. It might be worth being explicitly phrased out so.

L. 144: Are there any reference/evidence to support this rather strong claim? I don't know the SCM model well, but to my understanding this statement is not true… To me:

a. Kinetic energy should be conserved, I agree.
b. Why isn't mass conserved with NEMO4 (and the VVL)? I think it should be…
c. Energy is never conserved, at least due to the internal enthalpy that is never accounted for by the IFS (and that is only considered in NEMO with the VVL on). The manuscript does refer to it later in a footnote.

I'm happy to be proven wrong here, but I would suggest either proving me wrong and providing evidence for it or simply removing this slippery sentence. Incidentally, to my understanding, whether the interface is heat/momentum/mass conserving or not does not bear implications to the manuscript.

L. 171 – 174: please briefly introduce C_H (the same way it's dine for C_M above).

L. 175: well-posedness and eq. 8

L. 185: The manuscript is slightly confusing here, IMO. It is healthy for both stresses (eqs. (10a) and (10b)) to be different, because they are located at physically different interfaces (atmosphere – ice and ice – ocean, respectively). The way the manuscript is worded might make it seem like (10a) and (10b) being different is a model inconsistency. Which might not the authors' position, but the manuscript is somewhat ambiguous there. For clarity, I would distinguish “ocean” (which, in my opinion, can include both ocean and sea ice) and “*liquid* ocean” with a clear terminology distinction.

L. 217: I understand what the authors mean with the ice concentration already being accounted for by SI3 in *Q_i$, but I still thinks that explicitly writing $a_i * Q_i$ is more self-consistent, as Q_i has been introduced as the ice – ocean heat flux in the 100% ice cover case. The model can do what it wants (and it happens to “update” the boundary condition with some sea ice induced modulation), but I think writing out $a_i * Q_i$ is accurate as well and less prone to confusion.
I also recommend doing the same for the salinity flux and write $a_i * S_t$.
And I would split (14b) into two different equations (without or with VVL). In the non-VVL case (so, NEMO3.6), are you sure that there is no (1-a_i) factor in front of the (E-P) term?

L. 229-236: I would remove the sentence starting with “At each coupling window, ...”, because it’s covered more precisely from L.234 on. And I would move the sentence starting L. 230 (“This introduces a coupling lag”) after the sentence currently ending L.236. “At each coupling time step, the coupling variables *from one model* are averaged in time over the past window to obtain the interface boundary conditions *of the other model* over the next coupling window.”

L. 232: The footnote should be included as main body text there, IMO.

L. 243: ECMWF’s single executable coupling could at least be mentioned there. It’s sequential, atmosphere first. It’s much less flexible than OASIS, so less prone to investigations and tests. But it’s fast, which is why it’s been done in the first place. The relevant reference is https://doi.org/10.21957/rfplwzuol

L. 267 – 268: I’d remove the last sentence -- that’s just what a compiled code is.

L. 270: “Many aspects of the coupling setup **at run time** can be changed ~~at runtime~~ using this file” – editing the file during runtime is probably not safe. And that’s a good thing, IMO.

L. 272: “Whether an AOSCM experiment uses the parallel algorithm or one of the sequential ones **at runtime** can be controlled ~~at runtime~~ by modifying the LAG parameters in namcouple.”

L. 277: Is mentioning COCOA explicitly really needed? Maybe just stick to citing both papers together (Voldoire and Valcke)

L. 280: maybe “rewinding” in time?

L. 281: “basing the implementation on OASIS, with minimal changes in the ocean and atmosphere models, can facilitate its reuse in other climate models”

L. 291 – 292: This might be relevant in the conclusion paragraph as a perspective. Not here.

L. 318: Please specify the year at that point.

L. 335: and mostly, infeasible for 3D runs.

Fig. 6 and discussion: I think one remarkable point about these experimentations is that even after each coupling method differs one from another (e.g. around July 2nd 12:00), then they can roughly reconverge (i.e., be close one to another). It's not that obvious to me – one might have expected that once the coupling error gets bad enough, then all bets are off from the next coupling window, as the model's trajectory strayed away from the reference solution. But it seems to be able to catch up. Do you have a rough idea as to why? Could it be due to the SCM framework (i.e., the top and bottom BCs essentially acting as nudging)?

Fig. 6: bit of a detail, but I think having the diurnal cycle more clearly marked via the plot grid would make the figure more readable (e.g., put emphasis on ticks at 00:00).

L. 371: Could you provide any guesses at to what is causing the hyper-sensitivity of the turbulent heat fluxes to the SST? I'm wondering whether the point at which the fluxes diverge, but the SSTs don't, is close to neutral stability. That might translate into discontinuities in the bulk formulas via the stability functions.

L. 375: I don't understand the phrasing in the parenthesis – I think there should be two separate ones?

l. 393: To be fair to the atmosphere-first method, you can say that it's indeed pretty good for the SST but has neutral performance for atmosphere fields. The current phrasing might make it seem like it's worse for atmosphere fields.

L. 405: Specify the year.

L. 496: in the presence of sea ice

L. 503 – 507: It might be outside of the scope of the manuscript, but I think this point goes beyond the coupling problem. It is about having a physical core that the dynamical core can cope with. In the physical world, sea-ice albedo can be as jumpy as it wants to, but the model numerics do not work well with sharp transitions, and we use numerical models to represent it. It can be seen as a matter of adapting (potentially doing compromises to) the physics so that the numerics work OK. Unless we work with methods that can deal with less regular functions, like Discontinuous Galerkin. But that's another topic.