Interactive comment on "Quantifying the scaling penalty of sequential coupling: The sequential coupling cost (Cseq) metric for Earth System Model components" by Oriol Tintó Prims, Sergi Palomas, Simone Vacondio and Mario C. Acosta.

Line numbers refer to the PDF margin numbering of the discussion preprint (egusphere-2026-1661).

The paper presents a timely and useful contribution to the performance evaluation of coupled Earth System Models. It defines the sequential coupling cost (Cseq), the extra computational resources a sequentially coupled model needs to match an optimally configured concurrent setup, and it fills a genuine gap: the established CPMIP coupling cost was built for concurrent architectures and evaluates to zero for a single-binary sequential model. The derivation is clean, the metric is cheap because it reuses standard execution logs, and the three use cases (IFS-NEMO, NEMO-SI3 and OpenIFS-M7) nicely span an atmosphere-limited, an ocean-limited and a near-balanced case. I checked the central IFS-NEMO arithmetic and it is correct. I have a number of feedback points below, ordered roughly from most to least important, but none of them question the value of the method.

General comments

- **Positioning against your own prior work.** Two of you (Palomas and Acosta) are authors of Palomas et al. (2025, doi:10.5194/gmd-18-3661-2025), which already introduces a "partial coupling cost" and builds the optimal concurrent allocation N_conc = N_A(S) + N_B(S) from per-component scalability curves, which is the exact construction Cseq rests on; the motivating "13% on average" sentence at L24-L25 is in fact that paper's phrasing. Yet Palomas et al. (2025) is not cited. As I read it, the novelty of Cseq is that it turns this same concurrent-optimization machinery around to diagnose the sequential case, which CPMIP scores as zero. The paper would benefit greatly from citing Palomas et al. (2025), relating it to the already-cited Acosta et al. (2023), and stating in one short paragraph exactly what Cseq adds. As it stands, a reader familiar with your group's work will be unsure where the new contribution begins.
- **The OpenIFS-M7 target definition.** A second point, of similar weight, is that this use case defines its target throughput as the maximum the sequential setup itself can reach at 32 nodes (L231-L232, L246-L247). Because Cseq grows with node count, this is the choice that maximizes it, so the 46% and 51% headline figures are upper bounds rather than representative costs. I would ask you to say this plainly, and also to report Cseq for a fixed, science-relevant throughput so that the three use cases can be compared on a common basis.
- **A consistent overhead assumption.** Related to this, the headline Cseq values assume a concurrent coupler with zero overhead, whereas your own OpenIFS-M7 node estimates add a 10% overhead (L242-L249); the two are then not on the same footing. I would report Cseq under a single, clearly stated overhead assumption. In the same paragraph, the claim at L267-L269 that concurrent frameworks "typically incur only 1-5% overhead (Balaji et al., 2017)" is not supported by that reference: Balaji et al. (2017) report costs rising with resolution to about 25%, and your own Acosta et al. (2024) give a 5-15% range. Please correct both the figure and its attribution.
- **Cost and accuracy as two axes.** I would encourage you to frame Cseq as one axis of a two-axis evaluation of coupling. Marti et al. (2021, doi:10.5194/gmd-14-2959-2021) and the more recent Schüller et al. (2025, doi:10.5194/gmd-18-9167-2025) quantify the physical inaccuracy of coupling schemes with a Schwarz iterative method run to convergence, whereas Cseq and CPMIP quantify the cost. The strict concurrent-versus-sequential dichotomy at L26-L28 would also read better if it acknowledged iteratively converged coupling as a distinct third category.
- **The EC-Earth3 to EC-Earth4 comparison (Table 3).** This comparison should be framed as illustrative only. TM5 (about 300 km, 34 levels) and M7 (about 80 km, 91 levels) differ by more than an order of magnitude in degrees of freedom, on top of differing in model version and HPC platform, so the implied 8.25% versus 46-51% contrast is not like-for-like. You are already candid about this (L227-L229, L254-L256); I would simply make it explicit at the table.
- **A few citation corrections.** The 1 SYPD target is described as a milestone "for DestinE" (L155) but is actually a nextGEMS cornerstone, and its published reference is Segura et al. (2025, doi:10.5194/gmd-18-7735-2025). For OpenIFS, L128 cites the OpenIFS@home paper (Sparrow et al., 2021), where Huijnen et al. (2022, doi:10.5194/gmd-15-6221-2022) would fit better. The M7 module at L127 is attributed to Huijnen et al. (2022), but its canonical reference is Vignati et al. (2004), which you already cite elsewhere.
- **A worked recipe in the paper.** You already archive the Escalador library and the reproduction scripts openly on Zenodo, which is exactly right and lets others reproduce the three cases presented here. My suggestion

is complementary: a reader should also be able to apply Cseq to a new model from the paper alone, without having to work through the code. A short appendix giving the quantities you take from the logs, the USL fit and its functional form, the N(S) inversion and the N_conc optimization, and one fully worked example would achieve that and would, I think, most increase uptake of the metric.

- **Extrapolation uncertainty.** Finally, several results rely on evaluating the USL fit outside the measured node range: the IFS-NEMO ocean optimum of 30.6 nodes lies below the 100-500 node range, and the NEMO-SI3 sea-ice optimum of 3.7 nodes below the 4-node reference. Since Cseq depends directly on these extrapolated node counts, a sensitivity estimate or confidence interval per use case would strengthen the headline numbers.

Technical corrections

- Figure 5: the caption calls the ocean "the ocean component (green)", but NEMO is drawn in purple and green is the concurrent model.
- Figures 6 and 7: the in-figure values (45.9%, 50.6%) differ from the text (46.2%, 50.9%), and the IFS-NEMO case is 12.9% in Figure 4 but 13% in the text.
- line 15: "leading inefficiencies" should be "leading to inefficiencies".
- line 40: stray parenthesis in "EC-Earth (IFS() coupled to NEMO".
- line 67: "we provide model developers and with a standardized formulation", remove "and".
- line 94: missing closing parenthesis after "(Cseq".
- line 95: "Being T_A(N_A) the time it takes…" reads better as "Let T_A(N_A) be the time it takes…".
- line 202: "with at N_NEMO = 15.6 nodes", delete the stray "at" (or "with").
- line 217: component naming is inconsistent (OpenIFS, OIFS and IFS-M7 all appear); please standardize.
- line 222: "the aerosols component" should be "the aerosol component".
- line 282: "one-eight" should be "one-eighth".
- line 283: "it can be seen than the ocean…" should be "that".
- line 284: "drops by a 8%" should be "by 8%".
- line 285: "of a similar magnitude than the ocean one" should be "to".
- line 291: "than the previous two cases" should be "than in the previous two cases".
- line 292: "resulted between 40% and 50%" needs a word ("ranged between…"); also hyphenate "32-node allocation".

Recommendation. I recommend the paper for publication after minor to moderate revisions. The method is solid and useful; the main work is to position it clearly against your group's earlier coupling-cost work, to present the more favourable numbers on a consistent and explicitly bounded basis, and to correct a handful of citations.