

EGUSPHERE-2025-3883

“Applying Corrective Machine Learning in the E3SM Atmosphere Model in C++ (EAMxx)”

Aaron S. Donahue, Elynn Wu, W. Andre Perkins, Peter M. Caldwell, Christopher S. Bretherton, Finn O. Rebassoo, and Jean-Christophe Golaz

Response to Reviewer #2

The authors thank Reviewer #2 for their thorough and constructive review of the manuscript. We appreciate the many helpful suggestions, which have improved both the clarity and completeness of the paper. We have carefully considered all comments and revised the manuscript accordingly. Below, we provide a point-by-point response, together with the locations of the corresponding revisions in the revised manuscript.

Specific Comments

Comment 1: Since the authors argue that there are benefits from their model developments other than the attempts at nudging to high resolution, these benefits should be stated more clearly in the motivation of the work.

Response: We agree. In the original submission, the motivation emphasized the corrective ML application but did not clearly highlight the broader software-development outcomes of the project. In the revised manuscript, we added text to the Introduction stating that the study also developed software infrastructure that will support future efforts to couple ML with SCREAM.

Location in revised manuscript: Section 1, p. 2, line 35, paragraph beginning “*This study applies their corrective ML approach to SCREAM...*”.

Comment 2: Also, it is not clear to me whether the changes made to the SCREAM code are openly available (important with respect to the benefits claimed above) -- the section on Open data is very superficial.

Response: We agree that the original code and data availability statement was too brief. In the revised manuscript, this section was substantially expanded. It now includes links to the SCREAM and fv3net repositories, the specific code commits used in the study, a Zenodo archival software record, and the archived datasets used for analysis and figure generation.

Location in revised manuscript: Code and data availability section, p. 13, line 274, paragraph beginning “*We provide a Zenodo archival repository...*”.

Comment 3: What is the throughput of the tested models (high res, coarse and without nudging)?

Response: We appreciate this question. Throughput was not evaluated as part of this study. The main goals were implementation of the corrective ML workflow and evaluation of predictive

skill. Because the corrective ML approach did not produce the level of improvement we had anticipated, we did not extend the analysis to include a full throughput comparison across model configurations. Although, we did measure the overhead introduced by the inclusion of the corrective—ML model in EAMxx, see line 175.

Location in revised manuscript: This scope is reflected in the framing of the study and discussion of outcomes in Section 5, especially the paragraphs beginning “*This study investigated the transferability...*” and “*There are several possibilities for why corrective ML...*” on p. 9 to 10.

Comment 4: A detailed description of the models before ML is missing. Which parameterizations are active?

Response: We added a more complete description of the EAMxx/SCREAM configuration in the Methods section. The revised manuscript now explicitly identifies the spectral-element nonhydrostatic dynamical core and the major subgrid parameterizations, including P3 microphysics, SHOC turbulence and boundary layer physics, and RRTMG++ radiation. We also note that SCREAM does not include a deep convection scheme or a gravity wave scheme.

Location in revised manuscript: Section 2, p. 2, line 41, paragraph beginning “*SCREAM is a configuration of the atmosphere component...*”.

Comment 5: How do the authors deal with the spin-up periods of the models? What are the initial conditions? How do they ensure that the coarse and high resolution models simulate the same climate state? How long were the runs for training?

Response: We clarified these points in the revised manuscript. We now state that all simulations use identical model configuration and start date, and that initial conditions are generated from ERA5. We also added text explaining that, because tunable parameters are held fixed across resolutions, the untuned coarse and fine simulations would be expected to diverge over long timescales; however, this study focuses on a one-year period, and we expect closer agreement when the coarse simulation is nudged toward the fine-resolution target. In addition, Section 2.1 now specifies that one year of fine-resolution output was used for training and validation.

Location in revised manuscript: Section 2, p. 2 to 3, including the paragraphs ending with “*Initial conditions for all simulations are generated from ERA5...*” and “*We therefore expect the coarse simulations...*”; Section 2.1, p. 3 to 5, paragraph beginning “*For the training and validation discussed in this paper...*”.

Comment 6: A detailed investigation of the produced nudging tendencies is missing. Here I would expect some suitable figures. Do the nudging tendencies make physical sense, considering the better resolution of physical processes in the high res? How robust are they with different seeds or slightly different starting conditions? How large are the nudging tendencies in comparison with the coarse unnudged tendencies? What is the consequence of nudging with 3-hour average fields?

Response: We substantially expanded the manuscript in response to this comment. Most notably, we added a new figure showing the target nudging tendencies and the ML-predicted tendencies for temperature and specific humidity. We also expanded the Results section to

discuss the systematic underestimation of tendency magnitude, especially in the tropics, and its likely connection to degraded tropical performance. In addition, the revised manuscript includes expanded discussion of the use of 3-hourly averaged fields, both in the Methods and Discussion sections, and adds discussion of robustness across random seeds together with the effect of novelty detection.

Location in revised manuscript: Figure 2, p. 10; Section 4, p. 8 to 9, including the paragraphs beginning “*Results from a one-year free-running coarse-grid simulation...*”, “*As shown in Figure 2...*”, and “*To evaluate the stability and robustness...*”; Section 2.1, p. 5, paragraph beginning “*For the FV3GFS study, instantaneous 3-hourly snapshots...*”; Section 5, p. 9, paragraph beginning “*There are several possibilities for why corrective ML...*”.

Comment 7: What motivates the use of the variables used for validation of the ML model (Tab. 2)? I notice that no dynamics variables are present despite the nudging also being applied to the winds.

Response: We agree this omission was not ideal in the original manuscript. In the revised manuscript, Table 2 has been expanded to include eastward and northward wind. We also added text in the Results section noting that the wind comparisons showed little to no difference between the nudged and ML-corrected simulations, which is itself an informative result.

Location in revised manuscript: Table 2, p. 6; Section 4, p. 8, line 192, sentence beginning “*Not shown are the eastward and northward wind comparisons...*”.

Comment 8: There is no detailed description of the ML models trained, please add. Please also add a description of the out-of-sample novelty detector in the methods section.

Response: We expanded the manuscript to provide a fuller description of the trained ML models in Section 2.3, including their inputs, outputs, and network structure. We also added discussion of the out-of-sample novelty detector and its effect on robustness in the revised manuscript.

Location in revised manuscript: Section 2.3, p. 6 to 7, paragraph beginning “*Following Ai2’s prior work with FV3GFS...*”; Section 4, p. 9, paragraph beginning “*To evaluate the stability and robustness...*”.

Comment 9: How often is the mass clipper applied (in percent)?

Response: We expanded the manuscript to clarify how the mass clipper is used and to discuss its implications for conservation. We did not quantify its application frequency in percent in the present study, but we now note that its cumulative impact would be an important consideration for longer simulations.

Location in revised manuscript: Section 2.4, p. 7, line 151, paragraph beginning “*Special care is needed when applying corrections from the thermodynamic model...*”.

Comment 10: Can the authors make some sense of why they see improvements in some variables and not in others? Please add discussion here, e.g. connected to the extended discussion of the resulting nudging tendencies.

Response: We agree, and we added a more detailed interpretation to the Results section. The revised manuscript now explains that surface radiative fluxes improve most because they are learned directly from the fine-grid reference, whereas variables such as water vapor path and surface temperature depend more strongly on the accuracy of the predicted temperature and humidity tendency corrections. Because those tendencies are systematically underestimated, especially in the tropics, some fields degrade rather than improve.

Location in revised manuscript: Section 4, p. 9, line 199, paragraph beginning “*The pattern of variables that improve or degrade...*”.

Comment 11: Please add a table of all the trainings done, how many runs of them failed? (p. 8)

Response: Rather than adding a separate table, we revised the Results section to describe the multi-seed experiments more explicitly. The revised manuscript now states that three random seeds were tested, that some runs failed without novelty detection because inputs became out-of-sample, and that the novelty detector enabled successful completion of the full-year simulations.

Location in revised manuscript: Section 4, p. 9, line 206, paragraph beginning “*To evaluate the stability and robustness...*”.

Comment 12: Figs. 2, 3, 4 -- please describe clearly what is seen in the caption, not only discussion, and make sure that the labels in the figures are correct (e.g. Delta pressure for bias in pressure)

Response: We revised the captions for Figures 2 to 4 to make them more descriptive and more self-contained than in the original submission. In the revised manuscript, the captions now describe the plotted quantities and intended comparisons more clearly. Note that Figures 2 and 3 from the original manuscript are now Figures 3 and 4. Figure 4 from the original manuscript has been removed.

Location in revised manuscript: Figure 2 caption, p. 10; Figure 3 caption, p. 11; Figure 4 caption, p. 12.

Technical Corrections

Comment 13: The second sentence in the Introduction (p. 1) is not a complete sentence.

Response: Corrected. In the original submission, the sentence beginning “*For instance, predicting shifts...*” was a fragment. In the revised manuscript, this has been replaced by a complete sentence.

Location in revised manuscript: Section 1, p. 1, line 12, sentence beginning “*For instance, they can help predict shifts...*”.

Comment 14: P. 1, “This project aims to develop a computationally efficient machine learning based emulator for SCREAM” seems misleading, since it is not a complete emulator of the model that is developed, just the ML-based nudging.

Response: We agree. In the revised manuscript, this statement has been replaced with more precise wording indicating that the ML component emulates high-resolution SCREAM nudging

tendencies, rather than the full model.

Location in revised manuscript: Section 1, p. 1, line 21, sentence beginning “*This project aims to apply a computationally efficient machine learning...*”.

Comment 15: P. 7, first sentence in the last paragraph is missing “to”: “While it is possible avoid ...”

Response: Corrected.

Location in revised manuscript: Section 3, p. 8, line 178, sentence beginning “*While it is possible to avoid Python...*”.

Comment 16: P. 2, “both approaches have shown successful results, often outperforming state-of-the-art GCMs while only using a fraction of the computational resources” – this statement seems too simplified, please elaborate.

Response: We agree that this sentence was overly broad, and we removed it from the revised manuscript.

Location in revised manuscript: Removed from the second paragraph of the Introduction.

Comment 17: P. 2, first abstract in Section 2: Please add a reference for “maintains good performance across a number of high performance computing systems.”

Response: We added an appropriate supporting citation. The revised manuscript now cites Donahue et al. (2024) in this context.

Location in revised manuscript: Section 2, p. 2, line 43, sentence beginning “*The adoption of Kokkos in EAMxx unlocks...*”.