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 #1

The authors would like to thank Reviewer #1 for their in-depth review of our manuscript and the several suggestions which we believe have improved the paper. We have made every effort to address each of the reviewer comments and concerns in the revised manuscript. Please find below a list of the reviewer comments and our response.

Specific comments

Comment: Figures and tables: I strongly recommend to revise the captions of all Figures. There are some clear inconsistencies how you name the error metrics in the captions of the figures. e.g. Fig. 2, Fig.3 and Fig.4. Also the shown subfigures and what its is shown needs to be explained more in detail. In addition, I would suggest to remove sentences that are evaluating the results from figure captions.

Response: Thank you for the suggestion. We revised the captions of all figures to make them more descriptive, more consistent in their terminology, and less interpretive. We also clarified what is shown in each panel and improved the table captions so that the purpose of each table is clearer to the reader.
Location in revised manuscript: Figure 1 caption, p. 3; Table 1 caption, p. 5; Table 2 caption, p. 6; Figure 2 caption, p. 10; Figure 3 caption, p. 11; Figure 4 caption, p. 12.

Comment: GCMs vs. neural networks: The use of „model“ for both a GCMs and neural networks complicates the understanding of the manuscript for a broader audiences. In this case you need to find a consistent and clear nomenclature throughout the manuscript. Otherwise it is very hard to understand whether you mean a neural network or a GCM for a reader.

Response: This is an excellent point. We reviewed the manuscript to reduce ambiguity in our use of the word “model” and revised wording where clarification was needed. In particular, we now more explicitly distinguish between the GCM, the atmosphere component, and the ML models in key places throughout the manuscript.
Location in revised manuscript: Revisions are distributed throughout the manuscript. Representative examples include Section 1, p. 1 to 2; Section 2.3, p. 6 to 7; and Section 4, p. 8 to 9.

Comment: Interpretation of the results of the conducted simulation:

Your drawn conclusion from your very short results section are important for the community as they provide helpful information for follow-up experiments from other working groups. Nevertheless your results section needs more description of the results. The shown Figure 2 and 3 provide more details that need to be explained and compared, e.g. the bias in the area of the intertropical convergence zone that led to your conclusions or where the ML-corrected simulation is clearly beneficial, e.g. the shortwave and long wave radiation biases over the Southern Ocean. Also I would suggest to add a figure that shows zonal means in a latitude-height plot of e.g. temperature differences and humidity differences. This will help the reader to understand the drawn conclusion in your discussion and conclusion section.s

Response: Thank you for this suggestion. We substantially expanded the Results section to better interpret the simulation results. In particular, we added a new figure showing zonal-mean target and ML-predicted nudging tendencies for temperature and specific humidity, which helps illustrate the vertical structure of the bias and its strong tropical signal. We also added discussion connecting the underestimation of nudging tendencies in the tropics to the weakening of the ITCZ, and we expanded the interpretation of why some variables improve while others degrade.
Location in revised manuscript: 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 *"The pattern of variables that improve or degrade..."*; Figure 2, p. 10; Figure 3, p. 11; Figure 4, p. 12.

Comment: Citations: Below you will find some suggestion for citations that you may add. I feel it very important to pay attention on the way you are referring to your own work. You are saying, e.g. based on the „AI2's approach", but there you need to add a citation for people that are unfamiliar with previous work. The same applies when you are referring to „FV3GFS" studies or the neural networks you are applying. It would be good to give references in this case for readers that are unfamiliar with existing work.

Response: Thank you for these suggestions. We have added the recommended citations where appropriate and revised several passages so that references are provided when prior Ai2 work, FV3GFS, fv3net, and related ML approaches are introduced.
Location in revised manuscript: Representative additions appear in Section 1, p. 1 to 2; Section 2, p. 2 to 7; Section 3, p. 8; and Section 6, p. 12 to 13.

Technical comments

Comment: Line 3: „designed for heterogenous systems" feels unclear. Here I would like to suggest to add a little bit of guidance for readers that come from a different scientific

background, something along the line „heterogenous high performance computing systems“. Otherwise there may be some misunderstanding whether you still mean the „Earth system“ here from the previous sentence.

Response: Thank you, this is a good point. We replaced this wording with “heterogeneous computing architectures” to make the meaning more precise.
Location in revised manuscript: Abstract, p. 1, line 2, sentence beginning *“SCREAM is a kilometer-scale configuration...* d*esigned for heterogeneous computing architectures”*.

Comment: Line 5: Instead of a „few years“ I would prefer to say „decades“ as some European research projects recently run experiments with km-scale models over a few tens of years successfully, e.g. NextGems (see the Hohenegger 2023 below) or EERIE.

Response: Thank you. We revised this statement to say “a few years to a few decades” and added the suggested reference to acknowledge the broader range of kilometer-scale modeling efforts.
Location in revised manuscript: Abstract, p. 1, line 3, sentence beginning *“While the enhanced accuracy of kilometer-scale modeling...”*; Hohenegger et al. reference added in Section 1, p. 1.

Comment: Line 10: Please harmonize how you refer to „FV3“ or „FV3GFS“ (see line 31) throughout the manuscript.

Response: Thank you for catching this. We revised the manuscript to use “FV3GFS” consistently.
Location in revised manuscript: Abstract, p. 1; Section 1, p. 2; Sections 2 and 5 where FV3GFS is discussed.

Comment: Line 15: Here you may add some references as others draw a similar conclusion in the past, e.g. Hohenegger2023 as a recent example.

Response: Thank you for the suggestion. We added this reference.
Location in revised manuscript: Section 1, p. 1, line 12, paragraph beginning *“Accurate future climate projections are crucial...”*; reference list, p. 16.

Comment: Line 18: Here you may need to add a reference as well, e.g. the Satoh2019 SRM paper

Response: Thank you again for the suggestion. We added this reference.
Location in revised manuscript: Section 1, p. 1, line 18, sentence discussing the inability of coarse-resolution GCMs to resolve storms, clouds, and complex topography; reference list, p. 17.

Comment: Line 24: Here you may add a reference to Eyring2024, as they draw a similar conclusion

Response: Thank you for the suggestion. We added this reference.
Location in revised manuscript: Section 1, p. 2, line 24, paragraph beginning *"ML has the potential ...*"; reference list, p. 16

Comment: Line 29: Here you may add a reference to Hu2025, as this is an example that replaces an MMF setup in E3SM.

Response: Thank you for the suggestion. We added this reference.
Location in revised manuscript: Section 1, p. 2, line 25, sentence discussing ML replacement of specific physical processes of GCMs; reference list, p. 17.

Comment: Line 31: Please introduce the acronym FV3GFS and provide also a reference to a paper here e.g.

Response: Thank you for the suggestion. We introduced the acronym FV3GFS and added the suggested reference.
Location in revised manuscript: Section 1, p. 2, line 31, paragraph beginning *"This study builds on the success..."*; reference list, p. 18.

Comment: Line 45: I would suggest to rewrite this sentence as it may feel potentially slightly subjective. I agree that a throughput of more than 1SYPD is quite a gain in computation efficiency but this does not suffices a detailed model simulation evaluation to claim something as „unprecedented". I would suggest just saying that : „EAMxx achieved to simulate more than a 1 year per day in km-scale simulations (Taylor 2023),..."

Response: This is a good point. We removed the word "unprecedented" and revised the sentence accordingly.
Location in revised manuscript: Section 2, p. 2, line 51, sentence beginning *"While EAMxx achieves computational performance..."*.

Comment: Line 48: I would suggest to add a reference after „AI2's approach (e.g. Bretherton2022)" to give one example reference for people that are new to the topic of hybrid modeling. Also I would suggest to replace „highly accurate training data" with for example just „fine-scale training data" to help the readers get your main point (That you are learning from high resolution data which includes fine-scale structures that a GCM would not be able to simulate directly).

Response: We made both of these changes. We added a reference to the Ai2 FV3GFS work and revised the wording to "fine-scale training data."
Location in revised manuscript: Section 2, p. 2, line 54, paragraph beginning *"While EAMxx achieves computational performance..."*.

Comment: Line 49: This sentence may feel a little bit unclear for readers that are not familiar with the topic. Please give here one example what you modified or stress that you will explain the modifications in the following subsection 2.1 or 2.2.

Response: Thank you for the suggestion. We revised the sentence to make it clearer and now explicitly point the reader to Sections 2.1 and 2.2 for further detail.
Location in revised manuscript: Section 2, p. 2, line 55, sentence ending *"see sections 2.1 and 2.2 for more details."*

Comment: Line 55: I would suggest to write here a „coarse EAMxx simulation is run with inline.." instead of „coarse-grid model" to increase the clarity of the sentence.

Response: We rewrote the sentence accordingly.
Location in revised manuscript: Section 2, p. 3, line 62, paragraph beginning *"The corrective ML approach described in this study..."*, sentence containing *"Finally, a coarse-grid EAMxx simulation is run..."*.

Comment: Figure 1: This figure needs some major revisions. You need to highlight at least the different GCMs (e.g. coarse EAMxx, SCREAM) you are using for each step in such a central introduction figure. Otherwise it is very hard to understand for readers that are only scanning over the manuscript what „Coarse-grid simulation" is. Also try to use a similar offset between the headers of each step and the respective frame. The same applies to the length of the different arrows. Furthermore I would suggest to give more explaining details in the caption of the figure to make it easier to understand for readers. Something a long the line „Flow diagram for the

corrective ML approach, which uses ML predicted nudging tendencies (step 3) to correct coarse scale EAMxx simulations (step 4).

Response: Thank you for these suggestions. We revised Figure 1 to better identify the fine- and coarse-resolution EAMxx/SCREAM simulations used in each step, and we expanded the caption so that the workflow is easier to understand without relying on the main text.
Location in revised manuscript: Figure 1 and caption, p. 3.

Comment: Line 64: Please add a citation here, where interested readers can find this essential information about omitting a deep convection parameterizations regardless of resolutions!

Response: We added the citation to Caldwell et al. (2021).
Location in revised manuscript: Section 2, p. 2, line 48, sentence beginning *"Notably, the SCREAM configuration omits a deep convection parameterization..."*.

Comment: Line 71: Please give some details what Table 1 and Table 2 contains in a dedicated sentence for each table.

Response: We revised the surrounding text and the table captions to better explain what each table contains and how the listed variables are used.
Location in revised manuscript: Section 2.1, p. 4 to 6, especially the sentences introducing Tables 1 and 2; Table 1 caption, p. 5; Table 2 caption, p. 6.

Comment: Line 79: I would delete the „somewhat" between „is" and „different" as it does not add additional information for the reader. Please add a reference at the end of the sentence to make clear against which of the few FV3GFS studies you will draw a comparison in the following sentences.

Response: Thank you for the suggestion. We removed "somewhat" and added the relevant reference.
Location in revised manuscript: Section 2.1, p. 4, line 93, paragraph beginning *"The approach to the vertical remapping in this study is different than in the FV3GFS study."*

Comment: Tables 1, 2: Please add some descriptive information in the header of Table 1 and 2. E.g. What do you mean by „nudged runs" in this context? I guess these are the variables that you use to machine learn nudging tendencies to correct the coarse EAMxx. What is each column

showing? This will strongly help readers to understand what you do and what the purpose of these two tables is. Also I would suggest to illustrate the column headers in bold font.

Response: We agree that the original tables were not sufficiently descriptive. We expanded the captions to explain the purpose of each table and what each column contains.
Location in revised manuscript: Table 1 caption, p. 5; Table 2 caption, p. 6.

Comment: Line 90: Please homogenize how your are referring to the „prior FV3GFs work“ throughout the manuscript. What do you mean with „coarsened reference solution evolution“? Try to clarify what you mean here, please.

Response: Thank you for this suggestion. We clarified the wording and now describe this as a coarse-resolution 100 km simulation nudged to track the vertically interpolated and horizontally coarsened reference solution evolution.
Location in revised manuscript: Section 2.2, p. 5, line 106, paragraph beginning *“Similar to the work with FV3GFS...”*.

Comment: Line 95: This is a little bit redundant information about the nudging time step of 3 hrs. You may very well delete this sentence.

Response: We agreed that the original phrasing was awkward and revised the text so that the nudging timescale is incorporated more naturally into the preceding sentence.
Location in revised manuscript: Section 2.2, p. 5, line 106, paragraph beginning *“Similar to the work with FV3GFS...”*.

Comment: Line 97: Please say which „model“ you are meaning here, your ML model a coarse version of EAMxx?

Response: We revised the sentence to remove this ambiguity.
Location in revised manuscript: Section 2.2, p. 5, line 111, sentences describing interpolation to the EAMxx pressure coordinates at each timestep.

Comment: Line 106: You need to add a link with a dedicated reference to the fv3net that follows the standards of Copernicus journals here. Please add references also throughout your manuscript when you are referring to that package.

Response: Thank you for the suggestion. We added a Zenodo software reference for fv3net and cite it in the manuscript.
Location in revised manuscript: Section 2.3, p. 6, line 121, sentence beginning *"All models utilize Ai2 Climate Modeling Group's open-source fv3net Github repo..."*; Code and data availability section, p. 13.

Comment: Line 110 - 121: If these models are designed in prior published FV3GFS paper, please give references for each of the models. Something along the line „This model configuration is based on XYZ."

Response: We added references in Section 2.3 and expanded the description of the model configurations, including the dense network architecture used in this study.
Location in revised manuscript: Section 2.3, p. 6 to 7.

Comment: Line 120: I would suggest to write „physical consistency" instead of „physical realizability" as this wording is more common in climate modeling.

Response: Fixed.
Location in revised manuscript: Section 2.3, p. 7, line 137, sentence beginning *"We apply output limiters to ensure physical consistency..."*.

Comment: Line 134: Please explain what you mean with „mass clipper" in a few words otherwise it is hard to understand for readers that are not working on convection parameterizations, what you did here and why.

Response: We revised the sentence to clarify that the mass clipper sets negative values to zero.
Location in revised manuscript: Section 2.4, p. 7, line 151, sentence beginning *"A mass clipper, in which negative values are set to zero..."*.

Comment: Line 159: I would suggest to add a „to" between „possible" and „avoid" here.

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

Comment: Line 166: It would be great to show this underestimation of nudging tendency also visually in a potential supporting information figure. Are these biases occurring along the entire column or are they more pronounced on lower model levels? Also does these biases reflect some regional patterns? Such additional figures would help to improve this rather short paragraph.

Response: Thank you for the suggestion. In response, we added a new figure showing the zonal mean of target and ML-predicted nudging tendencies for air temperature and specific humidity over the one-year free-running simulation. This figure shows that the ML model generally underestimates the target tendencies and that this underestimation is most pronounced in the tropics.
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..."* and *"As shown in Figure 2..."*.

Comment: Line 167-172: I would suggest to elaborate a little bit, what you are showing in Figure 2 and Figure 3. Please clearly state what is show in these figures. Is it just the difference of annual means, that looks so different or is it the mean over all differences along the time dimension. Also you need to explain how you compute the bias (please use this term also in the respective headers of the subplots) and the RMSE. Are these quantities area-weighted, taking into account Earth sphericity, or just an unweighted average over all grid cells. This needs to be clearly stated before interpreting the results. In general I would suggest to highlight that your EAMxx simulation with ML nudging introduce some regional biases. Especially the general weakening of the Intertropical convergence zone and also the reduction of precipitation along the boreal and austral storm tracks deserves a clear mention in the text.

Response: We expanded both the figure captions and the Results section to make these diagnostics clearer. The revised captions now better describe what is plotted, and the caption for the spatial error figures now states that the RMSE is calculated as a global area-weighted mean. In the main text, we also added discussion of the regional biases, especially the tropical underestimation of nudging tendencies and the associated weakening of the ITCZ.
Location in revised manuscript: Figure 3 caption, p. 11; Figure 4 caption, p. 12; Section 4, p. 8 to 9, including the paragraph beginning *"As shown in Figure 2..."*.

Comment: Figure 2: You need to explain what you mean by „annual mean simulation error" as it is not introduced in the main text. Please explain all shown variables in the caption of Figure 2, e.g downwelling LW flux at the surface (first row). I would suggest to revise the second sentence as it may feel subjective given the regional biases.

Response: We revised the figure caption to more clearly describe the plotted fields and to reduce interpretive wording. Please note that in the revised manuscript the figure numbering has changed relative to the original submission, so the relevant spatial error figure is now Figure 3.
Location in revised manuscript: Figure 3 caption, p. 11.

Comment: Figure 3: Please revise the caption of this figures as discussed above. I would suggest to begin what is displayed here. You need to homogenize how you name the displayed difference metric, here you are writing „annual mean spatial patterns“, which may confuse readers. I would strongly suggest to use one name for the water vapor path throughout the manuscript (see Line 172) and especially in this Figure. Also I would suggest to leave out any model interpretation out of the caption of a Figure. You may very well discuss these things in the main text.

Response: We revised this caption accordingly. In the revised manuscript, the relevant figure is now Figure 4, and the caption has been made more descriptive and less interpretive. We also made the variable naming more consistent throughout the manuscript.
Location in revised manuscript: Figure 4 caption, p. 12; Section 4, p. 8 to 9.

Comment: Line 174-180: I like that you did these tests. Though, I would shortened this paragraph and move the largest portion of it into the supporting information or an appendix, as it is difficult to understand at this location. You may need to add more explanation of Figure 4, as it is very hard to understand, what this figure shows and also why you want to display it in your manuscript. Maybe I am missing something, but why are two random seeds using the novelty technique and one not? Why do you pick e.g. the downwelling LW flux at the surface? What is the effect of the novelty technique on other variables? If you want to keep this paragraph you need to expand the interpretation. E.g. Which dependence on random seed and the use of the novelty method is visible? Where do you see regional differences between the random seeds (I see relatively large similarities over most areas)?

Response: Thank you for this suggestion. We streamlined this discussion substantially in the revised manuscript. Rather than centering the discussion on a separate figure for the random-seed comparison, we now keep the focus on the main robustness result, namely that some seeds failed without novelty detection because of out-of-sample inputs, and that the novelty detector enabled successful completion of full-year simulations. We also note that the ML-predicted tendencies were similar across random seeds.
Location in revised manuscript: Section 4, p. 9, line 206, paragraph beginning *“To evaluate the stability and robustness...”*.
Note: The current revised manuscript still includes a Figure 4, but it is now used for a different main-results figure than in the original submission.

Comment: Figure 4: Please revise the figure caption. Explain clearly what is shown. Here again you use a different terminology as in Figure 2 and Figure 3. Also I would personally move this figure to the supporting information as it does not show new general results or helps with the general understanding. Maybe you can move the shortened random seed paragraph and the reference to this figure to the discussion section as this would showcase that you did some

sensitivity experiments? This may helps with the flow and improves the general structure of the paper.

Response: Thank you for this suggestion. The figure numbering and content have changed in the revised manuscript. The current Figure 4 now presents one of the main results rather than the earlier random-seed comparison, and its caption has been revised for clarity and consistency. The original Figure 4 has been removed.
Location in revised manuscript: Figure 4 caption, p. 12.

Comment: Line 182 - 187: This paragraph is very clear and I fully agree with what you are writing. Here you may very well add that you explored the transferability of ML parameterizations from one GCM to the next, which is an important and key step for future advances in Earth System Modeling as working groups will depend on integrating ML machinery from other groups. I am thinking here of coupled simulations where people add ML atmospheric and ML ocean parameterization together. So this project was a success in my eyes as you were able to conduct stable model runs for some random seeds despite your stated model deficiencies.

Response: We thank the reviewer for this helpful suggestion. We added text at the beginning of the Discussion to emphasize that this study investigates the transferability of an ML parameterization from one GCM to another, which we agree is an important issue for the broader earth system modeling community.
Location in revised manuscript: Section 5, p. 9, line 213, paragraph beginning *"This study investigated the transferability of an ML parameterization from one GCM to another..."*.

Comment: Line 189-190: The point that EAMxx has no convection parameterizations is a very essential and integral point of your paper. Please try to convey this also at the begin of your results section. This will also help you to address my general circulation point above. I would suggest to add also here in the discussion regions where a usual deep convection scheme is important and hence you would expect biases of your ML algorithm.

Response: Thank you for the suggestion. We added discussion in the Results section linking the strongest ML biases to the tropics and explicitly noting that EAMxx does not include a deep convection scheme, which may help explain why the tropical biases are especially large.
Location in revised manuscript: Section 4, p. 8 to 9, line 194, paragraph beginning *"As shown in Figure 2..."*.

Comment: Line 193-194: This sentence is a little bit unclear to me and needs to be revised. I guess you are stressing in the second part of the sentence, that EAMxx run on coarse resolutions

need some sort of auxiliary convective parameterizations as such processes can not be directly simulated.

Response: We revised this sentence for clarity. The revised manuscript now states that potential improvements in the ML model from the stronger signal may be too small to offset the larger biases in EAMxx associated with the absence of a subgrid convection scheme.
Location in revised manuscript: Section 5, p. 9, line 221, paragraph beginning *"There are several possibilities for why corrective ML was less successful..."*.

Comment: Line 196-199: Here you may add again a reference to one Ai2 paper, as people may be lost that are only scanning over the discussion and conclusion.

Response: We added a reference to the prior Ai2 work in this part of the Discussion.
Location in revised manuscript: Section 5, p. 10, line 234, paragraph beginning *"As discussed in the Methods section..."*.

Comment: Line 199: You may delete „anecdotal"

Response: We removed this word.
Location in revised manuscript: Section 5, p. 10.

Comment: Line 203: Which models are you meaning here, the GCM or the ML algorithm? I would call them by name, as this will help with the understanding of this paragraph.

Response: We revised the wording in this discussion to be more explicit about which component is being referred to.
Location in revised manuscript: Section 5, p. 10 to 11, especially the paragraph beginning *"Adoption of the corrective ML framework involved the development..."*, line 240.

Comment: Line 204 - 208: Please add a reference to your EAMxx github page here. It cannot hurt to remind readers that is openly accessible.

Response: We added a reference to the E3SM GitHub repository in this discussion.
Location in revised manuscript: Section 5, p. 10 to 11, paragraph beginning *"Adoption of the corrective ML framework involved the development..."*. Line 240

Comment: Line 212-215: Thanks for conducting these experiments. If you have results of these simulations available, you may include one figure in the supporting information that shows differences between the vertical and no vertical interpolation experiment. Again please pay attention to the use of „model“ here.

Response: Thank you for the suggestion. Because these experiments were preliminary, we chose not to add a figure, but we revised the text to make clear that the analysis was preliminary and that a more detailed investigation of the vertical interpolation issue would be appropriate for future work.
Location in revised manuscript: Section 5, p. 11, line 246, paragraph beginning *“Finally, as outlined in the Methods section...”*.

Comment: Line 218 and thereafter: Please add references in the conclusion section for the readers that only have a look over the conclusion section.

Response: We added references in the Conclusion where appropriate.
Location in revised manuscript: Section 6, p. 12 to 13.

Comment: Line 223: What do you mean by „doubly-periodic use cases“? You may add here an example to help to understand this sentence.

Response: We added a citation to Bogenschutz et al. (2025), which provides context for this use case.
Location in revised manuscript: Section 6, p. 12, line 262, sentence ending *“useful for debugging and hypothesis testing Bogenschutz et al. (2025).”*

Comment: Line 231: I would include here that Kochkov24 show that “it has accurate weather forecast and improved simulation skill of some climate variables.“

Response: We revised this passage so that the discussion of improved weather forecast and climate skill is more clearly tied to Kochkov et al. (2024).
Location in revised manuscript: Section 6, p. 12 to 13, line 270, paragraph beginning *“There is increasing interest in the earth system modeling community...”*.

Comment: Line 235: You may need to revise this section to fulfill Copernicus journal standards. Just replacing all links here with citations that point in your references to the software and data would be sufficient, I guess. This webpage is helpful in this respect (https://www.geoscientific-

model-development.net/policies/code_and_data_policy.html) or you may have a look over recently published GMD papers.

Response: Thank you for pointing this out. We substantially revised the Code and data availability section to better align with journal expectations. The revised version now includes archival software and dataset references, repository locations, relevant commit hashes, and data-access information.
Location in revised manuscript: Code and data availability section, p. 13, paragraph beginning *"We provide a Zenodo archival repository..."*.

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..."*.