Reply to Reviewer 2

This manuscript describes the process of retuning of version 1 of the atmospheric component EAM of E3SM climate model, which focused on parameters related to various cloud processes, and how the retuning impacted a range of quantities beyond the tuning targets, ranging from surface temperature in the present-day to aerosol forcing (via cloud adjustments) to cloud feedbacks.

General comments

The manuscript does a good job of explaining the reasoning that led to the strategy used in retuning, which relies on the bet that improving the representation of clouds will lead to improvements across the board. It's nice to see that the bet pays off.

Reply: We thank the reviewer for the positive comment.

There's rather a lot of detail in section 2, describing groups of parameters were tuned. There's really a lot of detail in section 3, which explains how the returned model behaves with respect to a wide range of emergent phenomena. There is so much detail, in fact, that the manuscript works much better as a description of what was done than it does as an explanation of what was learned. If the authors' goal is to document the strategy and its impacts they have succeeded, but if they aim to influence the ways in which readers undertake or understand model tuning they would be well advised to bring their ideas into sharper focus. Sharpening the manuscript will almost certainly involve relegating material to appendices or supplemental material.

Reply: We have considered the reviewer's request to reduce length but respectfully decided to stand by our instinct that to be sufficiently comprehensive about this challenging topic it is important to show all of these diagnostics. We also note that there is precedent in GMD for such technical details for model development and documentation. This paper attempts to carefully document the approach, strategy, and results of model calibration, including the strengths and weaknesses of the approach and results. Furthermore, since most of the recalibration is part of the E3SMv2, this paper serves as a documentation of improvements to E3SMv1 that lead to the new model.

A large proportion of the very many figures are of the form of Figure 4: six (well-constructed) maps showing the difference of v1 against observations, four maps showing the change induced by the four sets of parameter changes, then a map showing the aggregate change of the final retuning relative to the original. Readers are left to judge improvement by mentally subtracting the bias in the upper left plot from the change in the lower right plot. Could the lower plot be revised to show, for example, the improvement or degradation in the original bias as a result of the tuning?

Reply: A primary goal of this study is to assess the impacts of parameter changes on the simulated climate. We have considered the reviewer's request and experimented the suggested way to present results, but we found that this does not show the impacts of the parameter changes clearly. Therefore, we believe that showing the differences between the default and the recalibrated model is the best way to present the results.

Versions of the six-ma figure (e.g. Fig 7) with no observational constraint are harder for readers to assess.

Reply: Indeed there is no observational constraint for the global climatology of PBL decoupling strength (Fig 7) and cloud-top entrainment efficiency (Fig 8). However, they provide physical insights into the PBL characteristics and low clouds. By comparing them between different model configurations, we assess how the parameter changes affect PBL and low-level clouds, which is a primary goal of this study.

Tuning relies on the ability to measure improvements in simulations, normally relative to observations. It's remarkable that the authors spend essentially no time discussing the sources of their observational constraints, or how uncertainty in these constraint is or isn't considered as part of the tuning strategy.

Reply: We acknowledge that the uncertainty associated with observational datasets varies. We have tried to minimize the impact of observational uncertainties by using cloud simulators (Fig 5) and by accounting for the satellite sampling strategy (Fig 12). We have also added observational constraints, where available, in Table 6, and the text "Satellite observations summarized in Stubenrauch et al. (2013) and Neubauer et al. (2019) are also provided but we note that it is dangerous, and can be misleading, to compare model state variables with satellite retrievals without using a simulator since large retrieval and sampling uncertainties exist.".

The tuning strategy used by the authors is somewhat traditional. Comparisons to other approaches (e.g. the automated calibration to process-scale constraints used by HiTune, doi:10.1029/2020MS002217 or the formal inference discussed in the Clima project, doi:10.1016/j.jcp.2020.109716) would no doubt be welcome.

Reply: We have added a paragraph to discuss the potential of using automated machine learning approaches for model calibration in the end of the paper:

"It is natural to wonder if an equivalent or superior ESM calibration might have been achievable with less human effort or fewer computational resources via semi-automated machine learning (ML) methods that emulate or expand the workflow outlined in this paper. Indeed, emulating a complex model's parameter sensitivities following human constructed trial simulations to aid model calibration and uncertainty quantification would be an intriguing possibility. Several recent studies have shown successful application of ML methods in model calibration (Cleary et al., 2021;Dunbar et al., 2021;Couvreux et al., 2021;Hourdin et al., 2021). In theory, reinforcement learning (RL) with an appropriately formulated agent-based optimization system could be guided via its loss function formulation with skill metrics that optimize for the same patterns and mean state climate metrics that we prioritized in this study. In practice, however, this ML task faces a fundamental challenge that the cost of an individual agent-reward sample is performing multi-year climate simulations. The workflow outlined in this paper has the considerable advantage that experienced human experts make educated parameter interventions based on assessment of the simulation that discriminates desired effects in a nuanced way and tolerates certain unintended consequences. It is not clear how available ML methods could be infused with analogous physical foresight to make similar decisions, and thus logical to expect they would require more evaluation samples to succeed via brute force. Therefore, experimenting with clever strategies to increase reward density and to integrate physical knowledge from experts in the ML workflow would be a highly worthy long-term challenge."

More specific comments

Tuning of course involves the changing of specific parameters. The variable names in the specific computer code are perhaps too specific to be in the main text. This information, and indeed probably the original and changed values, could be summarized in one or more tables in an appendix.

Reply: A primary purpose of this paper is to document the approach, strategy, and results of model calibration, including the strengths and weaknesses of the approach and results. Furthermore, since most of the recalibration is part of the E3SMv2, this paper serves as a documentation of improvements to E3SMv1 that lead to the new model. Therefore, we place the tables in Section 2.

Line 113: the current term of art is "perturbed parameter ensemble".

Reply: We have changed “perturbed physics ensemble” to “perturbed parameter ensemble”.

The subsections of section 2 are labeled as tropical clouds, low clouds, etc. In practice each section might also be categorized according the scheme whose parameters are being tuned. Indicating this (e.g. “Tropical clouds and the deep convection scheme”) might guide readers’ attention.

Reply: We label the subsections by the cloud regime/type because there are multiple parameterizations that affect one cloud regime/type. While the reviewer’s comment is well-received, we choose not to introduce another layer in the sub-sections but we have revised the text to better introduce the recalibration process.

Section 2 describes the re-tuning in detail, including which parameters are re-tuned and why. General material (e.g. line 286-294) should be deferred or removed.

Reply: Line 286-294 describes how the skewness in CLUBB is formulated and re-tuned. We think this is a very important piece of information because retuning the skewness significantly reduces the outstanding bias of thin marine stratocumulus and overly bright shallow cumulus, as described in Section 2.2.

Line 220: The authors scale the temperature variance provided by one scheme by a factor of 2 before introducing it in another scheme. It’s not clear whether this is a reasonable physical assumption. It if is the choice should be justified; if it’s not the choice should be explained.

Reply: We have added the sentence to provide an explanation: “Based on sensitivity tests, a scaling factor of 2.0 was introduced to enhance the effect so that the simulated tropical clouds are in better agreement with observations (as discussed in Section 3)”

Section 2.2: how are the different cloud regimes identified in practice, during a simulation?

Reply: We used the vertical velocity at 500hPa and the lower tropospheric stability (Medeiros and Stevens, 2011). The geographical distribution is also very useful. We have added the text in Section 2.2.

One reason for showing six panels in Figure 4 and its many analogs is to highlight the geographic distribution of the impacts of parameter changes. The authors might ask themselves if maps are the best way to show these differences in all cases.

Reply: We do want to reveal the impacts of parameter changes on the simulated climate around the world, including both the positive and negative impacts. Therefore, we chose to show the global maps.

Line 395-400 could be edited for clarity and to remove general material, as could lines 486-490.

Reply: Because the readers might not be familiar with the approach, we think it is necessary to provide the description of the experiments.

Line 528: EIS is thought to control low cloud properties, not their feedback (sensitivity to surface temperature change).

Reply: We have changed the text to “EIS has traditionally been considered as an important cloud controlling factor affecting low clouds and low cloud feedback (Klein et al., 2017;Myers et al., 2021).” For instance, Myers et al. (2021) state that “The strongest individual components of the [low-cloud] feedback in almost all regions globally are those due to SST and EIS (Supplementary Fig. 6), owing to the large forced responses of these cloud-controlling factors".

The central point of lines 715-728 could no doubt be made more compactly and directly.

Because the readers might not be familiar with the approach, we think it is necessary to provide the description of how the ERFs are computed from the model.

Line 739-740 are bewildering.

Reply: We have changed the text to "Both longwave and shortwave radiation affect surface temperature and atmospheric cooling rates, which govern the hydrological cycle."

Figure 13 is hard for readers to interpret. Coding bias with shapes and variables with numbers is really quite unfriendly - bias would be better coded with size or shading, leaving shape to stand for quantities. But readers will also appreciate guidance in interpretation, since all the panels look very much the same to an unpracticed reader.

Reply: The Taylor diagram (Taylor, 2001) summarizes multiple aspects of climate simulations in a single diagram. It has long been widely used in atmospheric science, and its format is familiar and fairly standard. We have added the reference and also the text "While EAMv1_CLUBB and EAMv1_MP do not produce different results from EAMv1, we find that the meridional wind at 850 and 500 hPa (coded as number 4 and 7) in EAMv1_SGV and EAMv1_ZM are in better agreement with ERA-5 the normalized standard deviation reduces."

Line 831: Does the present analysis use the specific kernels of Zelinka 2012 and Pendergrass 2018? If so this should be made explicit. If the text refers to the ideas the original papers (e.g. doi:10.1175/2007JCLI2044.1) should be referenced.

Reply: We used the kernels of Zelinka et al. (2012a, b), Zelinka et al. (2013), and Pendergrass et al. (2018) for our analysis. We have revised the text to cite the papers: "In Figure 18, climate feedbacks diagnosed using the Pendergrass et al. (2018) radiative kernel reveal that the non-cloud feedbacks are invariant across different model configurations and that the variation in total climate feedback is due solely to the spread in cloud feedbacks as a result of our parameter and subgrid adjustments. Further decomposing the cloud feedback into its total, shortwave, and longwave components via cloud radiative kernels (Zelinka et al., 2012a, b;Zelinka et al., 2013) indicates that cloud feedbacks are weakened from 0.77, 0.35, and 0.42 W m^{-2} K^{-1} in EAMv1 to 0.47 (-39%), 0.20 (-43%), and 0.27 W m^{-2} K^{-1} (-35%) in EAMv1P."

Figure 18, especially panel a, is not particularly informative, since readers are asked to compare small changes in large numbers introduced by tuning.

Reply: Fig 18(a) shows that "the non-cloud feedbacks are invariant across different model configurations and that the variation in total climate feedback is due solely to the spread in cloud feedbacks as a result of our parameter and subgrid adjustments." Fig 18(b) then shows the decomposition of cloud feedback to identify which specific cloud feedbacks (i.e., shortwave or longwave; amount, optical depth, or altitude) are affected by the parameter changes. We find this analysis very insightful.

Reference

Cleary, E., Garbuno-Inigo, A., Lan, S. W., Schneider, T., and Stuart, A. M.: Calibrate, emulate, sample, J Comput Phys, 424, ARTN 109716, 10.1016/j.jcp.2020.109716, 2021.

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., Rio, C., Audouin, O., Salter, J., Bazile, E., Brient, F., Favot, F., Honnert, R., Lefebvre, M. P., Madeleine, J. B., Rodier, Q., and Xu, W. Z.: Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration Tool for Parameterization Improvement, J Adv Model Earth Sy, 13, ARTN e2020MS002217, 10.1029/2020MS002217, 2021.

Dunbar, O. R. A., Garbuno-Inigo, A., Schneider, T., and Stuart, A. M.: Calibration and Uncertainty Quantification of Convective Parameters in an Idealized GCM, J Adv Model Earth Sy, 13, ARTN e2020MS002454, 10.1029/2020MS002454, 2021.

Hourdin, F., Williamson, D., Rio, C., Couvreux, F., Roehrig, R., Villefranque, N., Musat, I., Fairhead, L., Diallo, F. B., and Volodina, V.: Process-Based Climate Model Development Harnessing Machine Learning: II. Model Calibration From Single Column to Global, J Adv Model Earth Sy, 13, ARTN e2020MS002225, 10.1029/2020MS002225, 2021.

Klein, S. A., Hall, A., Norris, J. R., and Pincus, R.: Low-Cloud Feedbacks from Cloud-Controlling Factors: A Review, Surv Geophys, 38, 1307-1329, 10.1007/s10712-017-9433-3, 2017.

Medeiros, B., and Stevens, B.: Revealing differences in GCM representations of low clouds, Clim Dynam, 36, 385-399, 10.1007/s00382-009-0694-5, 2011.

Myers, T. A., Scott, R. C., Zelinka, M. D., Klein, S. A., Norris, J. R., and Caldwell, P. M.: Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity, Nature Climate Change, 11, 501-+, 10.1038/s41558-021-01039-0, 2021.

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stier, P., Partridge, D. G., Tegen, I., Bey, I., Stanelle, T., Kokkola, H., and Lohmann, U.: The global aerosol-climate model ECHAM6.3-HAM2.3-Part 2: Cloud evaluation, aerosol radiative forcing, and climate sensitivity, Geosci Model Dev, 12, 3609-3639, 10.5194/gmd-12-3609-2019, 2019.

Pendergrass, A. G., Conley, A., and Vitt, F. M.: Surface and top-of-atmosphere radiative feedback kernels for CESM-CAM5, Earth Syst Sci Data, 10, 317-324, 10.5194/essd-10-317-2018, 2018.

Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Di Girolamo, L., Getzewich, B., Guignard, A., Heidinger, A., Maddux, B. C., Menzel, W. P., Minnis, P., Pearl, C., Platnick, S., Poulsen, C., Riedi, J., Sun-Mack, S., Walther, A., Winker, D., Zeng, S., and Zhao, G.: Assessment of Global Cloud Datasets from Satellites: Project and Database Initiated by the GEWEX Radiation Panel, B Am Meteorol Soc, 94, 1031-1049, 10.1175/Bams-D-12-00117.1, 2013.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram., J Geophys Res-Atmos, 106, 7183-7192, Doi 10.1029/2000jd900719, 2001.

Zelinka, M. D., Klein, S. A., and Hartmann, D. L.: Computing and Partitioning Cloud Feedbacks Using Cloud Property Histograms. Part I: Cloud Radiative Kernels, J Climate, 25, 3715-3735, 10.1175/Jcli-D-11-00248.1, 2012a.

Zelinka, M. D., Klein, S. A., and Hartmann, D. L.: Computing and Partitioning Cloud Feedbacks Using Cloud Property Histograms. Part II: Attribution to Changes in Cloud Amount, Altitude, and Optical Depth, J Climate, 25, 3736-3754, 10.1175/Jcli-D-11-00249.1, 2012b.

Zelinka, M. D., Klein, S. A., Taylor, K. E., Andrews, T., Webb, M. J., Gregory, J. M., and Forster, P. M.: Contributions of Different Cloud Types to Feedbacks and Rapid Adjustments in CMIP5, J Climate, 26, 5007-5027, 10.1175/Jcli-D-12-00555.1, 2013.