We appreciate the useful comments. Reviewer comments are in black and our responses in blue, with **bold** indicating where we have made a change to the paper.

The opinion paper by Carslaw et al. presents a well layed-out argumentation for leaving space for a PPE component in future model development. It presents a very interesting overview of PPE work in the field and I appreciate the effort of the authors to land on several recommendations for how to use PPEs. This paper serves its purpose to provide an "opinion" in the field of climate model development and beyond.

My suggestions for improving the paper are minor:

To the end of the abstract the authors mention, that PPEs should be prioritized when allocating computing resources. I am not sure what prioritize here means. Giving 1st priority to PPEs is to my opinion a too strong wording, since other methods and workflows using climate models may claim priority for good reasons. Preparing a bug-free, multi-purpose ESM model code may need considerable computing resources, often not appreciated by funding agencies neither. And, while having great potential, PPEs as such do not remove structural error or provide scenario simulations for different futures. Not all problems require a PPE. See also another "opinion" paper by Jones et. al, https://doi.org/10.5194/esd-15-1319-2024, on the use of ESMs and improved cooperation to develop them. A word of caution when mentioning priorities would be to the advantage of the paper.

We appreciate this comment. We decided to just delete this statement about priority.

Missing is possibly also a discussion of the challenges of PPEs. Why haven't PPEs been used more often? There are obstacles for that. The PPE implementation in models, efficient launch scripts, the large amount of data, the demanding handling of a lot of data, the waste of resources on implausible model variants, all these are challenges. But I support very much that model development teams should consider the use of PPEs in their model development workflow, as opposed to only one-at-a-time-testing of parameter choices for tuning and model improvement.

We have added some text on this in the paragraph on point 1 where we recommend operationalisation of PPEs. This is where we said that barriers to wider adoption have been tackled… We now say:

"*There have been several barriers to wider adoption of the PPE approach, such as the challenge of selecting appropriate parameters and perturbation ranges (which often requires input from multiple developers), designing the simulations, implementing the perturbations in models, defining efficient workflows and simulation submission scripts, and the large data volumes that are produced. However, we estimate that PPEs have now been developed across a wide range of model types by over twenty research groups in small research teams and large modelling centres, which has reduced some of the knowledge barriers. Barriers to wider operationalisation have been successfully tackled at several modelling centres with streamlined PPE workflows (Elsaesser et al., 2025; Yarger et al., 2024).*"

Line 225 invites for asking me to add another word of caution: "PPEs provide the only means to disentangle structural and parametric causes of model–observation biases". Single process investigation, varying one at a time parameter variation, has been used efficiently in the past to include more correct and important processes in climate models. Evaluation with multiple

observations has been shown to be useful to find structural uncertainty of models, without a PPE.

Yes, this is correct. We could add that it's the only way to *rigorously* disentangle the causes. Perturbing one parameter at a time may or may not identify the best new process to include. In some cases, the link between a model-observation bias and a missing process might be straightforward, but there are any cases where this is not the case. For example, in the aerosol world, too-low particle number may be caused by missing nucleation mechanisms or by incorrect sinks. But we agree that a PPE is not the only way.

We have rephrased this sentence to say "PPEs provide a very effective way to disentangle…"

Another "missing":

The importance of parameterisation documentation. When exploring model diversity and comparing model sensitivities across models, understanding the code differences and details of parameterisation choices has been a long standing challenge in all MIPs. It becomes even more important when doing multi-model PPEs. How to do such documentation efficiently is still a challenge. It might be, though, a positive side effect of organising multiple model PPEs, that such documentation becomes more clear, apparent and accessible for understanding model differences. The authors state in the conclusion: "We started this opinion piece by pointing out that there are several essentially competing efforts in climate modelling – complexity, resolution and initial condition ensembles. To this list we add perturbed parameter ensembles." I believe transparent model documentation should be added to this list, in particular when thinking also of human resources needed to do the modelling.

This is a good point, although we don't think the best place for this is in the coda of the paper.

We have added a sentence in recommendation 2 related to MMPPEs: "*Multi-model PPEs (MMPPEs) will require an improved level of parameterization documentation, although efforts to design MMPPEs may naturally bring this about.***"**