Response to Reviewer 1

We wish to thank this Reviewer for their positive comments and their suggestions to expand our discussion. We have made more detailed comments below in this respect.

Comment

I'm not sure if it belongs in the introduction, or discussion, but I am struck by the relationship between this approach and calibration approaches which use an emulator to relate model parameters to (potentially multiple) model outputs (c.f. Watson-Parris et al. 2022) in order to reduce model biases. Such emulators allow estimation of the sensitivity of particular outputs to given inputs (essentially feature importance, see e.g. Lee et al. 2013), and have recently been used with multiple observables (LWP, Nd, etc) to constrain ERF_aci (Regayre et al. 2023). I feel the authors' approach adds a useful step in relating the biases to observables, and this might help alleviate the difficulties in that work in choosing the 'best' observations to use for bias (or uncertainty) reduction. Generally, since the authors perform a small perturbation study I think it would be useful to more explicitly link the approach to model tuning and discuss how it can help improve the process.

Thank-you for highlighting these studies. We have now included some comparison to the study presented in Watson-Parris et al. (2022) in our introduction. We have also been following the work of the perturbed parameter ensembles such as that of Regayre et al. (2021 and 2023) with interest, and as a group have discussed potential ways to extrapolate our methods into this space. We have now included some discussion on this topic in the introduction and conclusion.

Line 54: 'Another way to evaluate and in some instances tune models is to explore parameter uncertainty (Lee et al. 2013, Regayre et al. 2021, 2023). In these cases the parameter space (the range of plausible values) and their impacts in global climate models are emulated with more simplified statistical models. This allows re-sampling over a range of multi-parameter values many times over what is possible with physically driven models. From these large samples, the uncertainty attributed to particular parameters can be identified and the best combination of parameter values can be constrained based on comparisons with observations. These methods present a powerful way of reducing uncertainty of climate models within known and quantified parameters and physical mechanisms.'

Line 77:'Of particular note is the application of machine learning to climate emulation, i.e. emulating the global response of complex climate models, as outlined in [**?**]. Climate emulation has typically used simple models to estimate what the response of the climate (usually temperature), may be to changes in forcings. These models tend to not capture spatially varying and non-linear processes well, whereas machine learning has been shown to do well in this space, but has been challenged by a lack of data for training purposes. [**?**] have now provided a dataset and some initial machine learning frameworks designed specifically for training models for this application, which may provide a new way to determine possible climate responses to changes in forcings, beyond that of the temperature.'

Line 603: 'This finding is particularly relevant to methods used to constrain models to observationally plausible values, such as that done in Regayre et al. (2021, 2023). In these studies, perturbed parameter ensembles (PPEs) have been used to sample distributions of many parameters, after which, observations are used constrain the model to internally consistent and plausible values. These studies, which provide an efficient and comprehensive way to both evaluate and tune model parameters, use huge arrays of data representing complex changes in the model. While alternative methods to determine feature importance have been implemented in these studies, an approach such as the one presented in this work would provide an efficient way to interpret the effects of the parameter tuning. We further note recent developments in SHAP in which multiple targets can be predicted and evaluated, potentially providing a significant advantage for studies using PPEs.'

Minor typos

L257: 'but' - 'by'
We have fixed the above as suggested.

L404: understating - understanding
We have fixed the above as suggested.