Review of "Calibrating Climate Models Using Inverse Methods: Case studies with HadAM3, HadAM3P and HadCM3"

General comments

The manuscript discusses using Gauss-Newton line-search algorithm in optimising/tuning the atmospheric and coupled versions of Hadley Centre model. The optimisation uses a multi-criteria target, that includes 5 additional constraints to the authors' previous work. Optimisation is done for HadAM3-model by perturbing 7-/14-parameters. Similarly, parameter optimisation for HadAM3P-model is done by perturbing 7-/13-parameters. Additionally, tests are done for only perturbing a subset of the parameters at once. The optimisation seems to work well for the 7-parameter cases, but less so when the number of simultaneous parameter perturbations is increased. The results show the the GN line-search algorithm is able to optimise the parameter values in the training set, but also produces atmospheric models that can be used in coupled climate runs.

The manuscript is well written, and the results are well presented and interesting. The authors have done a good job in presenting numerous aspects of both the optimisation process as well as how the optimised parameters impact the model in climate runs. I have quite a few comments, but they probably won't require too much work. I am suggesting a "minor revision" for the manuscript.

Specific comments

Eq.3: Could you elaborate on "e^i is the ith coordinate vector", the Nocedal and Wright book is not readily available.

P5 L15: "...perturb towards the middle...".
Perturbing always towards the centre is most likely a good idea, but would the algorithm converge if the optimal parameter values were at the edges of the "allowed" space? Just thinking about hypothetical cases where the default parameter values would be ill set. (Just interested in your thoughts about this, not necessary to add any text.)

P5 L22: "several steps", how is this defined? Is the number of steps dynamic or static? Linked to this, could you make sure that you include the total number of steps also when you talk about how many iterations it took for the algorithm to converge/stop (e.g. P10 L3).

Eq. 5: Isn't the additional constraint a double penalty for radiation?

P7: Did you try to run the 14-parameter case with the same step size for the 7 parameters as you used in the 7-parameter case? You are using much smaller steps in the 14-parameter case for the parameters that you later identify as being the most dominant in the cost function (ENT, RHC, CT, …). (Again, just interested in your thoughts.)

P9 L7: What are the "successful parameter sets"? And how do the ensemble members differ from each other?

P9 L31: "extreme limits". This is probably linked to the choice of always perturbing towards the middle? Have you tried/thought of an algorithm design, where you would lose the constraint of perturbing towards the middle and start the estimation closer to the default parameter values? This way the number of iterations might get smaller, or alternatively you could decrease the steps size (and algorithm termination criteria) and try to find the "exact" minimum of the cost function.

P11 L23-27: Not sure if this is a fair conclusion, the 14-parameter cases were much worse in performance than the 7-parameter cases (i.e. maybe a better constrained cost function could improve the 14-parameter converge at the cost of requiring more iterations/evaluations).

P14 L11-15: I don't understand why you are comparing against the control only? Why not do this comparison against each parameter sets own individual coupled runs?

P15 L11: Your cost function is area based, why would the extra-tropics and tropics offset one another?

P15 L11: Aren't CT and CW almost the same as ICE_SIZE?

P17 L30: I would argue against drawing any conclusions based on numerical/toy model experimentation. In my opinion, parameter estimation/optimisation in GCMs is definitely not a smooth problem!

P18 L1: "models that appear similar." Is this only in the target criterion sense? There have to be differences between the models in some fields, no?

Technical comments
P2 L19: double "parameters in the cloud scheme"

P4 L20: define *S* and *O* here already

P13 L1: "initial random" -> "initial extreme random"

P14 L4: "All four cases", a bit confusing, took me a while to understand it was 2 cases from 7-parameter cases + 2 from the 14-parameter cases.

P15 L19: "For these 6..." ?

P17 L23-24: "Given the sensitivity..." too long sentence, please rephrase.