Dear Professor Sun,

We have uploaded our revised manuscript. We apologise for the long delay, but we found the comments and discussions at NCAR in December particularly useful and interesting, and believe that the manuscript is substantially improved as a result.

We would like to draw your attention to two major changes to the manuscript.

1. Several reviewers questioned the structure of the paper and in particular the blending of discussion of model ensembles with constraints on the equilibrium climate sensitivity. We have now separated these two topics into two Parts, which we believe should make the paper easier to follow. We are presenting these two parts together as a single paper in order to make clear the similarities and differences between the underlying theory in both cases.

2. We now introduce a weighting scheme that can account for model dependency (Section 4.2). This again was something that several other commenters asked about and we realised after a bit of thought that it was in fact fairly straightforward to implement. The method, while very simple, does present a concrete approach to dealing with model dependence which was missing in the original manuscript.

Regards,

James Annan

Reply to Ben Sanderson

Thank you for your comments. This response is in addition to our previous comments made as part of the Interactive Discussion.

We would like firstly to draw attention to two major changes to the manuscript.

1. Several reviewers questioned the structure of the paper and in particular the blending of discussion of model ensembles with constraints on the equilibrium climate sensitivity. We have now separated these two topics into two Parts, which we believe should make the paper easier to follow. We are presenting these two parts together as a single paper in order to make clear the similarities and differences between the underlying theory in both cases.

2. We do now introduce a weighting scheme that can account for model dependency (Section 4.2). This again was something that several commenters asked about and we realised after a bit of thought that it was in fact fairly straightforward to implement.

As a consequence of point 1 in particular, the diff file is particularly unhelpful, as the latexdiff utility is unable to parse the manuscript correctly. It is really not possible to describe the changes in detail, as every section has been changed (and often completely rewritten) in order to separate out the two parts. However, the underlying argument and reasoning of the paper is unchanged.

There are the following significant additional changes:

A second demonstration of model dependence is now included in Section 4.1, which uses the discrete set of model outputs without the need for any parametric and distributional assumption such as a multivariate Gaussian distribution built around these outputs. While the latter approach is widespread and probably not unreasonable, we think it is attractive to be able to present a demonstration which makes a minimum of additional assumptions.

The equations for the simple climate model in Section 6 were previously garbled in their presentation. However, the model itself, and the results, are unchanged.

We hope that our previous response was adequate in terms of the specific points discussed at that time. In particular, we do not expect that accounting for model dependency will change existing climate model ensembles in any systematic way, because the current level of dependency is moderate and we do not have any prior expectation that the dependent models are atypical. We now include in the manuscript an explanation along the lines of our initial reply to reviewer Gab Abramowitz. In short, we would anticipate the performance of the ensemble mean to be insignificantly changed for any moderate estimate of model dependency and duplication. Conversely, in the event of a future ensemble being dominated by a small subset of highly replicated models, then accounting for model dependence could become much more important.

Reply to Reviewer #2

Thank you for your comments.

We would like firstly to draw your attention to two major changes to the manuscript.

1. Several reviewers questioned the structure of the paper and in particular the blending of discussion of model ensembles with constraints on the equilibrium climate sensitivity. We have now separated these two topics into two Parts, which we believe should make the paper easier to follow. We are presenting these two parts together as a single paper in order to make clear the similarities and differences between the underlying theory in both cases.

2. We now introduce a weighting scheme that can account for model dependency (Section 4.2). This again was something that several other commenters asked about and we realised after a bit of thought that it was in fact fairly straightforward to implement. The method, while very simple, does present a concrete approach to dealing with model dependence which was missing in the original manuscript.

As a consequence of point 1 in particular, the diff file is particularly unhelpful, as the latexdiff utility is unable to parse the manuscript correctly. It is really not possible to describe the changes in detail, as every section has been changed (and often completely rewritten) in order to separate out the two parts. However, the underlying argument and reasoning of the paper is unchanged.

In reply to your specific points (and in addition to our earlier reply of 29th September in the Interactive Discussion phase):

1,16: True but not really important for our argument. Even with IC ensembles the long-term climate change will be very similar. We have changed the wording here slightly.

1, 19: We don't think this would in principle be a problem, on the other hand, it may be difficult to distinguish this situation from one in which a ubiquitous error exists, especially when this error is embedded in a complex system with multiple compensating errors.

2, 11: accepted

3, 14-15: It is quite clear that the error of the ensemble mean is substantially larger than any realistic estimate of observational uncertainty or internal variability, at least for many of the more robustly observed climate variables. Of course we agree that for some poorly-observed variables, it might not yet be possible to demonstrate this.

4, 20: A good point and it's been changed from "value" to "magnitude".

5, 1-5: substantially reworded.

5, 19-20: Ok, but we think that the way we've presented it should be clear to most climate scientists.

6, 6-9: Deleted

10, 2: Possibly, but we are not clear how that concept applies to our example.

10, 32-33: CMIP5 is presented as a future test case.

12, 8-9: The priors are just subjective choices chosen to give reasonably acceptable results. The model does not have to relate well to the real world in order for the point to be made, however.

There are also the following significant additional changes:

A second demonstration of model dependence is now included in Section 4.1, which uses the discrete set of model outputs without the need for any parametric and distributional assumption such as a multivariate Gaussian distribution built around these outputs. While the latter approach is widespread and probably not unreasonable, we think it is attractive to be able to present a demonstration which makes a minimum of additional assumptions.

The equations for the simple climate model in Section 6 were previously garbled in their presentation. However, the model itself, and the results, are unchanged.

Reply to Gab Abramowitz

Thank you for your comments. This response is in addition to the reply we gave during the Interactive Discussion phase.

We would like firstly to draw attention to two major changes to the manuscript.

1. Several reviewers questioned the structure of the paper and in particular the blending of discussion of model ensembles with constraints on the equilibrium climate sensitivity. We have now separated these two topics into two Parts, which we believe should make the paper easier to follow. We are presenting these two parts together as a single paper in order to make clear the similarities and differences between the underlying theory in both cases.

2. We now introduce a weighting scheme that can account for model dependency (Section 4.2). This again was something that several other commenters asked about and we realised after a bit of thought that it was in fact fairly straightforward to implement. The method, while very simple, does present a concrete approach to dealing with model dependence which was missing in the original manuscript. To reiterate what we said in our earlier reply to you, however, accounting for model dependence in this way should not be expected to affect ensemble performance to any significant extent, not least because model (near-)replication and hence dependence is essentially unrelated to model performance. The change in effective ensemble size by down weighting related models is very small. However, in a hypothetical future iteration of CMIP which could contain a large number of replicates of a few models, this could in principle become a much more important matter.

As a consequence of point 1 in particular, the diff file is particularly unhelpful, as the latexdiff utility is unable to parse the manuscript correctly. It is really not possible to describe the changes in detail, as every section has been changed (and often completely rewritten) in order to separate out the two parts. However, the underlying argument and reasoning of the paper is unchanged.

There are the following significant additional changes:

A second demonstration of model dependence is now included in Section 4.1, which uses the discrete set of model outputs without the need for any parametric and distributional assumption such as a multivariate Gaussian distribution built around these outputs. While the latter approach is widespread and probably not unreasonable, we think it is attractive to be able to present a demonstration which makes a minimum of additional assumptions.

The equations for the simple climate model in Section 6 were previously garbled in their presentation. However, the model itself, and the results, are unchanged.

We have extended our discussion of previous work to include both Abramowitz and Gupta (2008) and Bishop and Abramowitz (2013). We had not previously realised how substantial the differences were between these two approaches.

We also refer the reviewer to our previous reply of the 29th September 2016, in the Interactive Discussion.