

Dear Gabriele Messori,

Thank you very much for the opportunity to submit a revised version of our manuscript. There is no need to apologize for the delay, we appreciate the time and care taken in handling our submission. Below we explain how we addressed the reviewers' comments, point by point. Our responses are shown in green.

Kind regards,
Britta Grusdt on behalf of all authors

Response to Reviewer 1

The manuscript "On combining climate models into weighted ensembles" addresses model ensembles with specific weights per model based on two approaches: individual performance weighting and ensemble performance weighting. It is argued that, as regarding ensemble performance weighting weights of individual models depend on the tuning of the whole ensemble, co-dependencies of the models are implicitly accounted for. The intention of the manuscript is to delineate the two concepts based on Bayesian probabilities particularly regarding considering inter-model dependencies. Also different methodologies inside the two approaches are analysed.

In general, the manuscript is well written and comprehensible. The purpose of the study is met. Nevertheless, it would be beneficial for modelers and other people dealing with climate model data but not proficient in Bayesian statistics, to give a brief introduction regarding the most important Bayesian concepts used inside the manuscript (prior, posterior, MCMC, Dirichlet distributions, ...).

To clarify the key Bayesian concepts we used in the paper, we added a brief section in the appendix (Appendix C) that provides an overview.

I recommend to publish the article on condition of adding the aforementioned explications as well as implementation of the following minor revisions:

Line 67, 88, Box 1, 95, 100, 101, 110, 117, 184, ... [general]: I would recommend to refrain from the term "prediction" when talking about climate models

We avoid the term "prediction" and write 'model output' or 'simulated data' throughout the manuscript now.

Box 1: it would be good, to include something like $i=[1,N]$ or $i=1,...,N$ in Box 1; in Equation 4 N is used but not defined, the first explanation of N is in line 136 (although it is quite intuitive what N should mean)

We updated Box 1 and explicitly specify the indices there.

equation 5: superscript v missing for y in $P(y|...)$

We corrected this.

Lines 114–120: please add reference(s) for this paragraph

We added a reference to a classical textbook on Bayesian data analysis here.

L140: please add more information and reference regarding Dirichlet distribution (if not done in the general explanations regarding Bayesian statistics)

As suggested, we added a new section in the appendix with a short introduction to the statistical concepts used in the paper, referencing the classical textbook on Bayesian Data analysis from Andrew Gelman et al. We added a bit more information about the Dirichlet distribution and, for illustration purposes, show some examples in a new figure.

Figure 1: a) although clear from the text, please add information that model 1 and 2 are hidden below model 1 in the two lower rows

We added this information to the caption of Fig.1.

Line 194: please inform the reader that i.i.d. = independent and identically distributed

We updated the text accordingly.

Lines 217/218: please include reference for "common pri distribution ... Inverse-gamma"

We added a reference to the Gelman textbook.

Line 266: Reference for ERA5 missing (question mark in brackets)

We added the reference.

Line 278: This (always same weight for M1 irrespective of number of M2 copies) only holds true for ensemble performance weighting (not for individual performance weighting) [as also can be seen in Figure 6]

That's right, we edited the sentence to make it clearer.

Figure 7: is each point in Fig. 7a a different location and Fig. 7b the spatial variance? From equation 8 and the discussion I deduce that this is not the case. So I don't understand the single points. In Figure 6 there is just one weight for each model and now we have plenty of them? Or is this the probability distribution of the weights again?

That's right. Here we show the sampled individual weights, i.e. the entries of the sampled weight vectors, while in Fig.6 we only show their mean values. We edited the caption of Fig. 7 to make this clearer.

Line 332: please write out in full SSP5-8.5

We wrote the full name and added the abbreviation in parentheses.

Line 333: please use (en) dash instead of hyphen for time span

We corrected this.

Figure 8: Please add unit to x-axis. Is this K per W/m² or K per doubling of CO₂?

We updated the label of the x-axis accordingly.

Figure 9: "Projected" instead of "Predicted" please, please write out in full SSP5-8.5

We changed this accordingly, but use the abbreviation SSP5-8.5 since it was introduced earlier in the text where we wrote the full name.

Line 351: I cannot find any "thin dashed green lines in Fig. 9"

We fixed this and now correctly refer to the shaded regions.

Line 368: "ECS as performance metric" this is no metric but a variable considered for weighting (which is then applied to temperature data) -> a basis for assessing model performance

We changed this accordingly.

Lines 393–396: This is a good point. The other way round, also holds true: models with very similar model structures can lead to different projections, e.g. if they are based on the same ocean model but the atmosphere is different, ...

That's right. We added a few sentences here pointing at this.

Please check: sometimes, blank spaces between commas ("e.g.,see") or dots ("vs.0.5820") and the following words/numbers are missing

We added missing spaces and hope that we didn't miss any.

Response to reviewer 2

I consider this paper a successful attempt to put clarity in the field of ensemble weighting.

What is probably reducing the value of this manuscript is the fact that what starts as a nice and well-presented and exhaustive theoretical framework, appears tailor made for the mere application to climate modelling.

In my view that remains a mere application, an important one but yet an application, whereas all the treatment presented applies to any kind of ensemble treatment for any kind of data.

In this respect I would change the title to reflect this aspect thus distinguishing between the general character of the topic and the application to climate models.

Our Response: Thank you for raising this point. We agree that the theoretical framework is general and could be applied to other kinds of data beyond climate data. However, we deliberately frame and motivate the paper around climate model ensembles, since this is the application context that shapes our choices throughout; in particular the model interdependencies, as well as the choice of ECS as an example. We believe that not mentioning climate models in the title might make interested readers expect a paper that goes beyond our contribution of a focused application to climate model data, that we believe, and as you also mention, is important in itself.

To address this point, we made edits in the abstract, the conclusion and in the discussion, noting that the framework is not specific to climate data and may be relevant for ensemble weighting problems more generally.

In ACP, other examples exist of publications that intended to put clarity on the topic of ensemble modelling regardless of the fields of application:

ACP - Est modus in rebus: analytical properties of multi-model ensembles
(<https://acp.copernicus.org/articles/9/9471/2009/>)

ACP - Pauci ex tanto numero: reduce redundancy in multi-model ensembles
(<https://acp.copernicus.org/articles/13/8315/2013/>)

ACP - Seeking for the rational basis of the Median Model: the optimal combination of multi-model ensemble results (<https://acp.copernicus.org/articles/7/6085/2007/>)

Many aspects of this paper are common with these publications although their fields of applications are different.

We had not been aware of this research. We cite these papers in the discussion paragraph on the broader applicability of the framework that we added.

Distinguishing the theoretical treatment from the application case would increase the value and breath of your research. Making the case that your approach as a vaster field of applications will increase that value and breath even further.

More technical aspects were exhaustively covered by the fellow reviewer, and I have nothing more to add as the paper is very well written. Attending to his comments will further improve the manuscript, its readability and accessibility by non-experts in Bayesian statistics.

We addressed the comments from reviewer 1 (see above).