Earth System
Dynamics
Discussions

Interactive comment on "A weighting scheme to incorporate large ensembles in multi-model ensemble projections" *by* Anna L. Merrifield et al.

Anonymous Referee #2

Received and published: 14 January 2020

This paper explores the usefulness of an established model weighting procedure (following Knutti et al. 2017, Sanderson et al. 2017) for incorporating large ensembles of single-model projections into multi-model projections. The model weighting method is shown to produce reasonable results three large ensembles ("SMILEs") are combined with an ensemble of CMIP5 model runs. The paper is generally well written and the results interesting. Some improvements explaining the methods would be useful, but overall I think pending minor revisions the paper should be suitable for publication.

Main comments:

1. The selection of predictors is not completely convincing. I appreciate that the main purpose of the paper is to demonstrate that the weighting method is plausible for the type of ensemble considered, not to explore all possible choices of predictors. Nev-

ertheless Appendix C shows that variability (SLP and SAT standard deviation) shows a weak past/future relation, and Fig 2a suggests weak or no past/future relation for DJF NEU SAT estimated forced trend. Are the results sensitive to exclusion of these predictors?

2. Terminology of "independence weight": it's confusing that both numerator and denominator of equation (1) are called "weights". A more intuitive use would be that a "weight" is a quantity that's larger when the model run is given more weight, i.e. a stronger influence on the results. The term "weight" is used this way for the overall weight (left hand side of eq. 1) and the "performance weight", but not for the "independence weight". Could a different name be used or if not then could a note on this terminology at least be made clearly in Sec 3 (around l.155-165)?

3. The explanation of dynamical adjustment in Appendix A could be clearer. The meaning of Ns, Na and Nr isn't clearly explained. The description refers to the "observational record" but the method is also applied to models, both past and future. Is this appendix meant to be a standalone description of the method or is it assumed the reader is already familiar with the references? I would suggest to improve this description for benefit of completeness and also so that the reader isn't obliged to go to the references to have a basic understanding of the method. It could at least be described how the weights (beta_i) are determined.

Comments & suggestions by line number:

18: "increases linearly": maybe say "changes linearly". It seems unintuitive (at least to me) to describe the weight as increasing when it's actually the reciprocal of the "independence weight" that gets multiplied by the performance weight. This makes sense after reading Sec. 3, but someone reading just the abstract could find this confusing.

20: "subsetted ensemble" –> "subsetted ensemble of one model run per model"

45: "more-than-representative uncertainty" - what does this mean? Please clarify

and/or give a reference for this concept.

62: "ensemble, first," –> "ensemble. First,"

97-98: Don't most reanalyses provide both SLP and SAT?

100: ERA-20C doesn't assimilate surface temperature. Do you know that it's suitable for evaluating SAT trends? This could be tested by comparison with an observational dataset (HadCRUT4?).

109: "representative distribution" - what is this? The distribution is well defined for each model by virtue of the ensemble size. But is this term meant to suggest it's "representative" of the true variability? If not (and I'm not sure how that would be known), suggest change "representative" –> "well defined".

113: put quotes around "macro" (similar to "micro" at l.117)

121: "preindustrial" misspelled

122: "conditions, " –> "conditions: "

141: "and model, " –> "and model; "

167: "definition of climate, " –> "definition of climate: "

170: "fit for purpose" –> "fitness for purpose"

178: "trend" –> "estimated forced trend" (and perhaps add that meaning of this will be explained below)

191: "idea" –> "assumption"

205-206: Perhaps clarify here that the SMILEs reinforce the relationship in the sense that model-mean values (3 data points, one for each SMILE) support the relationship. It's not because the relationship is evident within the SMILEs, which it should not be since the relationship is due to model differences.

222: "trends, " –> "trends: "

223-225: The positive relation for the SMILEs is only for 3 models, and the CMIP5 relation is very weak. Overall this suggests no relation (across models) between past and future estimate forced trend for DJF NEU.

226: Not it's "bolstered" - perhaps more accurate to say that it's "robust to"? The relationship looks essentially the same for both cases in both panels of Fig 2b.

229: "use is" –> "use them"

235: "SMILE" –> "SMILEs"

235: Are these distributions over gridpoints? That is, at each gridpoint in the domain of interest, a weighted or unweighted mean over models is computed, and this contributes one member of the distributions shown in the Fig 2 box-whisker plots. Please clarify. If not then I'm not sure what the "weighted distribution" is.

253: "tail broadly" –> "tail is broadly"

288: "function number" –> "function of the number"

300: "a weight" –> "an overall weight"

324-325: Seems an odd way to say this. The distance-based independence measure used in the weighting is a proxy for model structural differences. Consider rephrasing as something like: "models have some independence from one another while members of a SMILE have none (in the sense of model structural uncertainty)".

367-368: Perhaps qualify this by saying it's a modest narrowing (according to Fig 3ai).

371-372: Again, this is a modest shift. Fig 3ai shows the 95th percentile of weighted ALL is only slightly higher than for unweighted ALL.

406: "target month" - meaning the target year, for the month under consideration?

457: I think you mean Figure B3.

Figure B3 caption: perhaps note that the unweighted distributions are the same in every panel, being shown for reference.

488: "domain-averages" –> "domain-averaged"

489: The "emergent constraint" here is that the model's climatological bias is more or less unchanged from past to future. Perhaps useful to also describe it in this simpler way?