Geosci. Model Dev. Discuss., 6, C2157–C2162, 2013
www.geosci-model-dev-discuss.net/6/C2157/2013/

Geoscientific Model Development Discussions

Interactive comment on "A regional climate modelling projection ensemble experiment – NARCliM" *by* J. P. Evans et al.

J. P. Evans et al.

jason.evans@unsw.edu.au

Received and published: 11 December 2013

Our reply to the comments of reviewer 1 (A. Gobiet) are given below in italics.

Evans et al. describe a semi-objective method for designing a GCM-RCM matrix and the application of the method within the NARCliM project. The topic is highly relevant, since most downscaling initiatives base their GCM-RCM matrix design so far on rather pragmatic decisions and thereby sample GCM and RCM uncertainty in a non-optimal way. A growing number of publications recently investigated several topics relevant for model selection (e.g., model performance measures, model independence measures, climate change uncertainty range), but only few publications describe their combined application for model selection. For this reason, this paper is very timely and relevant.

However, one of the applied methods is not appropriate for the purpose (Bishop and Abramowitz, 2013 (BA2013)) and can lead to opposite results than desired (see specific comment no. 2). This is the major weakness of the study and has to be resolved before publication. One way to resolve it would be to use another technique to define model independence, which might lead to a different RCM and/or GCM selection. In the case that modifying the selection is not feasible at the current stage of the NARCliM project, the authors should at least investigate to what degree my concerns are relevant for their specific case and clearly communicate that the BA2013 method is not suited for model selection, at least not in its basic form as applied here. In any case, a major revision of the study is needed before it can be published. The paper is well written in principle, but it lacks clarity and details in some parts (e.g., comments 3, 4, 5). The authors should also improve in this respect.

Issues with the model selection method concerning the measure of independence are addressed in response to comment 2 below. We note here that the authors are not aware of any other method that explicitly considers model independence and can be calculated using widely available data. While limitations of the method are acknowledged and discussed below, the method is shown to perform well when applied to the model ensemble used in this study.

Specific Comments 1) Title: The title doesn't reflect the contents of the paper very well. The paper is not a description of the NARCliM project in general, but describes more specifically the design of the NARCliM GCM-RCM matrix. This should be reflected in the title.

The title has been amended to "Design of a regional climate modelling projection ensemble experiment - NARCliM"

2) P5122 L23, P5124 L11, P5126 L11, P5127 L17, P5128 L16: The method of BA2013was designed to assign weights according to independence and performance in a model ensemble, but not to select a sub-set of an ensemble. It is inappropriate for

this purpose, since it gives weights (which are apparently used to define ranks in this paper) according to the independence from the rest of the ENTIRE ensemble. However, this does not imply that a sub-set of few models with the highest weights are optimally independent from each other. A simple example might highlight this caveat: Assume you want to select 2 as inde- pendent as possible models out of a 5 member ensemble. For the sake of simplicity assume that all ensemble members have the same quality compared to observations (same error variance sigmaËĘ2 in BA2013). Further assume that models 1 and 2 (group A) are identical, models 3, 4, and 5 (group B) are identical as well, and the models in group A are independent from the models in group B. An optimal choice would obvi- ously be to select one model of group A and one model of group B. But what would be the result of your method? It would assign the largest weights to both models in group A (with regard to the entire ensemble, they are the most independent ones). I.e. the identical models 1 and 2 would be top ranked and selected. This is clearly not the desired result.

The reviewer has a significant point. Text has been added to caution the reader about this point and to refer to a publication (Evans, J.P., Ji, F., Abramowitz, G. and Ekstrom, M.: Optimally choosing small ensemble members to produce robust climate simulations, Env. Res. Letters, 2013.) that investigates the properties of small ensembles created using this independence coefficient. That publication shows that when applied to real climate model data this independence coefficient is much better at selecting models for small ensembles than performance based measures when statistical properties indicative of independent models are desired.

Text has been added where this independence coefficient is first mentioned at the end of P5122 "It is important to note that these independence coefficients were not designed for this purpose and it is possible to imagine an idealised experiment where they would not lead to selection of the most independent models (see Supplementary Material). However, when tested against actual climate model ensembles these independence coefficients do perform as desired and have been shown to select small

ensembles with the desired statistical properties (Evans et al 2013)."

Supplementary Material has also been added to explicitly explore the idealised thought experiment proposed by the reviewer (see attached). It is shown that the actual thought experiment proposed results in a non-invertible covariance matrix and hence cannot produce any independence weights. A variant which includes some random noise is explored and shows that even relatively low levels of noise are sufficient to select optimally independent models almost all the time.

3) P5124 L19 "independence rankings": Please describe how exactly you derive these ranks. I assume that you take the weights of the BA2013 method and rank according to these weights (rank 1 for the largest weight and so on), but this is not explicitly explained here, which forces the reader to speculate.

The text "Here we rank the models based on the magnitude of these independence coefficients."

4) P5124 L13 and L21 "...highest rankings that span the range..", "...models that best sample the range of future changes...": These are are rather vague formulations calling for a clearer explanation. Later in the paper it turns out that you apply these criteria in a purely subjective way. In order to help the reader, it would be good to mention this already here.

"in a subjective manner" is added to P5124 L13 and "subjectively" to P5124 L21

5) P5126 L6, Fig 2: Your statement that the "...overall RCM performance metrics increase gradually . . .." is not very well supported by the design of figure 2. You could, e.g., order the models in fig. 2 according to their "overall performance" in order to make that clearer. Also, please describe what you mean by "overall performance". Is it the average of "clim" and "impact"? Or do you also sort out models that perform very bad in only one of the two metrics?

This text is rewritten for clarity as "The models are then ordered from the best to the

worst model based on the clim metric (the impact metric provides a near identical ordering), and the differences in the metrics between neighbouring models is shown in Figure 2. It shows that the overall RCM performance metrics increase gradually from the best to the worst model, with differences between the models of generally less than 0.01. This gradual increase rises sharply at the 6th worst performing model, with differences greater than 0.015 in the clim metric. A similar decrease in performance is seen in the impact metric. Since these 6 worst performing models show a rapid decrease in performance they are excluded from further analysis."

6) P5127 L6 "they are still able to sample much of the range of behavior in the full ensemble for each event": This is obviously not the case for the SURFERS event, as you describe a few sentences before. More generally, your discussion of figure 3 might be misled by the issues I expressed in comment 2. The AB2013 method might in some cases not lead to the selection of independent simulations, regardless of model performance. This could also be an explanation for the model selection results in the SURFERS case.

As stated in the text the measure used combines both model performance and model independence and in the SURFERS case the selection does not include the models that produce very poor precipitation estimates after the storm peak but rather selects models that peak at different times. Reasons for the poor post-storm model performance for SURFERS but not the other cases is discussed in detail in Ji, F., Ekström, M., Evans, J. P. and Teng, J.: Evaluating rainfall patterns using physics scheme ensembles from a regional atmospheric model, Theor. Appl. Clim., 1–8, doi:10.1007/s00704-013-0904-2, 2013. The potential for this selection method to choose sub-optimally independent models is acknowledged and discussed in the reply to comment 2 and the Supplementary Material.

7) P5129 L1: You state that the identification of "worst models" is more robust than the identification of the "best models". This doesn't seem to be a trivial statement, please give arguments to explain why that is the case.

The text has been changed to "How best to combine such measures remains unclear, however the objective here is not to identify the "best" models to use in the ensemble but rather to identify any consistently poor performing models over the area of interest to remove from being considered as possible ensemble members. This identification should be relatively robust to the individual measures used in a comprehensive evaluation as any model whose estimates are far from the observations are likely to perform poorly across a wide range of metrics."

Please also note the supplement to this comment:
http://www.geosci-model-dev-discuss.net/6/C2157/2013/gmdd-6-C2157-2013-supplement.pdf

---

Interactive comment on Geosci. Model Dev. Discuss., 6, 5117, 2013.