

Interactive comment on “A regional climate modelling projection ensemble experiment – NARClIM” by J. P. Evans et al.

A. Gobiet (Referee)

andreas.gobiet@uni-graz.at

Received and published: 27 November 2013

General Comments

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.

C2021

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.

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.

2) P5122 L23, P5124 L11, P5126 L11, P5127 L17, P5128 L16: The method of BA2013 was 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 independent 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 σ^2 in BA2013). Further assume that models 1 and 2 (group

C2022

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 obviously 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.

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.

4) P5124 L13 and L21 "...highest rankings that span the range..", "...models that best sample the range of future changes...": These 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.

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?

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

C2023

performance. This could also be an explanation for the model selection results in the SURFERS case.

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.

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

C2024