

## ***Interactive comment on “On the meaning of independence in climate science” by J. Annan and J. Hargreaves***

**B. Sanderson (Referee)**

bsander@ucar.edu

Received and published: 17 September 2016

The submitted paper discusses a framework for the consideration of climate model independence. In their discussion, the authors note that recent papers discussing climate model interdependencies have not identified a clear definition of what is meant by the term, and furthermore there is no clear relationship between the vague concept of known structural relationships between climate models and the formal statistical definition of independent, or conditionally independent data as it pertains to our confidence in a given projection of the future.

The paper raises some important questions, but its focus is currently mixed between two concepts which I am not yet convinced should be considered in the same paper. Section 3.2 presents a framework for considering the co-dependency of climate models

C1

in the CMIP archive, while Section 3.3 provides a framework for assessing whether given model outputs represent independent constraints on a future projection given an ensemble whose members have an unknown structure. Although I understand that the authors are trying to make the point that the concept of conditional independence of data is universal from a Bayesian perspective – I don't yet feel like the manuscript marries these two concepts. Specifically, it is not yet clear what the aspiring Bayesian should do with information on model codependency once it has been confirmed.

The existing literature on this topic has to date been focused on analyses which relate inter-model distances to expected model relationships. As such, models exhibiting similar bias patterns have thus been considered potentially codependent and could be potentially down-weighted to account for this replication. The authors of the present study argue that such an approach could potentially erroneously downweight models which converge (through independent logic) to a similar 'truthful' solution.

In their consideration of model independence, they begin by considering why the 'truth-centered' worldview is inappropriate for an ensemble such as CMIP by noting that if the output of one model at a given location is known (as well as the true model), then this should change our expectation of another unknown model because we expect models to have biases of the same sign. Even if this is true, this is not a proof of the non truth+error nature of the ensemble, rather it is an assertion that we believe that the ensemble has a common bias in its present day state.

The authors then present their worldview, which is that codependent behavior should be expected in some cases from the archive (where there are clear examples of common code, or where models arise from the same institutions), and that the data from the models can be used to test that suspected dependency. In an example using the CMIP3 archive of historical simulations, the authors show that Euclidean distance between models from the same institution is less than the distance between two random models in the archive, and this can be seen as evidence of an a priori belief that the results of these models are therefore not independent. Furthermore, this knowledge

C2

can be used to produce a better prediction of the state of an unseen model from an institution which already has a model in the archive (compared a prediction based on the existing ensemble mean).

What the authors should make more clear is what should be done with this information once it has been acquired. Once we suspect that two models are somewhat related, and the data has confirmed this, what is the appropriate course of action to use this information? The authors could follow up with a case study (perhaps broadly along the lines of Tebaldi, 2005) of how a simple projection informed by CMIP would change if the assumption of independent, truth centered models was dropped and replaced with assumptions informed by their analysis.

Another more general concern is this philosophy raises as many questions as it potentially solves. Firstly, there is obviously a requirement to have a prior assessment of which models should be independent, in order to use the ensemble data to test that assessment. Models from the same institution represent one source of potential code-dependency, but the cross-pollination of model components grows ever more complex with each generation of CMIP archive.

The authors use CMIP3, implying that they consider successive generations of the archive to be functionally equivalent for conceptual arguments such as the one they present. However, in some ways, CMIP3 represents a simpler world - CMIP5 (and likely more-so CMIP6) exhibits a more complicated *mélange* as components and entire models are ported from one institution to another. Tabulating and formulating this information into expectations of cross-model interdependency is itself a gargantuan task so it must be considered whether it is a practical recommendation to consider the potential independence of every model pair in a rapidly expanding and incestuous model family.

The authors rightly acknowledge that an a posteriori downweighting of models with similar biases could potentially downweight our most likely models if this was the only

C3

action we took. However, I personally see the type of weighting done for interdependency and emergent constraints as somewhat tangential. The analyses presented thus far in the literature have found that the clearest evidence of common structure from model output comes from the consideration of distances in a high-dimensional space. As more unrelated variables are added to a metric, if two models continue to exhibit similar biases, one can be increasingly sure that the models are related eventually to a point where the distance between two models is a clear outlier in the distribution of distance from one of those models to the rest of the ensemble. However, for emergent constraints on future response to climate change (be it climate sensitivity or any other metric), there is little evidence that such combined multi-variate error metrics are of any use. If any strong emergent constraints do exist, then they are likely to be targeted measurements which exhibit a clear relationship to a given feedback in both their present day and in their future state.

A case in point is perturbed physics ensembles. The authors themselves have demonstrated in previous papers that such ensembles tend to exhibit common bias patterns, and when assessed in a multi-variate context such as a rank-histogram analysis, these ensembles appear under-dispersive in comparison to MMEs such as CMIP. However, such ensembles can also exhibit a greater range of climate sensitivity than the CMIP archive. As such, there is no single space into which the models can be projected which is informative about likely future response, and model dependency.

What is likely (and this is a failure of the existing literature, as well), is that interdependency is not a single property of a model pair, rather it is conditional on the question being asked. The set of all possible outputs from a climate model is high dimensional, and it is likely that only a subset of these outputs is relevant to a given aspect of future response. To take a trivial example: two models which share a sea-ice component would exhibit non-independent results for high latitude warming, but perhaps very little practical codependency for assessing rainforest dieback.

With this in mind, I don't see data mining of the inter-model distance structure as nec-

C4

essarily problematic as long as one is aware of what the distances represent (and I would be the first to admit that there is more to be done on this topic). With a carefully chosen metric, the existing literature has shown that the output data from a model can very nicely identify inter-model relationships – and I believe we would be foolish not to use these relationships directly in our analyses (rather than using them indirectly to confirm or refute prior suspicions of interdependency).

The second part of the paper then goes on to discuss how the results of a large ensemble of simple models can be used to assess the degree to which two potentially observable model outputs are conditionally dependent given knowledge of the model's climate sensitivity. Although this discussion is interesting, I would encourage the authors to either formulate it into a separate paper or to make a stronger case for how the concepts are related, as I feel it currently confuses what seems the primary issue of the paper – that how to formally treat model interdependencies in a small, structurally diverse ensemble.

In summary, the authors raise an important point that the existing literature has yet to provide a formal solution for addressing model interdependency which can be implemented into a well-defined mathematical framework. The paper identifies how model output data might be used to test dependency which might be usable in a formal sense, but in its current form does not demonstrate how this information could be practically used. I would encourage the authors to provide a clear case study of how such information would update a projection which might be made with the multi-model archive, and discuss more fully how their approach could be practical in the context of an ensemble without clearly defined prior expectations of model interdependency.

Ben Sanderson

---

Interactive comment on Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-34, 2016.