Earth Syst. Dynam. Discuss.,
doi:10.5194/esd-2016-34-RC3, 2016

Earth System Dynamics
Discussions
Open Access

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

G. Abramowitz (Referee)

gabsun@gmail.com

Received and published: 27 September 2016

I applaud the intent of this paper and its back-to-principles approach, as opposed to the pure pragmatism (potentially at the expense of principle) that does seem common in this field. It is also well written, to the extent that I have not offered any of the usual spelling, grammar or semantic corrections.

I'm not entirely sure how much it actually achieves, however. It borders on an opinion piece, in that it espouses a particular approach to framing dependence in climate science, without demonstrating in detail how the approach actually might work. Given there is not a great deal of work in this emerging area, this is not necessarily terminal for the paper at all, but at the same time it does not contribute a great deal new that can be of practical use, at least as it stands. I found it very hard to decide how to respond.

General comments:

I confess that I'm personally uncomfortable with a definition of model independence that does not require reference (via observations) to the system that is being estimated by the models. While I appreciate the formalism of starting from an absolutely unambiguous definition of statistical independence, it nevertheless seems counterintuitive to me to discuss independence of estimates of the climate system without any reference to it. There are clearly many other published attempts that do this (including my own), so I'm certainly not going to suggest this renders a paper unpublishable.

Even in the case of Abramowitz and Gupta (2008) and Sanderson et al (2015) - which I actually think are conceptually similar in their treatment of dependence - I feel that defining dependence in terms of model distances (without reference to observations), separately calculating model-observation distances, and then combining them to form weights, is not ideal.

My objection to raw inter-model distance (e.g. RMS) being the metric for independence is that we cannot a priori tell whether a given distance 'd' between two models is 'good' (suggesting they're independent) or 'bad' (suggesting they're dependent) if we have no observationally-based reference point. That is, these models are independent if 'd' is large relative to their distance from observations, and dependent if 'd' is small relative to their distance from observations. So while we can account for this in weights as the two papers above do (effectively defining conditional dependence), I feel we still do not have a 'clean' proxy measure for dependence in this context (i.e. distance in this sense does not equate to independence). This dissatisfaction is what eventually lead to the work in Bishop and Abramowitz (2013) and Abramowitz and Bishop (2015).

Next, despite the use of some examples in the paper, they are not concrete examples that illustrate how the principles espoused here can actually translate into a useful improvement in predictions, once dependence is accounted for. In fact, there seems no guidance in how to account for dependence once it is recognised. Without this, it's difficult to get a handle on exactly what the paper is prescribing we do, beyond adhere to statistical formalism.

I also wasn't clear why the paper covered two topics that in practice seem somewhat unrelated - independence of models making projections and independence of constraints. As it stands, this confuses the focus of the paper. It might well be better to stick to the first of these (which is what the literature review speaks to) and actually illustrate the kind of solution it might provide.

Specific comments:

On line 26 of page 3, the authors suggest that "this approach has the weakness that models that agree because they are all accurate will be discounted, relative to much worse models, without any allowance being made for their good performance relative to reality." Yet, at least as I understand it, the example they give on pages 8,9 & 10 - suffers from exactly the same problem, does it not?

I also feel that the above reference (to Abramowitz and Gupta (2008)) is not quite representative. In that paper - which I now disagree with for the reason outlined above - models were down-weighted if their outputs were similar and they were not close to observations, not simply if their outputs were similar, as suggested here. Indeed the second sentence (L26) is actually misleading, as Equation 2 and the discussion around it in that paper directly address this point and accounts for this issue.

P5, L7: What about the case when the multiple lines of evidence disagree? I would argue the result is not always more precise as suggested here.

Gab Abramowitz

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