Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2017-30-RC1, 2017

Geoscientific Model Development Discussions

Interactive comment on "Practice and philosophy of climate model tuning across six U.S. modeling centers" *by* Gavin A. Schmidt et al.

S. Sherwood (Referee)

s.sherwood@unsw.edu.au

Received and published: 28 February 2017

This review of model tuning practices is potentially a useful contribution to the literature. The complexity of modern models means tuning processes are complicated and can be opaque, but as pointed out by the authors, interpretation of model-data differences, and therefore model evaluation and improvement, depends crucially on how a model was tuned. I think this paper will be acceptable to the journal and a valuable contribution, once a few issues are addressed.

1. The authors need to clarify how their contribution relates to the 2016 BAMS review article by Hourdin et al. They state that theirs can be viewed as a "followup" specific to US centres, but do not spell out what they are adding. I think what they are adding is more detail on the tuning practices at these six centres—but

they need a clear statement. The issues discussion in Section 2 seems much too lengthy for a followup, unless there are important issues that were overlooked by Hourdin et al. It seems that many of these points were already made in the Hourdin et al. paper, or in another paper by Schmidt and Sherwood which is also frequently cited. I think the authors should shorten Section 2, summarise in Section 1 what they are adding to Hourdin et al., and indicate as appropriate within Sections 3-5 where they are repeating what was in Hourdin et al. vs. what is new.

2. I found the manuscript to be of uneven clarity in identifying whether the tuning practices are current and being used right now, or whether they only apply to existing, released model versions. Although some sections specified version numbers (e.g. GFDL), others (e.g. NCAR) did not, although at the end the NCAR section did discuss some issues that arose for CAM6—but with too little detail (e.g., "...with some tuning to those schemes, ENSO performance skill was enhanced.") Are any of the centres changing their practice? Is GFDL tuning climate sensitivity now that they know how to do so? Table 2 should note when the answers in the table hold, since in the future they may change. At least one of the authors (Golaz) has been outspoken in asking questions about tuning for climate sensitivity, but this manuscript remains strangely silent on what the US centres are planning (or currently doing), only arguing that this was not done in the past.

3. The text mentions model selection in the introduction, but I did not see any further mentions of this. Have any of the modelling centres ever discarded a working model version because of its climate behaviour (e.g. climate sensitivity) or any other interesting reason? If not, a statement to this effect would be nice.

4. I often hear grumbling about a hidden problem in GCMs being significant biases in their mean surface temperature, which are swept under the rug by using anomalies (itself a type of simple model calibration), and which some think

should be a significant factor in evaluating models. On the other hand, Hourdin et al. claim that global mean surface temperature is the "dominant shared target" for tuning efforts around the world. If so, isn't global-mean temperature a useless metric for evaluating models, since it only measures how hard the centres chose to tune for this particular target? Can you please say something about this, at least for the US centres? How hard do centres tune for this target, compared to other targets which may require compromising on global-mean T? What (if any) are these other, conflicting targets?

5. Related to (4), it would be nice for Section 4 (or, alternatively, Section 2) to give a better overview of the typical tuning sequence for a coupled model. For example, it seems that centres first tune the AGCM with observed SST to get the TOA flux (im)balance right (do they tune to get LW and SW separately correct?), typically by way of tuning things related to clouds, then tune the ocean (though much less is said about this and I am not sure what the target is), then probably retune the coupled model for global SST, ENSO / MJO, etc? Are the AMIP and CMIP versions of models used in CMIP5 tuned identically, or was there further tuning to the coupled model that is not retroactively put into the AGCM used in AMIP? Some of these details could be clarified also for the individual centres. It also looks like most centres that have aerosol indirect effects end up tuning those to be something they think is reasonable (which is a very important thing to know, probably the most important of all the information presented in this paper, since aerosol forcing in GCMs is a key source of information, even used by IPCC WGI Chapter 7 assessment of this, and many people may mistakenly believe this offers independent information!). Currently Section 4 summarises what is different between centres, but doesn't give this typical set of steps in taken.

Minor corrections

- 2:12-15. The fact that model behaviour depends on expert judgments about

model design is no different to any other modelling exercise, and has been the situation in climate modelling since day dot. Can you restate more precisely what new problem is brought on by recent developments? It seems like the new problem might be that modelling centres now have control over the climate behaviour of their model in ways that they did not before, and that the result could be that climate predictions begin to converge toward what modelling centres think is the most likely/plausible outcome even if it is wrong.

- 3:21-23. We don't know the true aerosol indirect forcing, so this needs to be reworded—do you mean to say the model didn't warm enough globally compared to observations until the critical radius was changed? That artefacts in the geographic warming pattern were produced in the simulation that were judged to indicate too-strong indirect effects, and/or that were ameliorated by making the indicated change to the critical radius?
- 3:24-26. This sentence is too hard to understand.
- 7:1-3. I assume you mean global, climatological (seasonal or annual mean) fields? Please specify
- 7:8. Please change "we" to "the DOE modelling group" or similar. "We" should refer to the authors, not the modellers at one centre.
- 9:18. "RFP" is introduced with no definition. This reviewer does not know what it means, which made the following text hard to review. I have no idea why the ratio of "RFP" to climate sensitivity is meaningful.
- 16:1-3. Run-on sentence.
- 16:18-19. Rephrase; models are not tuned by models, but by model developers.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2017-30, 2017.