Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-185-RC3, 2016

Geoscientific Model Development Discussions

Interactive comment on "Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model" *by* Daniel Williamson et al.

Anonymous Referee #3

Received and published: 11 October 2016

In this paper the authors consider the issue of climate model tuning - whether that be an ocean general circulation model (GCM) or an atmospheric GCM. I should state up front that, unlike the other reviewers, I am not a statistician. I am a climate modeller who is tasked with the development, freezing, and application of complex climate models that provide input to the CMIP process and IPCC assessments. As such I am not as familiar with the history and details of the history matching approach. What I can say, however, is that the philosophy and methodology presented in this study would seem to addresses the leading order issues associated with the enormous and complex task of climate model tuning. The authors have provided an approach that is not only sensible in principle but also applicable in practice. My initial impression was that

this would all become impractical for models with high spatial resolution. This concern was nicely anticipated by the authors and specific examples of how low resolution results can be used to constrain and guide the tuning of a model at higher resolution was demonstrated.

This is an important study that needs to be made visible to the climate modelling community. In my experience, it displays an understanding of the nature and role of climate model tuning that is well beyond that of most scientists who participate in the development of such models and practically all of those who use their output. As the reality of anthropogenic climate change becomes increasingly accepted, the need for more rigorous and credible results from large and complex climate models will be increasingly required to inform policy and support decision making. The present study provides a methodology that represents a significant step towards that goal. My recommendation is that this study be accepted for publication in the Journal of Geoscientific Model Development in essentially its present form. A few minor comments follow.

Minor Comments:

p.3, ll.18-21. This is an important point. It is often the case that during development, the modeller is attempting to assess the model's ability to reproduce a physical phenomenon or feature of the real climate system (eg, Madden Julian Oscillation, the Quasi-Biennial Oscillation, ENSO, etc.). Is it difficult to determine whether the present formulation of its physical parametrizations allows such behaviour for "some" combination of values of its physical parameters or whether the representation of physical processes in the parametrizations are inadequate and require further development. The iterative refocussing method would seem to provide a powerful tool to help decide such issues.

p.6, ll. 1-2. "We also note that the real ocean has never been in equilibrium and hence a tuning procedure that works by comparison to observations may not require an equilibrated ocean." It may be true that the real ocean has never been in equilibrium but if

a validation exercise against observations depended on its transient state, reproducing that transient state would seem to be a much more daunting task than what seems to be suggested in this passage of text. Perhaps I misunderstood the point that was trying to be made here.

p.8, ll.1-2. I agree that uncertainty in the observations against which climate models are assessed is critical but do we even have this information from the observational community?

p.8, ll.28-29. 'If errors can be "tuned out" with better choices of the free parameters, they are not structural at all, they are parametric'. I agree that this would be very informative but it is a necessary rather than sufficient condition for this conclusion to be valid. A tuning exercise that gets some metric within observational error for some range of free parameters is suggestive but does not guarantee that such agreement is obtained for the "right" reasons. Further investigation would still be required to support such a conclusion.

p.14, ll.12-13. "If the entire parameter space is ruled out using a certain metric, a structural error has been located." Again, this is a necessary but not sufficient condition. A potential issue/error in the estimate of observational uncertainty could also be the reason for such behaviour.

p.14, ll.19-22. I would also add that the final NROY space also nicely "defines/documents" all physical behaviour, for the set of metrics considered, of a particular model version (ie specific formulations of physics, the model resolution, numerics etc.). Currently, such behaviour is assessed from one set of model parameters and is used to drive decisions about further development of physical parametrizations. The current approach can be counterproductive if the issue is just parametric and not structural. The more complete description of potential model behaviour captured in the NROY space would allow such decisions to be made in a more rational and effective manner.

p.15, l.30. replace "at at" with "at".

p.20, Fig.4 It is suggested that there are lines with 7 different colours/patterns in this figure. I could only see 3 or 4. It might be better to show all 8 depths separately in addition to the continuous vertical profile. As it stands, it is not possible to see all of what is being described in this figure.

p.22, l.1 change "that fail 2 our more" to "that fail 2 or more".

p.36, ll.1-6. The history matching philosophy is one of identifying and then rejecting free parameter settings that are likely associated with unphysical model states or behaviour. As the authors correctly point out, following such a procedure, the NROY space is a residual of the exercise. As such it has passed necessary but not sufficient conditions in regard to the quality/plausibility of the underlying model. It is basically an efficient procedure to document what a specific model version is, and is not, capable of. From this perspective, it raises the question, is the parametric survey of model behaviour really a "tuning" exercise? The iterative refocussing approach discussed in this study would seem to be more of a tool to survey/discover an existing range of model behaviour associated with a specific set of frozen physics and numerics. In this regard, "tuning" is not parametric (ie connected to the specific values of physical parameters), but rather structural (ie connected to the decisions related to how we choose to represent/model physical processes in our climate models). To me, this represents an important advance in our approach to the development and application of such large and complex models.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-185, 2016.