

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

Daniel Williamson et al.

D.Williamson@exeter.ac.uk

Received and published: 24 November 2016

[a4paper,11pt]article

C1

Author’s response to Reviewer 3

November 24, 2016

We are grateful to reviewer 3 for the time they have taken to give this paper an extremely positive review. We feel it is very important that our target audience (geoscientific model developers and tuners) has been represented in the peer review of this paper, and we thank the reviewer for taking the time to engage with our ideas. We answer each of their minor points below.

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.

We agree and we thank the reviewer for highlighting this.

C2

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.

Our point is that optimising by finding an equilibrated state that matches observations, even though the observations are not those of an equilibrated state themselves seems difficult to justify. That it would be harder to reproduce the transient state may be true, but is not a good reason for the optimisation approach based on equilibrium. We advocate our approach that is based on ruling out rather than optimising, which we introduced straight away at the start of the next section.

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?

Rarely if ever. A point we devote a paragraph to in the discussion at the end of the paper. Our view is that routine reporting of uncertainty in observations and the gridded products on which they are built would be of enormous benefit, in particular to the modelling community. In some instances it may be relatively simple to provide, for example it may have been computed as part of the procedure for deriving a gridded product which incorporates many datasets, e.g. optimal interpolation based on kriging.

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

C3

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.

We agree and have softened the text at this point in the paper to reflect this.

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.

We have added a similar clarifier at this point in the text.

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.

We agree and have added a couple of sentences at this point in the paper to say that.

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

Changed

C4

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.

We have carefully considered this, as we agree with the reviewer that in a printed version the separation between lines is not clear. The figure currently follows the same format as the global mean salinity presented in figure 5, in which there are clear differences between the observation based datasets and the GO5 simulation. The fact that the lines in figure 4 are so close together demonstrates generally good agreement between the observational datasets at many depths. It is not helped by the fact that there is a large temperature gradient between the surface and the abyssal ocean. The image is produced as a high definition PDF so that in a digital version, the reader can zoom in to observe the behaviour of these different models/data sets. If the reviewer feels this figure needs revision, rather than plotting the 8 depths separately (we prefer to show the full structure of the profile) we could present the information in a similar manner to Fig. 6, right panel, with the x-axis scaled to reveal the detail near to EN3 (our target). We did not include this style of plot in the original submission because we wanted to capture the full range of solutions from both the first and second waves, and by zooming in many of these early wave simulations are off-scale for most of the depth range.

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

Changed

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

C5

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.

This is an interesting question, that speaks to the question of what is or is not 'tuning'. We agree entirely with the reviewer that to change the general approach to this activity and the more general activity of model development to include a full parametric survey of 'not implausible' model behaviour respecting key uncertainties would be an important advance in that area. In the final analysis, what we call 'tuning' maybe a semantic issue. For the foreseeable future at least, modelling centres will continue to develop their models and then adjust free parameters in order to provide submissions to CMIP-Next. Whether we call tuning the search for the best model for this submission, the locating of a representative set of models, or anything else, we believe our approach is at the very least an important part of this process and speaks to what is currently done in that field. We allude to this in section 7.

Ultimately, what tuning (the term currently in vogue amongst the climate model development community) involves and what the reported results look like will be determined by the community itself. We hope this paper can influence the direction of the discussion.

C6