

Interactive comment on “Overview of experiment design and comparison of models participating in phase 1 of the SPARC Quasi-Biennial Oscillation initiative (QBOi)” by Neal Butchart et al.

Neal Butchart et al.

neal.butchart@metoffice.gov.uk

Received and published: 12 January 2018

Summary: The paper lies out the rationale for QBOi, defines several experiments that will allow scientific questions about the QBO to be answered, and gives some detailed information on the participating models. Not being an outright expert on the QBO, my view is that the experiments have been well thought out and that the paper will form an adequate basis for QBOi. Weaknesses of the paper include that in a few places unnecessary options are given to the participants that may complicate the evaluation of the results. Also some forcings could be specified in more detail, in order to reduce ambiguity. Furthermore, I am not sure enough relevant detail is conveyed about the

C1

internal make-up of the models to allow analysts to understand why models behave the way they do. Maybe an additional appendix could be written, consisting of one-paragraph descriptions of the individual models, highlighting particular aspects of their make-ups that the PIs consider to be relevant for their simulation of the QBO, that go beyond just their representation of parameterized GWD. Finally, in some places I was unclear as to the function of this paper: The title suggests that the paper is about defining the experiments and describing the models, but in places the text talks about the performances of the models, which I consider out of scope for this paper. Such a discussion of model performance should be reserved for subsequent papers that conduct the scientific analysis.

>Again the authors thank the reviewer for their interesting comments. Unfortunately there appears to have been a slight misunderstanding. The protocol for phase 1 of QBOi has already been discussed and agreed by the community (see Hamilton et al., 2015 and Anstey et al. 2015) and in many cases the experiments have already been completed. Hence it is not appropriate to delete options from the phase 1 protocol (i.e., what the reviewer refers to as “unnecessary options” might actually be options that have been followed by some groups). The purpose of this manuscript is simply to document the protocol that is been used in phase 1 and give the scientific rationale used to agree that phase 1 protocol. Possible revisions and extensions of the protocol for the next phases of QBOi are already considered in the “closing remarks.” Further it was a deliberate decision for QBOi to be less prescriptive than CMIP in order to encourage wider participation. Again this point is already made in the second paragraph of the Introduction, but we don’t consider this paper to be the right place for a more detailed discussion of the relative merits of these different approaches. We agree that the QBOi approach does include ambiguities in some forcings (i.e., ozone) however this is mitigated for by the fact that the models are tuned to give their best QBO with that forcing. This point is now emphasised in the revisions to the subsection on ozone. The important thing from a QBOi perspective is that the same forcings (apart from the CO₂ amounts and SSTs), together with their concomitant tunings, are used across all

C2

the experiments performed by each model (see page 5).

>We don't consider an additional one-paragraph description of each model will provide significantly more information than is already given in the tables, especially as an additional table has now been added on dynamical cores (Table 7 in the revised manuscript). In our opinion these extra paragraphs would be just duplicating information that can be better obtained from the cited references for each model but with the risk that with the duplication possible mistakes and ambiguities are introduced.

>The function of the paper is quite clearly stated in the last three sentences of the abstract. While we agree that subsection 5.1 does diverge into a comparison of model performance (or at least parameterisation performance) and ideally merits a separate paper the co-author who performed the comparison is no-longer in a position to prepare such a manuscript. As this basic comparison of the underpinning model features will be critical for the analysis papers we therefore consider it is essential that this material remains in this manuscript. Neither of the other 2 anonymous referees had any criticism of including GWD material.

P1L2: Capitalize "General Circulation and Earth System Models".

>Lower case is correct here but we are happy for the sub editor to make the change if that is considered appropriate for the journal's style.

P1L4: "Verisimilitude" is an unusual word in this context. While I appreciate stylistic richness, perhaps consider replacing with something more widely understood, such as "realism".

>The first draft of this paper had "realism" but the consensus among the lead authors was this wasn't strictly correct in this context and "verisimilitude" was considered more accurate. "Verisimilitude" is also preferred to all the alternative synonyms given in the Thesaurus we consulted.

P2L9: has -> have

C3

>Unchanged

P2L24: has -> have

>Changed

Figure 1: Perhaps a sentence can be written about the Unified Model family. HadGEM2-CC produces a realistic QBO, HadGEM2-AO, HadGEM2-ES, and the AC-CESS models do not. What is it about HadGEM2-CC that produces this difference in behaviour? Higher lid, more levels in the stratosphere?

>It is quite beyond the scope of this paper to analyse the QBO or lack of QBO in the individual CMIP5 models or, indeed, model families (there are other families in addition to the Unified Model) shown in Figure 1. On the other hand we provide the more generic statement in Section 2, more appropriate in the multi-model context, that it is the introduction of GWD parameterization and increased resolution in the stratosphere that improves the ability of models to reproduce QBO-like tropical variability

P4L31: Cut out "fragile – which is to say"

>"Fragile" usefully emphasises the point that the ability of models to reproduce QBO-like behaviour is easily broken by a small change in formulating some aspects of models and therefore we have retained "fragile – which is to say."

P5L18: Insert "," after "experiments".

>Inserted.

P6L4: You say elsewhere that a 100-year simulation would be preferred. I suggest to remove the options of 3x30 years here and elsewhere. Otherwise you just complicate the analysis and might introduce problems with ensuring that the 3 ensemble members are sufficiently independent of each other.

> See comments above. The protocol has already been discussed and agreed by the community and it is not appropriate to change it at this stage in the project.

C4

Figure 2: I suggest not to fill in the area under the curves in red, blue and black. These areas are overlapping, impossible to see in places and difficult to distinguish. One way to simplify this is to recognize that black is the mean annual cycle (i.e. it's periodic). If you subtract this off the other two signals and display the difference, the figure may become easier to read.

>Figure 2 has been redrawn in the rather different style suggested by Anonymous Referee # 3

P7L4: For the purpose of clarity, perhaps spell out explicitly what the 1xCO₂, 2xCO₂, and 4xCO₂ mixing ratios actually are (e.g. 300 ppmv, 600 ppmv, 1200 ppmv, or whatever the numbers should be). Also what are the mixing ratios for other GHGs (CH₄, N₂O, etc.) for these experiments?

>Full details of the GHG gas forcings are given in Appendix A. For this part of paper we prefer to keep things simple and just refer to 1xCO₂, 2xCO₂, and 4xCO₂, as it is difference between experiments that is the relevant information.

P7L19: Insert "some" before "models".

>Inserted.

P7L25: I would replace "9-12 months" with 12 months. Otherwise you may get an unnecessary diversity of responses there. Also spell out which reanalyses are supposed to be used here.

>As noted above the protocol is already agreed and includes the 9-12 months option. Full details of the experiment setup are given in Appendix A where it is spelt out which reanalyses are supposed to be used here. Our approach is to keep the main text of the paper as simple and readable as possible and direct the GMD reader to the Appendix and Supplemental for the details.

P8L14: ditto.

C5

>Ditto.

P9L22: What is a "vertical resolution equivalent to the model resolution"? Would it not be easier to say that data are requested on model (hybrid-pressure, in most cases) levels? In this case surface pressure and a recipe for the calculation of model level pressure need to be provided, or full pressure fields for hybrid-height models.

>We have added "(i.e., with the same number of levels in the specified altitude range)" after "vertical resolution equivalent to the model resolution." As the review notes not all the models use hybrid-pressure vertical coordinates and we wanted to specify something that would apply to all the models.

P12L6: Experiment 5A is designed for coupled AOGCMs, so the statement is inaccurate, but maybe models without an option to couple to an ocean can just skip this experiment. (Experiment 5A talks about an "appropriate initialization" of the ocean: What is that? Please provide more detail.)

>The statement is in fact correct. ALL the experiments have been designed for atmosphere-only models. However what we say is that one of the experiments (Experiment 5) can also be performed with a coupled model, in which case the experiment has been renamed Experiment 5a. In the spirit of QBOi the aim is to encourage as wide as possible participation hence the project will also consider results from coupled models for Experiment 5. "Appropriate initialise" means the ocean should also be initialised but by whatever method is considered appropriate for that particular model. Again QBOi aims not to over prescribe the experiments.

P19L3: I do not understand why the ozone forcing for models without interactive chemistry is not prescribed. This will be an uncontrolled and unnecessary source of inter-model variations, compounded by the lack of any requirements to document any ozone forcing used. Since almost all models listed will not have interactive ozone, prescribing ozone, and perhaps also conducting a sensitivity experiment using a variant ozone climatology, might be interesting to include. It would be straightforward e.g. to require

C6

models without interactive ozone chemistry to use the CMIP6 historical ozone climatology. An option here is to request variant simulation where ozone is varied from CMIP6 to whatever individual groups prefer, to get an idea about the influence of this choice.

>The subsection on ozone has been rewritten but we emphasise that QBOi has deliberately chosen not to be over prescriptive. For phase 1 of the QBOi analysis this sensitivity to precise details of the ozone climatology was not considered important since all the models are tuned to give their best QBO. The importance was in ensuring that each model used the same ozone across all the experiments performed, as is already emphasised in Section 3. Sensitivity to ozone will be considered in the next phases of the QBOi as noted in the closing remarks of the paper.

P24L12: Again, I suggest to replace the two options with a simple requirement to produce a single 100-year simulation.

>Again we note that the experimental protocol has already been agreed and it is not appropriate to remove one of the options.

P24L23: To remove ambiguity and make everyone's lives easier, the CO2 volume mixing ratio reached in 2002 should be numerically stated here. (How about the other GHGs?)

>If the model includes other GHGs it is already stated a few lines earlier that the 2002 CMIP5 forcings should be used.

P25L15: Replace "9-12 months" with "12 months". >9-12 months is in the agreed protocol so no change has been made.

P26L9: Why not require the data on ECMWF model levels, and ditto for other fields? This way no information is lost to interpolation.

>Although input winds and temperatures on ECMWF model levels could have been used, it was an unnecessary complication since any information lost in interpolation is not important as the profiles are simply meant to be representative of typical observed

C7

profiles.

Figure 7, section 5.1, appendix B: This is moving into the territory of comparing model performance, which I think would be out of scope for this paper. This off-line comparison is actually an interesting scientific exercise itself, but does not quite fit into this paper, the thrust of which is on describing the QBOi experiments. The brevity of presentation afforded to it here also does not quite do it justice. Please consider expanding this into a compact stand-alone publication (which might be submitted to GMD and would be an interesting companion paper to this one).

>We agree a separate paper on the GWD might be valuable but the reasons for not doing this are given in the general comments at the beginning of this response. Also it is not really in the territory of comparing model performance but more in the territory of comparing how GWD is represented in models in the same way as we compare how resolution is represented in the models.

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2017-187>, 2017.

C8