Geosci. Model Dev. Discuss., 6, C650–C660, 2013
www.geosci-model-dev-discuss.net/6/C650/2013/

Geoscientific Model Development Discussions

Interactive comment on "An optimally tuned ensemble of the "eb_go_gs" configuration of GENIE: parameter sensitivity and bifurcations in the Atlantic overturning circulation" *by* R. Marsh et al.

R. Marsh et al.

robert.marsh@noc.soton.ac.uk

Received and published: 27 May 2013

Anonymous Referee #1

In this paper the GENIE model configuration "eb_go_gs" described in detail in Marsh et al., 2011 GMD, is tuned using the tuning method of Price et al., 2009. Tuning targets are observational 3D ocean temperature and salinity fields and 2D air temperature and relative humidity fields. Additionally, simulations are performed where the effect of small changes of individual tuning parameters on changes of the AMOC state is ana-

lyzed, and it is revealed that in the described model version, small parameter changes may lead to large differences in the AMOC solution. This paper neither describes a new model nor a novel tuning method. The main value of the paper is the presentation of an optimal parameter set that is to be used in future studies that use the GENIE "eb_go_gs" model configuration. Therefore, the manuscript is suitable for publication in GMD.

I have some criticisms concerning the tuning method and some other comments that I wish to be addressed by the authors prior to publication.

General comments

Since the main value of the paper is the recommendation of one parameter set for future studies, this should be better emphasized in the abstract, at the end of Section 4.2, and in the conclusions.

We will revise the text accordingly.

The tuning method does not account for uncertainties of the observational fields. Hence, the observational fields are the optimal solution. Hence, the annual-mean velocity field and stream function calculated from the observational T and S fields should result in an optimal annual-mean circulation (this can be tested by prescribing T and S and let the model calculate the velocities). I believe that this won't be the case, because the overturning circulation is to a great part determined by the density distribution in the deep-water formation zones, and deep-water formation does not occur throughout the year. Additionally, the surface density values at deep-water formation zones might need to be different from observations in such coarse models as GENIE in order to have dense enough surface waters for deep-water formation. The simplest solution to circumvent this problem is to exclude these regions (the Southern Ocean from about 90S to 50S and the North Atlantic from about 90N to 50N) from the tuning metric.

The referee makes an interesting point about the utility of observations as a tuning metric in a coarse-resolution model in which a more realistic overturning circulation may be obtained with less realistic property distributions at high latitudes. While there is a argument for removing dense water formation regions from the tuning metric, it is uncertain whether revising the metric in this way would yield optimal circulation states that were more or less consistent with observations, and to repeat the tuning would be a major undertaking. However, we will include a discussion of this issue in Section 3 of a revised manuscript.

The figure captions are generally too short and do not explain the figures well enough. Extend them such that the figures can be understood without having to read the main text.

We will extend the figure captions accordingly.

Specific comments

p.931, l.10 "The root-mean-squared (RMS) errors defined by these fields and corresponding output fields from "eb go gs" for the last year of a 5000-yr spin-up model integration (sTocn, sSocn, sTatm and sQdry) define four objective functions". You should take a multi-year mean of the model output to remove numerical inter-annual variability of the model.

Numerical inter-annual variability in GENIE is generally negligible, with occasional exceptions (e.g., where the AMOC is bistable). We will clarify and emphasize this in the revised manuscript.

p.932, Equations: Add a superscript i to sˆi_Tocn and Sˆi_Tocn and add to the text that N_Tocn is the number of model cells.

We will do so.

p.933, l.3 The process of determining the pareto-optimal solution should be explained in more detail and Deb et al., 2002 should be cited at this point.

We will provide more detail and cite Deb et al. as suggested.

Sec.4.1,first paragraph: This paragraph is more a methods discussion than a results discussion and should therefore be moved to the methods section.

This was also noted by Referee 2. We will move this paragraph, and possible also the second paragraph in Sect. 4.1.

Sec.4.1, second paragraph: How are the 90 individuals chosen?

For each of the 90 individual parameter sets, the 13 parameter values are each randomly sampled in the range between minima and maxima specified in Table 1. We will specify this in the revised manuscript.

Sec.4.1, third paragraph and Fig. 1: This paragraph and the figure are hard to understand. Improve the explanation of what is shown in the figure and what the interpretation is. What exactly is plotted? If objective function values are plotted, then the axis labels should be "f_Socn", "f_Tocn", etc. But these values are a function of the parameter. As I understand there is one cross per individual. So are the crosses some sort of average objective value of one individual? When are runs "successful"? You write that Tocn and Socn are reasonably well correlated. But what about the tail in the figure towards larger Tocn?

Objective functions are plotted. The crosses locate the values of two of the objective functions defined in equations (1)-(4), with each value corresponding to an individual model run. Runs are "successful" if the model completes a 5000-year simulation. Tocn and Socn are reasonably well correlated only towards the optimal values, and we will note the tail in the revised manuscript. We will also improve the figure and caption as requested.

p.934, l.28 "Specifically, we have selected points 18, 37, 49, 56, 74, as they were the : : :" Color the crosses of these points in Figure 1 in the same way as it is done in Fig. 3.

We will do this.

p.935, l.4 “Alongside the corresponding Marsh et al. (2011) configuration (using untuned

parameters),: : :”. Mention that with “untuned” you mean not tuned by the method described in this paper. I’m sure the “untuned” model version was quite heavily tuned “by eye” or some other method.

As outlined elsewhere, we will clarify the provenance of the GMD11 parameter set.

p.935, l.24 It should be “normalized” standard deviation. Explain that this is the std. dev. of the model divided by the std. dev. of the observation.

Yes, we will clarify that standard deviations are normalized, and how this is done.

p.936, l.10 and Fig. 4 Here you compare modeled sea ice with observations. It would be helpful for the reader if the observations also appeared in the figure, either as a separate panel, or alternatively, by drawing a line of the annual mean sea ice extent on top of the plots.

We will consider how to include observations for comparison – a line of annual-mean extent may be the easiest option.

p. 936, l.19 “The Antarctic circumpolar flow is strongest, and hence most realistic, on point 74 (Fig. 5e), : : :” Please add a value. It is quite hard to read the values from the figure.

We will specify ACC strength in each case, either in the text, or in an additional table.

p. 936, l.24ff You compare intensity and depth of the overturning stream functions to observations. However, it is never mentioned to what observations the AMOC depth is compared to. You could base your comparison on nutrient proxies or carbon isotopes (best would be if they were directly simulated in the model). Alternatively, you could compare the AMOC to the AMOC of CMIP5 models.

An observed depth extent for the AMOC, in the North Atlantic, is 2000-3000m (see Fig.

2 in Lumpkin and Speer 2007) – we prefer this inverse analysis of WOCE observations, and we will cite it in the revised text.

Reference: Lumpkin, R., and K. Speer (2007). Global Ocean Meridional Overturning. J. Phys. Oceanogr., 37, 2550–2562. doi: http://dx.doi.org/10.1175/JPO3130.1

p. 927, l.2 “: : : we judge point 18 to be marginally most plausible”. Add a sentence such as “We recommend to use this parameter set in future studies using this model configuration.” This is one of the major results of this paper and it should be clear for future GENIE users what parameter set to use in their studies.

We very strongly favour a recommendation to always use as large an ensemble as possible, and in particular, as a result of this work, to use as minimum all of the 5 identified optima wherever possible. Our position in this regard will be refined through further analysis of climate sensitivity across the 5-member ensemble (see responses to editor’s comments).

p. 939, end of paragraph 1: As above, add a sentence that this one parameter set is recommended for future studies using this model configuration.

Table 3: highlight column of point 18 by bold letters or by a border, since this is the final parameter set.

Again, we will give this consideration.

Figs. 3 and 4: Extend the Figure caption such that the figure can be understood without having to read the paper beforehand: Explain that the reference is the optimum; what are the points?; what is GMD11?

We will extend the captions as requested.

Figs. 5 and 6: I recommend a discrete colorbar with the same increments as in the figure. This makes it much easier to read the figure.

We will investigate this improvement in clarity.

Fig. 7: Residual of what? Why not write “Objective function value”?

We will clarify the meaning of “residuals” (equivalent to “objective function values”) in Fig. 7.

Fig. 8: Dimensions of the parameters are missing. The “x10ˆ6” is confusing, because it is not clear whether it belongs to the upper or to the longer panel. I recommend K_q [10ˆ6 mˆ2/s]

We will clarify the dimensions and units per x-axis.

References: See references in the manuscript.

We will update the references accordingly.

Anonymous Referee #2

This manuscript presents the results of a tuning exercise in which 13 parameters are varied and 5 pareto-optimal sets of parameters are presented and recommended for future use. The manuscript is generally well-written and I consider it appropriate for publication in GMD. I have a number of rather minor suggestions for the authors to consider.

I realise this paper is not claiming to make any particular methodological breakthrough, but it seems a little misleading that the advantages of the method used are described only in comparison to the Latin Hypercube approach of Edwards and Marsh: the cited Price et al 2009 has already described it as having comparable performance to two other efficient methods that were previously used (Proximal ACCPM and Ensemble Kalman Filter). Additionally, part of the methodological description seems misplaced at the start of the results section.

We will accordingly revise the text in Section 3 to emphasize the competitiveness of our Pareto-optimal tuning alongside other methods. We will move most of the text in

the first two paragraphs of Section 4.1 (Results) to Section 3.

The wording on p934 l10- is hard to understand. Please make clear on l10 that you are actually talking about the cost functions of Tocn and Socn rather than the variables themselves (as you clarify later). The claimed correlation does not seem at all clear to me, and the same applies to Tatm vs Tocn (but Qdry vs Socn is evident). Also, I don’t understand the distinction between "correlation" and "competition" made in this paragraph. Is competition here a negative correlation - and then "correlation" refers to a positive correlation? That seems even less plausible for Tocn vs Socn. I conclude that I don’t know what you mean.

We will clarify this sentence, to emphasize cost functions. By “correlation”, we refer to the reduction in both cost functions, towards the optimal value. By “competition”, we refer to the increase of one cost function with the decrease of another cost function, in the vicinity of optimal values in both. We will clarify this meaning, and review the extent of correlation and competition between the four tuning metrics, again revising the text accordingly.

It would be useful to mention how the GMD11 set performs against the cost functions used here. Also, surely it was optimised somehow - its performance seems rather good to have been selected arbitrarily.

We will endeavor to include this comparison, and note the provenance of the GMD11 parameter set (see also our response to one of the editor’s comments).

The fine resolution sampling in the neighbourhood of one point is probably the highlight for me in this work. I suspect the underlying explanation is that the model has two stable solutions in the region of the transition, with the selection of on or off being a quasi-random response to the initial shock. Is this also the authors’ interpretation? This would be easily testable with a different set of initial conditions, which I would expect to give a different pattern of switching (with a similar overall appearance).

Yes, we agree that bistability is the explanation for the “noise” in these limited regions of parameter space. We will consider further experiments to investigate sensitivity of AMOC state to initial conditions, and revise the text to acknowledge the likely importance of initial conditions, in this regard.

It would be nice to see XML files for the 5 parameter sets uploaded as part of the paper (SI) rather than relying on the stability of a personal web page.

We will provide the XML files as requested.

Editor

Dear Bob et al. – please go ahead and reply to the Referee comments now that both

are in.

In particular – whatever you can propose in terms of adding a little more detail and substance to the paper, as touched on by both Referees, would aid the final paper. For instance, Referee #2 found your fine-scale sampling of parameter space of interest and would like to see a little further detail/analysis to more fully understand what is going on here. There were also suggestions of developing the comparison between the tuning method and alternative methodologies.

As outlined in our responses above, we will revise the text to recognize the raised issues, but we may not be able to undertake fresh tuning, or further experiments, in the limited time available for preparation of a revised manuscript.

Also please clarify whether a single parameter set, or all 5 highlighted points, is your recommendation for future work. If 5 – in what way do these 5 provide a useful hold on model 'uncertainty'? For instance – seeing the spread (between the 5) of responses of e.g. AMOC strength and Arctic and/or Antarctic sea-ice cover, under a simple global warming scenario, would be instructive and help give a sense of whether the 5 members predominantly just represent slightly different initial distributions and circulations, or whether they span a wide range of dynamical sensitivities to perturbation.

We marginally favour the parameter set represented by “point 18”, while emphasizing that all five parameter sets are plausible. We will clarify this position in the revised manuscript. However, we will also undertake a limited range of experiments with the 5 parameter sets, along with GMD11, under doubled CO_2 (and possibly also 1% per annum rising CO_2) to investigate changes of the AMOC and sea ice extent, as suggested. Pending the results of these experiments, we will further augment and revise the manuscript. Based on recent experience with a prototype version of GENIE that features a dynamical atmosphere of minimal complexity, the expectation is that differences in response will be limited, but this is certainly worthy of investigation.

Lastly – the reference model – 'GMD11' actually generally scores 'better' in the Taylor diagrams than the 5 selected points. Why? As mentioned by Referee #1 – GMD11 is a key point of comparison and we need to have a summary description of how it was tuned. It is surprising that all this tuning effort does not produce an objectively 'better' model than (ad-hoc tuned?) GMD11. We need to have a little discussion of 'why'. I do not see this (apparent failure to create a better model tuning) as a series issue, and you many even want to argue that one of more of the 5 selected members are in fact better. For instance, the value of creating an ensemble of calibrated models might out-weight the availability of a single previous instance (GMD11), but then this does require more discussion and justification of the 5-member ensemble and what advantages it conveys in being used (instead of e.g. GMD11).

We agree that it is an oversight not to reflect on the determination of parameter values used in GMD11, and the lack of “improvement” in objective terms. We will, however, emphasize some more “structural” differences – such as stronger north-south asymmetry in sea ice (Fig. 4), stronger wind-driven gyres (Fig. 5) and generally more pronounced 2-cell AMOC (Fig. 6) in the new 5-member tuned ensemble. The virtues of establishing a small ensemble of plausible tuned models may be further emphasized by the results of new CO_2-forced experiments (as proposed above).

If you find it easier to revise the manuscript at the same time as you reply to the Ref-

erees, then please do, but I would prefer on balance to see the replies first if possible (we stand a better chance of a faster and overall smoother process this way).

We choose to reply first and proceed subsequently with manuscript revision.

Interactive comment on Geosci. Model Dev. Discuss., 6, 925, 2013.