

Interactive comment on “Global sensitivity analysis of Indian Monsoon during the Pleistocene” by P. A. Araya-Melo et al.

P. A. Araya-Melo et al.

pablo.arayamelo@uclouvain.be

Received and published: 22 August 2014

We appreciate the comments and corrections given us by the first referee. They clearly helped to clarify many points, hence, improving the scientific discussion and presentation quality.

1 Minor science points

- 1. 1623/25 Appears to contradict earlier statement that parameter dependence is smooth... It's also not clear to me at this stage why only one outlier appears to distort**

C1352

In fact we believe that it does not contradict the earlier statement. To see this, let us call $f(x)$ the function mapping x on the value of the simulator, at stationary state, obtained after a *very* long integration and sampled over a very long time. This is essentially the function that we are modeling with the emulator, and we know that a given experiment does not provide an exact estimate of this function (because limited sampling, and possible remnant effects of the spin-up phase, since we only integrate over 400 years). In general, the hypothesis is consistent. Perhaps one of the best, non-trivial result, is that the size of the nugget obtained by emulator calibration is consistent with what we know about model variability. Yet, it happens that 2 experiments don't fit the scheme, and it seems we have two options: either call into question the hypothesis about the smoothness of $f(x)$, or call into question these two experiments as good measures of $f(x)$. Given that 59 experiments work with the smooth $f(x)$ hypothesis, we found it more parsimonious to question these two experiments. The only thing we can say is that the simulator configuration looks OK. So, ideally, we should have launched these experiments with other initial conditions, or continued these two experiments over several hundreds of years, in order to inspect their statistics. We did not do that, but we have a couple of long experiments (thus, not those one in particular) that confirm what we could already suspect: 400 years spin-up, and in fact only one hundred years with the right orography, is a bit short for spin-up in the North Atlantic Ocean and it cannot be excluded that we have to face effects associated with multiple or meta-stable states in the North Atlantic convective region that have resulted in these experiments behaving as outliers. In fact, this state of affairs illustrates very well the spirit of this approach: at constant computing time, we have traded the benefits of very long experiments which almost guarantee ergodic statistics, against an ensemble of shorter experiments that provide a broad picture of the sensitivity of the system at the cost of picking out one or two experiments that appear to be anomalous. It is remarkable of the emulator is that it is capable of distinguishing what we are then tempted to call *outliers*.

C1353

2. **1624/10 How about other reasons for oceanic change - e.g. precip?**

That is a fair point. It is correct that, lacking repeated experiments with the same inputs (and different initial conditions), or a continuation of these experiments, statements about what actually happened with these experiments appear a bit speculative. The statement was softened as follows:

Experiments 11 and 40 have, however, low to moderate glaciation levels and there is no obvious reason of why their behavior should differ from the other experiments. We could be sampling some meta-stable state of the ocean circulation, possibly excited by the spin-up procedure. As noted by Hewitt et al. 2006, this can be relevant for understanding the palaeoclimate record. Incidentally, it is noteworthy that the emulator appears an effective approach to identify a region of the parameter space that could be the focus of further inspection. For the present purpose, we however chose to consider the global, consistent picture provided by the remaining 59 experiments. Consequently, we consider a new emulator, calibrated on these 59 experiments.

3. **1624/13-15 This is important but not very well justified or well written.**

See rewritten statement.

4. **1624/25 Er. . .because you removed them! Not relevant.**

Well, we think it is. Even with these two experiments out of the way, we had no guarantee that emulator error would be broadly Gaussian distributed : it could be terribly fat-tailed ! It is not, and it was relevant to check this as part of the validation procedure.

5. **1626/27 As far as I can tell from this figure your signal direction is consistent with Zhao et al. but not magnitude (approx. 3 times larger) - can**

C1354

you comment? (And what is the magnitude of Braconnot and Marti result?)

Zhao et al. give sensitivities of the order of 0.5 degree C and 30 mm/month over the regions of interest, for the *mid-Holocene*, that is, we are looking at an eccentricity of 0.018 in a non-glaciated state, and remember, mid-Holocene means perihelion reached in March, vs perihelion reached in early-January today. If we report these values on our Fig. 12, we get sensitivities of similar order of magnitude as them.

6. **1627/15 "linear" - How is this quantified - same reasoning as 1628/2? Refer to values in Table 2 for both?**

Fair point : this was clear from a figure that did not make it to the final version. Here is the corrected text:

The values of λ_ϵ obtained by optimization of emulator likelihood are generally large (Table 2). The implication is that the simulator response to obliquity is most likely linear.

7. **1628/21 "reverse" - do you mean the response is not monotonic? Is this shown some- where? Is it strange that they *exactly* cancel?**

Yes, we know,... this puzzled us for a while so we really looked at this in detail. The answer is yes, the the response is not monotonic and it does almost cancel (as strange as it may be), which shows the interest of a variance analysis vs inspection of marginal dependencies. This said, the response is very small and would not be detectable in climate records anyway, and we found it better to remove the paragraph.

C1355

2 Comments and questions for improving clarity

1. **At the moment the Emulator and Sensitivity Measures sections are quite heavy going for someone a little familiar with emulators, and not sufficiently intuitive or descriptive for someone new to them. They should either be moved to an appendix or SI, or expanded with more explanatory text. I have assumed the latter in my comments.**

We went for the latter option, see following response.

2. **1610/3 Useful to give Pleistocene dates**
changed in to

... relevant for the Late Pleistocene, ca. the last 1 Myr.

3. **1616/4 if you use the terms maxi-min and orthogonality here, mention them when describing them on p1615 too** : done
4. **1617/13 "input vector associated with the jth component of the experimental design" -> I think this would be clearer as "input vector (set of parameter values) of the jth ensemble member"**

Let \vec{x}_j be the set of input values of the j th ensemble member.

5. **1618 Is it possible to add some more explanations of choices, terms and notation?**

Yes, of course it is. The contribution of a professional statistician (Dr. Richard Wilkinson) and the suggestions of the reviewers have been very helpful in this respect. We followed the suggestions of the referee, and, in addition, we have prepared an animation to be uploaded in supplementary material that further illustrates the experiment design and the effect of length scales and nugget. The

C1356

revised version also provides more details about the function of the nugget and also what happens when length scales are increased.

6. **1619/13 Can you explain why you want to (i.e. why A&C recommend to) restrict the nugget amplitude? Again, an intuitive explanation of this section would aid the reader.**

New version:

In order to guarantee that the emulator is at least no less informative than would be linear regression, Andrianakis and Challennor (2012) recommend the use of a *penalised likelihood* as follows...

7. **1620 The empirical distribution part is not clear. Please give more information on how rho was obtained, what the resulting values indicate, and perhaps add some kind of figure.**

A figure is in preparation for the revised version.

8. **1620/13-15. Every time I try to read this sentence I fail. . .and then you lose me for the rest of the section. I think the concepts here are not difficult, but they are currently a bit impenetrable in the explanations. This section needs to be longer and with shorter sentences...**

We have attempted the following rewrite (in addition to extensive rewrite of the preceding paragraphs)

The quantity S_p is the loss in output variance that would occur by fixing p , all other factors varying, compared to a situation where all factors vary. More precisely, this is the *expected* loss, averaged over all possible values of x_p . This quantity is often scaled over the total output variance to produce what is classically referred to as the main effect index (Saltelli (2004), ch. 1). On the other hand, \bar{S}_p is the output output variance that occurs when factor p is varies, all other factors fixed.

C1357

9. **1621/17 This could be better explained: e.g. "These two are equivalent if the covariance between all input factors does not depend on the value of factor $p(x_p)$.**

We have to disagree here, they would not be equivalent, but we have accounted for the suggestion of giving examples, as follows:

The distinction is especially important when there is a covariance between input factors, in this case CO₂ and ice volume. For example, the main effect index associated with, say, ice volume, includes an implicit effect associated with the fact that CO₂ varies along with ice. By contrast, the total effect index does not include this effect. For this reason, we use the latter here.

10. **1622/13 IND could stand for India or Indian Ocean. How about using NI and IO, or NIND and INDO? or just call them the land and ocean boxes/regions?**

We completely agree. This is changed in the revised version.

11. **1623/3 Can you add a recap of the meaning of the scales (e.g. "length scale of the correlation between outputs from different simulations, for each of the five input factors") and nugget (e.g. something like "error due to simple modeling of this correlation" or "model structural uncertainty")? And maximization of the penalized likelihood? (e.g. "i.e. found by the best fit of the emulator to the ensemble outputs"?)**

This has been taken into account, and changed accordingly in the revised version.

12. **1623/10 First sentence could be clearer..**

Yes it could ! We propose

C1358

For $e \sin \varpi$, $e \cos \varpi$ and ice volume, the length scales λ are of the same order of magnitude as the range covered by the input factors. This is the ideal scenario : the space between two experiments is consistent with the decorrelation length of the simulator.

13. **1623/14 The "however" is confusing in this context. And "response to" makes more sense to physical scientists than "response in". . .? Why write "This is not a problem on its own"? Better: "This indicates that a sparser sampling of this factor would have been sufficient".**

Agreed and corrected.

14. **1623/25 "mean effect" is not explained here nor in its first appearance (1621/18). Something like "mean effect, which here shows the response to x and y while keeping z constant" would help. Convoluted is not really the right word. Waves / ripples / distortions in the contours? Departure from smooth linear gradients?**

We have omitted references to main (and not mean as we incorrectly wrote) effects and took the reviewer's suggestion to use "departure from linear gradients"

15. **1624/22 Either insert "standard" before "normal" or else remove "normalised by standard deviation". Give context to p value, e.g. "null hypothesis rejected at. . ."**

Now reads:

the normalised errors are compatible with a normal distribution based on the Shapiro–Wilk normality test, except for continental temperature (normality rejected with 97 % confidence);

16. **1625/3 "observation" is confusing -> how about "uncertainty of the simulation, i.e. how representative the 100 year simulation is of the mean model**

C1359

state".

Agreed and corrected accordingly.

17. **1625/7 Better to use residuals errors than residual uncertainties, I would say, because you are talking about the differences between known numbers here.**

Agreed and correctly accordingly.

18. **1625/10-12 Not particularly well-written. How about: "Thus, remarkably, the emulator calibration has successfully estimated model internal variability using only 100 year means."**

Agreed, and changed in the revised version.

19. **1625/16 The final sentence of Fig. 7 caption text is useful and could/should be re-used here. For clarity, reorder figures / text to match each other (precip first or temp first).**

Agreed, we added the sentence and re-ordered the figure.

20. **1626/14 Low glaciation looks the most symmetric in terms of contours - can you please clarify?**

We agree the sentence is a bit obscure, and decided to delete it.

21. **1627/9 I must be missing something but I'm confused - you seem to describe this as if it is a lag between maxima within the same year, but it is a phase shift across many years?**

We see where the confusion may come from, since we do not specify to which season we refer to. The text now reads: quote On this point, our analysis reveals that the maximum JJAS SST response occurs when perihelion is reached in May.

1627/10-13 I don't understand this bit. We feel the sentence can be misleading, so we decided to suppress it from the text.

C1360

1627/16-19 Delete (poorly written repetition of earlier discussion). Agreed. This is already changed in a previous comment. We added the following text:

As it can be seen from Table 2., the values obtained by the emulator for obliquity are generally large, thus, hinting a linear response.

1628/8 Seems odd to only cite such an old paper - add something more recent (e.g. emulated..?) too.

Well, we wanted to cite the first paper giving this value in HadCM3, which is what we are concerned here, more than the actual value of sensitivity.

1629/8-20 I think these explanations need expanding.

Figure 4. caption confusing and has errors: suggest "Diagnostic of emulator performance for sea surface temperature (left panel) and mixed-layer depth (right panel), showing mean and standard deviation of emulator prediction versus actual HadCM3 output.

Agreed, changed accordingly in the text.

Figure 6. Suggest rewriting, e.g., replace second sentence onwards with "The response to the input factors is generally smooth, except in the regions near experiments 11 and 40 (glaciation levels 3 and 7)

Suggestion appreciated, changes made in the text.

3 Corrections

Most, if not all, of the corrections pointed out by the referee were taken into account. Many of them were typos, while other were clarification of ideas on the text.

C1361

Regarding the comments/corrections on the different figures, they were all taken into account. Fig. 2 was re-plotted: sea-level pressure now showed, label on color table included. Color scales on Figs. 6, 8 and 12 were corrected, so now they are not saturated and correspond to the plotted figure. Also, contour increments were changed as well, they are now visually easier to interpret.

Interactive comment on Clim. Past Discuss., 10, 1609, 2014.

C1362