

Interactive comment on “The impact of structural error on parameter constraint in a climate model” by D. McNeall et al.

D. McNeall et al.

doug.mcneall@metoffice.gov.uk

Received and published: 26 August 2016

On behalf of the authors, I would like to thank both reviewers for their thoughtful, constructive and useful reviews. The review from anonymous reviewer 1 is posted below, with inline responses.

Reviewer 1 (Anonymous, denoted R1:)

R1: Review of paper: “The impact of structural error on parameter constraint in a climate model” by Doug McNeall et al.

Thank you for inviting me to review this paper. The paper is interesting and important as it addresses whether a component of a GCM can be calibrated for one part of the globe, but applied elsewhere. Climate models are heavily dependent on transferabil-

C1

ity of parameterisation of sub-model structure, and a knowledge of when this fails is important. I can see the aim of the paper, and it will be useful to have in the literature.

However there did seem to be a slightly excessive use of statistical terminology. That's fine if the statistics is of standard form, but that's not the case here as the methods utilised are more novel. Please ensure that the literature is cited sufficiently well that any part of this paper can be understood by calling upon the appropriate referenced papers.

Response: With a paper at the interface of climate modelling and statistics, finding the correct balance of technical versus general description will always be difficult. Our strategy was to write for a more general audience, but to include a comprehensive set of references to literature at this interface. The statistical foundations of Gaussian process emulators are fairly standard, having been used in computer experiments across a wide range of subjects. With that in mind, we might add the following reference as a general and instructional introduction to the subject, for non-experts:

O'Hagan, A. Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering & System Safety* 91, no. 10 (2006): 1290-1300.

Using emulators for climate science work is rarer, although a literature is building. Our paper uses standard emulators in a less standard way, in order to learn about the model and the climate. We can expand the literature review to include more examples of emulators being used in novel ways in climate science, in order to include more context for the reader. Some specific examples of related analyses from the climate science literature are:

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L. and Yamazaki, K., 2013. History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate dynamics*, 41(7-8), pp.1703-1729.

C2

Bounceur, N., Crucifix, M. and Wilkinson, R.D., 2015. Global sensitivity analysis of the climate-vegetation system to astronomical forcing: an emulator-based approach. *Earth System Dynamics*, 6(1), p.205.

Tran, G.T., Oliver, K.I., Toal, D.J., Holden, P.B. and Edwards, N.R., 2016. Building a traceable climate model hierarchy with multi-level emulators. *Advances in Statistical Climatology, Meteorology and Oceanography*, 2(1), p.17.

R1: Below are some comments that the authors might like to consider for a revised manuscript:

Overall points

The title is possibly too general. The emphasis is on DGVM modelling of forests, not general overall issues of structure.

Response: While we take the reviewers point here, we feel that the techniques used in the paper are sufficiently generalisable to be of interest to the wider climate modelling community. A key theme of this paper is that it attempts to improve the DVGM within the context of an Earth system model, which has it's own biases in climate simulation. An alternative title could be "The impact of structural error on parameter constraint in the land surface component of a climate model", but we welcome suggestions from the Editor.

R1: The Abstract needs to be something that can be read in isolation, such that the reader can obtain a strong idea what the paper is about.

To my mind, there is some repetition (e.g. three times says this uses "a history matching approach", and yet doesn't define what this actually is). Removing repetition can make space for more details. Extra description of the parameters changed would be helpful, rather than a vague "parameters that lead to a realistic forest fraction".

Response: The reviewer makes some good points here, however describing the individual parameters in the abstract and yet making it shorter might be challenging.

C3

The focus of the paper is on the techniques for learning about the parameters, rather than the parameters themselves. Perhaps a broad description of the types of systems the parameters help control might be appropriate? We agree that the abstract could be more compact, avoid repetition and perhaps offer a clearer description of history matching. With that in mind, we suggest the following as a re-write:

We use observations of forest fraction to constrain carbon cycle and land surface input parameters of the reduced resolution global climate model, FAMOUS. We use an ensemble of climate model runs to build a computationally cheap statistical proxy (emulator) of the climate model. We then use a "history matching" approach, comparing the emulated climate model output at various parameter settings, and ruling out as implausible those where the simulated output is judged statistically incompatible with observations. We use the emulator to simulate the forest fraction at the best set of parameters implied by matching the model to the Amazon, Central African, South East Asian and North American forests in turn. We can find parameters that lead to a realistic forest fraction in the Amazon, but using the Amazon alone to tune the simulator would result in a significant overestimate of forest fraction in the other forests. Conversely, using the other forests to calibrate the model leads to a larger underestimate of the Amazon forest fraction. We argue that this finding indicates a structural model discrepancy. We characterise this discrepancy, and explore the consequences of ignoring it in a history matching exercise. We use sensitivity analysis to find the parameters which have most impact on simulator error. Finally, we perform a history matching exercise using credible estimates for simulator discrepancy and observational uncertainty terms. We are unable to constrain the parameters individually, but just under half of joint parameter space is ruled out as being incompatible with forest observations. We discuss the possible sources of the discrepancy in the simulated Amazon, including missing processes in the land surface component, and a bias in the climatology of the Amazon.

R1: Reviewing this, I'm trying to really understand what the main thrust of this paper

C4

is about, in the statistical/algorithm sense. Can I confirm that the over-arching message is that quantity delta in Eqn (1) is important, can be characterised, and shows geo-graphical variation. To my mind, that is a powerful result. It basically says if (i) not enough process representation is introduced in to a model, then structure deficiency gets masked by parameter fitting, and (ii) doing so will create problems between different locations.

Response: That is a good summary of the main thrust of the paper. We would also like to highlight that it is not just missing process representation, but poor process representation (i.e. biases in other parts of the climate system) that can lead to errors if the delta in equation 1 is not taken into account. We would also like to highlight some of the novel techniques that we've developed to learn about discrepancy and its impacts. We will endeavour to make this clearer in the introduction to the paper.

R1: It would be nice to acknowledge that structural errors presumably also reduce confidence in any model for future projections, even when just at a single region where it performs well for contemporary periods.

Response: I suggest that we include this point in the discussion section. One advantage of including and estimating a discrepancy term is that future projections should acknowledge the uncertainty caused by the structural discrepancy. While this may lead to more uncertain projections, they should be more robust - that is, they should offer a more accurate estimate of uncertainty.

R1: Page 9, starting "Does this region represent". This is a critical part of the paper, discussing how in effect a standard best-fit might not always be appropriate. Can the discussion be led back to Eqn (1), and in particular the structural delta parameter? (Also line 1, page 9, I cannot see in a Table or diagram what the alternative potential values are, for comparison against the default inputs – apologies if I've missed something). Where are the local, or continent-scale, delta values given?

Response: At the moment, this just says "without using a structural discrepancy func-

C5

tion", but we agree with the reviewer that this could be much clearer. We will refer this straight back to equation 1, with the implications for mean and uncertainty of the discrepancy function (not) used.

The alternative potential values are a multidimensional cloud of points in parameter space, and therefore hard to summarise in a table (or even in a graphic - we are reduced to a two dimensional projection of the five dimensional space). The graphic (figure 4) has the space in normalised units - it might be clearer if we were to place the default parameters on this graphic. The model error in each forest at the input values indicated by figure 4 can be estimated by looking at figure 5, which shows the output of the model at this region.

R1: Details P2, line 10. Again, please give the reader some idea what "History matching" is, given other quantities such as "calibration" and "tuning" are defined at this point.

Response: We shall include an early, simple description of history matching, which may well be more unfamiliar than tuning or calibration to readers.

R1: Around lines P2, lines 25-29. It would be really nice to have more concrete reasons why emulators, parameterisations etc are needed. This usually comes down to two factors: (1), computational speed prevents very high resolution modelling, even if the processes are more fully understood. For example, parameterisation of convection. (2), we don't know what the values should be, and these may exhibit strong regional heterogeneity. The latter is more the case for this paper, with questions asked as to what are the appropriate number of plant functional types that should be in land surface models – and if the number is high, can for example EO provide the values.

Response: The reviewer makes a good point that we don't discuss the possibility of regionally varying parameters, and what that means for the current analysis. We shall include a section on regionally varying parameters in the discussion section, and expand the section on parameterisation accordingly..

C6

R1: Check notation is consistent throughout. P3, line 23, FAMOUS is described as a “climate simulator”. In the minds of the authors, is this different to a standard GCMs. Do they regard FAMOUS’s reduced resolution as removing it from being regarded as a full GCM?

Response: In the statistical emulator literature “simulator” is often used for computational process models in order to distinguish them from statistical models. We will make this clear, and review use of “simulator” and “model” in the paper in order to ensure consistency.

R1: Again, in Section 1.3, this is now the 7th or 8th time that “history matching” is mentioned – it would be good to help the reader as to what it is, even if it is only to provide a methodological citation at this point.

Response: See previous response.

R1: Cox (2001) is a technical note. Better to give a peer-reviewed reference?

Response: It is possible to cite well known papers that use TRIFFID (e.g. Cox 2000), but Cox (2001) is the the standard reference that outlines the technical details of TRIFFID.

Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A. and Totterdell, I.J., 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408(6809), pp.184-187.

R1: P5, line 1. I don’t understand the context of the sentence: “The Amazon region is not wet enough for a fully humid region to exist”. If this refers to the FAMOUS model, and in particular its atmospheric response, then this will make any DGVM fail if rainfall totals are too small. P5, discussion of beta parameter. In a similar vain to the comment above, is it OK to treat the atmospheric beta parameter as a “nuisance” parameter? Isn’t there a risk that errors in GCM-projected precipitation – for example - will affect best-fit parameters in Table 1?

C7

Response: FAMOUS has known biases, including a climatologically dry Amazon region, and this is indeed one of the strong candidates for low forest fractions in that region, as discussed later in the paper. However, in the Amazon region there are also possible confounding feedbacks between land cover and climate, making attribution of any biases more difficult. No climate simulation is perfect, and biases large or small are a common problem to be dealt with in any analysis. Our analysis offers new techniques to identify and characterise such biases, and the way that they might impact our estimates of the values of input parameters. The beta parameter is not correlated with any of the land surface parameters in the ensemble design, and so we felt justified in excluding it from analysis of the land surface parameters. However, it may well have an impact on climatology, and this could be the subject of a future study.

R1: P5, line 18. From code that is shared with other centres, TRIFFID has a rapid spin-up option to near-equilibrium. Does it really need 10000 years?

Response: The fast spin-up mode was used in the simulations - only the equivalent of 10,000 years for each decade was used in this mode. The climate simulations were the averages of the last 30 years of a 200 year run. We shall make this clearer in the text.

R1: Trivial thing, but it might be nice in Figure 2 to write as S.E.Asia (not SEASIA).

Response: This will be amended to be consistent with the other plots (a space added).

R1: Can I confirm that a reader could find all details of the emulator in the Roustant et al 2012 paper. So, for instance, what a “leave-one-out cross validation metric” is.

Response: Roustant et al. (2012) is very comprehensive in its mathematical description of the emulator, and the software package that it informs. Leave-one-out cross validation is not related to the emulator itself, but is a broader validation algorithm. We will include a suitable reference (e.g. Hastie, Tibshirani and Friedman (2001)).

Hastie, T., Tibshirani, R. and Friedman, J. 2001. *The elements of statistical learning*

C8

(Vol. 1). Springer, Berlin: Springer series in statistics.

R1: Figure 7 I find very useful as it allows assessment of the geographical differences, providing more information than the global parameterisation Table 3. There are quite a few statistical methods available to determine parameter importance and/or nuisance parameters. An extra sentence stating what additional benefit the FAST algorithm brings would be helpful – i.e. beyond just the Saltelli reference.

Response: The FAST algorithm is ideally suited to our situation in that a) it provides an accurate global sensitivity analysis, including main effects and interaction terms and b) is easily and cheaply calculated using the emulator and a convenient R package. We shall include a sentence to this effect in the section.

R1: Figure 8 is important as it shows how the Amazon has a difference response. Or put another way, a calibration of NL0 and V_CRIT_ALPHA for the Amazon could find a pair of parameters that would clearly be sub-optimal when applied to the other 3 regions. And vice-versa. I'd like to see more discussion around Figure 8, how it demonstrates the structural problems (i.e. very different responses to NL0 and V_CRIT_ALPHA, depending on location), and again – can this be related back to the delta parameter? This will also link better to the paper title, which is about model structural problems.

Response: Linking this clearly back to the structural discrepancy function at this point is a good idea. However, the discussion that the reviewer requests here is a large part of the later analysis (e.g. figures 10 - 12, and section 3.5). We could indicate the more detailed discussion in this later section in the text of this earlier section.

R1: Figure 13 is nice and clear, and in many ways it is a shame that the paper is so long in technical details before getting to that point. Obviously this is a slightly naïve comment, but could it simply be that the trees of the Amazon have evolved differently to those of Africa. This could possibly be due to different imposed climatologies that the trees have adapted/acclimated to. So one conclusion of this paper could simply

C9

be that any land surface model such as TRIFFID requires a parameter mask, or ancillary fields, that are different for different places. The paper hints at this, page 16, in “Causes of discrepancy”, where different rooting depths are considered. One future work extension might therefore be to include a root depth as a geographically-varying parameter, to add to those in Table 1? Would this then collapse delta down to zero for all locations?

Response: This interesting and useful idea should clearly be included in the discussion section.

Interactive comment on Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-17, 2016.