

Revised version of Global sensitivity analysis of the climate-vegetation system to astronomical forcing: an emulator based approach

Referee Report

February 3, 2015

1 General comments

The authors have done a great deal of re-writing and the paper is much better because of it. It is now clear to me that this paper intends to act as a road map for other emulator-based sensitivity analyses in the palaeoclimate literature and will be a valuable paper. I will be recommending this for publication, however, there are still major corrections that I feel need to be checked carefully before the paper is accepted. My principal issue is with the sloppiness and lack of clarity surrounding the rewritten mathematical sections. There are basic mistakes in the statement of statistical models that I would expect the experienced authors to catch before submission. There are undefined variables, and there are places where unclear word based definitions of key quantities are used without the required equations. Given that this is to be a road map for their field, I think it is important that all of this is corrected, checked by all authors, and again by the reviewers.

In particular, it is very important to keep the audience in mind. The new version, which has removed the step by step derivations of many quantities which were derived in other papers is definitely an improvement upon the first. However, it is now far too cryptic and for no good reason. The following structure will enable the reader to understand and to replicate the technology, which is what this paper is aiming for: Define the key quantities that you are interested in for this paper in terms of the simulator (a version of the V 's and T 's expressed in words in the new derivation). Explain the requirement for the emulator and introduce it as you have (with the statement of the statistical models suitably corrected). Then explain how it (the emulator) fits into expressions for the V 's and T 's, before writing down all of the integrals that have to be computed in order to do it (15)-(20).

2 Major comments

- Page 7 line 8 where the 27 member design is executed 3 times. 3×27 is the 81 runs I refer to in my first review. You have a finite budget with which to explore parameter space and the effects of initial condition uncertainty (else why would you only use 27 members). You have run the model 81 times, but only using 27 distinct parameter choices. In my view this is wasteful and is exactly the type of mistake that using Latin Hypercubes instead of full factorial designs is so good at avoiding. In particular, the 2 ensembles with different initial conditions need not all have been run at the exact same parameter choices and emulation could have

teased out the different effects. This point needs more discussion and justification. You nicely set out the goal of the paper to explore sensitivity to the parameters, yet 2/3 of your run budget is spent checking the effects of initial conditions or of a different model configuration. Only having 2 sets of initial conditions also seems insufficient. Other choices were available. For example, with an assumption that the initial condition uncertainty was constant in parameter space (one I believe you make in defining the nugget, though not one that isn't contentious), the variance could well be estimated by choosing a point in parameter space and running say 10 different settings of the initial conditions there. Experimental design is about more than just choosing the values in the design matrix. It would be unreasonable to ask for the experiment or the results to change, but I think this partitioning of your run budget should be highlighted and the penalty in terms of not exploring the design space as well as you might have, and not estimating the nugget as well as you could have, discussed. This will enable readers who might consider performing their own sensitivity analyses in this way to consider what they want out of their experiments when they partition their own run budget, rather than using this as a road map.

- Section 2.3.2 should come immediately after line 5 on page 8. This section explains what it is we are looking for and what the different quantities mean. It seems strange to talk about Gaussian processes and to have 8 subsequent numbered equations for all the quantities we want before knowing what these quantities are or why we want them. Explain what we are looking for and why. Then explain how we get it using emulators via the equations.
- Page 8 line 5. The expression for V_i and the subsequent interpretation in words are not enough for clarity. Please add explicit equations for V_i and the subsequently redefined indices you talk about in 2.3.2. Though I know what you mean when you say "the variance is with respect to x_i , I am a statistician and I have to think extremely carefully about this (or write the equations for myself). The readership of the journal should have it easier. Please then check that you mean page 10 line 10, and give an example or a further illustrative equation as it is not clear why these two should have the same value without the presence of non-linear interactions or correlated inputs. This is an important fact to check, but I'm not sure there isn't a - sign missing and, if not, I have to go through 3 layers of worded explanation and the new context mentioned on line 6 of page 10 in order to guess the correct equations to then prove it to myself. You don't want the reader to work that hard.
- Page 14-15. It is fine to point to Oakley and O'Hagan for a detailed derivation, but for the paper to be self contained, a little more detail and explanation is required. Remind the reader here what the V quantities we are using the GP to estimate here are (in terms of integrals of the simulator output over the input space) and include an extra symbolic step to show them where the emulator replaces the simulator in these expressions. This will allow the reader to understand what equations (15)-(19) are for and how they combine to give (20). The point of this isn't to prove that the mathematics is correct by deriving (15)-(19). It is to guide the reader carefully to the main results so that they understand them.
- Page 17 line 15 onwards. The main issue is that you will use incorrect emulator results for obliquities that you already know lead to the oscillation, but choose to ignore, in your sensitivity analysis, thus biasing your results. To say this even more clearly: You make an incorrect and simplifying assumption that oscillatory

behaviour cannot happen anywhere in the phase space, instead of conditioning your analysis on the event that the parameters are not such that the behaviour occurs. I am fine with you doing this given your explanation in the response to reviewers, as long as it is stated in this way up front. Instead of hinting at a generic design strategy, you should also say that the UQ literature has a way around this problem already that would build this conditioning in. History matching, (e.g. Edwards et al. 2011 and Williamson et al. 2014) can be used to find the region where the oscillation does not occur (most of the space), so that in a more careful study you could correct this bias, even if it were inappropriate to attempt to capture the oscillatory behaviour with a separate emulator (i.e. you could do the conditional sensitivity analysis, but you have chosen to spend your run budget in a different way).

- Every point in the discussion requires the qualifier "in the LOVECLIM climate model" somewhere, and this discussed. It reads as though we can actually conclude these as scientific facts about the world and this is definitely not the case. Your conclusions and results are valid for the model (conditionally, see previous comment), but to claim these as results of palaeoclimate significance without a disclaimer is too suggestive that the model is informative for the real world in all parts of the explored parameter space.

3 Minor comments

- Page 5 Line 2 Readers of this journal won't know what you are sampling nor what the space you are referring to is.
- Page 5 Line 20. e and ϵ are not defined, which means it is not clear why i_1 and i_2 are uncorrelated with i_3 . I believe the first version did define these, so those definitions need to go back in.
- Page 6: Line 12: "The monograph by (e.g. Santner et al 2003)...."
- Page 6: Line 15 "an Gaussian process" -- > "a Gaussian process"
- Page 7: There exist papers that have tackled design in constrained spaces before. Please at least mention this and perhaps cite
 - Dragulic, D., Santner, T. J., Dean, A. M. (2012). Noncollapsing space-filling designs for bounded nonrectangular regions. *Technometrics*, 54(2), 169-178.
 - Vernon, I., Goldstein, M., and Bower, R. G. 2010. Galaxy formation: a Bayesian uncertainty analysis. *Bayesian Analysis* 5(4): 619-846, with Discussion.

and the paper by Williamson and Vernon you mentioned in your response to my first review. The first of these three obtains maximin designs for regions with simple known constraints such as the ones you represent, whilst the latter two address the problem of arbitrary constraints, with the first using your sampling solution and the second addressing some of the short comings of this.

- Page 8, line 9 "Then, the main effect [of x_i] is also..."
- Page 8, line 13 change "that" to from and add "that" before the colon.

- Page 8 line 15. V undefined even though V_i is.
- Page 9. Equation (9), the first integral is missing a transpose I think, else, if $m(x)$ is not vector valued, write this as $m(x)^2$.
- Page 9 line 11. What other interpretation could MM' have? This sentence appears to be redundant,
- Page 9 line 12 inconsistent placement of the star next to x 's when compared with the equations on page 8.
- Page 9 line 13 "This is what we refer [to] in the following..."
- Page 10 line 8 and 9. You don't fix or vary i , but rather x_i or forcing i .
- Page 10 line 12. Having 2 indices for T is new and comes without explanation.
- Section 2.4.2 line 5, Y is $p \times n$ and \bar{Y} is $p \times 1$ hence the two arrays are non-conformable. You need an operator on \bar{Y} .
- 2.4.2 line 5, D is "non-rectangular diagonal" with this way of writing the SVD. Most programs save on storage by making D square diagonal, then making U or V non-rectangular (depending on whether n is larger than p).
- Bottom page 11 top page 12. Wilkinson, who I note is an author of this paper, makes explicit the approximation using the first n' eigenvectors by adding an error term. This is important because you don't assume the output lies in the space spanned by the first n' eigenvectors, but that there is a point in that subspace that approximates the output with an appropriate error. By appropriate, I mean you add an error term to the equation at the bottom of 11 and make its distribution explicit in the modelled as Wilkinson 2010 did. Though I leave it to you whether or not to make this error term explicit in your revision, the sentence at the top of page 12 needs to be made precise.
- Page 12 line 7. Gaussian processes should be standard for climate model emulation but are not yet close. Most of the main papers on climate model emulation for the expensive models at least, Rougier 2009, Murphy et al 2009, Sexton et al 2011, Williamson et al 2013, do not in fact fit GPs. Then there is the huge literature on pattern scaling (which is also being called emulation) that fits even simpler models. This is still a niche approach for climate models, but it is the right one.
- Page 12 line 7 $h(x)h$ should be $h(x)'\beta$.
- Page 12 line 7: To call this the prior mean, you should make explicit the fact that we are uncertain about β and that you are imposing the non-informative prior $\pi(\beta) \propto 1$. It may be easier not to use the language prior mean at this stage, or to say that you have a prior mean for the GP conditional on β
- Page 12 line 8 beta appears in words not symbolically.
- Page 12 line 16, again you don't quite have the prior predictive for y without either the priors for the parameters or making the conditioning explicit. E.g. you mean $y_k | \beta, \sigma \sim N(\cdot, \cdot)$

- Page 12 line 22: The joint pdf of β and σ^2 is proportional to σ^{-2} using the weak prior. This is not what you have written.
- Page 13 introduces new parameters which has implications for all of the conditioning on page 12. Everything you said on page 12 is now conditioned on fixed values of Λ_k and ν_k . The way you set up everything makes it seem as though the correlation function is known (and by extension any parameters). So the conditioning is implicit and OK. But once you make these parameters uncertain and estimable, it is important to be clear that everything you've said up to now is conditioned on fixed values of these.
- Comment that ν_k has been given no explicit dependence on x . As argued in, for example, Williamson et al. 2014 (Climate Dynamics) (and one of the Rougier papers, though which escapes me), the nugget is also a way of accounting for internal variability. Hence your repeat runs do offer some information on estimating ν_k , though you make the choice not to use it (another fact that requires comment).
- Page 14 line 7-8. Remove "if the product $u_k u_k'$ is interpreted as an outer product". This seems redundant (what other interpretation might there be?).
- Page 14 line 9, To make sure you don't introduce undefined parameters, please change this to a version of "Let $\nu_k = \nu$ and $\Lambda_k = \Lambda$ for $k = 1, \dots, n$ ".
- Page 15 and elsewhere. The use of n' when you have opted to use the $'$ symbol to indicate the transpose of a vector or matrix is not ideal. Either avoid $'$ unless indicating a transpose in your notation, or use something like A^T to indicate the transpose. Either is acceptable.
- Page 15 line 10. Please make the equation for C explicit. You want readers to be able to reproduce this technology and do it correctly, not have to work out some of the pieces for themselves.
- Page 16 line 6, clarify use of "output variable" to mean spatial-(temporal?) field for a particular output by example (e.g. temperature).
- Expand the first point of 2.6. The same length scale assumption is used in the independent approach, but you don't really make that explicit, though I think you hint at it. The main point is that far fewer emulators are required. The other point is that both approaches make this simplifying step, but you claim it is easier to check your version as there are few GPs to fit when doing the sensitivity analysis. It's a good point that should be made clearer. Perhaps this first bullet point should be two points?
- Page 16 line 12, change "implicitly assumes" to "imposes".
- Page 16 point 2. It is not clear why imposing this covariance structure in the emulator will reduce your posterior variance, nor what this reduction is relative to. Please clarify. To the last sentence of this point, I'm not sure how using this form of emulator exacerbates the dependency of the analysis on the choice of design any more than any other method that uses the design to fit the model using the type of weak prior you use. Sure, the dependency is there, but because you use a weak prior on Σ , not because of the particular features of the principal component structure and I don't see how these features make it worse.

- Page 16 line 17 "co -variance" -- > "covariance".
- Page 18 line 18: PCs
- Page 18 line 26. What does this sentence mean? Do you mean "outside" instead of "off". And why should this be true at all? Consider a binomial with 27 trials and probability of success of 0.01 (the chance we are outside of the 99% credible interval). The probability of at least one success is over 0.23!
- Page 22 line 27 "A variable responding linearity"
- Page 26 line 24 missing bracket