Geoscientific
Model Development
Discussions

[GMDD]

Interactive
comment

Interactive comment on "Global Sensitivity Analysis of Parameter Uncertainty in Landscape Evolution Models" *by* Christopher J. Skinner et al.
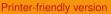
D.E.J. Hobley (Referee)

hobleyd@cardiff.ac.uk

Received and published: 17 April 2018

First of all, sincerest apologies for the extreme delay in providing this review.

In this submission Skinner and colleagues present a new approach to understanding the sensitivity of landscape evolution models using the Morris method, and using model functions in place of objective functions. This approach is semi-quantitative and somewhat subjective, but nonetheless has utility in assessing sensitivity for such models where computational demands may be prohibitive for a "proper" SA. They outline the method and apply it to the CAESAR-Lisflood model being used to simulate a pair of catchments. They illustrate that the approach works in this context, and that it highlights the key importance of selection and calibration of the sediment flux law above all

other parameters. They also discuss other aspects of the model utility, using these two cases as examples.

I enjoyed this manuscript; the approach seems simple, but given the dire state of past attempted SA in geomorphic modelling this is a very much worthwhile contribution to the literature. In my opinion it requires minor to moderate revision before acceptance, as detailed below. My primary concerns relate to lack of clarity in the methods.

Per GMD's review criteria: 1. I believe this paper sits within the scope of GMD, though I do not feel bet placed to judge this. It presents a novel approach to the sensitivity analysis of LEMs. 2. Both ideas and tools are novel. 3. The paper seems to represent a significant advance in the state of the art of sensitivity analysis within the field. 4. Assumptions are made clear, but description of methods needs further attention. As it stands, the method could not be understood in its entirety based only on the text. 5. Results support the interpretations and conclusions, assuming I have adequately followed a couple of opaque parts of the methods. 6. See 4. Significantly more methodological detail is needed. 7. Credit is given. Abstract could even put more emphasis on the "model function" aspect of this work, which seems novel and key to the approach. 8. Title describes paper 9. Abstract is concise, though needs a touch more definition of terms to make it crystal clear to the non-expert. 10. Presentation is good 11. Language is fluent and largely precise, though I have flagged up a few instances of imprecision related to the naming of model input and output information ("parameters"?). 12. Symbology is good, though the equation presented perhaps could be tweaked to enhance clarity. 13. Some clarification is necessary throughout. Structure is good. 14. References good. 15. Code is not supplied, but freely available via the net. However, the supplementary information is confusing, in that S2 does not appear to be referred to from the text. This needs to be thought through and resolved. Given the brevity of the paper and of S2, and the importance of the topic it discusses, it should probably be integrated into the main text.

I have not attempted to formally assess the fit of this manuscript to GMD, though I

believe it is appropriate. I have also not attempted to check in any way that the detailed requirements for publication in GMD (e.g. version numbers, adequate documentation) are all met, and leave this to the editor.

Most of my comments are best suited to a line-by-line approach. However, a brief overview is warranted:

* The main issue I had with the manuscript was related to lack of detail in the methods, and some sections where key concepts needed expounding on more. The inline comments detail this, but this is essential work. It is difficult to interpret the presented results for yourself because some crucial, detailed information is missing. * In particular, the methods are very opaque when thinking about the time component of the models, in that you basically don't talk about it. How is it determined when a model run is "done"? Is there some external constraint on total time to run for? Please provide more information. * The importance of the sediment equation choice. Skimming Andy's review, I think I largely agree with his criticisms on this front, though see below for detail in this review. I suspect a lot of this criticism is again coming from too brief a description of this part of the methods. You may be able to head off a number of my concerns simply by expanding, and taking a more pragmatic approach to why you've made these assumptions (i.e., this is how a lot of models are applied "in the wild", ignoring known geomorphic complexity, so this is how you've done it here; this is an illustrative study so almost the details of the actual geomorphology in those places don't matter; it's instructive to see if there's any influence from known imperfections in the model assumptions; etc) * I found it hard to keep some of your terminology straight, largely around which "parameters" or "metrics" you meant at points later in the text. Take more time to define things more clearly at first use, then be very careful to define those two terms as one of your other input/output classes whenever used subsequently (there's a lot of detail on this below). * A variety of stylistic/text things, though most of these a copyeditor will catch (e.g., rogue capitalisation) * Some concepts need to be introduced earlier in the manuscript, e.g., Morris method definition, model function. *

I think the importance and novelty of the Model Function approach in this context is quite underplayed, and could be brought out more.

In summary, I thought this was a succinct, neatly packaged study that achieved its stated objectives, and warrants publication in GMD once it has been expanded a little. I am of course happy to provide further clarification by GMD's discussion mechanism. I look forward to seeing it promoted out of discussion paper status soon.

Dan Hobley

Inline comments:

Abstract: I think the briefest of introductions to objective functions, and model functions would be appropriate inside the abstract, since they are the core of the paper.

31: "dominant" 47: perhaps "from Earth surface processes, for instance, . . ." 61: Surely "legitimated by theories" -> "directly physically constrained or measurable" 66: Probably add something to the effect of: SAs are key in scenarios where input parameters are tuned (i.e., link to the prev bit about parameterisation more firmly) 74: A bit more context to set the scene here: which fields in environmental sciences have been doing this well? Can you give a couple of examples? 92: capitalisation 116: "The study" – Ziliani ref has not appeared any time recently in the text. Rephrase this bit, and set the context for why you're about to discuss this specific study. 118: Perhaps "see below" for the MM Section 1.1 – I found this to be very clear. 130-131: Grammar issue to do with word "both" 127-141: Another major issue is that even with data in hand, it is challenging to derive meaningful metrics from them. E.g., a pixel-by-pixel map of topography is not a good optimisation target, since small error in the exact position of the channel gets magnified when in fact the model may be doing a good job overall. So integrative metrics must be found, and that in turn is subjective. There's also the related issue of stochastic processes inside landscape response being hard to capture in a spatially resolved way; see, e.g., work by Mary Hill (I see this gets discussed obliquely around ln 150, but be a bit clearer about it – i.e., give us the logic by which Hancock & Wilgoose

C4

(ln 154) started using statistical measures in the first place). 157: "the LEM SIBERIA" 160-161: "these measurements" not clear which measurements. Rephrase. Ln 164: For me, "/reliability" is redundant here. If I was you, steer clear of the implication that an accurate model is necessarily good for future prediction, as that's a slightly different thing. 172: "MM" You need to introduce the Morris Method in the intro somewhere to say why it's important, who has tried using it before, etc etc. Calling forward to the methods for most of it is fine, but the intro needs a little more. 172: Dash is probably a new para. 175: As for the MM, the intro also needs to define a "model function" before you get here. (Also, consistency of capitalisation!)

Ln 184: As above, please deal with the MM definition before we get here. 195-6: This needs a supporting reference. 2.2 – a big chunk of this stuff should be sat in the intro, as it introduces the method and its background. 244:"stochastically"->"randomly"? 246-7: "a number": how many? Is this constant? Do we step around the values in a random-walk fashion, or is this more like a ratchet to move through all possibilities? i.e., more detail here. 250: the Main Effect can't be both the mean and SD of the EEs, as the text here implies as written. Presumably it's the mean, but make it clear. The role and handling of the SD in general in this method is unclear, partly due to folding it into the ME here then talking later about parameter normalisation against the ME. This makes it difficult to parse whether the SD is normalised or not, and if not, why not. (cf lns 362, 377->, fig 3) Expand and take more time over the details here. ln 255: something not right with this equation – it doesn't read sensibly. Neither r nor j appears in the RHS of the equation. Feels like there should be subscript outside the bar symbology indicating how we move through j space. Check the equation, consider again if this is the clearest way you can show it, and be considerably more generous with the explanation. The fact that an EE is parameterised as "d" seems somewhat perverse... This would be a lot clearer if you gave a concrete example here, e.g., say what the data structure would look like for a given model output measure *and* Model Function. It's very opaque how model output measures and model functions differ. Make this explicit both here and especially later in 2.5. 271: repeated how? Presumably there are other

inputs to the model besides precip (e.g., land use, topo, etc etc) so either talk about them all here, or none of them. What's the source of the DEM in each case? 2.3.2: as at 271; more of this info is presented here, but it still needs more. 2.3.1 & 2.3.2: What e.g. Shrahler order are these streams? This needs saying to present a contrast with your different order measure, lns 285-287. Fig 1 caption: Add "Note difference in scale" 285-7: What is the justification for this division, which, AFAIK, is novel? Explain why an existing stream order method is not chosen. 303-304: "excluding... dune and soil development". Return to 2.1 and explain in more detail which processes are turned on and off in CAESAR-Lisflood. (Note this is not the same thing as which are being tested as part of the SA.) 305: Please give information on the typical run time of each model run on whatever rig was used to give the reader some context on what this information means. Ln 311: I believe the fragment about it being a qualitative method is unnecessary here- delete it. I'm not sure it is technically true, and not needed anyway. 312-318: "it's difficult to define what this means" I agree...! I get the gist of this, but you need to tighten it up and expand significantly. Delete the sentence about "broadly equal" – which isn't true, since I assume the values themselves are not equal, and say what you actually mean ("to have an approximately equal influence on the output, as defined by...?" Or something??) If you get it right, you can and should remove the "It is difficult to define what this means". Ln 316 – On what basis is it sometimes not appropriate to do this? Ln 317/8 – give us the specifics of the steps you used for Manning's n, to make this more concrete.

320-22: This could do with a lot of expansion, as it becomes key to the results and discussion. On what basis were these two laws selected from the large field of possibilities? The second sentence needs a lot more detail. This is the first time we're hearing that the MM can be used to tackle epistemic uncertainty in model set-up, rather than common-or-garden parameter uncertainty. Tell us a lot more about how this works and why you're doing it, then explain with more detail what exactly "as binary two-step" means. I presume you mean the choice of law becomes a parameter in the MM, but in that case, aren't you also switching in and out another big subset of the parameters?

So how is your data now comparable? Doesn't the parameter set have to stay static for the MM to work? I may have misunderstood, but if I have, it's a sign you need lots more detail on how the MM works. 327/8: Formatting of ref. Otherwise, I like this GS description.

Ln 340-344: This could be clearer. Surely some statistical methods have promise; after all, you're about to devise one. You need something in here to refine the scope of the methods you're talking about, e.g., do you just mean to exclude previously attempted methods? Is it the fact that single discrimination criteria don't work, and that's the problem? Etc Ln 344: say why these reviewed methods failed, since referring out the the 2001 work doesn't provide the context. Ln 349: You need to be clearer in defining that Model Functions are a new thing that you're creating in this work, and that this is a major contribution of this work. The introduction should reflect this (and thoroughly introduce the idea), as should the abstract. If of course it isn't your idea and it's coming from somewhere else, please make that clearer and add the relevant references more clearly. Also, as noted earlier, please explicitly discuss more clearly how model output measures differ from/work with/are part of Model Functions (See also 360-2). Ln 353: dash is a sentence break 360-362. Rephrase; this doesn't fully make sense. Do you really mean "versus"? I think part of this needs to read "was normalised to the highest ME for any parameter within the Model Function". Again, it's really hard to follow and keep straight the respective roles and interrelationships of: the parameters back in the equation (i.e., i, j), vs what you mean by "parameters", vs Model Functions, vs model output measures, vs MEs, etc etc. I'm fairly some of these are equivalent to each other, but it's very hard to keep straight, even when flicking back and forward back to the methods. Make your terminology bulletproof, crisply defined, and repeat yourself as needed for maximum clarity – as this is very hard to follow. As above, illustration by example would make this a lot clearer. 362/364: "aggregated". I'm conceptually uncomfortable with taking a mean (right? That's what you mean?) of numbers which have already been turned into the equivalent of percentages and scaled to each other. This will create some odd statistical dynamics, I think. I guess this is fine given the

method is already pretty qualitative, but it makes me uneasy. I'd invite the authors to reflect on this, and consider putting something in to reassure the sceptical reader. More concretely, how are you aggregating standard deviation measurements? Are these scaled alongside the MEs? If so, that will very quickly get confusing. Explicitly tell us what you are doing to handle these, and let the reader sort it out for themselves (cf, ln 250). 365-371: I think this means that, implicitly, these methods won't work to compare true transient behaviours? It's impossible to understand without telling us how the method treats time in general. Which leads to. . .

General comment: In the methods, the role of time in this approach is very unclear. Time elapsed is never referred to until you're talking about trimming off spinup. Is EE calculated continuously through time, and you stop at the best possible value? Are you comparing time series? Or just best fit at a time slice? This information must be present. In general, this is a symptom of your description of the key numerical methods being very brief. Be much more generous, with examples etc. as you work through from EEs to ME to Model Function aggregates. Take as much space as you need and let this "breathe" as much as you can.

Figure 3: I'm concerned (possibly unnecessarily?) about cross-correlation of the means and SDs here (cf, ln 362). If you normalise the SD's by the maximum ME, then you're building in dependence of one on the other. If you don't, how did you aggregate those SDs? Are these means of means of SDs for each EE cluster, or are you taking statistical measures of ME distributions themselves? You need more methods clarity to make this easy to understand and intuitive to interpret.

Fig 5: please add the abbreviation to each graph caption. Each sub-graph also needs a label a, b, c etc. Although the caption specifies it, a heading on each column of graphs giving the location would make the figure easier to read. Adding "(biased smaller)" and "(biased larger)" annotations above 1-2 and 4-5 for the GSS graph would prevent the reader having to flick back a number of pages seeking the description of what these are.

3.3: On what basis were SED, CVS and GSS selected for this section? Justify.

425: This is too terse to make it clear. Expand & rephrase to be more precise about what you mean. It's not clear where you are talking about the all the figures in the figure at once, when just one site vs the other, or when you mean variation between parameters in patterns shared across both sites.

General note: The subsections in section 4 are formatted differently to everywhere else.

431: "The implications. . . have been discussed". I don't feel like I've seen a discussion of the implications; the methods were very factual (rightly). You need to re-summarise whatever implications you are thinking of here. If you mean results, say so – though this would seem somewhat redundant.

435/7: "metrics". Your terminology has been imprecise throughout – we can now add "metrics" to "parameters", "Model Functions", "model output measures" and "MEs" as a list of terms it's really easy to lose track of and mix up. I would advise you to remove all mention of metrics and parameters from the manuscript, or at least define which thing you mean by it each time. So here, I think you mean model output measures, right? Get everything internally consistent.

450: "basin's". Nice points here.

463-5: "nonstationary". If I've followed (cc past comments on approach taken through time), I don't think the approach described in this paper as currently deployed would be able to analyse this kind of nonstationarity. You should say that explicitly here, though the point is well taken.

4.3: Again, points well taken, but this is more a literature discussion. This should be grounded more in your specific results. You've shown that this is the most important choice you can make in CAESAR-LF, so what does that mean? Should users be calibrating as for e.g. Siberia? Is there reason to expect this overwhelming importance

to be carried over to other sites? Does this mean calibrating SED well is by far the most important thing, so other effects are secondary? This section is negative/bear-ish, and doesn't outline any actual approaches, advice, or other opportunities to use the SA productively to get around this problem. How bad actually is the lack of constraint on SED? Even a statement along the lines of "this analysis suggests that detailed justification and calibration of model choices around sediment transport will lead to the most effective gains in model skill" would help. This seems like a key result here.

And leading on from this ->

Ln 528: "binary". I also have some issues here regarding the implementation of the SED equations – you say the choice was binary, but how then did you calibrate the internal parameters in each equation (cf, comments in methods section above)? This might be resolved by more clarity in the methods, but also, it seems like maybe you could have got inside this element of the model a bit more and explored those sub-parameters too, per Andy's comments. Are you simply applying the different equation with the same parameters in each? Or do you allow some kind of pre-optimisation to get the params to where they need to be in each field site? Additional material in the methods clarifying this is essential, and I'd also strongly advise you to refer explicitly to how this choice of sed law parameterisation is affecting the importance of the SED param, both here and in 4.3 above. I see material in S2 discusses this, yet isn't referenced from the text. I'd advocate pulling that material into the main text as part of addressing this comment.

Conclusion: please switch out MNR, GSS and VEG, IOD, etc etc for text equivalents to make this conclusion more stand-alone. I also feel like the influence of SED is a key result, and should also be mentioned in here. Otherwise, I like this as a summary.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-236, 2017.