**GMD Paper Review Response – Reviewer 1**

**Summary to both Reviewers**

The authors would like to thank both reviewers for their detailed and insightful comments. We believe we can readily address their concerns with the changes outlined below. Overall, we have made it clearer that this paper's purpose is as an illustration of how the Morris Method can be used to understand model sensitivity in complex, multiparameter distributed models, using the example of the CAESAR-Lisflood LEM.

In summary, the main edits we will make are –

1. Provide a clearer justification for our choices for the way sediment transport formulae (SED) were handled in the Morris Method. However, Reviewer one suggested an additional Morris Method test using a sub-set of parameters using only extreme values to better understand how fewer numbers of steps for SED in the method presented might influence our conclusions. As an alternative – but giving the same effect - we have assumed that changing SED is a 2 and 4 iterative step change within the method and divided each associated elementary effect by 2 and 4 accordingly, then plotted the data compared to our original outputs. These show a progressive reduction in the role of SED and we will add this test into the supplemental material.
2. Increase detail and clarity in the methodology employed.
3. Tighten up terminology. The term "model output measure" will be removed as is covered by "model function". We will include a brief glossary in the methods –
   - Parameter – Adjustable value within a model. The value is determined during model set-up. The value is often based on recorded values or adjusted during calibration.
   - Objective function – an error score between model outputs and observations used to evaluate model performance.
   - Model function – a measure derived from model outputs used to evaluate model behaviour in lieu of an adequate objective function.
   - Elementary effect – a value used as part of the Morris Method, indicating the change in function value (objective or model) resulting from a change of parameter value during a single repeat.
   - Main effect – the mean of the elementary effects from all repeats, for a specified parameter and a specified function.
4. Increase the prominence of the model functions within the manuscript, including mentioning them in the abstract and introduction.
5. Be more explicit with the 30-year timeframe used for the tests. This choice served two purposes – to keep model run times at a manageable level, and to show an operationally relevant timeframe where rigorous model assessment is required for decision making.
6. Define and describe the Morris Method earlier in the manuscript.
7. Make it clearer that this is a test to demonstrate the method and understand the behaviour of single LEM in more detail. No modifications were made to the model function and it is used in its 'off-the-shelf' form. It is not an effort to calibrate a model or to reproduce a real scenario or observed data. As such some things are excluded or conceptualised – starting parameter values are not optimised, parameters are applied globally, there is no representation of bedrock included in the model. This also justifies the choices of sediment transport formulae.

In addition to these changes, we will address further issues raised by the reviews, and also line-by-line changes. Below we list our responses to these, excluding any references to grammar or typological changes.

Reviewer comments are shown in black, *and author responses are shown in red*.

**Review 1 – A Wickert**

Skinner and coauthors' sensitivity analysis of landscape evolution models is a much needed addition to the geomorphic literature. Many evolution models often have been treated as sandboxes in which to experiment with quantified conceptual process understandings rather than as predictive models, with a few exceptions that include earlier work led by coauthor Coulthard using the CAESAR model. I also think that I have known of this work in progress for some time, as my lab donated some compute time at the request of coauthor Schwanghart.

Conceptually, I find the idea of a sensitivity analysis to be a very good approach. Often, landscape evolution models include competing processes designed to simulate the effects of real processes. Such additions typically include descriptions (qualitative and/or quantitative) if their impact on a landscape, but often do not include a more mathematical analysis of how these processes influence the solution space.

*Thank you very much for the comments. We are glad you appreciate the complexity of the task – there is a great deal of information in and behind the paper and choosing what to highlight and discuss was a repeated issue in the paper's formulation.*

The work of Skinner and coauthors is significant and publishable, but on reading their paper, I found multiple causes for concern. I could not find easy answers to the most major concerns, enumerated below.
• The sediment transport formula was the dominant source of uncertainty, but I believe that this may in part be because these sediment transport formulas were not appropriate for Tin Camp Creek (Australia).
– The grain size distributions for the rivers displayed significance of sand in the UK and a dominance of sand in Australia. Both the Wilcock and Crowe (2003) and the Einstein (1940's-50's) formulas are tested with coarse sand as the smallest grain size class. In the Australian case, about 50% of the sand is finer the grain size used to produce the sediment transport formula.
This is at the upper limit of the curve in Figure 6 of Wilcock and Crowe (2003), where their solution begins to bend more sharply but the data end. Therefore, there is great uncertainty and little constraint in the formulation.

*We are not attempting to fit the model perfectly to each environment, and the use of Einstein and Wilcock and Crow (W&C) was largely because these are the formulae available in the un-modified model as downloaded.*

*There is a general point about sediment transport formulae and their applicability – that they generally perform well on the data they were generated from, but much less so when applied to other circumstances (eg Gomez and Church, 1989 and others since). So in taking any sediment transport rule out of its 'comfort zone' we will encounter issues.*

*For the second point, looking at Wilcock and Crowe (2003) they use sand fractions of 0.0005 to 0.64 m (Figures 1, 2 & 3, and P121) in their 48 flume experiments from which their formula was derived. For our first catchment (the Swale) there is a pretty good agreement between the sediment grainsizes used and the W&C ones. In the second catchment – yes, there are finer sands and a greater proportion of*

*sand in those simulations than used in the formula development. However, we find the same sensitivity to sediment transport formula in both basins, despite the different grain size mixes. Sediment transport formula choice has a much stronger impact than grainsizes as shown in our Figure 3.*

*The issue of choice of sediment transport model and its applicability to any site is difficult to address as there is no general model. One can either calibrate to the site or develop a model specifically for the site which is impossible for most applications. This is no different to hydrological modelling where no model fits all situations. All model outputs are limited by this. We hope to get this point across in the paper.*

*The application to real basins – with their representative grainsizes is a nice thing to have – but the important finding is the similar sensitivity to sediment formula choice in two quite different settings/basins. Therefore, we argue that whilst our application of the W&C to the Australian example may be outside the limits of the initial W&C development – it does not mean that the overall finding is incorrect. There is certainly scope for a study/paper on the role of sediment transport rules in LEMs (see later comments).*

– The Australian example has a dominance of sand. Are there bedforms that appear in the river? If so, could you discuss the role of their form drag, which to my knowledge is not included in your model, and how it could affect sensitivity to choice of sediment transport formulations?

*To clarify, there are bedforms in the creek – and form drag is not included in the model. CAESAR-Lisflood will not generate realistic bedforms at this scale/resolution of application and these are not factored into the sediment transport formulas.*

– I find your discussion of sediment transport in section 4.3 to be unnecessarily vague. It is not unusual to see in the landscape evolution modelling literature a statement to the effect of "sediment transport formulas are problematic and it is a difficult thing so the error is probably there". Scientifically, this is unhelpful and in my opinion a little lazy. I think that here you have the opportunity to analyze why this is your major source of uncertainty, which is one way in which I hope this study can rise up above the others. Regardless of whether anyone trusts the form of your sediment transport formulations for the chosen grain size, form drag, etc., you have two mathematical formulations that must produce divergent outcomes for the set of provided hydrologic and topographic states. (I am presuming that over your 30-year time scales of interest, overall topography changes little.) Based on an analysis of these formulations, can you make a prediction of the factors that lead to this divergence?

*These are really good points, well made. There is much scope (in another study) to look at how different sediment transport formulae respond (in the different settings) in this model that would be incredibly instructive to the LEM community and this is something we are working on (but first we need to look at how all the model parameters interact and influence model behaviour).*

*While not a direct comparison as prescribed in this paper but in the context of model testing Coulthard and Hancock have examined geomorphic change by comparing CAESAR and SIBERIA models over millennia. These tests show a lack of divergence and an equifinality of final form.*

*However, we would also like to emphasise that the primary purpose of this paper is a methods paper, introducing a method to assess the sensitivity of LEMs to changes in parameter values. Therefore, a detailed analysis of different sediment transport formulae is beyond the scope of this paper and also beyond the scope of GMD and would be better placed in an Earth surface based journal. However, we agree that this would be incredibly valuable and a more in-depth test utilising more sediment transport formulae is planned. We see this paper as the introduction of a longer and larger set of experiments*

• Your discussion notes that the sediment transport formula's importance may be overstated due to the smaller number of options for this than for the other variables. However, you do not analyze this possibility. or whether this would even lead to sediment transport formula remaining the dominant influence. Could you argue how your conclusion about sediment transport formulas is (or is not) still valid, considering this? I make a suggestion below (530-533).

*The way we applied and dealt with the SED parameter in the Morris Method was selected because these were choices available in the version of the model downloaded. By applying it as if it were a parameter, an operator is presented with an either/or decision in which formula to apply – there is no in between (even with more choices of formulae) – and we want to understand the influence of that choice on the behaviour of the model. The Morris Method is subjective, but when applied correctly does provide useful information. A more detail response can be found below (530-533). Even if we had used a choice of 5 formulae, it would not be possible to arrange these in an order of iterative steps within min-max bounds, therefore any switch from one to another must be always be handled as a change of only a single iterative step.*

• Your premise is to test landscape evolution models, but the 30-year model run period is much shorter than most geomorphic models are used. Indeed, I wonder how much landscape evolution occurs, versus how much, over these time-scales, CAESAR can be thought of as a sediment-routing model with erosion or deposition being negligible (and therefore avoiding the nonlinearity in which changes in topography affect the long-term response of a LEM.) I think that this short time scale should be made explicit early in the paper. A discussion of how these results can (or cannot) be transferred to different time scales would be helpful as well.

*Yes, this is a much shorter timescale than typically applied, but establishing a dominance on a process or set of processes over shorter times will still provide us with insight into longer term processes. Lessons learnt over the short time scales will apply over longer time scales. The 30 year time scale was largely chosen of model run time convenience – but it is also a timescale that is relevant to contemporary management decision making. Decision making processes based on geomorphic modelling require a far greater standard of understanding and handling of modelling uncertainties than currently employed by the LEM community, and lessons can be learnt from the hydrological and meteorological modelling communities. We will explain this more explicitly in the Introduction (please see point 5 in the Summary).*

In addition to these, the paper would be improved by a careful set of proofreading. It is repetitive in several places and includes a number of issues in both grammar and style. The overarching issues here are:
• Many proper nouns are capitalized; why?
• Your abbreviations should be used with an "s" to indicate whether or not they are plural in a given instance
• The Morris Method is mentioned 2–3 times before it is defined or described. Its description should be more closely tied to its in-text mentions.

*This was also raised by reviewer 2 and we will make sure the method is defined earlier in the revised manuscript (see point 6 in the Summary)*

• You define the difference between an objective function approach and a sensitivity analysis at multiple points; reduce this to one.

*We will edit this down to one, clearer, definition (see points 2 and 3 in the Summary).*

• In general, many explanations are very "hand-wavey". Please do a thorough read-through to reduce the fluff and improve the density of new information. If this is not done, it will be hard for a reader to see what interesting new conclusions you have come to.

*We will sharpen up our descriptions.*

Line-by-line and section-by-section comments are as follows:

Abstract: Why 3 paragraphs? I think you can shorten and tighten this.

*Revised in line with this and reviewer 2's comments.*

18. em dash after models; comma after example. Sensitivity Analyses one example here of something that is capitalized for reasons I don't understand; I don't think that this is just UK English.

*Capitalised as it should be defining the acronym SA, but this is not included. Will edit accordingly.*

47. I do not believe that your above references cover glacial or aeolian processes. Disregard if I am incorrect (does CAESAR include aeolian processes?); add references or remove these notes if I am. They are unimportant anyway to the landscapes that you are studying.

*Will change to just "Earth surface processes" as suggested by reviewer 2.*

61. Comment: even few-parameter stream-power-based LEMs are quite heuristic. I do not think that your note here is unique to models with large numbers of parameters.

*Good point – will include this.*

65. Correctness: an analysis cannot investigate, but you can.

79-85. I appreciate this list!

*Thank you.*

122. Define MM here or list section in which it is defined.

*Will define earlier (point 6).*

128. Sentence fragment after comma

130. Incorrect in general: many landscape evolution models are not designed to be predictive over annual to decadal time scales. I find CAESAR to be quite unique in its time-scale flexibility, due (I believe) to its explicit integration of flow and sediment transport processes.

*Will edit to make it clear that this is an issue relating to the nature of CAESAR-Lisflood and relates to the initial motivation of the work, ie it is tempting for operators to use the model to simulate changes*

*over these shorter timeframes, but it has not been tested. This is a major issue if it is used for decision making purposes (see point 5).*

131. I don't see what you mean by "multi-dimensional approach"

*The phrasing is unclear. It means that the performance of the models needs to be assessed across all these timeframes. An LEM might produce reasonable behaviour when assessed at a millennial timeframe, but the behaviour at smaller timesteps could have no physical-basis. There could also be an element of equifinality where similar outputs from longer-term model simulations emerge after very different patterns of short-term behaviours. A model should be able to reproduce correct behaviours at all these timeframes and should be assessed on this basis.*

133. When is an objective function not a score between observed and simulated values? Or do you mean that we can have synthetic observations?

*Good point – will edit (also see point 3).*

149-152. I think that point-based measurements must account for all of the complexity in the system, but may not be able to distinguish the source of the measured parameter's value.

*A point-based measurement could never possible account for all the complexity in a system. For instance some changes may be spatially restricted – material is eroded, transported, and deposited wholly within the catchment so no signal of this change will ever cross the catchment outlet.*

155. What is a width function?

*According to Lashermes and Foufoula-Georgiou (2007), the "width function of a river network is a one-dimensional function which summarizes the two-dimensional branching structure of the river network. It represents the distribution of travel distances through the network and, under the assumption of constant flow velocity, the probability distribution of traveltimes. Thus its significance for understanding the hydrologic response of basins and the scaling characteristics of streamflow hydrographs is important". It is described in Hancock and Wilgoose (2001) as cited.*

*Lashermes, B., Foufoula-Georgiou, E., 2007. Area and width functions of river networks: new results on multifractal properties. Water Resources Research 43, W09405–W09405.*

155. Cumulative area distribution – of what area?

*Described in Hancock and Wilgoose (2001) as cited – "The cumulative area distribution (CAD) is a function defining the proportion of the catchment which has a drainage area greater than or equal to a specified drainage area. The CAD describes the spatial distribution of areas and drainage network aggregation properties within a catchment."*

156. It is not possible to use variables as an objective function. One needs... a function. This may include these variables, of course!

*These are not variables but values extracted from physical and numerical experiments, and compared as an objective function. Will edit this sentence to make this more explicit.*

162. , and so are (comma)

162. "more objective" is vague: do you mean more in quantity or more as in better?

*Better. Will change.*

165. "data" is plural. "data" is included twice as well, and the sentence is generally awkward.

167. Really? These data are not available? Not even in heavily-monitored experimental catchments? It is difficult to make sweeping statements, so I would ask you to prove this.

*Good point – will mention that some information can be gathered from heavily-monitored experimental catchments.*

175. which –> that. This is an important distinction, and is often overlooked.

176. rm "will": tense confusion

185-186. "Medium" and "small" are nearly meaningless; could you provide catchment areas?

*Will replace with area values.*

200. Is the rainfall time-series uniform in space or not? (I read later that it need not be, so please note this here, as this sets CAESAR ahead of other LEMs)

*The rainfall time-series is derived from RADAR rainfall estimates and vary spatially for the Swale. The data has a spatial resolution of 5 km and a temporal resolution of 1 h (see 2.3.1.). Rainfall is uniform for Tin Camp Creek and based on one rainfall station (see 2.3.2.)*

204-205. Note how erosion and sediment transport are calculated here, as this is central to your conclusions.

*We tried to limit our explanation of the model as its functionality is very well defined elsewhere and we have made no changes for this test. However, more detail on this aspect would be helpful and will include.*

205. What is an "active layer system"?

*The active layer system allows for the storage of sub-surface sediment data by keeping it in strata going from a bed rock layer (if used), a base layer, several buried layers, and finally a top, active stream bed layer (Van de Wiel et al, 2007).*

*Van de Wiel MJ, Coulthard T, Macklin M, Lewin J. 2007. Embedding reach-scale fluvial dynamics within the CAESAR cellular automaton landscape evolution model. Geomorphology 90: 283–301. DOI: 10.1016/j.geomorph.2006.10.024*

223-224. This is where singular vs. plural usage of acronyms can stand out.

240. What is the Design of Experiment? And how did it use R?

*We will remove the reference to design of experiment as we do not provide the wider contextual information (which relates to how different SA operate) so is not helpful. Instead we will make it clearer that the "sensitivity" package was used to perform the process described in the paragraph beginning at Line 244.*

250. is –> are

250. these constitute the Main Effect. (otherwise it is not clear how two things become one)

*This needs to be rephrased – there are two main effects, the mean and standard deviation.*

266. elevation drop is also called "relief". I understand it here, but "relief" may be more common.

*Will change to "relief"*

277. Grammar: Contrasting –> In contrast.

279. Comma after "Swale"

280. I thought that "rain gauge" is two words.

*Will change.*

282. Did you therefore use uniform precipitation for the Swale catchment?

*No, rainfall series are based on NIMROD composite RADAR rainfall estimates and vary spatially. The resolution is provided in Section 2.3.1*

Table 1: (3 and 4) Do you mean to say that you prescribed a lateral erosion rate that is constant? This seems strange to me. (9) In/Out of what? (13) Evaporation or ET? (14) Roughness for channel, hillslope, or both?

*CAESAR-Lisflood utilises a constant value for lateral erosion rate, which is different to the in-channel lateral erosion rate, and is part of the meander development section of the model. (9) the In/Out difference is the ratio between water input and water output and is used to decide the timestep the model operates at. (13) Should be ET. (14) global value (see point 7 in the Summary).*

311. It is definitively not qualitative, as it gives you numbers! Perhaps "quantitative but subjective".

*Good point, will remove (as also suggested by reviewer 2).*

320. "Laws" is really strong. "Formulas" could be better. Or formulae, as you seem to be leaning Latin. Except with Germanic-leaning capitalization.

*Yes, will change to formulae.*

Figure 2: Perhaps note the D range of the data with which the sediment transport formulas were created

*Will include a note on this in the manuscript (see point 7 in Summary).*

340-343. Section 1.3 doesn't include what you state here. I also find it hard to believe that topography and discharge should have no relationship to geomorphic change. You will need to provide some evidence.

*Should say "Section 1.2"*

343-344. This is no surprise: they studied equilibrium landforms, while you are studying 30-year time scales in which only extreme events can cause significant landscape change. In other words, your time scale removes the significance of topographic evolution and its associated feedbacks on the system.

*Long-term landscape evolution is disproportionally influenced by successive extreme events. The short and long-term dynamics are intrinsically linked.*

348-350. Yes, you have mentioned this (note my general comment).

370-371. How have you assessed that 10 model years is a sufficient spin-up time?

*It was just kept constant between the catchments. There is a wider issue here on the influence of spin-up time on model behaviour and uncertainties, another aspect we hope to look at as part of the LEMSIP work.*

381, 383, etc. Consider giving full parameter names where possible to help the reader follow the text.

*Originally we did, but the text became unwieldy as some names are long. We will seek a better balance in the revised manuscript.*

385. which –> that

Figure 3. Consider full model names; otherwise, you have converted numbers to codes, which readers will then have to cross-reference with your table.

*Same as above, but will seek a better option.*

Figure 5. Is this catchment-wide elevation change, in-channel elevation change, or otherwise? In addition, are re you certain that mean elevation change is the appropriate metric? I can imagine that significant spatial variation in aggradation vs. incision could occur, and wonder how much this may affect your results. In addition, tens of cm of incision over 30 years seems very rapid to me: could you comment on this?

*Catchment wide, sub-divided by the stream order definition – entire area under appropriate shadings.*

*We don't believe mean elevation change is necessarily a good metric but use it for illustrative purposes here to demonstrate that there are changes spatially on the 1600 DEMs generated by the simulations. Tens of cm of incision over 30 years is not uncommon in certain areas of both catchments – though changes in trunk streams in the Swale are controlled (in the field) by bedrock – which is deliberately not included in these simulations.*

433. Small "s" on "LEMS"

443. How much do you trust gauged suspended sediment discharge and the associated rating curves? I do not know that these are so straightforward either. And if you mean bedload + suspended load, then I would argue that the data generally do not exist.

*Indeed, will add a note here that even when such information exists it is not perfect and has its own uncertainty. Similar could also be said about the rainfall time-series used to drive the model.*

4.1 (general). This section indicates to me that solving the LEM problem may be impractical due to the amount of time-lapse spatially distributed data required. Could you comment on this?

*With present methods it might be. But that does not mean we should not try and advance and seek new methods and techniques to try and address this. LEMs represent a potentially powerful tool for understanding geomorphic impacts due to changing climate, land use, and flood risk interventions, which could be applied for decision making purposes. The rewards for solving the LEM problem are worth pursuing a solution.*

4.2. This section seems just to read, "we don't know how these models work and what the general rules are". It is OK to just write that! This seems to beat around the bush.

*We will tighten this section up.*

456-457. Environmental models can be transferable between catchments. For example, I would argue that a thermal model is very transferable! Please be clear in what you mean by "environmental".

*We are particularly referring to models of open systems which have variable parameters that are calibrated. The calibrated model cannot be directly transferred to another catchment, however similar they are, and will need a new calibration. The same would apply to a sensitivity analysis. Will tidy up the text to make this clearer.*

473. Your sediment transport formulas do not include thresholds. Please explain how this compares to thresholded models if you include this point.

*Wilcock and Crowe operates with a threshold – Einstein does not, but transport rates increase with a (loosely described) cubic function of stream power, in many ways mimicking a threshold in operation.*

475. Your formulas include one that performed well in the Gomez and Church test and one that was not considered. What is the basis for using GC 1989, therefore, to declare that sediment transport formulas are not good?

*Gomez and Church (1989) summarise that no formulae work well outside of the data upon which they were developed. We use this to illustrate a weakness of sediment transport laws that has been identified before.*

477. Both of the formulas that you have employed are based on theory, and fundamentally on the force balance on a grain via the Shields number. I would suggest to not simply call these "empirical", but to actually note where the boundary between theory and empiricism lies. Indeed, these may be, for better or worse, some of the more theory-grounded components of a LEM! (Perhaps 2nd to the hydrodynamics)

*Good points. We will edit the manuscript to reflect that the formulae are theory based but have empirical (fitted) values/exponents within them.*

479. How do you know that they were not intended to represent variations in flow conditions? This statement is inconsistent with the fact that the underlying experiments have been performed at a wide range of tau=tau$_c$ ratios. You should be more specific or remove this comment.

*Our wording is "full variation of flow conditions". It is true that the formulae are derived and tested over a wide range of flow conditions, but not the full range that might be experienced in reality.*

489-490. Non-stationarity in hydrologic models seems a bit off-topic here.

*We don't believe it is off-topic as it refers to the point made immediately preceding the sentence. The use of a single, deterministic, parameter set is no longer common practise in hydrology, and one reason is non-stationarity of physical conditions. It is one of many lessons the LEM community could learn from the hydrological community.*

482-493. Do you think that the issue of scaling and calibration should deserve at least its own paragraph, if not its own section?

*Yes – will change.*

507-514. Do you mean that LEMs should follow hydrologic models' approach to uncertainty estimation in general (there are many such approaches), or specifically Lisflood-LP, and why? In addition, this paragraph gives little information about what these approaches are and why they are good.

*We mean that LEMs should follow hydrological modelling approaches to uncertainty generally, and indeed there are plenty. The Lisflood-FP approach is provided as an example, one which is widely used, and makes the most sense for CAESAR-Lisflood for obvious reasons. Will rephrase this paragraph to make this clearer.*

523. But it is quantitative! I think you are again confusing "subjective" and "nonquantitative".

*Will change to subjective.*

530-533. This is a bit of a bombshell that you are dropping on yourselves at the end: so you are unsure that the experimental design fairly weights the sediment transport formulas compared to the other values? There seems to be an easy answer, though: just take the binary extreme values of the other variables, and compare a subset of the runs with only 2 states considered for each parameter?

*Yes, this is a bombshell, and on reflection one we have dropped on ourselves unfairly. The purpose of the test was to use the Morris Method to assess the impacts decisions on parameter values have on the behaviour of the model – in this case addressing SED as a binary choice is entirely appropriate and justified. The Morris Method is subjective, and its purpose is to guide an operator in calibrating a model by identifying which parameters impact the model the most. The choice of SED is binary in this version of CAESAR-Lisflood and has the largest impact on the model behaviours. The minimum and maximum extents of the other parameters were deliberately set wide (+/- 50 %), wider than would normally be considered in a sensitivity test (eg, recent UK Environment Agency guidance suggest varying Manning's n +/- 20% to test a model's sensitivity\*). In this sense we could almost argue the opposite, that the impact of the SED choice is understated.*

*In response to your suggestion, instead we have reprocessed the model results assuming that the two SED formulae are the min and max choices across 5 steps. Therefore, switching from one to the other is 4 iterative step changes, and we divide all associated elementary effects by 4 and replot the normalised data. Here we see that the model shows less sensitivity to the parameter change, with others, such as Manning's n and Grain Size Set overtaking it. This could be argued that it is truer reflection of SED's role in model uncertainty relative to other parameters, but we would argue that the information because less useful to the operator as it does not reflect the true impact of the decision on the model outputs.*

*We also reprocessed the data as if the choice were 2 iterative steps apart as well and will present these results as supplementary material. We will modify our discussion accordingly and refer to this new material, but we do not believe this alters our results and conclusions for the reasons provided above.*

*\*Hankin, B., Arnott, S., Whiteman, M., Burgess-Gamble, L., and Rose, S., 2017. Working with Natural Processes – Using the evidence base to make the case for Natural Flood Management.* Environment Agency Report – October 2017. Project Number - SC150005

534-543: I think that the compute time and number of models should be mentioned far earlier in the paper (methods/results), and then perhaps referred to here as a reason for your decisions.

*We will provide some indicative values.*