Reply to comments of referee 1 (Professor Beven)

MAIN REVIEW (GENERAL COMMENT)

RC:

I am afraid I gave up on this paper (after making quite a lot of comments in the manuscript) at the point where Figures 4-6 are introduced and demonstrate the irrelevance of sensitivity analysis in the chosen case study. Almost certainly in these cases the performance of the model has more to do with uncertainty in the input and output data, that is totally neglected, than the factors included in the analysis. This is indicative of the apparently naïve way the issues associated with sensitivity analysis are presented in the introductory sections which can only be described as poorly presented. In particular, there is no real recognition of the potential for complexity of surfaces with sometimes rapidly changing covariation, including changes of sign, of factors in producing the outputs (that can be concealed in plots such as Fig 4) - yet such behaviour is common for real model applications. Also, despite the discussions of the last 30 years, the authors still seem (surprisingly?) to believe in the possibility of an optimum calibrated model.

> REPLY. Indeed the authors have been trying to follow a quite widely accepted idea that SA (of parameters) has a value. We also agree that model performance often depends on the data uncertainty than on the parametric uncertainty, however studying this relationship was the the objective of this paper. We agree that taking into account covariation as well would be the right thing to do.
>
> And yes, we believe in the possibility of an optimum calibrated model, as most people who do "SA of a calibrated model". We appreciate the need to assume multi-model representation of reality, the equifinality principle, and we have even contributed (modestly) to developing multi-model approaches and UA - but our experience with practitioners is that they typically want to use a single deterministic model rather work with multiple ones. We have to continue explaining them that assumption of a single optimal model could be misleading, and importance of explicit account of uncertainty.

RC:

The authors recognise that nearly all past intercomparisons of SA methods have suggested that different methods give different results, and that the same method might give different results when used with different outputs. So it is here too. This is not therefore unexpected, so where is the value in this paper, or in continuing to explore further SA methods as they suggest.

> REPLY. Indeed, this study confirms what has been demonstrated earlier, and in this respect its value is limited. The main idea here was to present "practical experience" (see the title) and to have a multiplicative effect by several models, several methods, several case studies. It seems though it was not enough.

RC:

Are the results really ever used to decide parameters "on which more resources can be put to ensure their higher accuracy". How would you actually do this for the conceptual models used in the paper, when it is effective values of model parameters that are needed to give good predictions? That would be a much more interesting paper.

REPLY. Agreed, the statement "on which [parameters] more resources can be put to ensure their higher accuracy" in relation to conceptual models is wrong. It seems though it was not enough.

RC:
As it is I cannot suggest that this paper is suitable for publication.

REPLY. Clear; we accept this. Our intention was to present the “Practical experience and framework”, and we realise (especially taking into account the other referees’ comments) that in the present form the study cannot be considered as a “research paper” in HESS.

- - - - - - - - - - - - - -

THE FURTHER TWENTY FIVE SPECIFIC COMMENTS MADE DIRECTLY IN THE MANUSCRIPT BY THE REFEREE

Comment #1
Page 2 Line 1
There are various definitions of UA proposed by different researchers (Cacuci, 2005; Pappenberger and Beven, 2006; Tong, 2006; Saltelli et al, 2008; Bastin et al, 2013). UA gives a qualitative or quantitative assessment of the uncertainty in the model results. The results can be qualitatively expressed in a graph showing the spread an ensemble of values or a distribution, as probabilistic flood maps, etc
Comment:
This paper is not about UA?

REPLY. We see SA as part of UA.

Comment #2
Page 2 Line 5
Due to the complexity and non-linear nature of hydrological models, it is hard to use analytical methods to study the uncertainty of hydrological models
Comment:
But you can use semi-analytical methods for local SA.

REPLY. Yes, indeed, but we had limited resources to cover all known methods.

Comment #3
Page 2 Line 6
Therefore, non-intrusive, sampling-based methods are commonly used, generally referred to as Monte Carlo Simulation (MCS), which can be seen as the simulation of a system that encloses stochastic or uncertain components.

Comment:

What does this mean in this context? And is not the point that MCS is only one way of investigating the output response surface.

REPLY. The only thing we wanted to say that MC is that sampling is an alternative to analytical methods.

Comments #4

Page 2 Line 15

“how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input” (Saltelli et al., 2008).

Comment:

No! Because you cannot actually apportion for a complex response surface - that requires additional (and not necessarily realistic) assumptions such as linear variance decomposition

REPLY. We agree, in general case it is so. However referring to “different sources of uncertainty in the model input” (Saltelli et al., 2008) we did not say “separately”.

Comments #5

Page 2 Line 17

(One may notice that this definition is not comprehensive, since uncertainty not only comes from model inputs but also from parameters, so for this reason, we will use the term “factor” instead of “model input”.)

Comment:

But Saltelli et al. include parameters as “inputs”. “We will use”??? Saltelli et al already use factors.

REPLY. Agreed, we could have referred to Saltelli et al. (Reason for explicit mentioning this is that in many papers it is not always clear if “inputs” include “parameters” or not.)

Comments #6

Page 2 Line 19

The main aim here is to identify the degree with which changes in various factors (**manifesting the corresponding uncertainty**) influence a change in model output.

Comment:

Is this really correct? - for a SA you do not need to specify any prior on the factors, so you are not actually taking account of that uncertainty, only looking at outputs in response to variation in the inputs.

REPLY. What we wanted say is just that “changes” manifest “uncertainty” without saying that uncertainty is expressed probabilistically (or one can say we assume a discrete distribution).

Comment #7

Page 2 Line 26

On the other hand, for distributed hydrological models, whose parameters can be gathered from the field, SA can help to target the most important parameters, **on which more resources can be put to ensure their higher accuracy**.

Comment:

This is standard fare – but is it correct? How would you do this when what is needed is effective value of the parameters? And how do you know that other parameters are not important when response surfaces are complex?

REPLY. We agree this statement does not fully reflect the mentioned complexity.

Comment #8

Page 2 Line 28

it is not yet really known what **is the optimal parameter vector**, and hence it is possible that sensitivity is investigated considering non-feasible parameters values.

Comment:

But THERE IS NO OPTIMAL PARAMETER VECTOR – that depends on both data period and evaluation measure or measures. That discussion has gone for 30 years!!! If you believe in an optimum then why not just evaluate sensitivity around that optimum????

REPLY. We appreciate this comment, but presented our opinion in answering the "general comment" above. (Perhaps using the term "optimal" is not very fortunate.)

Comment #9

Page 2 Line 31

SA is conducive to UA, and the main difference between their aims lies in that SA tries **to explicitly apportion the uncertainty of the output to the different factors**.

Comment:

See earlier comment – you cannot be sure that this is correct for any complex surface.

REPLY. Agreed. To be revised.

Comment #10

Page 3 Line 1

whereas UA provides a more general and often more detailed and rigorous account of model uncertainty.

Comment:

In what sense – can surely be based on exactly the same samples?

REPLY. We would like to stay with this (quite general) statement. Discussing SA vs UA further would pull us away from the main theme of the paper.

Comment #11

Page 3 Line 8

SA is typically categorized into Local Sensitivity Analysis (LSA) and Global Sensitivity Analysis (GSA).

Comment:

Surely needs to come before any mention of MCS.

REPLY. Agreed. To be corrected.

Comment #12

Page 3 Line 9

LSA concentrates on the sensitivity of factors at particular points in the factor space, for example, **around the vector of the calibrated parameters.**

Comment:

See earlier comment.

REPLY. Agreed.

Comment #13

Page 3 Line 12

A simplest expression of local sensitivity is the first-order partial derivatives of output to the factors. Define a model y = f(x), where y is the output of the model; x is factor of the model. The sensitivity of the factor (S) is defined as: $Si = \Delta yi\ \Delta xi$ (1) 15 where i is the i-th factor of the model. (Note, that in quite many studies instead of model output y the model error is used, e.g. Root Mean Squared Error or Mean Absolute Error.) Higher value of Si indicates higher sensitivity of the factor. Such measure of sensitivity is often called Sensitivity Index (SI). Figure 1 shows the expression of sensitivity of a model with two parameters (factors).

Comment:

You do not use this so why does it need repeating here.

REPLY. Agreed. To be corrected.

Comment #14

Page 3 Line 19

If we randomly sample several points in the whole parameter space, and obtain Si for each sample point.

Comment:

Oh come on!! So if you average large positive and negative values you will get zero. Are you really so unaware of the issue in SA?

REPLY. Agreed. This is a text-book issue which should have been mentioned.

Comment #15

Page 3 Line 28

It is demonstrated that different methods result in different ranking of factors, thus solid conclusions about the sensitivity of the factors are impossible to draw.

Comment:

Ok so not completely unaware – but what do these studies imply for your study??

REPLY. We are aware of the fact that solid conclusions about SA results may be hard to draw. Instead, we give suggestions rather than conclusions about how to choose appropriate SA methods based on several aspects (effectiveness, efficiency, convergence, implementation, etc.).

Comment #16

Page 3 Line 30

The result demonstrates **qualitative** SA methods are more efficient than quantitative SA methods, whereas quantitative SA methods are more robust and accurate.

Comment:

You need to at least say what is involved in a qualitative SA?

REPLY. Agreed. To be corrected.

Comment #17

Page 5 Line 20

They are sorted from best to worst, in which the first group produces the best 10% results (e.g. the results with least 10% model error), the second group produces the best 10%-20% results and so on.

Comment:

Can also be from largest to lowest for any output variable.

REPLY. Agreed. To be formulated better.

Comment #18

Page 6 Line 26

However, Razavi and Gupta (2015) have pointed out that they may suffer from scale issue, that is, the selection of the step size may influence the results **due to the complexity of response surface of the model**.

Comment:

Indeed!! See earlier comment.

REPLY. Agreed, to be addressed.

Comment #19

Page 7 Line 17

The advantage of meta-modelling is that by **simplification of the original complex model,** the overall running time is considerably decreased; the trade-off is a possible loss of accuracy.
Comment:
No. none of these simplify the complex model – they only interpolate the output response surface between the known values (with or without uncertainty). In doing so they might get quite the wrong local sensitivities since each method is constraining the gradient in some way.
Reply:

REPLY. Agreed. Had to be formulated better.

Comment #20
Page 7 Line 23
Different SA methods have different concepts and principles behind them, and, accordingly, the Sensitivity Indices may have **different meaning** and metrics.
Comment:
So why are you comparing them rather than accepting that they might produce different outcomes?

REPLY. Please see comment #15.

Comment #21
Page 7 Line 25
The evaluation of SA methods' effectiveness is aimed at finding out whether the relative Sensitivity Indices, ranking and screening of parameters **have sense** and indeed can be used in SA.
Comment:
But how do you know when you make no evaluation of the real nature of the surface in a complex case?

REPLY. Good point; a better explanation required. Of course we run the models for 10000 times in the first place. But maybe the provided analysis of the results is not deep enough.

Comment #22
Page 8 Line 5
Unlike assessing the accuracy of a 5 hydrological model, which can be compared with the **observation values**, for sensitivity there are no 'observations' to be compared with.
Comment:
But even then your observations may be significantly uncertain

REPLY. True; this statement requires will be given better explanation.

Commnet #23
Page 8 Line 6
h. To start somewhere, we will initially randomly sample a large number (say, 10,000) parameter (factor) vectors and run the model for each of them.

Comment:
That is not large for more than 4 or 5 factors?
Reply:

REPLY. Indeed - but we have start somewhere...

Comment #24
Page 8 Line 7
The **RMSE of the model output** will be plotted against parameter values as a scatter plot which will provide a rough image of the sensitivity of each parameter.
Comment:
RMSE? – but you have just said you have no observations to compare against? Since this is your reference it needs to be explained much more clearly.

REPLY. Indeed, perhaps not very well formulated. Here "observation" means there is no observation for sensitivity itself, not the model output. Though observations are highly uncertain in some sense, still they are the best references for evaluating model output. But when doing SA, you don't have such "observations" as references.

Comment #25
Page 10 Line 20
The model was run 10,000 times; the scatter plots of the $ERMSE$ against parameters for the **three models are shown in Fig. 4-6.**
Comment:
So given these figures why is SA relevant at all?

REPLY. The reason why we run the models 10000 times is to have a preliminary assessment of the model response and thus to draw the conclusions about the parameter sensitivity as a references for evaluation. This methodology have been reported in previous study as in Wagener et al., 2001, Hall et al., 2009, Pianosi and Wagener, 2015, etc. We will provide more explanation.

We would like to thank Professor Beven for the attention given to this paper and the comprehensive review, and pointing out the deficiencies, unclarities, and providing valuable suggestions.

We will be evaluating our options concerning extension and deepening of this research (which requires resources), and the target audience of this paper.