Hydrol. Earth Syst. Sci. Discuss., 10, C2226–C2239, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C2226/2013/

Hydrology and Earth System Sciences Discussions

HESSD

10, C2226–C2239, 2013

Interactive Comment

Interactive comment on "Technical note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models" *by* J. D. Herman et al.

J. D. Herman et al.

jdh33@psu.edu

Received and published: 4 June 2013

This comment contains author responses to Referee 2, Dr. Nossent. Author responses are shown below in plain text. Referee comments will be shown in italicized font.

The technical note "Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models" presents an application and comparison of the Sobol' sensitivity analysis method and the Morris screening sensitivity analysis method on a fully distributed hydrological model (the Hydrology

Laboratory Research Distributed Hydrologic Model). The focus of the comparison is mainly on the computational demands needed to achieve similar SA results.

We thank Dr. Nossent for his time and thoughtful review. We are fortunate to have someone of his expertise contributing to the quality of our work.

In general, the discussion paper addresses an increasingly important issue: research on sensitivity analysis (and indirectly uncertainty analysis) for fully distributed hydrological models. Due to the over-parameterization of the latter type of models, SA and UA remain very often challenging. However, I don't see the need to label this paper as a "Technical note". The power and drawbacks of both the Sobol' method and the Morris method are well known and the results presented in the paper primarily confirm this knowledge.

While we thank the reviewer for considering our contribution worthy of a full technical paper, our goal in this technical note is to provide a concise performance comparison of the Morris and Sobol methods. We will follow up with a more detailed study to explore the time-varying diagnostics made possible by the work presented here.

Overall, the paper is well written and clear. By additionally taking into account the good quality of the presented research, I tend to advise minor revisions for this discussion paper. However, there are a number of specific comments formulated below (sometimes detailed because I have done a big part of my PhD research on this topic), for which I would like to receive an answer (either in the paper or to me) and therefore I advise major revisions.

Thank you. We will be sure to take all of the detailed comments into consideration in our revised manuscript.

General: The manuscript does not describe the real purpose of the sensitivity analysis, nor are the obtained results linked with a certain objective. Although the paper tries to provide general insights in the use of the SA techniques for distributed models, formu-

HESSD
10, C2226–C2239, 2013

lating a certain objective is of the utmost importance (especially for the comparison of the SA results for the 2 methods). Otherwise, researchers might encounter so called type III errors (Saltelli et al., 2008): "right answers are sought for wrong questions".

Our objectives in comparing the Morris and Sobol methods are essentially twofold. First, we aim to compare the ranking of parameters, i.e., those that are sensitive and insensitive. Additionally, we compare the quantitative degree of sensitivity between the two methods, i.e., how sensitive the parameters are relative to one another. We will clarify these objectives in the introduction of the revised manuscript.

p4277, L8-9: It might be useful to add a number of other applications (Factor Fixing, Factor Prioritization,: : :) for sensitivity analysis (in particular in view of my first specific comment).

The application stated in this sentence ("identifying the key parameters controlling model performance") is intended to be a general statement of the purpose of sensitivity analysis, rather than a statement of the different applications which are possible. We will describe several additional applications in this introductory paragraph of the revised manuscript.

Also a description of the distinction between global and local SA techniques could be useful, since you start talking about global SA on p4277, L14.

We agree that a short description of local sensitivity analysis would be beneficial here. This will be added in the revised manuscript.

p4278, L3: Besides the reference of (Tang et al. 2007b), I would also add the work of Yang (2011), who presented a detailed comparison of different SA methods. Yang, J., 2011. Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. Environmental Modelling Software, 26(4): 444-457.

We will add this citation to our literature review.

p4279, L7: Although it is mentioned in section 4, I would have expected the total num-

HESSD
10, C2226–C2239, 2013

Interactive Comment

ber of studied parameters in section 2.1 or 2.2.

Agreed; we will move this information to the end of Section 2.1.

p4280, L9-11: This sentence should be moved, as you start giving general information on the Sobol' SA in this paragraph (or you should not mention any explanation on the first-order index p4280, L12, L15-16). Add an explanation why you are only using the total sensitivity index. I assume that you are only interested in the parameter rankings (and it is probably also related to the fact that the outcome of the Morris method corresponds with the total sensitivity index). Nevertheless, this should be mentioned, especially in view of my first specific comment.

Yes, this is correct. We are primarily interested in parameter rankings and thus the total indices facilitate the comparison between methods. We believe that the existing short discussion of first-order indices is also beneficial, but we will add an explanation to clarify our choice of total indices for this study.

Also with respect to the findings on the computational demand, it is necessary to add this explanation: despite the high computational demand of the Sobol' method, you receive a lot of additional model information "for free" (e.g. based on the first-order index). Hence, for a general view on the use of both methods, the first-order indices should be highlighted.

p4287, L18-21: It is clear that for the given configurations, Sobol' requires more time. However, as mentioned before, besides the total sensitivity indices, also other information can be retrieved from the same model evaluations (e.g. the first order indices). It is for example also possible to use the random samples and their resulting model evaluation in a sort of GLUE approach to optimize the model and perform an uncertainty analysis. This is not possible with the model evaluations used in the Morris method.

The Sobol method indeed provides valuable additional information, including first-order indices as well as an ensemble of model evaluations for further study, which is men-

HESSD

10, C2226–C2239, 2013

Interactive Comment

tioned in the second comment. We have observed and utilized these additional benefits in prior studies. In this short technical note, our focus is the comparison of parameter sensitivity between the Sobol and Morris methods; we believe that the first-order indices, while valuable for diagnostic insight, may detract from our core discussion in this particular study. We will highlight in our discussion the additional diagnostic value of the Sobol method.

p4280, L17-18: This requires more explanation or at least a reference, e.g.: Homma, T. and Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering System Safety, 52(1): 1-17.

Thank you, we will add this citation.

p4281, L3-4: Using the general formulations for the output mean and the output variance in the computation of the Sobol' sensitivity indices, can have a big influence on the convergence and the accuracy of these indices. Saltelli (2002) introduced an alternative formulation for the mean, which performs well in many cases. Additionally, W. Bauwens and I have studied the convergence of the Sobol' indices by applying different formulas for the square of the expectation value (4 different formulas) and the total variance (3 different formulas) (conference paper online available and peer reviewed journal paper in preparation): Nossent, J. and Bauwens, W., 2012, Optimising the convergence of a Sobol' sensitivity analysis for an environmental model: application of an appropriate estimate for the square of the expectation value and the total variance. In: R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.) (2012): International Environmental Modelling and Software Society (iEMSs) 2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting, Leipzig, Germany. http://www.iemss.org/society/index.php/iemss-2012-proceedings. ISBN: 978-88-9035-742-8, pp 1080-1087 Nossent, J. and Bauwens, W., 2013, Improving the accuracy and convergence of a Sobol' sensitivity analysis for an environmental model (in preparation for Environmental Modelling Software)

We have applied the traditional implementation of the Sobol method defined by Saltelli et al. (2008), pp. 171-172. This implementation has been used extensively in the environmental modeling literature and in other fields. However, we will be sure to point readers to alternative formulations with a series of citations in Section 3.1 of our manuscript.

For the first order index, the equation applied for the square of the expectation value has the highest influence, for the total sensitivity index, the equation applied for the total variance highly determines the convergence. Finally, also Bessel's correction should be applied for the equation of the total variance (divide by n-1 instead of n).

See above; the implementation presented here is extensively supported by prior literature.

p4281, L8: The use of N for the number of samples in both matrix A and B is somehow confusing. Mostly, this is denoted as $2N$.

p4281, L19: I'm not sure if the formulation $N(p+1)$ is the result of a confusion related to the previous remark or not, but it is wrong. The first-order and total sensitivity indices can be computed based on $N/2*(p+2)$ model evaluations (with your definition of N) (Saltelli, 2002). That is almost half your number.

p4283, L14: The numbers in Table 1 for the Sobol' SA are wrong. This is related with the comment on p4281, L19.

These three comments correctly identify an inconsistency in our manuscript. The sample size N should refer to the size of matrix A and B separately, not combined (p. 4281). With this definition, the number of model evaluations required to compute first and total order sensitivity indices is indeed $N(p+2)$. We will correct these definitions along with the values in Table 1 in the revised manuscript.

p4282, L3: The sampling technique that is applied for the Morris method should be mentioned in the paper (also related to the formulation on p4282, L23-24, since in

Campolongo et al., 2007 an alternative sampling technique has been introduced). Additionally, it might be useful to mention the distribution of the parameters from which you are sampling.

Please refer to the response to Anonymous Reviewer 1 for a more detailed discussion of the different sampling approaches available for the Morris method. In this study, we have used the traditional approach of Morris (1991) for generating random trajectories through the parameter space. This is an excellent point, and we plan to add a short discussion of this issue to point readers to recent literature with alternative approaches.

p4282, L20-26: This formulation is wrong or at least confusing. I would advise not to use "first-order index" or "total order sensitivity index" when discussing the Morris method. Mu and mu are both related to the total order sensitivity index (as they are also used to obtain parameter rankings), but they cannot be designated as such. For sure, mu is not a representation of the first-order index. In fact, mu and mu* are equal for monotonic functions. The absolute values in the computation of mu* are added to avoid that the different elementary effects cancel each other out (in case of non-monotonic functions). Interactions between parameters are somehow represented by the sigma value of the elementary effects and can occur for either monotonic or non-monotonic functions. This also shows that mu is not a representation of the first-order index.*

This is a valid point. The original manuscript contained a misinterpretation of the μ statistic, and this will be fixed in the upcoming revision. In this study we focus on the μ^* statistic, which as the reviewer notes is related to the total-order sensitivity index provided by the Sobol method.

p4283, L8: I assume you used a uniform distribution on [0,1] and a linear transformation to get the samples for the Sobol' SA?

Correct; we will add this explanation to Section 4 (Computational Experiment).

p4283, L10: What values did you get for the RMSE? Since you are using random samples of the parameters, the simulated values might deviate from the observed values. If the RMSE value becomes too large (which is possible since the RMSE is not a normalized measure), the variance estimation with the numerical integrals might become inaccurate. This problem was addressed by Sobol' (2001) and studied for environmental models by W. Bauwens and me (conference abstract and the earlier mentioned paper in preparation): Sobol', I.M., 2001, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation. 55 (1-3), 271-280. Nossent, J. and Bauwens, W., 2012, Application of a normalized Nash-Sutcliffe efficiency to improve the accuracy of the Sobol' sensitivity analysis of a hydrological model, Geophysical Research Abstracts, 14, EGU General Assembly 2012, Vienna, Austria, April 22-27, 2012.

The issue of model performance in sensitivity analysis depends on the input ranges chosen for sampling. In this study, we base our parameter ranges on prior work (Van Werkhoven et al., 2008) and in consultation with the National Weather Service to ensure that the ranges reflect what modelers would use in practice. We recognize the importance of this issue, and we will add a clarifying discussion to the manuscript where the parameter ranges are introduced, along with the citations provided by the reviewer.

p4283, L19: Did you check the evolution of the sensitivity indices? This can be of great value to assess the convergence.

Due to the computational cost of the model, we were not able to perform a full convergence analysis with increasing sample size. However, the confidence intervals derived from the bootstrap method (see below) for the Sobol indices for the $N = 6,000$ sample size provide assurance that the indices have converged.

p4283, L20: Provide the confidence intervals.

We monitored the 95th percentile confidence intervals to ensure convergence. In gen-

HESSD

10, C2226–C2239, 2013

Interactive Comment

eral, we considered confidence bounds acceptable if they represented less than 5% of the sensitivity index value. Clearly, this is not possible for very small sensitivity values, as these are statistically indistinguishable from zero. However, the most sensitive parameters were able to meet this criterion at the $N = 6,000$ sampling level. The $N = 1,000$ sample size did not provide acceptable confidence bounds, underscoring the need for a larger sample. Providing the full set of confidence intervals for the distributed model would add length beyond the scope of our technical note without altering our findings, but we will clarify the measures used to monitor confidence intervals in Section 4 of the revision.

p4283, L22: Did you try larger sample sizes for the Morris method?

The goal of our study is to determine whether the Morris method is capable of producing results similar to Sobol with far fewer samples. We were satisfied with this comparison for sample sizes between 20 and 100, and felt that larger sample sizes would be superfluous given the high computational cost of this distributed model.

p4284, L9: It might be useful to additionally emphasize that the sum of the total sensitivity indices is given per cell. One would expect a value higher than 1 for the (total) sum of the total sensitivity indices, but in this case this is still split up over 78 cells. This could be confusing.

We will emphasize this point both in the text and in the caption of Figure 4.

p4284, L24-25: I think this is a very strong statement that cannot be justified completely. It is possible, but not sure.

The statement in question is: "The RMSE metric is most sensitive to errors in peak flows, so the sensitivity indices in Fig. 3 can be interpreted in the context of the several high-flow events shown in the hydrograph in Fig. 2.". It has been established in prior work (e.g., Van Werkhoven et al. 2008) that the RMSE metric will be most affected by errors in peak flows. From this, it stands to reason that the RMSE for the aggregated

HESSD

10, C2226–C2239, 2013

period will be primarily attributable to the handful of high-flow events that occur during the period, although perhaps not exclusively so.

p4285, L21-22: The results on the convergence of the Sobol' SA can be correct in the way you have performed and discussed the analysis here. However, a number of concerns should be taken into account: (1) The actual sample size is half the one you mention. (2) The use of other equations for the square of the expectation value and the total variance could reduce the sample size required to achieve convergence. (3) The slow convergence is related to the very low values of the sensitivity indices (due to the high number of parameters, every parameter only contributes little to the total variance). This is logical, since a higher number of decimal numbers should become stable to achieve convergence and hence the numerical integration requires more samples. In particular, inferring parameter rankings based on these small, very similar values is very difficult. My colleagues and I also observed this slower convergence for small values of the sensitivity indices: Nossent, J., P. Elsen and W. Bauwens, 2011, Sobol' sensitivity analysis of a complex environmental model, Environmental Modelling Software, 26(12), 1515-1525.

Points (1) and (2) have been addressed above. Point (3) is an important insight, and we will be sure to emphasize this more in our discussion section, along with the relevant citation.

p4286, L17: It is known that in general Morris like screening methods are resilient to type I errors (non-influential factor is erroneously defined as important), but can be prone to type II errors (an important factor is classified as non-influential) (Saltelli et al., 2008). The use of mu should overcome the latter problem. The bunch of dots in the lower left corner of the figures of the bottom row of figure 5 confirm the strength with respect to type I errors: non-sensitive parameters are identified and Morris method is particularly suitable for Factor Fixing. However, despite the clear trends in both the top and bottom row of figures (Fig. 5), the non-linear trends (discussed on p4287, L2) in the top row and the outliers in the bottom row show that it is much harder to categorize*

HESSD
10, C2226–C2239, 2013

the sensitive parameters. I find it harder to identify the cluster of highly correlated parameter ranks near the most sensitive parameter on these graphs (mentioned on p4287, L11).

The issue of categorizing the sensitive parameters involves an important distinction. In the bottom row of Figure 5, the most sensitive parameters (approximately ranks 1-100) are highly clustered with few outliers. This indicates that the method of Morris can successfully identify the set of sensitive parameters, along with the set of insensitive parameters. However, the nonlinearity shown in the top row of plots in Figure 5 indicates that the method of Morris cannot properly distinguish among the most sensitive parameters; they are correctly identified as sensitive, but quantifying their sensitivities relative to one another is much more difficult. This result aligns with the common use of the Morris method as a screening method, rather than a quantitative interpretation of sensitivity. This point is discussed on pp. 4286–4287 of the original manuscript, and we will further emphasize it for the revision.

I believe that in particular a discussion of the outliers in the upper left corner of the figures of the bottom row is necessary. At this moment, these points are partly hidden behind the box with information. Depending on the purpose of the SA, these outliers might be important.

There are several outliers in the bottom row of plots in Figure 5, which is to be expected in a study of this size. As the reviewer has noted, these outliers fall primarily in the upper-left, where the method of Morris attributes higher rankings to certain parameters than does the Sobol method. Importantly, however, the outliers primarily occur between ranks 100-1000 (approximate). Thus, these parameters most likely have very low sensitivity values, even for the method of Morris where they are given an erroneously high ranking. We see from Figure 5 that the method of Morris provides the best rankings for the most and least sensitive parameters, but the rankings can become muddled in the intermediate range. This does not detract from its strength as an efficient screening method, however, and we will emphasize this point in our discussion. We will also move the legend boxes to avoid partially obscuring data points.

p4287, L12-14: This statement is correct based on the results shown in the paper. However, the results for a lower sample size for the Sobol' method are not handled. This makes the conclusions somehow biased towards the use of Morris method. I don't doubt the strength of Morris method, in particular for reasonable (high) sample sizes, but we (W. Bauwens and I) found out that the Sobol' method can also yield reasonable SA results and parameter rankings with limited sample sizes (paper under review): Nossent, J. and Bauwens, W., 2013, Evaluation and comparison of sensitivity analysis techniques for a complex, over-parameterized environmental model (submitted to Environmental Modelling Software) Therefore, it might be interesting to put the results into perspective and communicate the findings in this way.

For many combinations of models and simulation periods, it is true that the Sobol method provides valuable information with limited sample sizes. In our study, however, we found that the bootstrap confidence intervals on the Sobol indices were not acceptable for the $N = 1,000$ sample size, meaning that a smaller sample size would likewise yield inadequate confidence. This could be due to the complexity of the model, the degree of parameter interactions, or both–but for this particular study, we only obtained acceptable Sobol results at the $N = 6,000$ sample size.

In our paper we also suggest to address future research on the combination of Morris method and Sobol' SA. Morris method could be applied in a first stage to identify non-influential parameters. Due to the resulting reduction of parameters (Factor Fixing), the Sobol' SA would require less model evaluations to obtain the full amount of information on the model and its parameters. The results presented in this paper highly support this suggestion.

Yes, we agree that using the method of Morris as a pre-screening technique is a promising idea, particularly for highly parameterized distributed models. However, in our experience it can be very difficult to properly identify non-influential parameters and assign fixed values to them without influencing the subsequent Sobol analysis. It is only safe to remove parameters from the analysis if it is absolutely certain that the insensitive parameters are not involved in any interactions.

Technical corrections

p4277, L29: In general, it is better to use input factors instead of parameters when discussing the use of SA methods. Although mostly parameters are meant, the definitions for SA are also valid for input variables and the expression "input factor" covers both (Saltelli et al., 2008).

We would prefer to use the word "parameter" to maintain consistency with our prior work and to clarify the fact that the values being sampled are truly model parameters. However, we acknowledge the value of the term "input factors" as a more general description of values which can be sampled during sensitivity analysis, and we will be sure to add a clarification about this in our methods section.

p4285, L25: Figure 4 is very small and difficult to read.

Figure 4 was designed to be a full-page landscape figure. The discussion format does not allow landscape figures, but the editorial staff has stated that the landscape format will be permitted in the final version of the paper. This should solve the font sizing issue.

We will also correct the following list of recommended minor changes in our revision:

p4278, L4: Replace "significantly" by "quasi linear". This is more precise in this case. p4279, L24-25: This is a repetition of a part of the introduction. Try to reformulate this. p4280, L15-16: I suggest to change this sentence to "The first-order index is a measure for the fraction of the total output variance caused by the parameter i, without interactions with other parameters." The last part is essential, since interactions between e.g. parameters i and j can also be interpreted as "caused by the parameter i". p4281, L15: Reformulate 'the parameter sets are modified' p4282, L4: Replace "The

HESSD

10, C2226–C2239, 2013

Interactive Comment

method of Morris" by "It" p4283, L14: You have 2 times a Table 1 (the first one is also the Table in the supplement) p4283, L16: Shouldn't it be "the latter value" instead of "these values"? p4285, L9: I assume this should be "Fig. 3" p4290, L33: The reference (Saltelli, 2008) should be (Saltelli et al., 2008), as there are more authors of this book: Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S., 2008. Global Sensitivity Analysis: The Primer. John Wiley Sons, Ltd, Chichester, 304 pp.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 4275, 2013.