Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-215-AC1, 2020

Hydrology and
Earth System
Sciences
Discussions
Open Access

Interactive comment on "Simultaneously Determining Global Sensitivities of Model Parameters and Model Structure" *by* Juliane Mai et al.

Juliane Mai et al.

juliane.mai@uwaterloo.ca

Received and published: 29 September 2020

Reply to Anonymous Referee #1
Review received and published: 3 July 2020

Dear Reviewer,

Thanks a lot for your thorough review and the valuable suggestions. We will reply below in detail to your comments. Your comments are *italic*; our replies are highlighted **bold**. The **line numbers in red** are referring to the revised draft.

Best regards,
Julie, James, and Bryan

Summary

The authors introduce a novel sensitivity analysis method, called Extended Sobol' Sensitivity Analysis (xSSA), that advances upon existing procedures in several ways:

1. *it can provide insight into the sensitivity of individual model structure choices;*
2. *it can clarify the relation between parameter and structure sensitivity;*
3. *it can account for cases where model parameters are present or absent in different model structures; and*
4. *it is much faster than alternative methods.*

The main novelty of xSSA is that it estimates parameter/process sensitivity inside a flexible modeling framework (Raven), which allows the sensitivity estimates to be recombined through weighting. On any given timestep, the simulated states and fluxes *can thus be based on multiple different parametrizations of the same process, depending on how the weights are set. The authors test xSSA against two cases where analytical estimates of sensitivities can be derived (one case where each parameter only occurs once in all possible flux parametrizations, and one case where parameters are shared between multiple flux parametrizations), and in a real-world application of the Raven framework in a single watershed. They find that xSSA converge to the analytical solutions in both test cases, while current methods are only able to converge in the first test case. The real-world test case is used to showcase why process-based SA can be useful. I've read this paper with great interest. Model structure uncertainty is receiving considerable attention and this extension of existing SA methods to take advantage of modern multi-model frameworks is a welcome and timely contribution. Overall, the paper is easy to read but I have outlined various comments that can help the authors clarify their message. In general I think all the required information is there but some polishing would make the manuscript much more accessible for readers who are not so well-versed in Sobol' SA and Raven as the authors are.*

Thanks a lot for your interest and your positive evaluation of our manuscript.

General comments

The results section relies heavily on understanding of the Raven functions. It would be very helpful if the authors expand on the model description in section 2.1.2 or appendix C, by including the actual equations or descriptions of each parametrization.

We agree that this is a very valuable information. The details can all be found in the Raven documentation (Craig, 2020). We copied the according information and provide this now in the Supplementary Material. We decided to not include this in the manuscript or appendix to not artificially inflate the manuscript and distract from the actual message of the work. We attached the Supplementary Material at the end of this response letter for reference.

The results section relies on an understanding of each process to interpret model sensitivities. It is not entirely clear to me what each process includes. Can the authors clarify this by briefly explaining what each process in Figure 1C and further figures includes? For example, how does process 8 (potential melt) relate to process 5 (snow balance)? I don't think these explanations need to be very long, but it would be good if they include a bit more detail then the 1-3 words they currently get.

Thanks for this really good suggestion. As mentioned above, we now include the description of the processes and options used in the Supplementary Material. We also added a flowchart of the model structure used here to the Supplementary Material (Figure S1). The processes and functions are labeled according to the usage in this work (see circled labels such as M, N, etc.). This should make more clear how the processes are interlinked with each other.

I would encourage the authors to be careful with words such as "appropriate" and "important" in the manuscript. To some of the community, "appropriate process representations" might mean "process representations that are an accurate mathematical description of the real-world". To the authors (I believe) this instead means "equally sensitive, so equally good choices" (e.g. L538). Similarly, "important" seems synonymous with "high sensitivity" in this manuscript, but I don't think having high sensitivity over an arbitrarily wide parameter range necessarily dictates importance for matching a specific set of observations. Therefore I would strongly recommend the authors to go through the manuscript and either define such words clearly, or avoid ambiguity by being more specific in each case where such words are used (e.g. change "soil and surface processes are of secondary importance for streamflow prediction, ..." to "simulations are less sensitive to soil and surface processes, ..."; L556).

The reviewer is correct in the interpretation of our interchangeable use of "importance", "sensitivity" and appropriateness. We agree that it might lead to too much leeway of interpretation for a reader. We went through the manuscript and hope that we reduced this ambiguity.

line 521 ff. The analysis of model parameters (Fig. 5A) shows that the most sensitive ones are [...]

line 586 ff. The strong impact of these processes (together with the input adjustments) highlights the sensitivity of streamflow regarding snow and melting processes in this mountainous, energy-limited catchment.

line 589 ff. This demonstrates that soil and surface processes are of secondary sensitivity regarding streamflow. Their sensitivity may increase if the uncertainty of the snow and melting processes can be reduced, i.e. by narrowing parameter ranges during calibration.

line 610 ff. Evaporation (dark blue) is [...], expectedly, less sensitive during winter. Snow balance (medium green) and potential melt (orange) are sensitive as long as snow is present (Nov to May).

Line by line comments

L32. It might be more accurate to refer to "input (forcing) uncertainty" as "data uncertainty" or "observational uncertainty" to acknowledge that uncertainties are also present in model evaluation data such as streamflow observations. See e.g. McMillan et al. (2012).

We agree. We only focus in this work on the input uncertainty but at this part of the introduction it should be made clear that data uncertainty is the third source of uncertainty. We have made the following adjustment:

line 32 ff. Model structural uncertainty is commonly recognized (e.g., Gupta et al., 2012) as one of the three key components of hydrological model uncertainty, along with parameter uncertainty (Evin et al., 2014, among many more) and data (e.g., input forcing or observational) uncertainty (e.g., McMillan et al., 2012).

L47. It would be helpful to the reader if the authors could summarize the Baroni method in one or two sentences.

We agree. We rewrote major parts of that paragraph in the introduction and hope that the additional details given are now more helpful to follow the line of arguments. Following another reviewer's suggestion we also renamed the "Baroni method" with "discrete values method (DVM)" throughout the whole manuscript.

line 50 ff. To date, there have been limited attempts to simultaneously estimate model parameter, input, and structural sensitivities. One notable attempt is introduced by Baroni and Tarantola (2014) using a Sobol' sensitivity analysis based on grouped parameter. In that study, groups of soil and crop parameters, the number of soil layers, and a group of parameters to perturb inputs are investigated. These groups of parameters are pre-sampled

and a finite set of parameters for each of the four groups is chosen and each set is enumerated. The sensitivity analysis is then based on those enumerated sets. This means, rather than sampling each individual parameter like in a classic Sobol' analysis, an integer for each group acting as a hyper-parameter is sampled. The model is then run with the associated pre-sampled parameter set. While the approach may be generally applicable to arbitrary structural differences, in their testing, Baroni and Tarantola (2014) varied only in how the model was internally discretized (i.e., in the number of soil layers). The soil and crop parameters were always used for the same soil and crop process. The major limitation of this method is, however, that individual parameters need to be mutually exclusive and can only be associated to one type of uncertainty. The method hence limits the groups that can be defined, for instance, overlapping group definitions are not possible. The method will be referred to as "discrete values method (DVM)" in the following and will be contrasted to the method developed here to examine this limitation in more detail.

L53. "The method introduced ..." some text is missing here.

We deleted this. It was a remainder of a previous version. We are sorry about that.

L94. This special Raven property is a bit unclear to me. What dimension are the simulated fluxes weighted over? Is this a weighted average across multiple parameter sets, model structures, something else? – If this property is critical to the functioning of xSSA I think it should be explained in more detail here. Perhaps an example can be added.

The weighted average is for the estimates of the different process options. Let's assume there are three infiltration options. The first derives an infiltration of 1.0 [mm/d], the next 1.5 [mm/d], and the third 2.0 [mm/d]. The model would pro-

ceed with an estimate for infiltration of 1.35 [mm/d] ($= 1.0\times0.5+1.5\times0.3+2.0\times0.2$**) if the weights are 0.5, 0.3, and 0.2, respectively. This is performed for each process with multiple options at each time step (basically any time infiltration needs to be obtained during the simulation).**

We added the following additional explanation and hope this is more clear now:

line 105 ff. **[...] may be calculated via the weighted average of simulated fluxes generated by individual process algorithms; other flexible models may be revised to accommodate such analysis. The weighted averaging means that at each time step each option chosen for a process would derive an estimate for the flux, in [mm/d], and the weighted average of these estimates would be used for the next step.**

L103. I appreciate what the authors are going for, but "unconditional parameter sensitivity" is too broad a statement. The answers to questions A-D will be conditional on the catchment(s) being considered. It would be good to acknowledge that somewhere in lines 99-103.

Absolutely. We made the following adjustments and hope that it is now more clear that we indeed mean only "unconditional" regarding model structure and nothing beyond this.

line 111 ff. **The xSSA method allows us to efficiently estimate not only the global sensitivity of model parameters independent and hence unconditional of the chosen model structure [...]**

[A.]Unconditional parameter sensitivity: *Which model parameter is most influential independent and hence unconditional of model option choice?*

For example, which model parameter is overall the most influential

given all possible model structures (available in the modeling framework)?

L205. I'm somewhat confused about this statement. One does not need to run 12 fixed model structures but instead needs to run a single flexible structure that contains all the options that are present in the 12 models. How does this reduce the number of computations required? [...]

The runtime of running the 12 models independently would be only the same as the runtime of the single flexible structure if the time it takes to read inputs, to initialize the model, and to write model outputs would be negligible. This is certainly true for the two benchmark models but is not the case in most hydrologic and land-surface models. Most of these models do, for example, usually not allow the users to reduce the amount of model outputs written. Raven is highly optimized regarding I/O and initialization. The runtimes for three individual models of the 108 Raven models are 51.786s ($M_1 - N_1 - O_1 - P_1 - Q_1$), 52.695s ($M_2 - N_1 - O_1 - P_1 - Q_3$), and 51.985s ($M_2 - N_1 - O_1 - P_2 - Q_3$) each for 100 runs while the runtime of the single model with the flexible structure is 53.342s for 100 runs. This yields runtime savings of about 99% for using the flexible model structure with weights over running the individual models:

$$\left[1 - \frac{53.342}{(51.786 + 52.695 + 51.985)/3 \times 108}\right] \times 100\% = 99.05\%$$

[...] As far as I understand, it's still the same elements that are being tested. If the authors mean that all elements can be tested independently (implying that if and how they are connected to other elements can be ignored), than why would they need to be part of a model structure at all? Why not test each element in isolation and recombine the results through the proposed weighting? This could result in even further computational savings in cases where the same parametrization can be used in multiple processes (quite common in bucket models, possibly also in physics-based models that discretize snow/soil into multiple layers).

This is true but we think it will be pretty unlikely to beat the runtime improvement of 99% as shown above with the approach suggested by the reviewer.

L212. Caption of Figure 1. "The three processes are connected through A.B+C (C.D+E) ..." Text in the brackets should read (D.E+F).
L212. Caption of Figure 1. "Processes A (D) and C (E) ..." (E) should be (F).

Thanks for spotting this. This is resolved now.

Figure 1 caption. **The three processes are connected through $A \cdot B + C$ ($D \cdot E + F$) to obtain the hypothetical model outputs. Processes A (D) and C (F) have two options, process B (E) has three.**

L213. Which numerical scheme does Raven use to solve its model equations?

We would like to refer to the publication that introduced Raven (Craig et al., 2020) (end of Section 3.2 therein) for details on all numerical schemes supported in Raven. The default is the Ordered Series approach which was used here. Besides that Raven supports the explicit Euler and iterative predictor–corrector method for solving a set of ODEs (Snowdon, 2010). Additional details can also be found in the Raven manual (Craig, 2020). We have added the following information to the manuscript:

line 259 ff. **For the case study used herein, Raven is applied in lumped mode and the models are solved using the ordered series numerical scheme defined in Craig et al. (2020, (end of Section 3.2 therein)).**

L262. "forcings" → "forcing"?

Done.

L269. Why were only 20 years of data used if 56 are available? Wouldn't more data give a more complete assessment because a wider range of conditions is (likely) covered?

This is probably true. We however think that a 20 years simulation period is already covering a wide range of conditions. The reduction of the simulation period from 56 to 22 years (including the two years used for warm-up) was mainly to reduce the runtime of the whole analysis. The 22-year setup took about 22 hours and would have been 56 hours for the full period. We decided that the gain in results does not justify the longer runtime. A 20-year simulation period is indeed a very long period used for sensitivity studies: Markstrom et al. (2016) used 3 years of warmup and 11 years of simulations, Mendoza et al. (2015) used 2 years of warmup and only 6 years of simulations, and Cuntz et al. (2016) used 16 years of simulations.

L308. It took me a while to figure out that these numbers are: # of models x (# of parameters +2 × K), mainly because the order of operations is reversed compared to L307 (which gives # of parameters first and # of models second) and because the operation K × (N+2) from L303 has already been completed. I'd suggest to clarify this.

Thanks for pointing us to this. We have made the following adjustment and hope it is easier to follow now.

line 331 ff. **Out of the 12 possible shared-parameter benchmark models (Eq. 2) there are 4 models that contain 3 parameters, 5 models contain 4 parameters, 2 models consist of 5 parameters, and 1 model has 6 parameters. Hence, 72 000 ($= 4 \times (3+2) \times 1000 + 5 \times (4+2) \times 1000 + 2 \times (5+2) \times 1000 + 1 \times (6+2) \times 1000$) model runs would be required if $K = 1000$ reference parameter sets would be used.**

L350. The authors use analytically derived Sobol' scores for their shared-parameter model setup. Can these derivations be made part of the appendices or can the authors provide a reference to a paper that provides these?

We are deriving these values following the example provided in Saltelli et al. (2008) in example 5 described on page 179 and following. We added this to the manuscript:

line 386 ff. **All analytically derived indexes are obtained by following the descriptions in Saltelli et al. (2008, page 179 ff).**

L358. Single-sentence paragraphs look strange. Suggest to merge with preceding paragraph.

Done.

L364, L366. I was under the impression that the shared-parameter models was being tested. Why do these sentences refer to parametrizations A, B and C instead of D, E and F?

We are sorry for that. The reviewer is absolutely right. We adjusted the text to:

line 392 ff. **[...] model runs are required for the 7 process options D_1, D_2, ..., F_2 and the 4 weight deriving random numbers r_i. For the analysis of processes (sensitivity metric D) the model needs to be run $(3+2) \times K$ times to obtain sensitivities of the 3 processes D, E, and F.**

L429-441. I find this section difficult to follow, in part because it was not clear to me that the Baroni method uses a regular Sobol' approach. The only mentions of Sobol' so far (I believe) have been in relation to xSSA and the mention of Sobol' analysis on L434 threw me off. I'll repeat my earlier comment that a brief description of the Baroni method would be very helpful in understanding these results.

Agreed. We hope the revised section in the introduction (line 50 ff.) clarifies this now.

L435. "This contradiction cannot be resolved." Is it part of the Baroni method to include a single parameter twice? In my (admittedly limited) experience with the regular Sobol' method, one would include any parameter only once, regardless of how many times it occurs in the model processes being considered. This would mean that processes cannot be assessed individually if they share a parameter (which the authors already mention) but getting into this situation in the first place requires that one is looking to investigate processes, not parameters. I think the authors can make their reasoning stronger by repeating here that investigating process sensitivity requires a different approach then parameter sensitivity in cases where parameters are shared between processes.

The point is that the Baroni method (now Discrete Values Method DVM) indeed does not investigate the sensitivity of individual parameters either. We wanted to highlight the difference to this existing method. Most parameters certainly only appear in individual groups but especially when several process options are investigated (Baroni and Tarantola (2014) did not do this) several parameters will appear in several process options. For example, porosity is likely a parameter in each process option related to soil processes. Let's say we have two process options and found that option 1 depends on parameters x_1, x_2, and x_3 while option 2 only depends on two parameters, again on x_1 and an new parameter x_4.

Group 1			Group 2	
x_1	x_2	x_3	x_1	x_4
0.1	**5.0**	**10.0**	**0.2**	**6.0**
0.2	**2.0**	**20.0**	**0.4**	**7.0**
0.3	**3.0**	**15.0**	**0.3**	**8.0**

Even without knowing how exactly the Sobol' method works, the problem becomes clear when a value for parameter x_1 has to be picked. Is it 0.1 or 0.2 for

the first set (first row in above table)? This leads in any method to problems; not even only for the Sobol' method.

We followed the reviewers advise and emphasized again that a method that is applicable for shared parameters is needed when analyzing the sensitivity of process options and processes.

line 464 ff. **This contradiction can not be resolved in a method that does not allow for shared parameters. Shared parameters occur often in several process options of the same process but also across processes and hence need to be considered when analyzing process options and processes in flexible frameworks.**

L451. It might be good to add a reference to sensitivities of non-additive models not summing to 1. I seem to recall this is discussed in Saltelli et al. (2008) for example.

Yes, that is a good idea. We added two references there.

line 481 ff. **We do not expect the process sensitivities to sum up to 1 which is anyway not achievable with non-additive models (Sobol and Kucherenko, 2004; Saltelli et al., 2008).**

L461. "hence" → "this"?

Done.

line 493 ff. **The errors converge to zero in every analysis and this proves that [...]**

L499. Suggest to delete "and hence most sensitive"

Absolutely! Thanks for spotting this.

line 530 ff. **A sensitivity analysis regarding model parameters is often performed prior to model calibration to identify the most sensitive parameters which are in turn the parameters that [...]**

L509. It might be instructive to adapt the x-axis in Figure 5B, so that it shows which parameters (x-axis in 5A) are included in each process option in 5B. This could clarify whether process sensitivities can be traced back to particular parameters.

We in general agree with the reviewer. The information however is given in Table C1. We do not want to make the figure more complex than it already is. We however added the reference to Table C1 to the figure caption:

Figure 5 caption. **[...] The Sobol' sensitivity index estimates are determined also for (B) 19 process options and (C) the 11 processes. The information which parameters are used in which process option and process can be found in Table C1. [...]**

L515. "Same" → "The same"

Done.

line 547 ff. **The same holds for the two options of the evaporation process (dark blue bars) [...]**

Figure 5. It might be worthwhile to change the orientation of these plots so that the Sobol' scores are on the x-axis and the parameters/parametrizations/processes are on the y-axis, so that these are easier to read. I currently need to tilt my head back and forth to read the results in 3.3.2 and compare them to the axes in Figure 5.

Thanks for this suggestion. We did that and Figure 5 appears now as shown below.

figure-1.pdf

Fig. 5. Results of the Sobol' sensitivity analysis of the hydrologic modeling framework Raven. (A) The sensitivities of 35 model parameters (see Table C2) and 8 parameters r_i that are used to determine the weights of process options are estimated. The Sobol' sensitivity index estimates are determined also for (B) 19 process options and (C) the 11 processes. The different colors indicate the association of parameters and process options to the eleven processes. Parameters x_{29} and x_{30} are associated with several process options and are not colored but gray. The Sobol' main and total effects are shown (dark and light colored bars, respectively). All sensitivity index estimates shown are originally time-dependent and are aggregated as variance-weighted averages (Eq. 23 and 24). The average weights over the course of the year are shown in Figure 6.

L525. "The latter serves as a consistency check of the implementation." Can the authors clarify what they mean here? – upon reading further, it might make sense to swap this sentence with the one immediately after it.

We are sorry that the line of arguments was a bit mixed up here. We rearranged the order and provided a bit more information. It now reads as:

line 556 ff. **The zero sensitivity is expected since the SNOBAL_SIMPLE_MELT option does not require any parameters (see Tab. C1). Model outputs of such options do not change for different model runs and hence have a zero variance which leads to a zero Sobol' index. Such settings and parameters that are a priori known to yield zero sensitivities are beneficial in sensitivity analyses as they act as a consistency check of the implementation (Mai and Tolson, 2019).**

L527. I admit I'm a bit confused that model outputs of process representations do not change in different model runs. Because this process representation is connected to the rest of the model, and there are changes in the contributions of other processes as a result of different parameter values, wouldn't it be expected that the model states change as well, and as a consequence, that the contribution of this particular process to overall simulations changes too? [...]

The process outputs of other processes might change- especially when a parameter is in the current group that also participates in other groups (e.g., parameter x_{29}). That process options, processes and parameters are not independent of other parts of the model can be seen through the interaction effect that is derived by $ST_i - S_i$ of the respective parameter, process option or process. The analysis however shows how much the process impacts the overall model output (here streamflow). Therefore, it is a desirable behavior that also other process outputs might change. But one knows that these differences are only caused by the change of the parameter/process option/process currently analyzed.

[...] Without knowing with SNOBAL_SIMPLE_MELT (Q_2) actually does, I assume that even if it has a constant melt rate, it is still constrained by snow availability and thus cannot produce a time-invariant flux. I would expect such a case (no parameters in a given process, but influenced by other parameters by virtue of being part of a bigger model) as showing in a 0 Sobol' main effect, but a non-zero Sobol' total effect. Can the authors clarify this?

The mentioned snow balance process option does not contain any parameter. The process output of SNOBAL_SIMPLE_MELT over time t is $\mathbf{Q_2(t) = Q_2(M_{potmelt})}$ where $M_{potmelt}$ is the potential melt at time t. The potential melt, i.e., the calculation of available energy at the snow surface, is another process (i.e., T_1) because it is used for other options and in other places of the model. In other hydrologic models the snow balance and potential melt are usually not separated but Raven allows the user to mix-and-match different approaches with each other. The snow balance Q_2 itself has hence no parameter x. The potential melt $M_{potmelt}$ is an input. This means when all parameters associated to SNOBAL_SIMPLE_MELT (means none) are changed, nothing in the model outputs $Q_2(t)$ ever changes because literally nothing is changed. The output of the SNOBAL_SIMPLE_MELT is not constant over time though. It is just does not change. That is the reason for our statement in the comment before that this is a very helpful consistency check for the implementation of the analysis. The interaction effect is zero because none of the independent variables is derived by any of the other processes and process options. The output of Q_2 is hence always constant even if other parts of the model are changed.

L538. "The three infiltration options are equally sensitive and hence equally appropriate." Logically, only one or none of these infiltration options is appropriate (in the sense of accurately representing the real world). I also doubt that high sensitivity automatically indicates high appropriateness. I suggest to rephrase this sentence.

We adjusted the manuscript to the following:

line 547 ff. The three infiltration options are equally sensitive and hence are all able to achieve the same amount of variability in simulated streamflow time series. This similarity is an indicator that the choice of the infiltration option will therefore not influence the model performance.

L538. "quickflow" → should this be "infiltration"?

Indeed. Thanks for spotting this. We adjusted this to:

line 570 ff. The three infiltration options are equally sensitive and hence are all able to achieve the same amount of variability in simulated streamflow time series. This similarity is an indicator that the choice of the infiltration option will therefore not influence the model performance.

L545. Can it be said that rain-snow partitioning is a forcing correction function? It does not change the water balance, only the phase and thus by extent, the timing of liquid water availability.

Yes, that is correct. We slightly adjusted the phrasing in the manuscript:

line 578 ff. Technically, potential melt T as well as rain-snow partitioning V and precipitation correction W are handling inputs to the hydrologic system and can hence be regarded to quantify input uncertainties or, in other words, are forcing correction function and do not change the water balance within the model.

L556. "This demonstrates that soil and surface processes are of secondary importance for streamflow prediction, ..." Is this true? As far as I understand, the SA only shows that impact of parameter changes on the variability of the simulations. I don't think relatively low sensitivity automatically indicates low importance for accurate streamflow

simulation, because (1) no simulations have been compared to observations; (2) parameters ranges might be wider during this SA than their "real" range of values and thus much of this variability might occur in regions of the model output space that are far away from the observations. I would recommend slightly more careful phrasing, like used in L559.

We absolutely agree with the reviewer and rephrased this to:

line 589 ff. This demonstrates that soil and surface processes are of secondary sensitivity regarding streamflow. Their sensitivity may increase if the uncertainty of the snow and melting processes can be reduced, i.e. by narrowing parameter ranges during calibration.

L677. These are not author contributions.

The author contributions have been adjusted to the following:

line 702 ff. JM set up the analyses, implemented the sensitivity analysis based on groups of parameters, implemented the proper sampling of weights used in this study, wrote main parts of the manuscript, prepared all figures and tables; JRC contributed to the writing of the manuscript, implemented the weighting of process options in Raven, provided ranges for the parameters included in the analysis, helped to setup the model with the selected options and resolved inconsistencies in Raven detected by earlier versions of the sensitivity analysis, and helped with the hydrologic interpretation of the results; BAT contributed to the writing of the manuscript, provided feedback on the manuscript and the setup of all experiments including the benchmark models as well as helped with the hydrologic interpretation of the results.

References

Baroni, G. and Tarantola, S.: A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study, Environmental Modelling & Software, 51, 26–34, 2014.

Craig, J. R.: Raven: User's and Developer's Manual v3.0, http://raven.uwaterloo.ca/files/v3.0/RavenManual_v3.0.pdf, 2020.

Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, W., Jost, G., Lee, K., Mai, J., Serrer, M., Snowdon, A. P., Sgro, N., Shafii, M., and Tolson, B. A.: Flexible watershed simulation with the Raven hydrological modelling framework, Environmental Modelling & Software, p. 104728, https://doi.org/https://doi.org/10.1016/j.envsoft.2020.104728, 2020.

Cuntz, M., Mai, J., Samaniego, L., Clark, M. P., Wulfmeyer, V., Branch, O., Attinger, S., and Thober, S.: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model , Journal of Geophysical Research: Atmospheres, pp. 1–25, 2016.

Evin, G., Thyer, M., and Kavetski, D.: Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity , Water Resources Research, 50, 1–26, 2014.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resources Research, 48, https://doi.org/10.1029/2011WR011044, 2012.

Mai, J. and Tolson, B. A.: Model Variable Augmentation (MVA) for Diagnostic Assessment of Sensitivity Analysis Results, Water Resources Research, 55, 2631–2651, 2019.

Markstrom, S. L., Hay, L. E., and Clark, M. P.: Towards simplification of hydrologic modeling: identification of dominant processes, Hydrology and Earth System Sciences, 20, 4655–4671, 2016.

McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrological Processes, 26, 4078–4111, 2012.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, Water Resources Research, 51, 716–728, 2015.

Saltelli, A., Ratto, M., Andres, T. H., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis. The primer, John Wiley & Sons, Ltd., 2008.

Snowdon, A. P.: Improved numerical methods for distributed hydrological models, Ph.D. thesis, University of Waterloo, Waterloo, Ontario, Canada., 2010.

Sobol, I. M. and Kucherenko, S. S.: Global Sensitivity Indices for Nonlinear Mathematical Models. Review, WILMOTT magazine, pp. 2–7, 2004.