Title: Global Sensitivity Analysis of the distributed hydrologic model ParFlow-CLM (V3.6.0)
Author(s): Wei Qu et al.
MS No.: gmd-2022-131
MS type: Methods for assessment of models

First of all, many thanks for the comments and suggestions, which have enabled us to improve the quality of our manuscript considerably. We totally agree with most of the reviewer comments, such as 1) the time period (we will extend our simulation from 2017 to 2021); 2) the partial difference (we use now the absolute value); and we have corrected the manuscript point by point accordingly. Furthermore, we still believe that our manuscript is interesting for the community to improve the understanding of parameter sensitivities of the integrated hydrologic model ParFlow-CLM. Although, the applied LH-OAT approach is not a novel in itself, we are the first ones, who applied this for ParFlow-CLM, as there are limited information about the SA of ParFlow-CLM. In the following, the comments are marked in black and our answers in blue.

It will take several weeks to months for us to run the model again to extent the time period to 5 years in order to get a stable GSA result, therefore, we only reply the questions here, and the discussion as well as the new version of the manuscript will be submitted later.

Referee1:

In the manuscript by Wei et al, the authors apply global sensitivity analysis (GSA) using Latin-Hypercube One-factor-at-A-Time method to the distributed coupled model ParFlow-CLM. In general, I think that applications of GSA to complex and spatially distributed model are welcome contributions, given the challenges related to the long model run time and the high dimension of the parameter space.

However, I would recommend to reject the manuscript, as 1) I believe that it does not align with the aim and scope of GMD, 2) the objectives and novelty of the manuscript need clarification, 3) the implementation of the GSA has flaws and the interpretation of the sensitivity indices (based on provided equation) is erroneous, 4) the model calibration need clarification 5) the manuscript lacks a discussion of the results in light of the very large existing literature on sensitivity analysis and 6) the manuscript is unprecise at many locations. Please find below detailed comments.

1) & 2) We agree with the referee that we do not pursue new global sensitivity analysis (SA) approaches. However, we think that this is still an innovative manuscript, as we aim to fill the knowledge gap on the global sensitivity of the distributed hydrologic ParFlow-CLM model to different output variables.

3) We have changed the implementation of the GSA correspondingly, and now the absolute of the partial effect was used in the manuscript.

4) It is a misunderstanding to use the word "calibration" in the text, our aim is focus on the SA to the model output variables, so we will not specify a model calibration process. But we changed our text and the figure correspondingly to show the output range of the simulated runoff and clarified that we will not do model calibration in this study.

5) We will put more discussion in the manuscript when our simulation is done, and a new version of the manuscript will be submitted later.

6) We describe the text more precisely followed the comments of the referee as following.

1) The manuscript does not match the aim and scope of GMD, in that it does not describe a new model or new model developments, and it does not introduce a new method for assessment of models. Specifically, the GSA method used is published elsewhere (van Griensven et al., 2006) and the manuscript does not describe a novel framework for application of GSA to a complex and spatially distributed model. It actually does not describe the strategy used to apply the GSA method. The text p10 L236 suggests that the parameters were considered as being spatially homogeneous, but this

would need to be better explained in the manuscript as this seems to be a very strong assumption. Previous works have examined the application of GSA to spatially distributed model and contributed to address this challenge (a few examples are: Herman et al, 2013; Rouzies et al., 2021; Smith et al., 2022; van Werkhoven et al., 2008). It is not clear how the manuscript may relate to such previous works.

As we respond above, we applied an existing GSA method to ParFlow-CLM model, which was not studied anywhere else, so we believe that our manuscript still matches the aim and scope of GMD.

The assumption of spatial homogeneous parameters does not influence which parameter control the output variables estimation, and the study of Jefferson et al. (2017) also showed that even the hypothetical domain was updated to better represent actual site conditions, important parameters remained the same. Therefore, we use the spatial homogeneous parameters for our study.

2) The objective and the novelty of the manuscript need clarification.

- Sensitivity analysis was previously applied to the ParFlow-CLM model, and in this respect, the authors cite the studies of Jefferson (2015, 2017) and Srivastava et al., (2014). It is not clear how the manuscript relates to these previous studies.

These publications are the only existing SA for the model of ParFlow-CLM. That's the reason why we listed these studies, but I will compare our results with their findings more precisely in the new version of the manuscript.

Here are the findings of these three publications:

Srivastava et al. (2014) studied the sensitivity of the seven hydraulic and vegetation parameters (i.e. hydraulic conductivity, specific storage, porosity, wilting point, LAI, SAI, and aerodynamic roughness length) to the ParFlow-CLM simulated evapotranspiration and streamflow for three different bedrock conditions using Morris method.

Jefferson et al. (2015) applied a synthetic study about the sensitivity of 19 vegetation parameters to the ParFlow-CLM simulated energy fluxes (i.e. latent heat, sensible heat, and ground heat) for the plain and tilted-v domain for 144 hours using active subspaces method, and they reported that the leaf reflectance parameters are not sensitive, that's also why we didn't include these 9 parameters.

Later on, Jefferson et al. (2017) studied the sensitivity of photosynthesis and stomatal resistance parameters on the transpiration rate with the active subspace method, they reported that the transpiration is controlled by the stomatal resistance term in land surface models.

- In addition, the authors mention as an objective "to test the transferability of the results to regions with other topographies and climates" (p3 P80-81). It is well known from many past studies that sensitivity indices can vary tremendously across places (an extensive review is provided in Wagener and Pianosi, 2019). Therefore, we know a priori that it is very risky to extrapolate sensitivity analysis results beyond the study location. I also think that the analysis performed by the authors in this respect lacks breadth, because it is only based on three study sites.

We agree that the transferability of the GSA results to regions with other topographies and climates was studied, but these studies (Wagener and Pianosi, 2019) are about different models which are listed in the following table:

	Model	SA Method	Site Number	Parameter number	Time period
Rosero et al. (2010)	3 versions of Noah Land Surface Model	Latin Hypercube Monte Carlo Sampling	9 sites USA	32	45 days

Van Werkhoven et al. (2008a)	Sacramento Soil Moisture Accounting Model (conceptual rainfall-runoff model)	Sobol's sensitivity analysis	12 sites USA	14	1 year
Confalonieri et al. (2010)	rice model WARM (rice yield)	Morris and Sobol'	5 sites MED*	11	29 - 31 years
Ben Touhami et al. (2013)	Pasture Simulation model	Morris method	6 sites European	28	4 - 7 years
Shin et al. (2013)	4 conceptual rainfall-runoff models (IHACRES, GR4J, Sacramento and SIMHYD)	Morris and Sobol'	5 sites Australian	4-13	5 years
Hartmann et al. (2013)	3 lumped rainfall-runoff models (Hymod, HBV, and Sacramento Soil Moisture Accounting)	time-varying Sobol sensitivity analysis	12 sites USA	8-17	10 years

* Mediterranean region

In our study, we only focus on the ParFlow-CLM, and the transferability of the GSA to different catchments are complicated, that's the reason why we have such kind of synthetic study to investigate the topography or climate separately.

3) The implementation of the GSA has flaws and the interpretation of the sensitivity indices is erroneous.

- The sensitivity index is defined in Eq. 6-7. The index is actually different to the one introduced in the original method (Eq. 7 in van Griensven et al., 2006). In van Griensven et al., the index corresponds to the average of the absolute value of the partial effect, while in the manuscript it is directly the average of the partial effects. The problem of not using absolute values is that compensations between the partial effects can occur when the model response is non-monotonic. This has been discussed for instance in Campolongo et al. (2007) with respect to the Morris method and it is common practice to use the absolute value of the partial (elementary) ieffect to avoid this compensation problem (e.g. Eq. 1.49 in Saltelli et al., 2008).

We agree with the referee, and the problem is solved by using the absolute value of the partial effect.

- The interpretation of the sensitivity index (e.g. p5 L128, p11 L247, p11 L254-255) is not correct. The sign of the sensitivity index of Eq. 7 does not provide reliable information on the direction of change (as I mentioned earlier, the relationship may be non-monotonic) and a sensitivity index equal to zero does not mean that the parameter is not influential (as there can be compensation effects when absolute values of the partial effects are not computed).

We have changed the equation by using the absolute value of the partial effect and the explanation in the manuscript accordingly.

- The authors only use a subset of parameters for GSA (12 parameters), while the number of parameters is much larger in such a complex model. In particular, it has been shown that land surface models can include a large number of hard coded parameters with an empirical basis (that are therefore largely uncertain) and many of which can have a large impact on the model output. This is documented for instance for the NOAH-MP land surface model in Cuntz et al. (2016) and Mendoza et al. (2015). Therefore, the authors would need at least to discuss this issue in the manuscript.

We agree that there are a lot of parameters such kind of complex model, as we already discussed about the three existed publications, they discussed that several parameters are not sensitive, in addition, it is computation intensively to do such kind of SA for ParFlow-CLM, that is the reason why we didn't include all of these parameters, as we will run the model again, it is also possible to include all these parameters.

- Information on the setup of the GSA is missing. In particular, it would be crucial to know the sample size used and whether this sample size is large enough to obtain robust sensitivity indices estimates (as explained in Sarrazin et al., 2016). An indication on the sample size is also important from a methodological point view to understand which resources are required to run the analyses.

We mentioned the sample size of the GSA in in line 178 – 180, but we now give a more detail description about the setup of GSA in the manuscript.

The reason why we choose LH-OAT for our GSA of the ParFlow-CLM parameters is due to the relatively small sample size, which is required for LH-OAT, and as a rule of thumb, such kind of multiple-starts perturbation approach requires around 10 to 100 times model runs per input parameter (Wagener and Pianosi, 2019). In our case, we have 12 parameters, and our sample size is 30*(12+1) = 390 model runs, so it is enough to get the stable ranking for the GSA. In addition, as explained in Sarrazin et al. (2016) that screening convergence is reached when the partial difference of the lower-sensitivity parameter is below 0.05. Our study also fulfills the criterial.

- The GSA only covers about two years of simulations, which may be too short to obtain reliable and period-independent sensitivity indices estimates. I refer for instance to Shin et al. (2013) when the impact of the length of the simulation period is discussed). The sensitivity indices may take different values over another simulation period. I also did not understand what the time period selected is. In Sect. 2.5.2 the period is "from 2016-07-01 to 2018-31-12" is reported, while in Figures 4-5 "from 01.01.2021 to 31.01.2022" is reported.

We agree that it is confusing for reader of the different time period between the GSA and the model performance in Figure 4 and Figure 5.

And we now run our simulation again and keep the time period consistently from 01.01.2017 to 31.12.2021, which covers 5 years' time period. It agrees with the study of Shie et al. (2013), that the length of data period required to characterize the sensitivities assuredly is a minimum of five years.

4) The model calibration criteria are unclear and I think that the interpretation of the calibration results are not complete.

- Based on which criteria was the "best simulation" (p9 L223) identified?

The best simulation means the best fit parameters from our 390 runs of the GSA. But we now plot the result in corresponding to the range of the simulated runoff, as our purpose is not focus on the calibration. We would like to show the model performance in our catchment based on the SA results, so there will be no calibration in our manuscript, and we will explain it in a more clearly way in the new version

- "This indicates that there are still inaccuracies in the model parameters" (p10 L235): Mismatch in the simulated water balance could come from both issues in the parameter values (due for instance in part to the fact that the bias may not be considered as a criterion to select the best simulation), but also the model structures.

We agree with the referee, that the mismatch can be due to the inaccuracy of the setup of the model parameters, the model structure, as well as the uncertain of the measurements, but our aim is to see whether our model can generate reasonable runoff data for our catchment, in order to show the applicability of the ParFlow-CLM in our Stettbach catchment rather than a real model calibration, so we will delete the description of calibration and explain our simulated range of runoff based on our SA correspondingly. A clarification and more detailed description can be found in the next version of the manuscript.

- The authors would need to better explain the objectives of calibrating the model before performing GSA.

As described above, our intention was not to do a model calibration. In a revised manuscript we will describe this clearer.

5) The manuscript lacks a discuss section that would discuss the results in light of the very large existing literature on sensitivity analysis.

Yes, we add now a real discussion in the manuscript, please see it in the new version of the manuscript.

6) At numerous locations, the text is unclear to me and I provide examples below:

- p1 L25-26 "help us to reduce the computational demands of completing multiple simulations of expensive domains": this needs clarification

- p2 L47 "Thus": I did not get this logical link between the two sentences.

- p2 L51-52 "In addition, the identification of sensitive parameters should also help to reduce the danger of non-unique solutions, i.e. equifinality": to me the issue of equifinality arises because there are limited data/information available to constrain the model structures and parameters.

- p4 L122 "This is more efficient than LH": I did not get why and what is meant by LH. The GSA method could directly be applied using LH without building a tailored sample?

- p16 L329-332: The two last sentences of the manuscript are fuzzy. Therefore, the conclusions of the manuscript are not clear.

A more clearly conclusion will be rewritten based on the new simulation in the next version of our manuscript.

Referee 2:

General comments

This paper applied a global sensitivity analysis, named LH-OAT, to a distributed hydrological model ParFlow-CLM at the Stettbach catchment. The sensitivity analysis was carried out for 12 parameters, 3 different slopes, and 3 different meteorological conditions. The results investigated the sensitivity of various simulation variables to the above input factors (i.e., parameter, slope, meteorology). The paper does not develop new models or methods, nor does it fully evaluate previously published models, so I don't think it fits the scope of the Geoscientific Model Development journal. The innovation of this paper is limited, and the paper is not well written. Therefore, I suggest rejection.

Major comments to the authors

Lack of innovation. My main concern of this paper is lack of innovation. The authors applied an existing global sensitivity analysis method to an existing hydrological model and a small catchment. May authors clarify what the innovation this paper has in comparison to existing publications. In addition, sensitivity analysis results differ from catchment to catchment and are also dependent on the parameterizations of different hydrological models and simulation periods. Findings of this paper are based on one case study and one model, so they are likely not useful to other readers. How can the findings of this paper benefit broad readers?

As we explained before, we agree with the referee that we do not pursue new global sensitivity analysis (SA) approaches. However, it is believed that this is still an innovative manuscript, as we aim to fill the knowledge gap on the global sensitivity of the distributed hydrologic ParFlow-CLM model to different output variables.

Justifications of LH-OAT. Section 2.2. Eq. 6 of this paper is different from Eq. 7 of van Griensven et al. (2006). The latter uses an absolute value. Therefore, I hope the authors can justify why Eq. 6 is used and why it makes sense in this paper. In addition, please justify why 30 LHs are selected. I personally think they are not sufficient.

As we explained before, we will now use the absolute value of the partial difference for the new version of our manuscript.

Unfortunately, the presentation of this paper hasn't met the publication criteria yet. I provided a list of places where I think there are syntax errors. I hope the authors can make improvements.

Are the 12 parameters and all the meteorological inputs, such as precipitation, are spatially uniform over the catchment?

We have a relative a small catchment, therefore, our meteorological inputs are homogeneous for the whole catchment, in addition, spatially distributed parameters would probably increase the model performance, but are not really necessary for our SA study of ParFlow-CLM.

Minor comments to the authors

Please check syntax errors of the following sentences: Lines 42, 207-209, 235-236, 256-259, 261-263, 301, 318-319, 321-322, 325-326.

It will improve in the new version of our manuscript.

Please use a consistent format for date throughout the paper.

It is done.

Line 161. You mean 2900m*2700m?

Yes, we now use 2900 m * 2700 m instead of 2900*2700m.

Figure 3. Please check the temperature y-axis stick label of the second subplot.

We just changed it, and now it is consistent with the other two subplot.

Please use meaningful and easily understood terms to represent model output variables. For instance, in Table 2, the first column, these short names are hard to understand, and I need to refer to the descriptions below the table and go up and down multiple times. You can directly use their full name in the table. The same suggestions to Figures 6 and 7. In addition, I think using a table full of numbers is not recommended. Instead, you can consider plotting them in a 2D format, in which the partial effects are differentiated by colors.

We will update the Figures and make it easier to read.

References

Ben Touhami, H., Lardy, R., Barra, V., Bellocchi, G., 2013. Screening parameters in the Pasture simulation model using the Morris method. Ecol. Model. 266, 42–57.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Model. Softw. 22 (10), 1509-1518.

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., Attinger, S. and Thober, S., 2016. The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model, J. Geophys. Res., 121(18), 10,676-10,700.

Confalonieri, R., et al., 2010. Sensitivity analysis of the rice model WARM in Europe:exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. Environ. Model. Softw. 25 (4), 479–488.

Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Brielmann, H., Weiler, M., 2013. Testing the realism of model structures to identify karst system processes using water quality and quantity signatures. Water Resour. Res. 49, 3345–3358. https://doi.org/ 10.1002/wrcr.20229.

Herman, J. D., Kollat, J. B., Reed, P. M., and Wagener, T., 2013. From maps to movies: high-resolution time-varying sensitivity analysis for spatially distributed watershed models, Hydrol. Earth Syst. Sci., 17, 5109–5125.

Jefferson, J.L., Gilbert, J.M., et al. 2015. Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. Comput Geosciences 83, 127-138.

Jefferson, J.L., Maxwell, R.M., et al. 2017. Exploring the Sensitivity of Photosynthesis and Stomatal Resistance Parameters in a Land Surface Model. J Hydrometeorol 18(3), 897-915.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G. and Gupta, H., 2015. Are we unnecessarily constraining the agility of complex processâ based models, Water Resour. Res., 51(1), 716–728.

Rouzies, E., Lauvernet, C., Sudret, B., and Vidard, A., 2021. How to perform global sensitivity analysis of a catchment-scale, distributed pesticide transfer model? Application to the PESHMELBA model, Geosci. Model Dev. Discuss. [preprint].

Rosero, E., Yang, Z.-L., Wagener, T., Gulden, L.E., Yatheendradas, S., Niu, G.-Y., 2010. Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season. J. Geophys. Res. 115 D03106.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis, The Primer, Wiley.

Sarrazin, F., Pianosi, F., & Wagener, T., 2016. Global sensitivity analysis of environmental models: Convergence and validation. Environmental Modelling & Software, 79, 135–152.

Shin, M., Guillaume, J.H.A., Croke, B.F.W., Jakeman, A.J., 2013. Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R. J. Hydrol. 503.

Smith, J. D., Lin, L., Quinn, J. D., and Band, L. E., 2002. Guidance on evaluating parametric model uncertainty at decision-relevant scales, Hydrol. Earth Syst. Sci., 26, 2519–2539.

Srivastava, V., Graham, W., et al. 2014. Insights on geologic and vegetative controls over hydrologic behavior of a large complex basin - Global Sensitivity Analysis of an integrated parallel hydrologic model. J Hydrol 519, 2238-2257.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R., 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models, Journal of Hydrology, 324, 10 – 23.

van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y., 2008. Rainfall characteristics define the value of streamflow observations for distributed watershed model identification, Geophys. Res. Lett., 35, L11403.

Van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2008a. Characterization of wa- tershed model behavior across a hydroclimatic gradient. Water Resour. Res. 44 (1).

Wagener, T, Pianosi, F, 2019. What has Global Sensitivity Analysis ever done for us: A systematic review to support scientific advancement and to inform policy-making in earth system modelling, Earth-Science Reviews.