Comment on gmd-2021-425

Fanny Sarrazin (Referee)

Referee comment on "How to perform global sensitivity analysis of a catchment-scale, distributed pesticide transfer model? Application to the PESHMELBA model" by Emilie Rouzies et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-425-RC1, 2022

The manuscript entitled 'How to perform global sensitivity analysis of a catchment-scale, distributed pesticide transfer model? Application to the PESHMELBA model' aims to provide a methodological contribution to the application of sensitivity analysis to spatially distributed models, taking as an example a pesticide transfer model. The study investigates the application of three different global sensitivity analysis (GSA) methods, namely the Variance based Sobol' method based on Polynomial Chaos Expansion, the Feature Importance measure based on Random Forest, and The Hilbert Schmidt Independence Criterion (HSCI), that has been little used in previous hydrological and transport models. It examines the sensitivity at the local scale (homogeneous unit), as well as the landscape scale.

This is a welcome contributions, since performing a sensitivity analysis of spatially distributed model is challenging, because simulations are typically computational expensive, while the number of parameters is large. Many past sensitivity analysis applications have focused on lumped model representations. However, I think that a number of points of this manuscript need clarification and the novelty of the study needs to be better explained. I summarize my main points below, before providing detailed comments.

Main comments:

1) The method section focuses too much on the technical details of the sensitivity analysis methods (which are not methods, but methods taken from previous studies). The

methodology (how these methods are used) is not well explained, which I think should be more the focus of the paper.

2) Some clarification are needed regarding the setup of the case study (Section 2.2).

3) The manuscript lacks a discussion of the methodology and results with respect to previous studies. This would help to clarify the novelty of the study. In particular:

3a) The authors highlight that this is the first sensitivity analysis applied to the PESHMELBA model (e.g. L588 L26), but sensitivity analysis was applied to other pesticide models (e.g. Dubus et al., 2003; Hong & Purucker, 2018...). The manuscript lacks a review on previous sensitivity analyses (local, global) applied to pesticide models.

3b) It is also not clear to what extent the methodology for sensitivity analysis proposed in the manuscript is new compared to previous sensitivity analysis studies. In this respect, previous studies have also proposed to use first a computationally cheaper sensitivity analysis method (method that requires a relatively low number of model simulations, such as the Morris Elementary Effect Test) to screen non-influential inputs, before applying a computationally more expensive method (e.g. Sobol' Variance Based method) based on the subset of influential inputs (e.g. Garcia et al., 2019; Vanuytrecht et al., 2014). This could be discussed in the manuscript.

4) I wish to point out that the PESHMELBA model, as well as the code to compute the HSIC sensitivity indices are not publicly available, but are available upon request from the corresponding author (Code and data availability section). To advance open science (and to comply with the GMD guidelines?), I think that it would be valuable to make these resources openly available, especially since the paper has a methodological focus.

Detailed comments:

L21 p1 'simple enough to ensure flexibility': More explanation is needed here. This is vague and I am not sure what is meant by flexibility.

L30-31 p2 'catchment-scale model [...] afforded': Specify that this is spatially distributed models.

L61-73 p3: Also note the recentstudy of Smith et al. (2021).

Section 2.1: a presentation of the model parameters is missing. How many uncertain parameters that needs to be estimated are there? What are the different categories of parameters (e.g. soil, pesticide, vegetation etc., as I can read in Table 2). Parameters are only introduced much later in Section 2.5 (Table 2), which makes it difficult to follow Section 2.2 that describes the selection of the parameter values. The reference to Table 2 in the caption of Table 1 does not flow well.

Section 2.2:

- Why performing the experiment on a virtual catchment and not a 'real' one?

- I understand that the simulation experiment considers the application of the fungicide at the beginning of the winter period. Is this realistic?

- Why performing the experiments over a 3-month winter period? This is a very short time period.

- A justification for the soil moisture initial condition (hydrostatic equilibrium L157) is missing.

Section 2.3-2.4: I think that section 2.3 provides too many technical details that are not necessary to understand the methodology and analyses presented in the paper. The authors recognize themselves that this section could be skipped L183-184. My suggestion is to report only the main equations used to compute the sensitivity indices, while details on the derivation of these equations (that were taken from previous papers and that are therefore not really a contribution of this paper, if I understand correctly) can be moved in the supplements/appendix. I am mostly referring to the description of the Sobol' and HSIC methods, while I think that the description of the random forest method in Section 2.4 reads very well. The main equations and references of Section 2.3 can be combined with the summary of the GSA methods provided in Section 2.5, to provide the reader only with the information that are needed to understand the methodology and the analyses, while avoiding unnecessary repetitions between Sections 2.4 and 2.5. In addition, I think that an overview of the methodology (why do you need to use the GSA methods?) is needed before introducing the specific GSA methods.

Equation (17):

- The sensitivity index for a given input is the average of the first order indices estimated for the different model outputs, weighted by the outputs variance, am I correct? This paper aims to help applying these methods, therefore I think that interpreting the equations in simple (intuitive) terms, would improve readability and clarity. It is very nice to have the formal mathematical proof for the equation, but the proof does not have any practical implications and could be moved into the supplements/appendix (this is an example of how this section could be simplified, see my previous comment).

- Only first order indices can be estimated for multidimensional outputs? In Figure 10 I see that also the total indices are calculated at the landscape scale. How was this done?

Equation (24): If Xi and Y are not independent, the value of the dependence measure estimated for a given bootstrap resample (that is in a way obtained by randomly attributing values of Y to each value of Xi, if I understand correctly) will tend to be larger than the dependence measure estimated for the original non-bootstrapped sample? Why?

Section 2.4: The GSA workflow is not well explained in the text. In particular, the references to the sample sizes used are confusing. I read that 1000 points are used for PCE (L382), 4000 points for HSIC (L391), that 1000 points were derived from the 4000 points used for HSIC and that 1000 points are used for RF. It is only by looking at Figure 5 that I finally understood that these numbers are linked: 4000 points initially used for HSIC and then based on HSIC screening 1000 points are selected for all subsequent analyses. However, I am still a bit unsure why it is written L374 that 'a variance decomposition method was first used', isn't it HSIC?

L416 p17 '100 replications were used': Why using 100 replications for bootstrapping? 1000 bootstrap resamples are typically used (e.g. Archer et al., 1997; Yang, 2011).

Table 2: I believe that the LAImin and LAIharv are missing. The Table would also need to include an additional column that specifies at which spatial level the parameters are defined (e.g. soil horizon, plot/VFS). It took me a while and a bit of digging in the manuscript to get this information. I would also add the value of the standard scenario in Table 2, this would further improve readability.

Section 2.5: this section does not clearly explain that the vegetation parameters and hpond are considered for vineyard plots and VFSs separately. As already mentioned in my previous comment, I think that the parameter should be clearly introduced in Section 2.1, which would improve readability and clarity.

Section 3: As mentioned in my main comments, the manuscript lacks a discussion of the methodology and results with respect to previous studies, which could be highlighted in an additional discussion section.

P463 'It is commonly stated that […]'. This sentence needs to be better justified. A reference is missing (e.g. Wagener & Pianosi, 2019). It can also be that many parameters are influential, but have only a small impact on the output except for a few parameters (e.g. five or six) that dominate the output variability.

L566-568: Could you explain more why is it more costly to assess the sensitivity analysis at the local scale compared to the catchment scale? From Eq.17, it looks that anyway the catchment scale indices require the calculation of the local scale indices.

Minor edits:

L47 p2 and L569 p26: replace 'computational price' by 'computational cost'

L53 p2 'covariance': this is a technical term I suggest to specify that it refers to input interactions-.

Section 2.3.3: clearly state that this section refers to regression trees (as a classification tree can also be considered).

P18 L450: I would replace 'led' by 'leads' (present tense), since these simulations were actually not performed.

P18 L453 'hydrodynamical parameter': do the authors mean 'soil parameters' (as reported in Table 2)?

Table 2: There is an issue in the table header.

Figure 7: It is difficult to read the confidence intervals for the bars that have a dark brown color. I suggest revising the color scheme.

Appendix B: veget_LAI_min_1 appears twice.

References:

Archer, G. E. B., Saltelli, A., & Sobol, I. M. (1997). Sensitivity measures, ANOVA-like techniques and the use of boostrap. *Journal of Statistical Computation and Simulation*, *58*(2), 99–120. https://doi.org/10.1080/00949659708811825

Dubus, I. G., Brown, C. D., & Beulke, S. (2003). Sensitivity analyses for four pesticide leaching models. *Pest Management Science*, *59*(9), 962–982. https://doi.org/10.1002/ps.723

Garcia, D., Arostegui, I., & Prellezo, R. (2019). Robust combination of the Morris and Sobol methods in complex multidimensional models. *Environmental Modelling and Software*, *122*, 104517. https://doi.org/10.1016/j.envsoft.2019.104517

Hong, T., & Purucker, S. T. (2018). Spatiotemporal sensitivity analysis of vertical transport of pesticides in soil. *Environmental Modelling and Software*, *105*, 24–38. https://doi.org/10.1016/j.envsoft.2018.03.018

Smith, J., Lin, L., Quinn, J., & Band, L. (2021). Guidance on evaluating parametric model uncertainty at decision-relevant scales. *Hydrology and Earth System Sciences Discussions*, 1–30. https://doi.org/10.5194/hess-2021-324

Vanuytrecht, E., Raes, D., & Willems, P. (2014). Global sensitivity analysis of yield output from the water productivity model. *Environmental Modelling and Software*, *51*, 323–332. https://doi.org/10.1016/j.envsoft.2013.10.017

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*, *194*(February), 1–18. https://doi.org/10.1016/j.earscirev.2019.04.006

Yang, J. (2011). Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. *Environmental Modelling and Software*, *26*(4), 444–457. https://doi.org/10.1016/j.envsoft.2010.10.007