Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-437-RC1, 2022


# Comment on gmd-2021-437

Francesca Pianosi (Referee)

---

Referee comment on "Uncertainty and sensitivity analysis for probabilistic weather and climate-risk modelling: an implementation in CLIMADA v.3.1.0" by Chahan M. Kropf et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-437-RC1, 2022

---

The manuscript presents an interesting advancement to the CLIMADA risk modelling platform to enable rigorous assessment of uncertainty propagation through risk models. This new functionality will be of interest to a broad range of CLIMADA users as well as serve as inspiration for developers of other modelling platforms to implement similar advancements. I think the manuscript is overall well structured and well written. The discussion and outlook section is in my opinion particularly thoughtful and provides a very good account of benefits and limitations of current approaches to uncertainty quantification. I think this manuscript and the underpinning work offers a practical contribution to accelerate the uptake of good practices in the risk modelling community, as well as a contribution to foster further discussion on the need of better handling of model uncertainties. I would thus recommend the manuscript for publication after minor revisions. Below are some suggestions for improvement which I hope the authors may consider in preparing a final version.

MAIN POINTS

[1] sampling of hazard component

I understand that the hazard component in the CLIMADA platform is based on directly loading a hazard event set (p 4 line 85). I suppose CLIMADA users will use different approaches to generate the event set, sometimes relying on data, other times relying on model simulations. For example, in the case of flood risk assessment, people typically use dynamic hydrological-hydraulic models to produce maps of river flow peaks or flood inundation depths over the modelled domain. In such cases, I understand that the current implementation of CLIMADA does not allow to explicitly include the input factors of the dynamic hazard model into the uncertainty and sensitivity analysis (UA/SA). In other words, users cannot sample the input factors of the hazard models and propagate uncertainties, instead they will have to define ways to perturb the hazard model outputs

directly. I think in many cases this may be quite difficult as it may not be obvious at all how one can perturb spatially-distributed outputs (for example flood depth maps) in a way that is physically consistent and respect plausible spatial patterns. The authors only touch upon this on p. 6 line 133 where they say: "this modelling choice should be made based on solid background knowledge". Honesty I find this statement a bit simplistic - reality is very often we do not have that knowledge and defining a reasonable way to sample spatially-distributed variables can be the most critical and time-consuming aspect of setting up the UA/SA. I am not suggesting the authors should solve this issue (far from it!) but I think they should point out this is a very important and critical step, and a big area for future research and development. In my opinion this question - how do we sample complex and spatially-distributed variables in a meaningful way? - is one of the key research questions that the UA/SA community will need to work on if we want to move on to the next stage of applying this type of techniques to complex models.
Related to this point, I also wonder how difficult it would be to enable users to load into CLIMADA an ensemble of perturbed event sets, in place of applying perturbations within the platform? If this was viable, users could do the sampling and Monte-Carlo simulations of the hazard model outside CLIMADA, and then use the platform to combine the hazard samples with exposure and vulnerability samples and perform the final calculations of sensitivity indices.

[2] choice of sample size

p. 7 line 159: "In general, it is recommended to use..." I disagree. GSA literature shows that the sample size needed to achieve reasonable approximation of sensitivity indices significantly vary with GSA method but also, for the same method, with the case under study (see for example Sarrazin et al. 2016). For variance-based methods, "adequate" sample size can range from 1,000 up to 10,000 (or more) times the number of input factors (see for example Figure 5 in Pianosi et al. (2016)) so I would really avoid giving readers a "one-fit-for-all" recommendation, which may be misleading. As pointed out in Sarrazin et al. 2016, a better approach is to start with the sample size that one can afford to generate reasonably efficiently, and then check the robustness of the estimated sensitivity indices. If the key conclusions about the input ranking or screening are unambiguous despite the uncertainty in estimated indices, fine, otherwise one should either generate more samples of fall back to using a more frugal GSA method. This is indeed what the authors themselves do when they calculate confidence intervals and check that they do not overlap significantly. So I would suggest to revise this paragraph bringing in this discussion and potentially anticipating or referring to the later description of how confidence intervals can be used to estimate robustness of the results.
p. 7 line 182 and Section 2.3.5 - following up on my previous comment: I would also suggest to add some more information about the confidence intervals, specifically: (1) how are they derived? no need to go into the details but at least say in one sentence what's the key idea/methodology (bootstrapping?) to derive them. (2) how can they be used? again I would briefly explain to the reader how confidence intervals should be interpreted and used (I am thinking something like the discussion of Figure 4 in Noacco et al 2019 or even shorter).
I insist on this point as in my experience the choice of the sample size is one of the most confusing for GSA users, especially when doing GSA for the first time, so I think it is important to give sound and clear advice on this!

OTHER MINOR POINTS

p. 2 line 31: "In practice, the quantification of risk with climate risk model..." A recent paper that also makes this point and shows how GSA can be used for the evaluation of impacts models, particularly when fit to historical observations may not be a sufficient criterion, is Wagener et al (2022)

p. 2 line 48: "an analytical treatment is often not possible", I agree though when it is possible it should be considered as the primary route to SA. I think a useful reference here may be Norton (2015) which covers some of the analytical approaches to SA

p. 2 line 50: "uncertain input parameters" Here and everywhere else, the authors use the term "parameters" to refer to the uncertain inputs that are varied in the uncertainty and sensitivity analysis. I find this terminology potentially misleading, as some may interpret the term "parameters" in a narrower and more specific sense (for instance in dynamic systems terminology people tend to distinguish the model "inputs" into initial conditions, boundary conditions, forcing inputs, parameters - hence using "parameters" in a very specific sense). Indeed on page 4 (L. 108) the authors mention more generic "input variables and parameters definition". I would suggest either clarifying what the term "parameters" refer to or, as commonly done in the GSA literature, use the term "input factors" instead.

p. 6 line 133: "Note that the choice of the variation range ..." A reference with some good examples of this problem that could be added here is Paleari and Confalonieri (2016).

p. 8 line 191: "long single impact computation time". What are the factors that control computation time in CLIMADA? I understand that the platform does not execute any dynamic hazard models but instead it directly loads a hazard event set. Hence, even if the event set was generated with repeated executions of a (expensive) hazard model, the complexity of these calculations should not affect CLIMADA computing time. Am I right? If so, then I suppose CLIMADA computing time should be mainly controlled by the chosen spatial resolution - again, is this correct? I think these points would be worth clarifying.

p. 8 line 209: "Note that it is perfectly valid to use different sampling..." I am a bit confused by this remark. Yes it is possible to use different sampling strategies but why would one want to do that? My take on this is that the direction we are moving towards is quite the opposite, that is, the GSA community is increasingly focusing on UA/SA methods that work on the same generic input-output sample, so that we can reduce the effort of generating the sample (which is often the computational bottleneck of this type of analysis) and make the most of it for both uncertainty quantification and attribution. See for example discussion of "given-data approaches" in the introduction of Borgonovo et al (2017).

p. 11 line 296: "the distribution of uncertainty... is bi-modal" In terms of GSA, this is interesting as the use of variance-based indices with output distributions that are multi-modal or highly skewed may be a bit critical. In fact the underpinning assumption of variance-based sensitivity analysis is that variance is a good proxy of output uncertainty (in other words, variance is a good statistic to synthetically characterise the shape of the output distribution). This assumption is perfectly fine for symmetric distributions whereas it becomes more and more questionable with multi-modal or highly skewed ones. In such case a different GSA method may be more appropriate (see for example discussion in Pianosi and Wagener 2015). I am not suggesting that the authors perform any further analysis but maybe they may consider making a comment here or think about this in future developments.

p. 13 line 314: "strong correlations" should be "strong interactions"

REFERENCES

E. Borgonovo,X. Lu,E. Plischke,O. Rakovec,M. C. Hill (2017) Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box, Water Resources Research, https://doi.org/10.1002/2017WR020767

Noacco, Sarrazin, Pianosi, Wagener (2019) Matlab/R workflows to assess critical choices in Global Sensitivity Analysis using the SAFE toolbox, MethodsX, https://doi.org/10.1016/j.mex.2019.09.033

Norton (2015) An introduction to sensitivity assessment of simulation models, Environmental Modelling & Software, 69, https://doi.org/10.1016/j.envsoft.2015.03.020

Paleari and Confalonieri (2016) Sensitivity analysis of a sensitivity analysis: We are likely overlooking the impact of distributional assumptions, Ecological Modelling, 340, https://doi.org/10.1016/j.ecolmodel.2016.09.008

Pianosi and Wagener (2015) A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, Environmental Modelling & Software, https://doi.org/10.1016/j.envsoft.2015.01.004

Sarrazin, F., Pianosi, F., Wagener, T. (2016), Sensitivity analysis of environmental models: Convergence and validation Environmental Modelling & Software, 79, 135-152.

Wagener, Reinecke, Pianosi (2022) On the evaluation of climate change impact models,

WIREs climate change, https://doi.org/10.1002/wcc.772