Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-154-AC1, 2020

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "The impact of hydrological model structure on the simulation of extreme runoff events" *by* Gijs van Kempen et al.

Gijs van Kempen et al.

gijs.vankempen@hotmail.com

Received and published: 24 August 2020

We would like to thank the anonymous referee for their feedback. We appreciate that the reviewer generally liked the methodological approach. Below, we respond to the points raised by the reviewer.

Main points:

- The parameter ranges are taken from the original FUSE paper and applied in different climate zones. I am not convinced that the parameter space is fully (or sufficiently) sampled using theses ranges. For very different regions than the ones where the models were intended and developed for the ranges might be quite different and a stop in increase of change using the Kolmogorov-Smirnoff test might not indicate that the

Printer-friendly version

Discussion paper

space as sufficiently sampled, but could also be that there is a region of the parameter space that is not considered at all by the study set up.

We are not quite sure if we understand the point of the reviewer in this aspect. The upper and lower parameter boundaries are generally based on physical and conceptual understanding, and should in principle capture all values that these parameters could reasonably take, independent of climate or catchment type. As such, we do not doubt that the parameter ranges as provided by the FUSE paper are the right starting point for the sampling. Concerning the sampling itself; yes, given the high-dimensionality of the parameters and the relatively limited parameter sample size, there will be regions in parameter space that are unexplored. That is; there will be quite some space in between the samples. The latin hypercube sampling strategy, however, ensures that we sampled over the full parameter range and that there are no 'overlooked' regions or corners.

- I am also not fully convinced that the very same parameter range should be applied for the catchments that can be found in different zones, hence I cannot understand why in the synthetic test these ranges should be the same and not a plausible range known from or tested in real catchments from these zones

We believe that applying the same parameter range to different climates is well-justified. Most of the hydrologic model parameters are determined by catchment properties (landscape, geology, land use) and not by the climatic conditions. There can be a large variation of different catchment properties within the same climate zone, and therefore one can not beforehand limit or stretch the parameter range based on climate only. Of course, there are some relations between catchment properties and climate; elevation and/or slope can for instance influence climate but also catchment storage properties, vice-versa climate can influence the catchment through rain-induced erosion or through vegetation processes. However, this is difficult to predict or translate to generalities and depends on long soil formation processes and historical climate conditions.

- How much do the additional snow routine parameters potentially influence the plausible parameter ranges of the other parameters? I would argue that that could change quite a bit and again would expect some kind of evaluation for instance by using real catchments from the respective regions.

From a conceptual point of view, there is no reason to assume that snow routine parameters influence the parameter ranges of the other parameters. These ranges are determined independent of the snow process. Of course, when one would calibrate a model, it would make a difference for the final parameter values coming out of the calibration if snow parameters were included or not, but the parameter ranges of the other parameters would not be adapted for the calibration procedure.

- How much could using the same parameters in the snow routine effect the results? The very same degree-day was used despite the different climate zones. for snow influenced catchments the snow routine is crucial and varying for instance the degreeday will have large differences in the simulations. Please discuss

Degree day parameters not only depend on climate, but also on many local circumstances (such as the distribution over north and south facing slopes or wind conditions that are not necessarily specified in the Koppen-Geiger classification). As such, we think it is cleanest to keep the degree day parameters fixed across the different climates. It is a valid point, however, that the snow parameters were not sampled, whereas the other model parameters were. Also sampling the snow parameters would probably further broaden the uncertainty bands around the simulations. We will add a clarification and discussion of the treatment of snow parameters.

- One of the objectives of the study is to link extreme event via their return periods to their sensitivity to model structure if the extreme events are simulated. The authors use daily data and daily simulation, however, often very large events occur at shorter time scales. How could the approach be extended to these or would that shift the return periods very much? I assume that might be particularly relevant for arid zones.

Printer-friendly version

Discussion paper

Indeed in arid zones, extreme events are often related to flash floods which last for a few hours only. It would require higher temporal-resolution climate model output in order to be able to simulate such events. This would be computationally quite challenging, given also the localized and convective nature of the rainfall that triggers such flash floods. Our return-period method does allow for relatively easy translation from daily to hourly, but we are limited here by the possibilities on the climate modelling side. Currently, we implicitly assume that the 24h mean would also be among the highest if a flash flood occurred within those 24 hours. This is of course not necessarily the case, we will add a note on this to the discussion.

- The extreme events were selected by using the minimum and maximum, for many studies on extreme values (particularly low flows) a moving average is used to avoid effects of oscillations etc. in these ranges. Maybe that would also solve some of the problems with the hard-coded threshold?

Thank you for this suggestion. A moving average is indeed an option that we will investigate in order to see if this increases the robustness of the results. We expect, however, that it might not completely resolve the hard-coded threshold issue, since these periods are rather persistent.

- Extreme values are looked at only in terms of timing and maximum/minimum simulated streamflow. Other parts of the events might be interesting as well (event volume, deficit, duration etc.), while I see that that is not the focus of this study, I would appreciate a couple of words on these and how easy or difficult the proposed method could be extended to these characteristics.

Thank you for this suggestion. Indeed, extremes can be defined in many different ways, and min/max discharge or timing are only two of many. Event volume is generally a bit more challenging because it requires the definition of a start and an end of the event - equations exist for this but the parameters of these equations might be climate / catchment dependent. We will investigate this suggestion.

Minor comments

- The terms "drought" and "low flow" are not clearly distinguished. While one (drought) can lead to the other, low flow is a seasonal characteristic of the flow regime. Maybe use instead of simply drought the term "hydrological drought" but since the study is really about low flows, why not fully leave out the term drought?

We agree with the reviewer that the terms were used interchangeably. Indeed in the formal definition, drought is used for anomalies while low flows are a seasonal characteristic. Since we are looking for the most extreme low flows (basically negative anomalies in low flows) this could again be perceived as a drought. But to not further complicate the text, we will replace drought with low flow throughout the manuscript.

- Form: the results part is slightly mixed with discussion parts (referring to other studies). Then a synthesis follows and then, when the reader would expect conclusions, a new discussion part starts. While it is interesting in a way, I would propose to change the order. A reader that is looking only at specific parts can easily find them without having to go through the full paper. The discussion bits in the result part could together with the synthesis become the first part of a discussion before going into the discussion about limitations of the study setup.

Thank you for the suggestion. We will adapt the structure accordingly.

Line by line comments

All textual suggestions will be incorporated.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-154, 2020.