Response to Referee #1

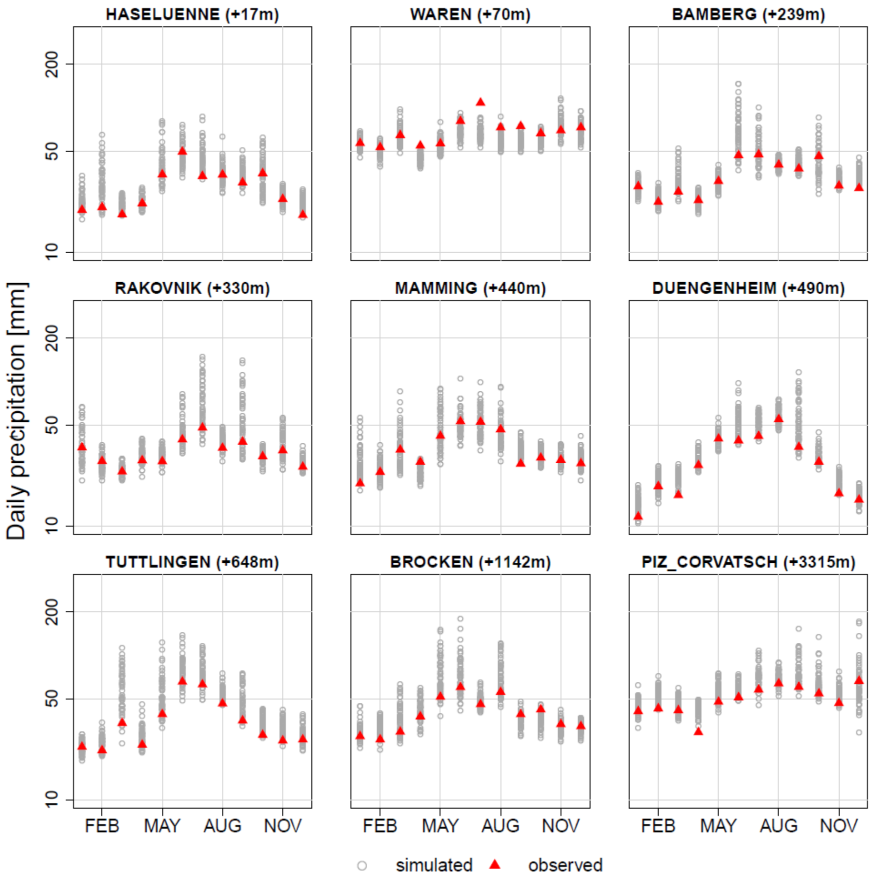
First of all, we would like to thank the referee for the time and effort put into reviewing the manuscript. This response carefully addresses all the comments. We further attach a change tracked version of the manuscript.

The authors present a modelling exercise where they compare and analyse flood damages under three different assumptions regarding the spatial distribution of return periods of flood peaks throughout the catchment. They apply a complex modelling chain that goes from synthetic weather generation to flood damage estimates. The chain is composed of a weather generation algorithm, a rainfall-runoff model, a 1-D flood propagation model, a 2-D spatial inundation model and a damage estimation model. Their emphasis is on the comparison of the risk curves obtained throughout the catchment for different assumptions of spatial dependence of flood return periods, taking as reference case the option of modelled dependence. The authors present results of their analyses for the entire Elbe catchment and for its division in 29 sub-basins. They infer conclusions on the over estimation or underestimation of flood damages with respect to the reference case.
The topic is relevant for the audience of NHESS, the objectives are clearly identified, the methodology for the analysis is adequate and the conclusions are relevant and correctly supported by the results and discussion. The analysis clearly shows the striking differences between the options of independence and complete dependence with respect to the reference condition of modelled dependence, and therefore the objective of the paper is clearly achieved. Therefore, I believe the manuscript deserves publication in NHESS.
SPECIFIC COMMENTS The authors are addressing a formidable task. They are reporting results obtained with a complex modelling chain that probably took years of work under the constraints imposed by the length of a research paper. It is only natural that some parts of their work or the models used have necessarily been left unexplained. I am suggesting a few points where I believe the reader would benefit from some additional details, such as the following:

We would like to thank the reviewer for his/her positive feedback and kind remarks.

a) On page 7, lines 161-163, the authors report that the weather generator was calibrated for the region of the Elbe basin with observed data and "was shown to capture precipitation extremes well". I think the entire analysis is dependent on the quality of the time series produced by the weather generator, particularly regarding spatial correlation and seasonality of extremes, which are very challenging. I think the paper would benefit from a more elaborate discussion of the spatial dependence of the extremes produced by the weather generator as compared to observations.

We thank the referee for this suggestion. We agree that the quality of the time series produced by the weather generator is significant for the entire analysis. In response to this comment, we provide the figure below which compares the simulated (grey circles, each per one simulation realization) and observed (red-filled triangle) daily extreme precipitation (99.9th percentile) at 9 selected stations (out of 528 stations) over the simulation area. It shows that daily precipitation extremes and their seasonality are captured very well by the weather generator at the locations of the individual stations. We currently carry out a spatial analysis of the weather generator performance and together with the at-site validation, this will be a self-consistent study on the weather generator. We do observe that the spatial rainfall performance is more challenging for the weather generator than capturing the local statistics as also known for other generators discussed in the literature (e.g. Serinaldi and Kilsby 2014). It seems that the spatial dependence of very strong rainfalls is somewhat overestimated which will presumably translate into the dependence of discharge peaks. However, we believe that for the purpose of the presented analysis, the overestimation of spatial rainfall dependence is not critical. The results of modelled dependence are located between complete dependence and complete independence for high return periods. With an ideal weather generator, they would be closer to the complete independence. Thus, our estimates for the difference between the assumption of complete dependence and modelled dependence can be regarded as conservative. Hence, the major conclusion challenging the assumption of homogeneous return periods still holds. This will be added to the revised version of the manuscript.



b) The topology of the model should be explained better. On page 7, line 186, the authors state that overtopping flow is calculated from the 1-D diffusive wave model. On line 170, they say that the runoff is routed by the Muskingum method and aggregated at the basin scale. This leads me to think that the possibility of overtopping is not contemplated in the reaches modelled by Muskingum. From Figure 1 I gather that the 29 sub-basins under analysis are actually composed of smaller units, which are the ones simulated in SWIM, but this is not mentioned in the text. I think the manuscript would benefit form a brief discussion of which rivers are included in the overtopping analysis and which criteria were used to delineate the SWIM basins.

In this study, regional flood model (RFM) considers SWIM sub-basins in the calculations. Only for a better representation, the risk results are presented for grouped sub-basins (29 sub-basins), which aggregate several original SWIM sub-basins. In SWIM model, the entire catchment area is first subdivided into sub-basins with average area in a range of 10 to 100 km^2 and definitely not larger than 100 km^2. Additionally, in response to overtopping analysis, the reviewer is right, the overtopping is considered only at the main river network and higher order tributaries that have a drainage area of 600 km^2 or more. This river network is explicitly modelled with the 1D-diffusive wave hydrodynamic model. The flood routing in smaller tributaries with drainage area below the above-mentioned threshold is done by the Muskingum routing within the SWIM model. These will be added to the revised version of the manuscript.

c) The operational definition of return period should also be discussed in detail. I first thought that the return period referred to peak flow and was estimated from the 10,000 year simulation in each location. However, on page 9, line 248, the authors say that they refer to a T-year flood event as resulting in the T-year damage. In addition to peak flow, flood damage is also affected by hydrograph volume, which is very relevant to determine the extent of the inundated area, at least for flash floods. Perhaps the authors might consider a brief discussion of this issue.

We thank the reviewer for pointing this out. Flood damage depends not only on the flood peak but on the hydrograph shape, floodplain hydraulics (e.g. dike overtopping and inundation patterns), exposure and vulnerability of affected elements. 10,000-year simulation of the risk chain enables us a large sample of damages.

From this sample, we derive an empirical frequency distribution for the probability of damage which is then used in the estimation of flood risk. Hence, we refer to a T-year flood as a flood resulting in the T-year damage in this study. We denote here the term “T-year flood” in a different way based on damage return period as compared to the traditional way based on the peak flow return period. This will be added to the revised version.

TECHNICAL CORRECTION From the formal standpoint, the paper is very well written, correctly organized and adequately illustrated with figures. Comparison of Figures 4 and 6 is handicapped by the fact that the return period is shown in natural scale in Figure 4, while it is shown in logarithmic scale in Figure 6. I would suggest using the same type of scale on both figures, if possible.

We thank the reviewer for this comment. In fact, in Figure 6, bottom-right panel, we illustrate the risk curves for entire catchment (1:29) on a logarithmic scale. Therefore, this comparison is still possible and we would prefer not to change these figures. We provide risk curves on a linear scale in Figure 4 for a better understanding of the damage ranges.

Reference

Serinaldi F., Kilsby C. G. (2014): Simulating daily rainfall fields over large areas for collective risk Estimation, Journal of Hydrology, 512, 285-302, doi: 10.1016/j.jhydrol.2014.02.043.