Anonymous Referee #1

The main objective of the paper is to develop a new methodology to generate flood hazard maps. Compared to other studies, the new methodology considers the exceedance probability of multi-model combinations based on forecasted peak discharges derived by a set of hydrological models. In addition, the type of building (purpose & structure) is considered to define the hazard at an object. This is important in terms of prioritization for e.g. early warning and emergency planning.

Authors: We do sincerely appreciate the detailed and very good comments from the reviewers', and we have acknowledged their contribution in improving the quality of this paper. We hope that we have now addressed all the reviewers' comments satisfactorily in the revised manuscript.

After recommendation from other reviewer, we have updated the title

"Buildings hazard maps with differentiated risk perception for flood impact assessment"

Please find our point-by-point responses to the comments

Reviewer: General comments:

- The paper is very interesting and the main objective to create hazard maps or doing impact assessment with a transparent declaration and consideration of uncertainties is desirable. Especially the presented approach using confidence intervals of a hydrological forecast ensemble is interesting and has potential. However, there are doubts about some major points in this study.
- It is questionable whether the approach presented in this study "inherently communicates the underlying uncertainties", as stated in the conclusion. Looking at the final map presented in figure 9a), the meaning of Scenario I, II and III is not directly derivable and the coherence of all classifications and the different combinations of hydrograph scenarios with building types is not easily comprehensible. Even if the names of scenarios in the map would be changed to "high exceedance probability" (=S I), "average exceedance probability" (=S II) and "low exceedance probability" (=S III) – what would already improve the understanding - it's still questionable whether the multi-model combination presented is the right way to deal with uncertainties in hydrological forecasting for flood impact assessment. To combine the exposure for different confidence interval hydrographs in a new scenario defined by the same exceedance probability is not very elegant. According to the output of the ensemble members, the M50% confidence interval hydrograph as used in the paper (= best-model = median) is the hydrograph that is forecasted as the most likely one and therefore, to define scenarios with low, average or high exceedance probabilities based on multi-model combinations referring to different confidence intervals is misleading. I try to exemplify this issue on table 3: The way the M%-hydrographs from the ensemble forecast are used would imply that the higher the confidence interval, the lower the exceedance probability of an event. But this is not the correct way to implement the confidence intervals here. At the time of the forecast and according to the model ensemble, it is as likely that a M25% or a M75% (= 50% confidence interval) event appears (when I understood it right that for example the lower 80% confidence interval figure 2

corresponds to the M10% hydrograph -> 80% confidence interval means that 80% of the forecasted cases will also be in this range and 20% not -> 10% at the lower and another 10% at the upper end of the scale). In general, the terminology used in this study is also misleading, as for example the 50% confidence interval discharge is used as the median (= M50% = best model), whereas the 50% confidence interval would correspond to the M25% discharge and M75% discharge -> 25th and 75th discharge percentile or 0.25 / 0.75 quantile.

I think that two (hydro-)statistical approaches were unconsciously mixed – the one of confidence intervals of a model ensemble forecast and the general probability of discharge to exceed a certain value (extreme value statistics are not any more relevant at the time of the forecast). When applying the method with the confidence intervals correctly, it appears that scenario I (defined as high exceedance probability) is the most unlikely scenario according to the model forecast. Therefore, the multi-model combination is not working as supposed.

Authors: Thank you for pointing out the error, as suggested we have replaced the term percentile in the revised manuscript and corrected Figure 3. We also agree that the exceedance probability is not the right choice of word to describe the scenarios, hence we have replaced it by “risk perception” and have added the scenario description in Figure 9.

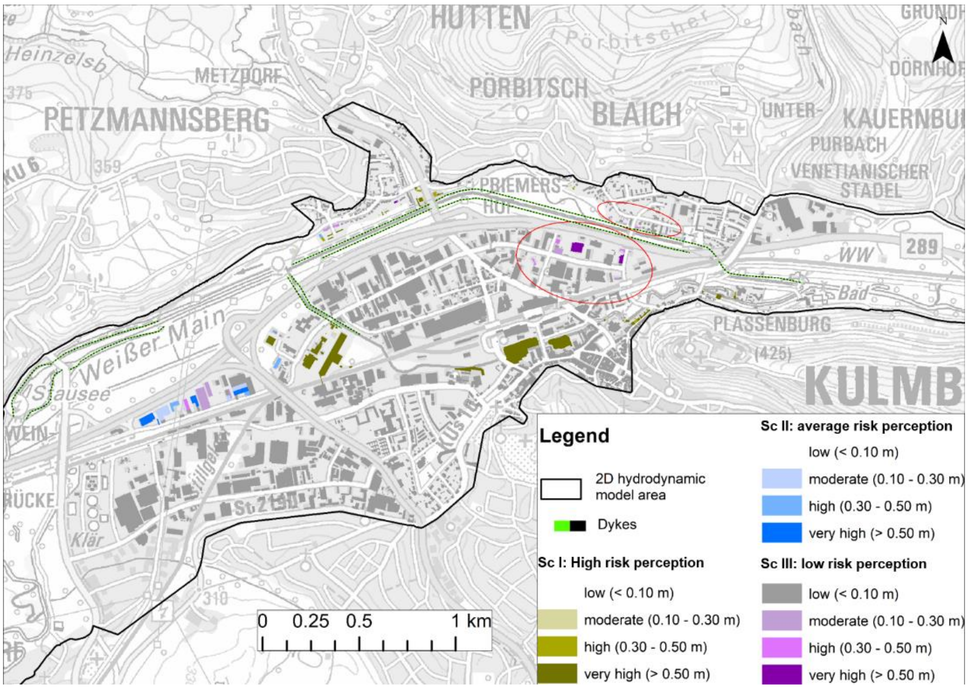


Figure 9(a) Multi-model combination

In addition, we stated that these maps offer an alternative way to communicate the underlying uncertainties. It will be interesting for authors to measure how the scientific community and practitioners will receive the methodology. In conclusions, Page 21: Line 12-14

“In summary, we have presented a new methodology for flood impact assessment using a multi-model combination in the form of building hazard maps with differentiated risk perception. These maps offer an alternative way to communicate the underlying uncertainties in forecasting models and are ready-to-use for decision-makers in the field of flood risk management.”.

- In addition, I don’t see the a fundamental novelty of this approach compared to the cited study by Zarzar et al. (2018), who already presents the use of a multi-model ensemble framework based on hydrological ensemble members for visualizing flood inundation uncertainty. I agree with the authors that the use of confidence intervals in the case of a high number of hydrological forecast members can support a transparent declaration of forecast uncertainties.

Authors: We have now differentiates the work done by Zarzar et al. (2018), they have used an average of HD model raster outputs to obtain the percentage of ensemble agreement. Whereas, we have exported HD models on buildings as urban features and proposed a new methodology of multi-model combination. In Introduction, Page 3: Line 1-4

“Zarzar et al. (2018) have used a multi-model combination framework consisting of hydro-metrological and HD models to visualise flood inundation uncertainties in which they have used an average of HD model raster outputs to obtain the percentage of ensemble agreement.”

- Besides all of this, to my point of view the display of the inundation pathways and extend is necessary for a use in early warning systems, emergency planning or flood impact assessment. The approach of a building hazard map doesn’t provide enough details to intervene (where does the water come in contact with the building, where are the “weak spots” in the river- and floodplain-system).

Authors: We acknowledge the importance of the inundation extent; we have proposed to use hazards maps in addition to other layers of information, such as inundation map for planning. In conclusion, Page 22-Line 1-5

“Finally, the output of the framework can be extended to hazard maps uploaded in a web-based GIS system to improve visualization, along with providing layers of additional information, such as inundation pathways and weak spot in the river and floodplains to provide sufficient details to intervene (aid in planning). This additional information would enhance the usefulness to different target users, such as planners, decision makers and flood forecasting agencies.

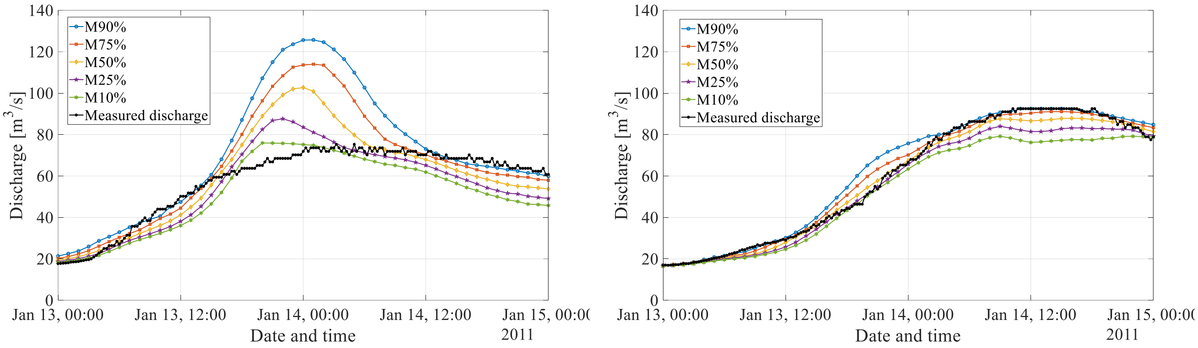
Reviewer: In addition to these major points, there are some further remarks in the specific comments.

Specific comments

- Section 2.1, 2.2 + supplement tables: It is mentioned that the particular focus of this study is on the development of the post-processing part (classification and multi-model combination, mapping). Therefore, the two parts of the framework that are already developed and explained similarly in Beg et al. (2018) and Bhola et al. (2018a, 2018b) should be shortened, as for example a detailed explanation of the model LARSIM, FloodEvac and HEC-RAS is not needed to understand the context of this study. Nevertheless, Figure 3 helps to understand the setup of the multi-model approach. It should be better explained, where for example the M10% hydrograph can be found in such a graphic (consistent use of confidence intervals). It would be helpful if you could add the forecast ensemble for the virtual station that was used as upper boundary condition and maybe mark the used hydrographs.

Authors: We agree with the reviewer comment regarding shortening the details, however, additional details are provided in order to reproduce the methodology, as suggested by the editor.

In the revised manuscript, we have added data labels in Figure 2 to clearly present the percentiles. These two stations where used as the input boundary conditions upstream of the model. The virtual gauge is nothing but the addition of these two hydrographs, we present here the data at gauges upstream of the city.



(a) Ködnitz **(b) Kauerndorf**

Figure 1. Hindcasted flood event of January 2011: measured discharge hydrograph along with 95%, 90% and 75% percentile discharges for gauges a) Ködnitz and b) Kauerndorf (Data based on Beg et al., 2018).

- Section 2.3: The main literature cited (Krieger et al., 2017) is specific to Germany and is not peer-reviewed. In which way are the classification methods for buildings and hazard types comparable with international, peer-reviewed literature? E. g. Jonkman et al. (2008), Dutta et al. (2003), as well as Thieken et al. (2008) and Kreibich et al. (2010) on german scale, differentiate loss estimations of residential/industrial/commercial etc. buildings due to different vulnerability, whereas here these types are represented in the same class here (III).

Authors: Thank you for providing the studies, we have compared their classification to ours in the revised manuscript. We also stated further break down of class III in conclusion. In methodology, Page 8: Line 6-14

> "There are various classifications of land use features available in literature. Dutta et al. (2003) have used direct and indirect damages as the basis of their classification and classified their study area in residential and non-residential categories. Jonkman et al. (2008) have classified urban features in residential, businesses, commercial and public property and agricultural to estimate flood loss. Furthermore, vulnerability was the basis of classification in residential (Thieken et al., 2008) and Industrial & commercial sectors (Kreibich et al., 2010) in order to estimate flood losses. We have used damage potential of a building as a basis for classification in order to focus on the flood impact assessment. Building damage potential is required for a variety of flood mitigation planning activities

including flood damage assessment, multi-hazard analyses and emergency measures (Shultz, 2017). The buildings were classified into four classes based on their function following the recommendation of the German standard for risk management in urban areas in the case of flash floods (Krieger et al., 2017)."

And in Conclusion, Page 21: Line 30

"In addition, other classification methods for buildings and hazard types should be evaluated, especially to further dissect the impact of class III in commercial and industrial."

- p. 8, lines 17 + 18; figure 4; p. 14, line 12: Check the numbers given on p. 8 – they do obviously not match the numbers per class in figure 4 and also not with the statement on p. 14, that the most buildings were classified in the classes II and III. I'm also doubting the usefulness of building class I, as there are parks included (which are not buildings) and there is only one member. Furthermore, it's questionable if the applied classification method in the context of the study makes sense, as relative to total number of 2695 buildings, creating separate classes with 1 and 20 (based on figure 4, I would assume that this is class IV) buildings respectively, lead to underrepresentation of these classes. Based to that, it's not surprising that for example the low exceedance scenario is similar to the M50% and M75%, as the buildings are mostly distributed in the classes combined with these hydrograph scenarios.

Authors: Thank you for pointing out the error, we have corrected in the revised manuscript. The building class I belongs to buildings that are in open green area such as small park and garden building. In methodology, Page 8: Line 25-26

"There are a total of 2695 buildings in Figure 4 of which 1, 958, 1716 and 20 buildings were classified in classes I, II, III and IV respectively."

Regarding the building classification, Page 8: Line 13-14

"The buildings were classified into four classes based on their function following the recommendation of the German standard for risk management in urban areas in the case of flash floods (Krieger et al., 2017)."

We acknowledge that depending on the aim the classification can be adjusted. Page 8: Line 17-18

"In any case, we acknowledge that the number of classes/criteria can be changed/adapted depending on the aim of the forecast."

We have also added in Page 8: Line 26-27

"The nature of the data in this case study lead to differentiated representation of the classes . It should be noted that, the classification aims at creating classes based on damage potential, and not on generating clusters with similar sizes."

- Figure 8 + 9, p. 15 line 1 - 3: In this context, you should maybe explain somewhere, how you assign the hazard for the same building (finally in figure 9) that is in a moderate hazard zone for scenario I but then rise to a very high hazard in scenario III (is the potential in the forecasted event to have very high hazard at a particular building somehow considered?).

Authors: Thank you for your comment. This was done based on table 3., in Results, Page 15: Line 4-7

" The main objective of the combination is to differentiate the impact of water depths on building classes. Therefore, to design the combinations, a high percentile was assigned to the buildings with a high damage potential class. Each scenario 5 presents a certain risk perception, which can be adjusted depending on the need of stalk holders. The hazard maps for the three scenarios are shown in Figure 8."

Table 1. Scenarios of multi-model combinations based on risk perception.

Scenario	**Risk perception**	**Building class**			
		I	II	III	IV
I	High	$M_{10\%}$	$M_{10\%}$	$M_{25\%}$	$M_{50\%}$
II	Average	$M_{10\%}$	$M_{25\%}$	$M_{50\%}$	$M_{75\%}$
III	Low	$M_{25\%}$	$M_{50\%}$	$M_{75\%}$	$M_{90\%}$

- "p. 16, lines 4 – 11: In this part, the time issue in real-time assessment of the framework is discussed. As well in the introduction it is mentioned, that flood forecasts might be restricted to computational time of the models. Please provide information about the lead- and run-time of the hydrological forecast model, the computational time for the HD model with each hydrograph and, therefore, how much time would be left for authorities to intervene. This is evident for early warning and emergency planning. It would of course also be interesting how the offline maps from Bhola et al. (2018a) would perform compared to the modelled confidence hydrographs.

Authors: We have provided information on the run-time of the entire map production. In Discussion, Page 20: Line 12-16

"In this study, a 50-member ensemble forecast was used from Beg et al. (2018) where the entire process took 25 minutes in a 3-core desktop in parallel mode to generate a forecast of 12 hours. Various percentile discharges were then run simultaneously in the HD model, which required 30 min to simulate a 12-hour event on an 8-core, 2.4 GHz (Intel E5-2665), including the initial start (Bhola et al. 2018a). Post-processing of the model results would consume an additional 15 min. Therefore, real-time hazard maps are delivered to decision makers in 70 min."

Technical corrections

- In general: As already mentioned, the use of the term confidence interval discharge in combination with the M% HD outputs seems not to be correct. Please replace confidence interval with "percentile" when referring to M%-levels – or change these levels accordingly.

Authors: The term is changed to percentile in the revised manuscript.

- p. 1, line 21: It would maybe help if you would explain a bit more in detail, what in this study is meant as multi-model combination. Based on this first explanation, one would assume that the building classification is not part of the multi-model combination and this consists only of the HD and hydrological model (which, according to figure 1, is not the case). The term is also used in various ways: multi-model, multi-model combination, multi-model ensemble combination, multi-model combination scenarios. E.g. in the first sentence of the discussion section, the context is given only to hydrology. If this is the case, then figure 1 should be adapted. This can be a bit confusing.

Authors: We have used multi-model combination consistently in the revised manuscript. The term is changed to percentile in the revised manuscript. Building classification is part of the entire framework but independent of multi-model combinations. We have focused this study on the multi-model combination. In Methodology, Page 3: Line 17-18.

> "The particular focus of this study is on the development of the framework of a multi-model combination in the post-processing component"

- p. 8, line 17 + 18: As mentioned, check the numbers for each class and compare with figure and other statements.

Authors: Thank you for pointing it out, we confirm that the numbers provided are correct in revised manuscript.

- p. 11, line 11/12: As mentioned in special comments, to my point of view the confidence intervals cannot be used to describe exceedance probabilities in the way it was done here.

Authors: We have corrected the term and used percentile consistently.

- p. 11, line 22: You refer to post-event information that "no serious damage was reported" -> p. 14 line 4: "figure 7c is in agreement with the post-event information" -> 7c = M50% discharge scenario. According to figure 6, in this scenario 126 buildings are exposed, 67 classified with high hazard -> How does that fit to "no serious damage was reported"?

Authors: This is explain in the discussion. The high hazard at M50% is due to their proximity to the Mühl canal. In Discussion, Page 17: Line 8-12

> "It should be noted that few buildings show very high hazards due to their proximity to the Mühl canal (**Fehler! Verweisquelle konnte nicht gefunden werden.**a). Even though there was no over-topping of water from the canal, the buildings near the canal were assigned the highest hazard, starting with a discharge of $M_{10\%}$. Up to a discharge of $M_{50\%}$, no inundation in the city centre was observed as the dykes were not breached."

- Figure 7, 8 and 9 b-d: Legends (and building numbers in 9b) are not readable. Also the maps themselves are at the edge of readability. The red circles in figure 9 are not in the legend and are never mentioned in the text (see comments below)? Is figure 9d really necessary?

Authors: The figures have been resized so the legends are readable, in addition all the figures are provided to the editorial so the final print will be of high-quality and readable. We have omitted figure 9d from the revised manuscript. The red circles are defined in the figure caption as well as in the text.

- p. 13, line 12 – 15: This error should be eliminated by either using water depth derived from water surface elevation [m a. s. l.] minus ground level of building [m a. s. l.] or by removing the river channel elements from the dataset you used to assign to the buildings. Check out Bermúdez and Zischg (2018).

Authors: Thank you for your comment. We have added this limitation of our study to the article. In Discussion, Page 17: Line 10-12.

> "Even though not implemented in this study, this misrepresentation could be overcome by removing the river channel elements from the dataset used to assign to the buildings as in Bermúdez and Zischg (2018)."

- p. 14, line 17 – 22: Please add the information that you in this specific case compare the red circles from M50% and the multi-model map. ID 1393 is not recognizable in figure 9d. p. 16. Line 3 + 4: What is the computational time of the whole framework? What's the lead-time used here?

Authors: We have added the information and stated the computational time of whole framework. Please see in specific comments.

Publication bibliography

- Bermúdez, María; Zischg, Andreas Paul (2018): Sensitivity of flood loss estimates to building representation and flow depth attribution methods in micro-scale flood modelling. In Nat Hazards 92 (3), pp. 1633–1648. DOI: 10.1007/s11069-018-3270-7.
- Dutta, Dushmanta; Herath, Srikantha; Musiake, Katumi (2003): A mathematical model for flood loss estimation. In Journal of Hydrology 277 (1-2), pp. 24–49. DOI: 10.1016/S0022-1694(03)00084-2.
- Jonkman, S. N.; Bockarjova, M.; Kok, M.; ˇBernardini, P. (2008): Integrated hydrodynamic and economic modelling of flood damage in the Netherlands. In Ecological Economics 66 (1), pp. 77–90. DOI: 10.1016/j.ecolecon.2007.12.022.
- Kreibich, Heidi; Seifert, Isabel; Merz, Bruno; Thieken, Annegret H. (2010): Development of FLEMOcs – a new model for the estimation of flood losses in the commercial sector. In Hydrological Sciences Journal 55 (8), pp. 1302–1314. DOI: 10.1080/02626667.2010.529815.
- Thieken, A. H.; Olschewski, A.; Kreibich, H.; Kobsch, S.; Merz, B. (2008): Development and evaluation of FLEMOps – a new F lood L oss E stimation MO del for the private s ector. In D. Proverbs, C. A. Brebbia, E. Penning-Rowsell (Eds.): Flood Recovery, Innovation and Response I. FLOOD RECOVERY, INNOVATION AND RESPONSE 2008. London, England, 02.07.2008 - 03.07.2008. Southampton, UK: WIT Press, pp.315–324.