***Comments of Reviewer 1 (Brenden Jongman) and responses***

*We thank you very much for the detailed review and the constructive suggestions to improve our article. We revised the manuscript carefully according to your review comments. Below you can find the detailed responses and revisions.*

**Major comments**

Comment 1: In my opinion, some more effort could be put in the fitting of the results into the broader scientific field. The paper rightfully refers to a range of literature on flood damage model comparison and uncertainty evaluation. It is for example very similar to the paper by de Moel and Aerts (2011) and Jongman et al. (2012), who also compare different flood damage models amongst each other and with empirical case-study data. Right now, these studies are only mentioned in the introduction. It would be good if the authors could try to embrace this more in the entire paper: what are you doing differently than existing studies? What is the added value of these results? How do the uncertainty results compare?

*Response 1: In contrast to the mentioned studies, this work does not focus on a model-intercomparison and aims rather on the general applicability and transferability of different loss models to other geographical regions. De Moel and Aerts (2011), for example, investigate different sources of uncertainty within flood damage modeling. Thereby the contribution and quantitative range (uncertainty) of the single components (e.g. water depths, damage models) on the total damage outcome is of importance. In their work the damage models are, however, based on different asset values and are therefore not really comparable with each other. Thus the difference in the damage outcome is a combined effect of different underlying asset values and shapes of the depth-damage curves. Additionally, their aim was not to evaluate the "best" damage model by comparing the outcome with observed loss data. Their goal was rather to identify the relative contribution and uncertainty of the different components to the overall estimates.*
*The work of Jongman et al. (2012), again, compares different models with regard to recalculating observed losses best. Apart from different quantitative and qualitative comparisons between the models their study focused also on different economic sectors (industrial, residential etc.). They applied different models for two study areas, one in Germany and one in the UK, even if they were designed for other remote European countries or even for the US. Nevertheless, like in the study of de Moel and Aerts (2011) they used for each original depth-damage curve a corresponding maximum damage value that was assigned to each land use class before. Thus regional differentiation of underlying asset values in the area under study is, in contrast to our study, not taken into account. Since models applying the same original asset values to a number of economically distinct regions lead to substantial over- and underestimations they conclude to use regional asset values in future studies, as carried out in the study at hand.*
*The main goal of our study was to analyze to which spatial extent damage models can be transferred from one (similar or related) region to another. Since it is still an open question under which conditions this is possible we focused on a local scale where local asset values and observed losses were available. The task was to apply loss functions from a neighboring region/country to find out whether a spatial transfer of functions derived in related and closer regions is possible. Within the risk analysis in Sect 3.4 we illustrated how uncertainty of damage estimates could be reduced by previous model validation. This was, for example, not analyzed in the work of de*

*Moel and Aerts (2011) or Jongman et al. (2012) who only demonstrated the maximal range caused by different loss functions (and asset values).*
*We stressed now these fundamental differences throughout the manuscript by opposing it more often to the papers of de Moel and Aerts (2011) and Jongman et al. (2012).*

Comment 2: The implications of the results could be emphasized better in the abstract, results section and conclusions. Right now the conclusion that 'more attention should be paid to flood loss assessments' by using 'more loss data' is in my view not strong enough. The results can be used to give more specific implications of this study: can we apply stage-damage functions to other areas? If yes, how can the results found here be used to make this application better, i.e. how do we transfer and adapt models? If your results show that we should not use the models in different areas, how will we move forward?

*Response 2: We modified particularly the second half of the abstract and conclusions as well as some parts in the result section to give more specific advice by means of the results obtained. In detail, we mentioned that a model transfer without validation might be critical. We recommend therefore that in case of missing data for model calibration and validation loss functions should at least be selected from related and more similar regions (in the sense of building and flood event characteristics) to improve the reliability of damage estimates. But a rigorous improvement of model estimates can only be achieved by collecting more and comprehensive loss data for model calibration and validation!*

Comment 3: It is not always quite clear what the goal of the study is. I understand that you try to assess whether models give accurate results when applied in different geographies than they are designed for. It seems to me that an important part of this assessment is in fact uncertainty assessment: the differences in the models are in fact the result of parameter definition (steepness of function etc), which partly result from the geography of design but for a large part from pure uncertainty in these parameters. This uncertainty assessment is addressed specifically in various parts of the paper, for example in figure 6. But then in other paragraphs, for example page 3509, you state that 'the focus of this study was not to evaluate the uncertainty of flood risk curves'. I understand that full uncertainty assessment is not possible, but attributing all differences to 'model plausibility' (i.e. regional focus) is also not realistic. Please discuss this point further.

*Response 3: As mentioned above, the goal was not only to transfer models to other regions, but also to find out whether models derived for similar regions yield more reliable loss estimates. The uncertainty assessment within a flood risk analysis in Sect. 3.4 was introduced to illustrate the range of damage estimates which occurs solely due to the fact that different damage models (in combination with the intrinsic uncertainty of asset values) were applied. An important part is thereby to show, how this range of damage estimates is reduced when only models are selected which were validated in the previous part. Thereby other uncertainties are not addressed as, for instance, no other flood frequency statistics or hydraulic models were applied which would be necessary for a comprehensive uncertainty analysis in flood risk assessments (see Merz and Thieken 2009). Since we kept the same underlying specific asset values in our study region in all loss assessments (in contrast to the*

*work of de Moel and Aerts (2011) or Jongman et al. 2012), this range can solely be explained due to the selection of the applied damage model.*

Comment 4: The model comparison is not quite clear.
The first reason is, that you extensively discuss the specific model FLEMOAT (~2 pages), but only marginally discuss the other 3 main models applied here (one paragraph). You don't fully describe what these others models are based on: are the curves designed for a certain house type or are they general? Can they be compared this easily with the extended FLEMO model, or should we take certain things in consideration when doing this?
Second, you make model combinations that are not explained clearly on forehand. In the results section (p. 3508, line 16), the reader suddenly sees himself confronted with 57 model combinations. Maybe you can explain this in the methodology and already mention it in the introduction?
A third point is the use of asset values. You seem to compare stage-damage functions of different models, but use the same asset values in the comparison. It is important to discuss that damage models are developed as a combination of depth-damage functions and corresponding asset values. While the ICPR model has quite steep damage functions, the asset values linked to these relative function are very low compared to other models, which is why the model generally gives an underestimation of losses. By taking only loss functions and not using model-specific asset values, you lose part of the comparability. This is an important element of your study, that should be discussed throughout the methods and results.
Finally, connected with the previous comment: in the paper (e.g. p.3507) you discuss uncertainty due to asset values. Somehow you use a range of asset values, but it is not made clear enough in the methods, result discussion and conclusions what this uncertainty is. Please elaborate further.

*Response 4:*
*1) The reason why we discussed the FLEMO model longer is that it contains more parameters (than the other models) which have additionally to be modified so that the model can be applied to the Austrian study region. Otherwise the reader would not know how the required input data were derived. Nevertheless, we extended also the description of the other (more simple) models as far as possible. Furthermore we mentioned also a few words about the HOWAS database where these models were derived from.*
*2) We now extended the last two paragraphs of the introduction (see also comment no. 8) but also the result section by explaining how the model combinations for the uncertainty assessment were developed.*
*3) What you mention here is carried out in the study of de Moel and Aerts (2011) or Jongman et al. (2012), for example, which you would like to embrace in the entire paper. But our aim was not to investigate the combined uncertainty of damage estimates due to different functions and underlying asset values. Instead we use identical site-specific asset values for all relative damage functions to isolate the single effect of model choice (and the effect of the underlying basic data from different geographical regions). We only introduced a range of asset values as also the mean specific asset values may be associated with intrinsic uncertainty. But this range is assigned to all functions to compare only the different shapes (and origins) of the applied loss models. In fact, the asset values that are combined with the loss function should reflect the local/regional economy and building stock as good as possible. Using asset values from very different regions does not make sense.*

*4) Although it is referred to the paper of Cammerer and Thieken (2013) to get more details about the uncertainty of asset values, we explained in Sect. 2.2.3 how the range/uncertainty of the applied asset values was derived. Both in the introduction, conclusion and other parts we stressed more frequently what this asset uncertainty means.*

Comment 5: I feel the comparison with empirical data is not always discussed into enough depth. Since you compare the results and empirical losses directly, you have to 100% sure that everything that is included in the modeled losses is also in the reported losses, and the other way around: e.g. direct losses, indirect losses, structural damage, content damage, the damage due to contamination etc. Are these fully consistent? It would be great if you could discuss this further.
Also, the discussion on causality of the results is still not wide enough. You suggest on several points that the FLEMO model is better because it includes contamination (e.g. page 3503) but it is not discussed in enough depth how this shows from the results. Together with my comments on the use of asset values, your conclusions are therefore not always compelling. I would suggest to improve this by adding more discussion on the model characteristics and the reasons for comparative differences.

*Response 5: In Sect. 2.2.2 we described in detail where the (independent) loss data for the model validation stem from (disaster fund). There, it is explained that the official loss data were obtained as anonymized and separated building loss data (object level but not georeferenced). Due to privacy protection and also general bias in loss databases (e.g. Downton and Pielke 2005) we aggregated the direct losses on buildings and contents to the whole study area. As the reported damage is not further divided in damage to building fabric and movable items, we assumed that the share of damage to household contents amounted to 30 % of the total loss in cases where it is indicated that also household contents were damaged. Thus it was possible to extract only the direct structural damage on buildings which was used to perform the bootstrapping procedure to receive a confidence interval of the reported structural building loss in the study area. Thus both, the loss estimates as well as the real losses, only contain losses of the building fabric. Therefore, a comparison of estimates and damage record is allowed.*
*In Sect. 2.2.3 we added some new information about the cost type of the asset values (see also comment no. 9 of reviewer #2) to clarify that it is justified to use the underlying asset values (replacement costs) for the damage modeling as the observed loss represent compensation payments (replacement costs).*
*In Sect. 2.3.1 we described the data basis for the derivation of the new loss functions. In the second paragraph we clearly mention that we only used the (relative) damage of buildings from this dataset (and no damage of household contents for instance). As this dataset contains information about contamination it is possible to derive loss functions which differentiate between contamination (and precaution), too.*
*In connection with a third (independent) survey, described in the third last paragraph of Sect. 2.3.2, it is possible to get information about the precautionary behavior and contamination level in the study area as these two facts are not indicated in the reported losses of the disaster funds. By means of this questioned information in the study area it is, however, possible to run the different functions in order to account also for potential damage reduction/increase due to precaution or contamination in the study area.*
*Due to the use of these independent loss data which describe the same type of damage (direct structural damage of buildings) our damage comparison*

*(modeled/reported) is definitely justified (see also new sentences in the second paragraph of Sect. 2.2.3). Nevertheless, we tried to make more clear (in the appropriate sections) which type of damage is addressed.*
*Concerning your second argument that FLEMO is better, we cannot agree with this statement. On page 3503 for instance, you mentioned that the FLEMO models are better because it includes contamination. The only fact what we explained here is that the loss characteristics differentiated by contamination underpin the importance of considering the effect of contamination to building damage. This was already stated in previous studies where loss estimates were more reliable when the effect of contamination was included (e.g. Prettenthaler et al. 2010). But we did not evaluate the performance of FLEMO at this point. With regard to the model comparison or best model performance (Sect. 3.3) it is shown that the best results (run 23a) in the German-wide dataset were obtained with the linear model as far as contamination is considered. With regard to the Bavarian subset the best model (run 23a) is represented by the polynomial function which also considers the effect of contamination. For both models the exact over- or underestimation is also mentioned in the manuscript (p. 3506). Therefore it is not clear for us why the FLEMO model should be the best. The results (and corresponding interpretations) in this chapter do not justify this conclusion. Please see also the answer to your similar comment no. 14.*

Comment 6: I miss a clear overview figure or table that shows the reader easily how the results of the different models compare, and which is 'better'. Right now a table is included with yes/no as to whether the estimates are in the significance interval. It would be great if this could be extended to a more continuous scale, which could support the discussion on the model discussion.

*Response 6: For that we introduced exactly Table 4 to show all modeling results, indicating the exact model estimate (in k€) and additionally the overall performance evaluated by the confidence interval. In the lowest line we introduced now the confidence interval to compare the simulated loss with the observed loss. Based on these evaluations we used a further figure (Fig. 5) to show graphically the comparison of the different models (and again their location within the bootstrap interval). However, there only the models of the Bavarian subset (and the three standard functions) are shown as only these models performed well (as indicated in Table 4) and to not overload this graphic with redundant information. We think that Table 4 and Fig. 5 should be sufficient to compare and to show the performance of all models.*

Comment 7: The section titled 'conclusion' is currently more a summary. The entire first paragraph and parts of the subsequent paragraphs describe again what you did. In my opinion it would be much better for the paper if it would get a real conclusion in which the results are put in context and the implications are made clear, without summarizing the methodology.

*Response 7: We shortened the first and second paragraph considerably (by deleting the more summarizing parts) and extended the results and recommendations in the third paragraph.*

**Smaller comments**

Comment 8: As mentioned previously: a quick introduction of the comparison method in the introduction would be useful. This should include mention of the models that will be compared, and the fact that you only look at depth-damage functions and not asset values.

*Response 8: You partly mentioned that in comment no. 4, too. As replied there, we extended and modified particularly the last two paragraphs of the introduction. There we explained in more detail which models were selected and that we focus on the impact of different damage functions, particularly from different geographical regions to investigate the overall transferability.*

Comment 9: Page 3488: explain better what the depth-damage functions are based on, i.e. what they represent: repair/replacement costs of structure, content, contamination, etc: this is not always the same in all models. Also mention how they are linked to asset values, that are very different in all models.

*Response 9: We think your suggestions should be better implemented on the next page, directly after the listing of the different models as it does not really fit on the previous page. Now we mentioned on page 3489 that there are also differences in the damage-functions regarding the cost type (repair/replacement), differentiation between relative and absolute approaches as well as empirical and absolute functions etc. by giving examples. But we do not want to go too much in detail as the differentiation of the loss models by empirical/synthetical, absolute/relative approach etc. is already done by Merz et al. (2010) or Jongmann et al. (2012) on which is referred for further reading.*

Comment 10: Section 2.1: is there any information on sectoral losses (e.g. residential vs. commercial)? This would be interesting information for the comparison of models.

*Response 10: For Tyrol some indications about losses to specific infrastructure (like telecommunication, energy supply etc.) are available, but there is no general division into different sectors such as residential, industry, infrastructure, agriculture etc. Nevertheless, we added the share of losses in the private sector, to get a rough hint of the damage distribution of this flood event.*

Comment 11: Section 2.2.1: some parts of this section link to the results of your hazard validation. Maybe you can be more concise here, and move some of this section to section 3.1, especially the last part of the paragraph?

*Response 11: We also thought about this issue before submission. However, we think it is an important information in this section what simulation runs were carried out for the validation and also afterwards. Thereby not only the two different versions "22a" and "23a" should be mentioned, but also explained how they differ. Nevertheless, some parts in this section or the last sentence of the last paragraph were moved to the validation section (3.1) as it contains indeed some results.*

Comment 12: Section 2.3.2: it was not directly clear here that you discuss 'general' damage models (very briefly) and a localized specific damage model (FLEMOAT). This difference can be made more clear.

*Response 12: Actually, this is already mentioned at the end of the introduction now. Nevertheless we introduced another sentence in this section to clarify that we discuss more simple damage models and also a multi-factorial model.*

Comment 13: You use ‘surveys_GR’, ‘surveys_BY’ throughout the paper. These seem technical variable names you use in the modeling. It is not reader friendly and not directly clear what the difference is and how this can be interpreted. In my view it would be better to describe these two different samples in words (e.g. refer to ‘the Bavarian sample’) and discuss how this changes the results

*Response 13: We replaced the names of these subsets throughout the text and tables by the whole and longer names now.*

Comment 14: Page 3503, second paragraph: here conclusions are drawn about which model is better, and about inclusion of contamination: however, it seems this is done by comparing the models mutually, not with empirical data. How can you say which one is better and draw causal relationships at this point?

*Response 14: We cannot agree that here premature conclusions are drawn about the “best” model and that the models are mutually compared. In this paragraph it is only statistically analyzed how the basic empirical loss data (from which the later loss models are derived) differ in the two subsets when dividing up in different water classes, contamination types etc. But at this point there is no conclusion drawn about which model is the best. It is only stated that contamination, for instance, crucially influences the loss pattern in the corresponding subset. Thus it is reasonable to include this parameter for example in the extended stage-damage functions which is already suggested in other studies (e.g. Prettenthaler et al. 2010; Nicholas et al. 2001). In this paragraph the difference between the two subsets become clear (larger variation of the loss data) but the different models are compared in Sect. 3.3 for the first time.*

Comment 15: Page 3504, second paragraph: similar to previous comment, it seems to me that these conclusions should be drawn after comparison with observed losses (next page), right?

*Response 15: As also suggested by reviewer #2 (see comment no.13) we moved the sentence with the two premature conclusions in this paragraph to the conclusions part.*

Comment 16: All figure captions can be extended. Right now most of the figures cannot be understood by the reader without going back to the methodology and/or result description. It would be better if you could provide a longer caption that fully explains the content of the figures.

*Response 16: We extended the captions of the figures a little bit to make more clearly what is shown there.*

Comment 17: Figure 4 has way too much information and is not informative as it is. I would suggest splitting it up in different figures, or choosing another way to represent such a vast number of graphs.

*Response 17: We agree that there were too much curves included in this figure. We split therefore both plots in two further subplots as suggested by reviewer #2 (see comment no. 16).*

Comment 18: Abstract line 21: 'loss assessment in the future'

*Response 18: Done.*

Comment 19: Line 7 p. 3491: 'mainly located' – be more explicit

*Response 19: We indicated now the percentage of the area located in the federal state of Tyrol.*

Comment 20: Line 17 p. 3491: remove 'between 1971 and 2006'

*Response 20: Done.*

Comment 21: P. 3506, line 23: explain what 'estimate the reported loss well' means.

*Response 21: We added that good model estimation depends on the fact whether the simulated loss falls within the confidence interval or not.*

**References cited:**


de Moel H, Aerts J (2011) Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates. Natural Hazards 58 (1):407-425. doi:10.1007/s11069-010-9675-6

Downton M, Pielke R (2005) How Accurate are Disaster Loss Data? The Case of U.S. Flood Damage. Natural Hazards 35 (2):211-228. doi:10.1007/s11069-004-4808-4

Jongman B, Kreibich H, Apel H, Barredo JI, Bates PD, Feyen L, Gericke A, Neal J, Aerts JCJH, Ward PJ (2012) Comparative flood damage model assessment: towards a European approach. Nat Hazards Earth Syst Sci 12 (12):3733-3752. doi:10.5194/nhess-12-3733-2012

Merz B, Thieken A (2009) Flood risk curves and uncertainty bounds. Natural Hazards 51 (3):437-458. doi:10.1007/s11069-009-9452-6

Merz B, Kreibich H, Schwarze R, Thieken A (2010) Review article "Assessment of economic flood damage". Nat Hazards Earth Syst Sci 10 (8):1697-1724. doi:10.5194/nhess-10-1697-2010

Nicholas J, Holt GD, Proverbs D (2001) Towards standardising the assessment of flood damaged properties in the UK. Struct Survey 19 (4):163–172

Prettenthaler F, Amrusch P, Habsburg-Lothringen C (2010) Estimation of an absolute flood damage curve based on an Austrian case study under a dam breach scenario. Nat Hazards Earth Syst Sci 10 (4):881-894. doi:10.5194/nhess-10-881-2010