The authors wish to thank the editors and reviewers for their time and effort for reviewing our manuscript. We hope that the changes have improved the manuscript to a level that is suitable for publication, and we look forward to your response.

Reviewer 1

General comments

Major weaknesses in the documentation of the data as well as in the presentation of the calibration of the loss function, have a considerable effect on the replicability of the study.
We appreciate your comment. In the new version, further explanations have been added. Please see the highlighted changes below.

In addition, the presented results are not discussed or framed in the context of existing studies, which makes it difficult to see the advantage of the presented method in comparison to similar approaches.
We are grateful for your suggestion. As mentioned below, a detailed comparison has been added to the new version.
Please see the changes in section 5.

Specific comments

Data description

In the documentation of the data used in the manuscript, several information are missing or not accurately described, which makes it difficult to fully understand each step of the analysis.
As explained in the next specific answers, data description has been made more detailed and clear; suggested technical corrections have been implemented, and an explanation on how information is combined has been added.

An overview table of the empiric data used for the model calibration could help to get a better understanding of the data set in terms of distribution and sample size.
Many thanks for your suggestion. In the new version, distribution and size of the empirical data utilised for model calibration have been shown in *Fig. 3*. It has also been presented in the *caption of Figs. 5 and 6.*

It remains also unclear what building values were used to calculate the relative damage. In L8 on p.5 the author states to use "mean depreciated value" while in L13 p.5 it says "average market values". Values that represent the actual cost of the building based on material and labour can differ considerably from market values depending on the demand for housing in a certain area.
We are very grateful for your comment. The sentence has been amended.
Please see L25 on p.4: "The recorded damage is compared to the average market values of the residential properties, as reported by the cadastral map for the semester preceding the flood event."

In addition, the spatial matching of the damage values and building properties (L13-L17 on p. 5) should be outlined more clearly including Figure 2. This includes a description on how the damage records were aggregated on building level and which assumptions have been made in case damage records were not available for all units in a building. In Figure 2 the authors should explain what the points and building shapes mean and what we can learn from that.

The processing of raw data and the spatial aggregation process is now described in more detail. The Figure caption now explains in detail what the points and shapes are.
Please see L30 on p4- L3 on p5 and the caption of Figure 2.

Calibration and validation of FLF-IT

To avoid confusion, I would suggest moving the part that explains the cross-validation procedure (L12 14 on p.6) in front of the bootstrapping and calibration part (L24 on p. 5 to L6 on p.6) so it is in chronological order.
We appreciate your suggestion. As a matter of fact, cross-validation procedure was related to the model validation which is one step after model calibration. In the new version, to avoid any confusion, the model calibration and the model validation parts are totally separated from each other.

It should also be stated how many samples were pulled out of the data set for each bootstrapping iteration. This is closely linked to the Data description section, where the overall size of the original dataset, the size of each subsample for cross-validation and the size of resampled dataset after bootstrapping should be stated. This can also help to explain the Number of samples in Table 1, which is unclear in the current version of the manuscript.
The overall size of the original dataset used for model calibration (613 samples) is presented in *L5 on p5, L10 on p6, Table 1, and the caption of Figure 3.*
Number of samples utilised in model validation are also added. *Please see L22-25 on p7*

Regarding the RMSE and MAE it should be stated if the percentage values are the original unit coming from the relative damage or if the RMSE and MAE were normalized. In case the values were not normalized it is not possible to assess the predictive performance of the model without knowing the distribution of relative damage in the original dataset. Therefore, either the distribution of relative damage records in the original dataset should be provided or the RMSE and MAE should be normalized.
We are very grateful for your comment. Distribution of the relative damage records is depicted in *Fig. 3*, and it is presented in the caption of *Figs. 5, and 6*.
The distribution of damage ratios and the magnitude of errors according to some sub-classes of water depth is also discussed in *section 5.2*

In addition, I would recommend to slightly restructure Table 3 by showing the 95% confidence interval with the lower and upper boundaries in the second column instead of spreading it over column two and three.
Corrected. *Please see Table 3.*

Discussion

Given the fact that the application of depth-damage functions is a quite frequently addressed topic in flood research (see Merz et al. 2010 and Hammond et al. 2015), I would highly recommend to discuss the results of this manuscript in the framework of existing flood loss functions to highlight the unique and novel character of this study. This discussion should also include a critical evaluation of the study and the limitation of the study design. For example in L1 f. on p.8 the authors state that "Results of these validation tests illustrate the importance of model calibration, especially when the water depth is the only hydraulic parameter taken into account [: : :]." However, without the comparison with an uncalibrated function it is not possible to proof that predictions of calibrated loss functions are significantly better that uncalibrated ones. Since the loss function was calibrated on a single event in Italy using a single building type, the limitations in terms of a temporal and spatial transfer should be addressed as well.

We appreciate your suggestion. In the new version, a detailed comparison has been added, the unique and novel characters of this model have been discussed, and the limitations of this study have been mentioned. In this version, section 5 which is related to results comparison and model validation has been changed substantially.
Please see the highlighted parts of section 5.

Furthermore, the novel characters of this model were mentioned before in *L27 on p1 & L17 on p10*. The limitations were also mentioned before in *L32 on p10*

Literature

P.2 L14: Jonkman (2007) provides a very detailed definition of (in)tangible and (in)direct flood damage and should be added here.
Added.
Please see L19 on p.2

P.8 L4: Merz et al. (2013) and Schröter et al. (2014) showed that additional damage influencing factors considerably improve the damage predictions and therefore should be added here.
Added.
Please see L9 on p.9

Technical corrections

P.1 L1: "Floods and storms": Damage caused by storms is actually not covered in this study. Therefore, I would recommend to include numbers for flood damage only.
Corrected. Now it refers to floods only. The following numbers were already related to flood inundation.
Please see L2 on p.2

P.2 L1 & P.3 L11f: "medium flood probability", "high flood probability". These are rather soft terms to describe flood probability. If available, I would recommend using numeric flood probabilities (e.g. "1% change to get flooded in any given year")
Corrected. They have been changed to probability in terms of return period.
Please see L5 on p.2; & L26 and L27 on p.3.: "exposed to a flood probability of once every 100 to 200 years" and "return period between xxx and xxx years".

P.2 L17: "I-O models": write full name the first time a new term is mentioned
Corrected as "Input-Output models".
Please see L23 on p.2

P.4 L10: "10 thousand": 10,000 or 10^4
Changed to "10,000".
Please see L25 on p.3

P.4 L17: "125 mm of rain". Please provide timespan "e.g. 125 mm of rain in 48 hours"
Corrected.
Please see L4 on p.4: "with an areal mean of 125 mm of cumulated rain over 72 hours flowing in the Secchia catchment."

P.4 L21 & L27f: "6.5 thousand hectares": convert into m^2 or km^2 to improve comparability with other values provided in this section.
Done.

Please see L8 & L15 on p.4

P.4 L30: “bi-dimensional”: 2-D
Done.
Please see L16 on p.4

P.5 L1: “one-meter resolution”: a one-meter resolution
Done.
Please see L17 on p.4

Table 3: “(in EUR m)”: Million? 10^6 EUR
Corrected.
Please see Table 3.

P8. L21: “takes empirical data of damage and depth”: According to the Data description section, the water depth was modelled and not empirically measured.
This sentence was related to FLFA and not FLF-IT. However, in order to avoid any confusion, the sentence was amended.
Please see L13 on p.10: “The FLFA approach takes data of damage and depth.”

References

Hammond, M. J., Chen, A. S., Djordjevi´c, S., Butler, D., & Mark, O. (2015). Urban flood impact assessment: A state-of-the-art review. Urban Water Journal, 12(1), 14-29.
Jonkman, S. N. (2007). Loss of life estimation in flood risk assessment; theory and applications. TU Delft, Delft University of Technology.
Merz, B., Kreibich, H., & Lall, U. (2013). Multi-variate flood damage assessment: a tree-based data-mining approach. Natural Hazards and Earth System Science, 13(1), 53-64.
Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article" Assessment of economic flood damage". Natural Hazards and Earth System Sciences, 10(8), 1697-1724.
Penning-Rowsell, E., Johnson, C., Tunstall, S., Tapsell, S., Morris, J., Chatterton, J., & Green, C. (2005). The benefits of flood and coastal risk management: a handbook of assessment techniques: Middlesex University Press.
Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., & Merz, B. (2014). How useful are complex flood damage models? Water Resources Research, 50(4), 3378-3395.

We very much appreciate your suggestions. Additional references were provided in the revised manuscript. Please see the highlighted references.