The authors would like to thank the Editor and reviewers for the positive and encouraging feedback. The additional comments were found to be very helpful in improving the quality of the manuscript and will be acknowledged. We have responded (in black) to each comment (in blue). Please note that the page and line numbers cited in our responses refer to the clean version of the manuscript.

Reviewer #1:

As a reviewer of the previous version, I do appreciate the improvement made in the paper. However, I remain dubious about the 'preparedness' part. It still reads/looks as a (quick) addition, and not a key, well-thought contribution in the paper. In fact:

- the Introduction does not mention preparedness at all (and then no background/SOTA on it);

- the Methodology has not description about the methods for the preparedness analysis, except Tables 1/2 (which are not explained in the details of preparedness).

The preparedness contribution is squeezed into Sec. 3.3.1 that mixes methodology and results. As a result, the quality of this contribution is very low.

I would remove it, or I would revisit the paper (from the structure to content), in order to reflect how this aspect is embedded in the study.

The authors would like to thank the reviewer for acknowledging the improvements done in the revised manuscript. We understand that the presentation of "preparedness" compromises with the continuity and clarity of the manuscript. As suggested by the reviewer we have removed the term "preparedness" unless in places where it means a general term rather than a conceptualized indicator. We have removed the section titled "Effect of preparedness". As understanding the role of adaptation strategies such as emergency measures success (ms) and knowledge about emergency action (ke) is an important finding of the study, we have re-written the section with more clarity. We hope that this re-written section provides more clarity and suits the flow of the manuscript.

P15/ L 321- 351:

"*Further, we analyse to what extent these adaptation strategies are helpful in reducing* $rloss$. *Through feature selection and Bayesian Networks we identified emergency measures success* (ms) *and knowledge about emergency action* (ke) *for companies and private households respectively, as the significant variables within the Markov Blanket of* $rloss$ *(see Tables S2 and S3 for details on the questions and responses). We applied the FLEMO*$_{flash}$ *model to derive predictive densities of* $rloss$. *Results were summarized using the median and associated uncertainty (25*th *and 75*th *percentiles) for selected combinations of hazard, exposure, and vulnerability conditions, rather than displaying the full predictive densities (Figure 6). For clarity of interpretation, Figure A1 illustrates step by step how predictive densities are derived from the posterior distributions using kernel density estimation based on 1,000 resampled values. Figure A2 provides an overview of the posterior and predictive densities across varying levels of measure success under same conditions of water depth and number of employees.*

For company building, the $rloss$ *Markov blanket consists of water depth (wd), number of employees (emp), and emergency measures success (ms). The incurred loss increases with increasing water depth but combining the water depth and number of employees, the relative loss experienced by companies with 9 to 12 employees is higher compared to those with 38 or more employees. We further investigate the role of successful implementation of emergency measures in reducing loss. For example, when considering only water depth and number of employees, companies with 38 or more employees, inundated by water depths ranging from 1.85 m to 2.40 m above the ground, experience a* $rloss$ *of 0.38. However, with successful implementation of emergency measures (ms=3), this* $rloss$ *decreases to 0.20, representing a substantial 47% reduction. On the other hand, when emergency measures were* perceived to be ineffective ($ms = 1$), the $rloss$ increases to 0.50, marking a 32% increase in estimated losses.

For household buildings, considering hazard and exposure with a water depth of ≥ 2.3 m above ground and no contamination, the predicted $rloss$ is 0.22. However, when residents had clarity on emergency actions ($Ke \geq 5$) this loss decreases to 0.05, reflecting a 77% reduction. Conversely, when residents did not have clear knowledge on emergency action ($Ke \leq 2$), the $rloss$ rises to 0.27, a 23% increase. Similarly, for household contents, we observed that when considering only water depths (hazard) of height 2.3 m above ground, the $rloss$ is 0.38. With ($Ke \geq 5$), this loss drops to 0.17, a 55% reduction. In contrast, with ($Ke \leq 2$), the $rloss$ increases to 0.44, representing a 16% rise. These findings highlight the potential of successful implementation of emergency measures and knowledge about emergency action in reducing flash flood losses (Barendrecht et al., 2020; Berghäuser et al., 2023; Bubeck et al., 2013; Sairam et al., 2019; Surminski and Thieken, 2017). While residents with high knowledge about emergency actions are more likely to take effective emergency measures, thereby reducing the severity of flood loss (Kreibich et al., 2005; Sairam et al., 2019), the knowledge about emergency action (ke) variable as considered in this study, does not capture which exact actions respondents undertook. Though the specific actions likely varied across respondents, empirical evidence indicates that having clear knowledge of emergency action generally contributes to reduced flood losses, consistent with previous findings (Kreibich et al., 2021)"

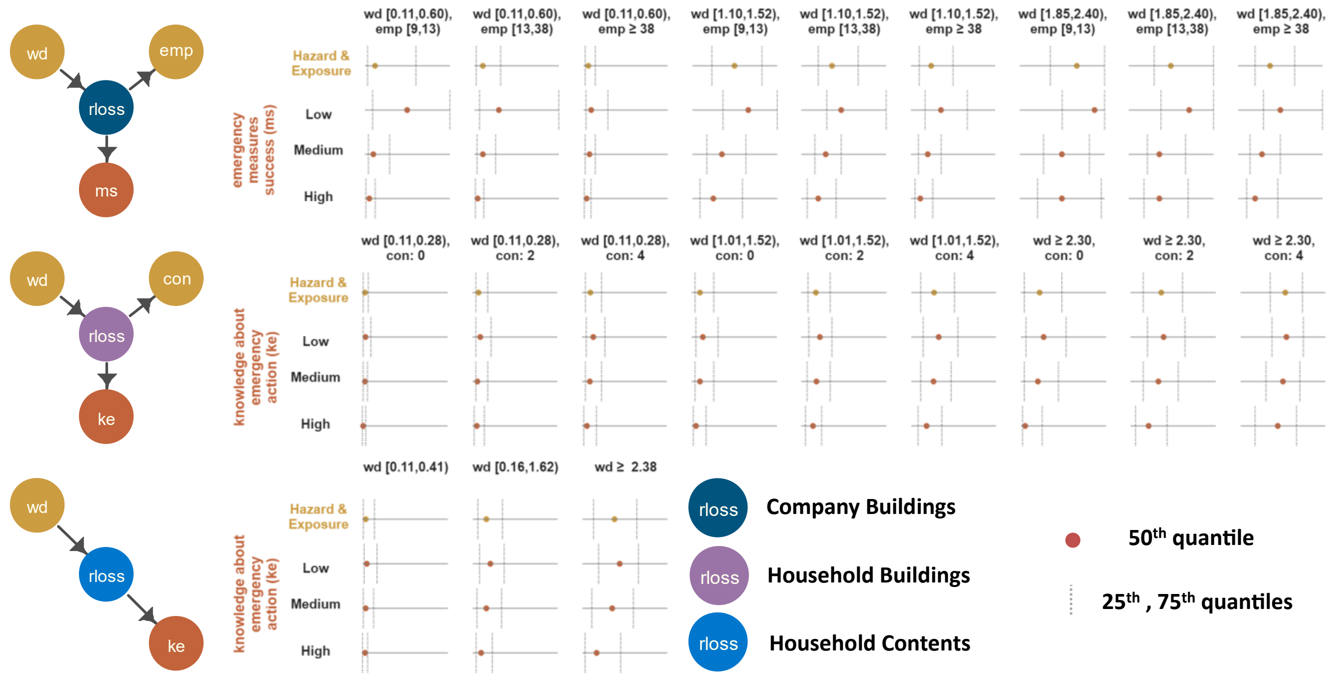

Figure 1. FLEMOflash application for company buildings, private household buildings and contents, considering relative loss Markov blankets. The first row in each panel shows the probabilistic predictive density of relative loss on the interval [0,1] based on the specific scenarios of hazard and exposure combination. The second to fourth rows in each panel illustrate the changes in relative loss with different levels of emergency measures success (ms) and knowledge about emergency action (ke) for the given hazard and exposure combinations.

Reviewer #2:

I would like to thank the authors for their thorough and timely response to my comments, as well as for the revisions made to the manuscript, which in my view provide an appropriate justification for the modelling choices adopted and the results obtained.

Nevertheless, I still have some reservations regarding the data underpinning the analysis and their representativeness with respect to the flash flood phenomena. While the use of the slope criterion allows for the inclusion of a larger number of data points, it may not necessarily ensure the quality or suitability of the dataset for this specific context. In addition, I remain uncertain about the inclusion of events related to heavy rainfall. The relatively low damage values estimated by the model (even in the absence of mitigation measures) seem to support these concerns. This aspect, however, does not preclude the publication of the article, since, as already mentioned, the manuscript clearly explains the modelling choices adopted.

We sincerely thank the reviewer for the constructive feedback and positive evaluation of our revised manuscript. We appreciate the reviewer for the insightful comments.

The slope-based approach was adopted as a simple and pragmatic criterion for point-based samples, allowing the inclusion of a broader empirical dataset while maintaining consistency across study sites. We fully acknowledge, however, that this criterion may select cases where the flash flood situation was not particularly severe. As highlighted in the conclusions, future developments should complement this approach with more physically based hydrological modelling and additional indicators. Specifically, we noted

P16/L365-369:

"*While FLEMO$_{flash}$ already provides a tool to support risk analyses, and impact-based forecasting, future developments could further strengthen its applicability by integrating complementary hydrological indicators (e.g., basin concentration time), incorporating building-level susceptibility factors (e.g., construction materials, structural condition, floor count), and expanding the empirical database by including high loss observations and more diverse geographic regions*."

In the methods section (P4/L100-101), we have already added "*Other metrics, such as river basin concentration time, may indeed provide a more process-based characterization of flash flood potential*." to highlight the limitation of considering only slope criterion.

We agree that we have used a rather broad definition of flash floods. However, the three flood events used in this study are already mentioned in section 2.1, of which one is a heavy precipitation event. We did not use rainfall as a criterion to distinguish flash floods, and only heavy precipitation events which qualified the slope criteria were used in this study. We have added a disclaimer to this effect in the methods section

P4/L103-104:

"*The cases with longer warning lead times in the sample are likely to be due to warnings of high precipitation than to flood-specific warnings*."

Nevertheless, I would suggest that the authors adopt a more cautious tone when discussing:

-(i) the model's ability to identify "the important variables and underlying processes that govern flash flood losses". In fact, the model essentially indicates that damages depend on flood intensity, exposure, and preparedness — aspects that are already well-established in the literature.

We agree with the reviewer in using a more cautious tone when discussing the models' ability to identify important variables. It is true that the model essentially indicates that damages depend on flood intensity, exposure, and preparedness which have been well-established in the existing literature. However, using a machine learning ensemble approach helps us do quantitative comparison of factors that influence loss, and select the most significant variables from a list of 20 variables considered in this study. This final selected list of variables is used for Bayesian Networks, which in turn help draw probabilistic dependencies between different variables.

In Section 3.1, we clearly highlight and acknowledge the role of existing studies stating that our results are in line with the previous findings. We have rephrased sentences that mention the "identification" of flash flood loss drivers as a novelty and highlight how the quantitative machine learning approach helps identify which of the variables are "more" important.

We have made the following changes in response to the suggestions made by the reviewer:

P3/L74-76:

"*We use machine learning based feature importance to select the most important variables from our dataset. The performance of FLEMO$_{flash}$ model is compared against conventional SDF models. Finally, we illustrate the loss processes with the CPT and Markov blanket in controlling the extent of loss reduction*."

P6/ L120-121

"*To derive the most significant drivers of flash flood losses from our list of variables, this study adopts a data-driven feature selection approach to the empirical data*."

P8/L 195-197:

"By quantitatively comparing the varied drivers of flash flood losses, these results also emphasize the importance of multivariable loss estimation models that capture the interplay across these drivers and their influence on losses."

- (ii) the predictive capacity of the model (for example, in the abstract and conclusions), given that the model relies on a limited number of variables and is based on data that are difficult to estimate in a predictive phase at the individual-item level (e.g., preparedness, number of employees, income, contamination, inundation duration).

As mentioned in Section 2.4.2, the model predictive capacity presented in this study pertains to an evaluation of model's performance with respect to available data and established metrics (e.g., MAE, CRPS, MBE). We acknowledge that these metrics reflect outcome constrained by available data and model choices.

As a word of caution, we have added the following part to the conclusion of the manuscript:

P16/L369-371:

"*It is important to note that the model's performance and predictive capacity, as presented are specific to the empirical dataset and survey variables available for FLEMO$_{flash}$, and the results should be interpreted within the context and limitations of the underlying data*."