Reviewer #1

R1.C1: *This paper provides a detailed description of the development of an expert based flood content damage model called INSYDE. It seems to be a follow up paper on the structure damage version of INSYDE, a model that seems to have found quite widespread use in the literature. The paper is well written and describes the development process well. The methods are solid but not very innovative and have been around in the grey literature for a long time (e.g. US Army Corps of Engineers). This paper goes in quite some detail describing the methods and adds much needed validation and is therefore definitely a useful addition to the scientific literature. That being said I have concerns about the validation results and more importantly the analysis of the results*.

Reply: We thank the Reviewer for the positive assessment of the manuscript. While we acknowledge the existence of various kind of damage models in the literature, we want to underline that INSYDE-content explicitly applies the synthetic approach to the context of content damage in residential buildings, a domain in which models are scarce. The proposed model was indeed designed following a component-wise, probabilistic and "what-if" logic, which builds on and extends the original INSYDE framework (Dottori et al., 2016). Its transparent and modular structure allows the users to clearly understand all assumptions and input variables, facilitating local adaptation based on available knowledge and data. In this respect, INSYDE-content overcomes limitations of both empirical models (typically fitted to specific case studies) and other synthetic models, which often function as black boxes without allowing modifications or explicit access to internal assumptions. While this aspect was briefly introduced in the original manuscript (L80-83), we will revise the paper to clarify and strengthen this point in the revised version.

R1.C2: *Figure 4 shows that for detached and semi-detached houses the variation in observed damages is much larger than the variation in predicted damages. My first impression is that the model always roughly predicts the same damage regardless of the circumstances (the blue dots are a nearly horizontal line). I think it may not be so bad because the log-log scale masks some of the variation. However, more information is required so readers can actually tell the model performance. For example, I currently cannot see if the variation in observed values is just based on some large outliers or whether there is some more fundamental problem whereby the observed losses have much more variation than the modelled losses. Also is there even any correlation between modelled and observed losses?*

Reply: We thank the Reviewer for the valuable comment. As also discussed in Molinari et al. (2020), claim data at the building scale are often affected by significant uncertainty and potential bias. For this reason, traditional flood damage models, as well as INSYDE-content, are typically more reliable when applied at aggregated spatial scales, rather than at the level of individual buildings.
This intrinsic variability in the observed data makes the interpretation of building-scale validation results particularly challenging. In our original manuscript (L368-374), we emphasized the importance of providing uncertainty ranges in the predicted damage values as a way to enhance the informative content of the model compared to purely deterministic approaches.
Nonetheless, we acknowledge the Reviewer's request for a clearer assessment of model performance. In the revised version of the manuscript, we will include standard error metrics (calculated based on the median of the predicted values per building, compared to the observed losses) to complement the visual inspection of the plots. However, we will also add a caveat to caution against over-interpreting these metrics, since observed values should not be considered absolute ground truth in damage modeling, due to the inherent limitations in claim data quality.
As also noted by the Reviewer, the apparent flatness of the predicted damage values for detached and semi-detached houses is partially explained by the use of log-log axes, which compress the visual perception of variability. Furthermore, from a theoretical perspective, a limited dispersion in predicted

losses is to be expected for these building types, given their relatively homogeneous characteristics and the shallow inundation depths recorded during the two flood events. In contrast, greater variability is observed for multi-family residential buildings, which generally exhibit broader heterogeneity in both exposure and vulnerability, leading to capturing a damage prediction variability of the same order of that for observed losses. In the revised version of the manuscript, we will therefore include comments on these aspects regarding the interpretation of the results shown in Figure 4.

R1.C3: *I understand that there is unexplained uncertainty in the model predictions as indicated by the uncertainty ranges in figure 4. However, if the model typically predicts more or less the same mean how do I know such a complicated model adds any value compared to a simple mean value as prediction?*

Reply: As also noted in our response to R1.C2, the apparent lack of variation in the predicted values is partly an effect of the log-log scale used in the plots, which visually compresses differences in magnitude. Moreover, for the two case studies presented, the variability in observed inundation depth was relatively limited, leading to a correspondingly limited spread in the predicted central estimates.
However, we believe that the added value of our approach lies in several aspects beyond the mean prediction. First, INSYDE-content is a probabilistic, component-based model that explicitly propagates uncertainty from input variables to output damage estimates. The resulting prediction intervals offer critical information to decision-makers, allowing them to understand the sensitivity of outcomes to input assumptions and to assess risk under uncertainty, an aspect that cannot be captured by a simple mean-based model.
Second, beyond estimating damage, the model provides a transparent and flexible structure to represent damage mechanisms explicitly. This feature allows, for instance, for scenario testing and the evaluation of building-scale mitigation strategies, which would alter specific input parameters and therefore result in different damage outcomes. This functionality makes INSYDE-content particularly suitable not only for risk estimation but also for supporting risk reduction planning.
In the revised version of the manuscript, we will better highlight these aspects, including a clearer explanation of how the probabilistic design and modular structure of the model enhance its applicability and usefulness beyond average damage prediction.

R1.C4: *Also very common error metrics are missing such as Mean Absolute Error, correlation coefficient or R2, so it's nearly impossible to assess how the model is doing from the information presented in the paper. Not all these metrics are needed but at least more information. Table 4 only gives an aggregated comparison, so basically gives a bias value. In one region there seems to be some bias but the authors do not really explain where this bias might be coming from. Lastly, I would expect an in depth analysis and discussion of the model performance in the paper based on the validation. That analysis is missing, making the validation not very useful in its current form.*

Reply: As discussed in our response to R1.C2, our original decision not to include commonly used error metrics was driven by the recognition that claim data at the building scale are themselves often affected by considerable uncertainty and potential bias. Consequently, they cannot be considered a definitive benchmark for validation purposes in the traditional sense. For flood damage models, quantitative validation against building-level claims must be indeed interpreted with caution, as it often reflects discrepancies not only in model performance but also in data quality and reporting practices (Molinari et al., 2020; Di Bacco et al., 2024). That said, we understand the Reviewer's request for a more detailed assessment. In the revised version of the manuscript, we will therefore include a set of standard error metrics, computed using the median predicted value per building, to provide a clearer quantitative basis for comparison with observed claims, and we will add a broader reflection in the discussion section on the limitations and interpretability of validation results in the context of flood damage modeling.

R1.C5: *Some of the input variables for the model validation seem sampled whereas others seem observed and the current text is very unclear about what is sampled and what is observed. This makes it even more difficult to interpreted the validation results.*

Reply: We agree that a clearer distinction between observed and sampled input variables can be useful for better understanding the setup of the validation exercises and interpreting the results. To address this, in the revised version of the manuscript we will include a summary table listing the model input variables used in the two validation case studies, clearly indicating for each variable whether it was directly observed or sampled from distributions.

R1.C6: *The word "to" in the title doesn't read well, maybe you can replace it with "for"? Or another solution.*

Reply: In the revised version of the manuscript, we will update the title to improve readability. The new title will be: "INSYDE-content: a synthetic, multi-variable flood damage model for household contents".