Response to Referee 2:

Below, we respond (**R**) to the more general and specific comments (**C**) of the reviewer. The changes in the revised text are mentioned in the response letter in *italics* along with line numbers referring to the revised manuscript (which is not uploaded yet). The revised manuscript includes one more figure and an additional subsection, hence the figure/section numbers of the preprint [*preprint: figure/section number]* are included in the response letter, when they differ between both manuscript versions.

General comments

C2.0: The manuscript addresses the topic of flood losses to microbusinesses in low- and middle-income countries and introduces an application of Conditional Random Forests for feature selection and Bayesian Networks to model flood losses, bridging gaps in existing methods that primarily cater to larger firms or macro-level analyses. In addition, it provides empirical evidence for key drivers of flood losses (e.g., water depth, building age, monthly revenue) and evaluates the models' transferability to another city (Can Tho), highlighting its potential for broader regional application. The approach is good in general; however, the manuscript should provide more details for readers to understand and reproduce this approach in other case studies.

R2.0: We thank the referee for taking the time and effort to provide comprehensive feedback on our manuscript. In the revised version of the manuscript, we have implemented almost all the suggestions, otherwise, an explanation is provided in this response letter. Due to the length of the manuscript, we did not explain the individual pre-processing steps applied on the survey datasets in detail, but rather provided a brief overview, as mentioned in **R1.3** in the first response letter. We had also removed parts of the methodology section (e.g. the flowchart) in the manuscript to keep it as short as possible, however, we included these parts again in the revised versions of the manuscript and the Supplementary Information (e.g. flowchart in Fig. S1).

Specific comments

C2.1: Provide more explanation about machine learning techniques of "Bayesian Network" and "Conditional Random Forest" and briefly explain why these methods suit the study.

R2.1: Indeed we have not explained the models in-depth due to the already rather large extent of the manuscript. As suggested, we have added brief explanations for both models and pointed out the suitability of the chosen methods compared to other ML-based approaches.

In the revised version of the manuscript, we modified and extended following explanations of Conditional Random Forests in Sect. 3.1.1 (lines 180:187):

“The set of candidate predictors presented in Table 1 exhibits a moderate to high degree of multicollinearity, for instance, the flood-related features are strongly correlated to each other. For this reason, Conditional Inference Trees were applied to account for these correlations during feature selection. Conditional Inference Trees were initially introduced by Hothorn et al. (2006) and extended by Strobl et al. (2007) to an ensemble of trees, a so-called Conditional Inference Random Forest (CRF). Each tree is grown only by a subset of features, which were identified before as significant based on their p-values (Hothorn et al., 2006). By this approach predictive features are identified, despite their potential collinearity to other candidate predictors. The choice of an unbiased version of the permutation-based feature importance method – namely Conditional Permutation Importance - further reduces the chance of biased importance scores for correlated features (CPI, Debeer and Strobl, 2020).”

In the revised version of the manuscript, we added following explanations to Bayesian Networks (BNs) in Sect. 3.2.2 (lines 210:216):
“They are better regional transferable than other ML-based models, such as regularized linear regressions, since BNs can be updated with new data and applied on incomplete information. Furthermore, they have the benefit of explicitly representing the dependency structures, quantifying uncertainty and the possibility of including expert knowledge alongside data. In more detail, the dependency structure of a BN represents (assumed) causal relations between variables, these dependencies can be set based on knowledge or logical conclusions. For example, the relation that less flood water infiltrates into a building if structural measures were taken beforehand is represented in the BN graph of Fig. 3 by a negative correlation (rho: -0.14) between both variables.“

However, the manuscript does not deal with the conceptual basis of Bayesian networks - the Bayes theorem - or other aspects of them. For a more in-depth understanding, we would like to refer for CRF models to the work of Strobl et al. 2007 and Levshina 2020; for non-parametric Bayesian Network flood loss models to the work of Paprotny et al. 2021. In particular, the benefits of BNs briefly mentioned in the manuscript (lines 211-212) - "e*xplicitly representing the dependency structures, quantifying uncertainty and the possibility of including expert knowledge alongside data*" - are described in more detail in the latter study.

CRF:

Levshina, N.: Conditional Inference Trees and Random Forests, in: A Practical Handbook of Corpus Linguistics, edited by: Paquot, M. and Gries, S. Th., Cham: Springer, 611–643, https://doi.org/10.1007/978-3-030-46216-1_25, 2020.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T.: Bias in random forest variable importance measures: Illustrations, sources and a solution, BMC Bioinformatics, 8, 25, https://doi.org/10.1186/1471-2105-8-25, 2007.

BN:

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Wagenaar, D., Castellarin, A., Carisi, F., Bertin, X., Merz, B., and Schröter, K.: A probabilistic approach to estimating residential losses from different flood types, Nat. Hazards, 105, 2569–2601, https://doi.org/10.1007/s11069-020-04413-x, 2021

C2.2: Case study selection: The manuscript justifies the choice of Ho Chi Minh City and Can Tho as case studies based on their high flood risk and economic importance. However, additional clarification is recommended (e.g., explain how these cities represent other flood-prone urban areas in Vietnam or Southeast Asia).

R2.2: This is a good point to further detail the choice of both cities in regard to their economic and topographic similarity to other flood-prone urban areas. The rapid economic developments in Southeast Asian states in the last decades are partly driven by the operations of micro-, small- and medium-sized businesses. The economic sector in urban areas of Vietnam, not only in HCMC and Can Tho, consists to a large share of those smaller businesses. The following sentence illustrates this aspect (Sect. 2.2.1 *[preprint: Sect. 2.1]*, lines 124-125) "The presented shares of the business sectors in the HCMC survey are representative for entire Vietnam (General Statistics Office 2018)" - HCMC is economically representative to other urban areas in Vietnam. Moreover, most of Vietnam's economic centres are located on flat terrain with access to the coast or a major river. Thus, they are similar flood-prone as HCMC and Can Tho, however, their type of flooding might differ (fluvial, pluvial, coastal). In the revised version of the manuscript, we adapted the following sentences to make the topographic similarity of HCMC to other urban areas more explicit (Sect. 2.1 *[preprint: Sect. 1]*, lines 79-81):
"*Similar to other SE-Asian metropolises, HCMC lies in a river delta area close to the coast. These densely populated, flat, riverine and coastal regions experience regular flooding in particular during the rainy season (Garschagen, 2015; Tierolf et al., 2021; Nguyen et al., 2021).„*

C2.3: Defining sample size: The manuscript mentions the number of surveyed microbusinesses but lacks a detailed rationale for the sample size determination (What statistical considerations or sampling techniques were used to determine the sample size?).

R2.3: In fact, we did not adequately explain the different sample sizes in the manuscript and have therefore added the following sentences to Sect. 2.2.1 *[preprint: Sect. 2.1]* (lines 126:129):
"In order to achieve a reasonable representation of HCMC, we selected the districts with the most frequent flood risk and also heterogeneity in socio-economic conditions. Within each district, the shophouses were chosen randomly. The sample size in each district was not chosen based on statistical considerations, but on recommendation from local experts."

C2.4: Ethical application for surveys: Ethical considerations for conducting surveys need to be explicitly addressed to ensure transparency and compliance with research standards.

R2.4: Thank you for the very crucial comment. The survey resulted in anonymous data from flood affected businesses. To connect the data of the interviews with flood characteristics, the location of the interviewed businesses was recorded. We decided to show in a GIS map (**R2.6**) only the rough geolocations of the microbusinesses to protect the anonymity of the interviewed businesses. The data stored and handled is conform to data privacy and data protection regulations.

C2.5: A flowchart summarizing the methodological workflow (from data collection to modeling and validation) can improve clarity for readers unfamiliar with machine learning or Bayesian methods.

R2.5: We moved the flowchart (Fig. S1) to the revised version of the Supplementary Information to not further extend the manuscript length. We linked the main steps, visualized by the flowchart, to the text in the data and methodology section.

C2.6: A GIS map that shows locations of surveyed microbusinesses can provide clear context and can be combined with flood hazard maps to give an overview of hazard and impact, which can be suitable for visualization and dissemination.

R2.6: We added a GIS map in the revised manuscript (Fig. 1) showing the rough locations of the microbusinesses surveyed in HCMC. The entire urban area of HCMC as well as suburban districts are located on low-lying terrain, thus, in combination with many open channels, the majority of HCMC's districts are facing a high flood risk. We tried to illustrate this aspect in the map by adding information about the low-lying areas alongside the rough locations of the microbusinesses. Besides the figure, we added the following text to the revised manuscript to explain this aspect (lines 134:137):

"*Figure 1 visualizes the rough locations of the microbusinesses surveyed in HCMC. However, their exact geolocations are not shown to protect the anonymity of the interviewees. Furthermore, the map shows that all surveyed microbusinesses are located on low-lying terrain often in short distance to a larger channel or tributary river.*"