EGUSPHERE-2024-1779 | Research article

Submitted on 12 Jun 2024

Multi-hazard susceptibility mapping in a karst context using a machine-learning method (MaxEnt)

Hedieh Soltanpour, Kamal Serrhini, Joel C. Gill, Sven Fuchs, and Solmaz Mohadjer

Handling editor: Mario Parise, mario.parise@uniba.it

We sincerely thank the two reviewers for the time and effort invested in evaluating our manuscript. We greatly appreciate their constructive comments and thoughtful suggestions, which have helped us improve the clarity, methodological precision, and overall quality of the study. We are also grateful for their positive remarks regarding the scientific relevance and contribution of our work.

In this document, we address each reviewer's comment point by point. For clarity, each comment is followed by our response and a description of the corresponding revisions made in the manuscript. Minor corrections (e.g., spelling, reference inconsistencies, and figure colour adjustments) are indicated directly as revisions. For comments requiring clarification or further explanation, detailed responses are provided. Please note that all line numbers refer to the **revised-tracked changes manuscript**.

1. Reviewer 1

General evaluation: The paper presents the methods adopted and implemented for machine learning of susceptibility mapping in a two component multi-hazard complex, using MaxExtent and on this basis I would suggest that the Scientific significance is Good to excellent - 2-1; the Scientific Quality is good, methods are clearly explained and reproducible, with the exception that the groundwater levels cannot be related to the ground levels (see Fig. 6 e depth to water on the fig and stated as groundwater level in caption, and datum for ground levels is not presented), which may also impact the interpretation and results, therefore my ranking would be Good to Fair 2-3; The Presentation Quality is Good (2) albeit two styles of writing are identifiable in the text, but both are readily interpreted by the reader.

Revision in manuscript: We confirm that the variable assessed in the model corresponds to depth to water table. Accordingly, Section 3.4.4 has been revised to clarify this definition, and the terminology has been standardised throughout the manuscript. All figures, charts, and captions (Lines 386, 404, 530, 703, 762 Figure 5 e , Figure 8e, Table 2, Table 3) have also been updated to consistently refer to "depth to water table" for clarity and transparency.

Comment: With respect to the presentation there are numerous cases of differences between cited reference dates and those reported in the reference list; not all references cited, e.g. Kim and Nam 2018, Pazzi et al., 2018; line 509 Radosavljevic and Anderson rather than et al.; are in the reference list; some references need correcting, e.g. Mokhtari or Mokhrarai.

Revision in manuscript: All in-text citations and reference list entries have been carefully cross-checked and corrected. Missing references (e.g., Kim and Nam, 2018; Pazzi et al., 2018) have been added, author name inconsistencies (e.g., Mokhtari/Mokhrari) have been corrected, and citation formats have been harmonised throughout the manuscript to ensure consistency with the journal guidelines.

Comment: There are some minor typos/ spelling issues, e.g. line 45 partially = partly, line 53 multi-hazard forming zones = zones that host multi-hazards? Line 80 forming landscape = landscape context? Line 174 the exposure of groundwater to rather than and easily eroded

...; line 178 bottom = base; line 232 Quaternary Alluvium; line 238 could be improved to ... which were mainly attributed to karst collapses ....; line 240 happened = occurred; line 250 for = to; Methods section - should this be written in the past tense?

Revision in manuscript: All indicated spelling and wording corrections have been implemented throughout the manuscript (e.g., “partially” revised to “partly”; “happened” revised to “occurred”; clarification of phrasing in Lines 45, 104, 181, 238, 248, and 274). In addition, the Methods section has been revised to ensure consistent use of past tense when describing the modelling procedures

---

Comment: Should the title for section 2 be edited to The multi-hazard environment of karst terrains? Line 591 should flattens out to be falls to? Line 607 should this read with increased thickness of alluvial deposits?

Revision in manuscript: The title of Section 2 has been revised to “The multi-hazard environment of karst terrains.” Minor wording improvements have also been implemented (e.g., Line 591 revised to “falls to”; Line 607 clarified to “with increased thickness of alluvial deposits”) to improve clarity and readability.

Comment: Figure 8 e again groundwater conditions need to be made clearer. Line 684 ... means of a classification ...Line 717 groundwater conditions need to be defined more clearly.

Revision in manuscript: In the revised manuscript, we have clarified throughout that the variable represents depth to water table, rather than general groundwater conditions. All relevant sections of the text (including Lines 684 and 717) have been revised accordingly, and Figure 8e (as well as other figures and captions where applicable) has been updated to explicitly state “depth to water table”. This clarification ensures that the variable is interpreted correctly as a measured vertical distance (in metres) rather than a general hydrogeological condition.

Comment: The content is relevant for NHESS and its international audience, and the title clearly reflects content albeit the word "the"might sensibly be changed to "a" as karst contexts can be very different. The abstract provides a full and clear overview of the content. Line 23 appears to be missing some words e.g. .... scenarios "in resilience procedures" ... or similar.

Revision in manuscript: The title has been revised to read “Multi-hazard susceptibility mapping in a karst context using a machine-learning method (MaxEnt)”. In addition, we made the incomplete phrase completed “Incorporating karst-specific multi-hazard scenarios *into resilience planning* processes supports disaster risk reduction efforts by raising the awareness of citizens, protecting elements at risk and facilitating decisions on disaster prevention. ”

2. Reviewer 2

Reviewer General Evaluation

In their article „Multi-hazard susceptibility mapping in the karst context using a machine-learning method (MaxEnt)“, Soltanpour et al. apply the Maximum Entropy model (MaxEnt) to characterize flood-triggered sinkholes in the Val d'Orléans karst region of France. The authors argue convincingly that karst terrains are underrepresented in the multi-hazard literature despite their inherent potential for hazard interactions. This topic is of clear interest and relevance to the readers of NHESS, and the approach of integrating official flood hazard maps as a contributing factor for sinkhole susceptibility mapping represents a meaningful methodological contribution. The manuscript is well-written, logically structured, and demonstrates careful conceptualization. I did enjoy reading it. In light of the comments below, I recommend publication after major revisions**. My main criticism is that the authors tend to overstate certain conclusions and claims throughout the manuscript, particularly regarding the interpretability of their model outputs and the validation of the model itself. I encourage the authors to adopt a more cautious tone in several places, which would strengthen the credibility of their findings.**

As a disclosure, while I have sound knowledge of karst systems and profound experience with ML applications in general, I have not personally applied the MaxEnt algorithm nor conducted hazard susceptibility assessments; my comments on these specific aspects are therefore to the best of my knowledge. Please also note, that I used assistance from Claude Opus 4.5 to structure my thoughts on the manuscript, generate text for this review and to identify minor corrections (wording, commas etc.). I bear full responsibility for the content of this review myself.

Reviewer General Assessment:

The authors identify elevation as the dominant predictor for flood-triggered sinkhole susceptibility (jackknife test, response curves). However, given the limited elevation range in the study area (82–126 m) and the consistent east-to-west decrease in elevation, I am concerned that elevation may serve as a spatial proxy rather than a causally meaningful predictor. The model may effectively learn „location within the study area“ rather than a true elevation-sinkhole relationship. I hypothesize that similar results could be obtained using distance-to-boundary or other location-describing features. This does not invalidate the susceptibility map per se, but it undermines the interpretability of conclusions regarding elevation as a driving factor. I strongly encourage the authors to discuss this limitation and alter statements attributing causal importance to elevation. In general, I would like to encourage authors to be careful of where causality can be inferred from the analyses and where it cannot. In my view, the data hardly allow any conclusions to be drawn about causality, but rather about presumed connections or correlations.

A similar concern applies to the groundwater response curve, where probability declines sharply beyond 15.5 m. To me, there is no obvious hydrogeological mechanism that would explain why erosion and dissolution processes would cease at higher groundwater levels. Could groundwater level act as a proxy for another underlying factor (e.g., subsurface structure, aquifer properties) that is not explicitly included in the model? Please discuss.

The authors state that one objective is „to evaluate the potential for the MaxEnt model to be used in multi-hazard mapping“ (Objective 3). While the application is successfully demonstrated, a true evaluation of MaxEnt's potential would require comparison with an established baseline method. The authors themselves cite Perrin et al. (2015), who applied weight-of-evidence in the same study area. Why not include this as a benchmark? Without such comparison, Objective 3 remains only partially addressed.

Section 4.1 „Model Validation“ is misleadingly titled. I would argue it is more of an evaluation than a validation of the model. The statement that „validation data includes both training and testing datasets“ (Line 529-530) is inconsistent with standard ML terminology and contradicts Line 492-495. Please clarify the evaluation and validation strategy in general. I recommend renaming this section (e.g., „Model Performance Assessment“) and clarifying the data split procedure. Please also explain whether AUC was calculated on the test set only (30%) or on combined data.

Authors Note to the Editor:

The concerns raised above - particularly regarding (i) the interpretability of elevation and groundwater depth as potential proxy variables, (ii) the distinction between correlation and causality, (iii) the scope of Objective 3, and (iv) the terminology and structure of the model evaluation section - are addressed in detail in the point-by-point responses below since the reviewer has repeated them in the specific comments . Where appropriate, the manuscript has been revised to adopt a more cautious tone, clarify methodological terminology, and explicitly discuss limitations.

Specific Comments:

Comment: Lines 63-79: Consider to remove or substantially shortening this part. A general elaboration on multi-hazards is not beneficial for the readability of this study. If you keep these paragraphs, use metric units (not inches).

Revision in manuscript: Revised accordingly. **"**30 inches" was changed to "76 cm".

Comment: Objective 3 – Baseline comparison (Lines 147-149): The authors state that one objective is „to evaluate the potential for the MaxEnt model to be used in multi-hazard mapping." While the application is successfully demonstrated, a rigorous evaluation would benefit from comparison with an established baseline method. The authors cite Perrin et al. (2015), who applied weight-of-evidence in the same study area and produced a susceptibility map. Why not compare MaxEnt performance against this existing approach? Without such comparison, Objective 3 remains only partially addressed, and claims about MaxEnt's suitability for multi-hazard mapping are difficult to substantiate.

Response: While algorithm benchmarking is one way to assess model performance in susceptibility mapping studies, Objective 3 was not intended to evaluate MaxEnt relative to alternative algorithms, but rather to assess its suitability for representing a multi-hazard susceptibility scenario in a karst context, specifically flood-triggered sinkholes. Our evaluation therefore focused on whether MaxEnt can **(i)** integrate different types of hazard data, including both continuous and categorical variables, **(ii)** produce stable and interpretable outputs within the data (e.g. response curves and variable importance), and **(iii)** produce a continuous susceptibility output, allowing fine distinctions in susceptibility between different areas.

Although Weight-of-Evidence represents an established approach for karst susceptibility mapping, it is based on different modelling assumptions and objectives. A direct quantitative comparison with a multivariate, regularised machine-learning framework such as MaxEnt is therefore not straightforward in a multi-hazard context. A direct comparison would require redefining the hazard, using the same predictors and spatial resolution, and aligning modelling assumptions across both approaches. Without this, any differences in results would reflect differences in problem formulation rather than the performance of the algorithms themselves.

We have clarified the wording of Objective 3 in the manuscript to make explicit that the evaluation refers to applicability and interpretability within a multi-hazard framework, rather than algorithmic benchmarking, and we acknowledge that comparative assessments against established methods represent a valuable direction for future work.

Revision in manuscript: objective (3) has now been revised in the "Introduction Section" to clarify that the focus is on exploring the applicability of the MaxEnt framework for representing a multi-hazard susceptibility scenario in a karst context, rather than evaluating its performance relative to alternative algorithms.

In addition, we have adjusted the "Conclusion" to ensure that the claims remain aligned with this clarified scope. We have also added a brief statement indicating that future work may explore comparative benchmarking against alternative modelling approaches under harmonised conditions.

Comment: Figure 1 (p. 6): I don't see any bidirectional arrows as indicated. I also think that some of the labels in the soil refer more to the limestone/bedrock and are therefore in the wrong place. Karst should also be better marked and labeled to make it clear that this is where the actual dissolution processes take place.

Response: The feedback mechanism was intended to be illustrated by the dashed arrows linking (4) secondary flooding back to (2) primary flooding. However, we agree that the bidirectional feedback is not sufficiently clear in the current version of the figure. We have revised Figure 1 to clearly show this feedback using clearer opposed arrows and improved visual distinction, and update the caption accordingly.

Revision in manuscript: The caption of Figure 1 has been revised to explicitly clarify the bidirectional feedback mechanism. The following sentence was added: "The dashed bidirectional arrow between stages (4) and (2) shows the feedback mechanism whereby initial flooding can exacerbate subsequent flooding events." In addition, the figure labels have been modified to clearly distinguish the soil/alluvial cover from the underlying karstified limestone.

Comment: Lines 492-495 (Data splitting): The authors use a 70/30 random split for training and testing. However, given the strong spatial clustering of sinkholes along the Loire River (58% within 1 km, Figure 4), random splitting may result in spatially proximate sinkholes appearing in both training and test sets. This spatial autocorrelation can inflate AUC values. Did the authors consider spatial blocking or leave-one-cluster-out cross-validation? I recommend acknowledging this limitation.

Response: : We acknowledge that sinkholes in the study area are strongly clustered along the Loire River, which means that a random train-test split cannot fully guarantee spatial independence between the two datasets and may lead to slightly optimistic AUC values. We ran the model 50 times to check that results were stable and not due to a lucky split. The AUC is moderate and consistent, which suggests a real spatial signal rather than overfitting. However, because the data are spatially clustered, we acknowledge that the AUC reflects within-area discrimination rather than fully independent predictive skill.

Revision in manuscript: A clarification has been added at the end of Section 3.5 stating that, due to the spatial clustering of sinkholes along the Loire River, a random 70/30 split does not fully ensure spatial independence between training and testing datasets. The manuscript now clarifies that the reported AUC reflects within-area discrimination rather than fully spatially independent predictive performance.

Comment: Section 4.1 „Model Validation“ (Lines 519-533): This section title is a bit misleading. True validation would require independent ground truth data (e.g., sinkholes that occurred after model training, or expert-validated susceptibility zones). The statement that „validation data includes both training and testing datasets“ (Line 529-530) is confusing and contradicts the earlier description of the 70/30 split (Lines 492-495). I recommend:

1. Renaming this section to „Model Performance Assessment“ or „Model Evaluation”
2. Clarifying whether AUC was calculated on the test set only (30%) or on combined data.
3. As I understand AUC in MaxEnt measures discrimination between presence locations and background points (pseudo-absences), not true presence-absence classification. Please include and discuss this aspect.

Response: We acknowledge that in much of the hazard susceptibility and risk-mapping literature, ROC-AUC derived from data splitting has traditionally been referred to as “model validation”. However, we agree that under a stricter interpretation, true validation would require independent or temporally distinct ground-truth data, which are not available for this study.

To avoid ambiguity we have therefore renamed this section to “Model Performance Assessment” and have clarified that AUC values are computed using the held-out test data (30%) averaged over 50 replicate runs. We also explicitly note that, in the MaxEnt presence-only framework, AUC quantifies the model’s ability to discriminate presence locations from background points rather than true presence-absence classification. This clarification have been added to improve transparency and to guide interpretation of model performance.

Revision in manuscript:
1. Section 4.1 has been renamed to “Model Performance Assessment”.
2. The description of the data split has been clarified to specify that AUC values correspond to the mean test AUC (30% withheld presence records) averaged over 50 replicate runs.
3. We also added clarification that, in the MaxEnt framework, AUC measures discrimination between presence locations and randomly sampled background (pseudo-absence) points rather than true presence–absence classification. Corresponding adjustments were made in the text and in the caption of Figure 6.

Comment: Lines 584-593 (Elevation response curve interpretation): This is one of my main concerns. The authors identify elevation as the dominant predictor and interpret the optimal range (90–105 m) as causally meaningful for flood-triggered sinkhole susceptibility. However, I am skeptical of this interpretation for the following reasons:

1. The study area has very limited elevation variation (82–126 m, i.e., only ~44 m range).
2. Elevation decreases consistently from east to west across the study area (as noted in Lines 394-395).
3. There is no compelling hydrogeological mechanism explaining why sinkholes should occur preferentially at 90–105 m elevation per se.

I hypothesize that elevation serves as a spatial proxy for location within the study area rather than a causally meaningful predictor. The model may effectively learn „areas in the central-western part of the valley“ rather than a true elevation-sinkhole relationship. I would expect similar model performance if elevation were replaced by „distance to eastern boundary“ or similar location-describing features.

This does not invalidate the susceptibility map, but it undermines the interpretability of conclusions attributing causal importance to elevation. I strongly encourage the authors to:

- Discuss this limitation explicitly.
- Temper statements such as „Elevation appears to be the main variable“ (Line 714) and similar claims throughout the manuscript.
- Consider testing whether a location-based variable (e.g., X-coordinate or distance to a reference point) yields comparable model performance.

Response: Given the relatively limited spatial extent of the study area, the influence of spatial autocorrelation and the use of variables (predictors) acting as spatial proxies cannot be entirely eliminated. As the primary objective of this study is to produce a susceptibility map identifying areas with different likelihoods of flood-triggered sinkhole occurrence, we fully agree that the data and modelling framework do not allow causal conclusions to be drawn. We have therefore modified the manuscript to avoid causal language and to explicitly acknowledge this limitation, emphasising that the results describe associations and presumed connections rather than mechanistic relationships.

Revision in manuscript: We have revised the manuscript throughout to avoid causal language and to clarify that the MaxEnt model identifies statistical associations rather than direct physical causation. Specifically, terms such as "driving factor" have been replaced with "most influential predictor" (e.g., Page 8, Lines 210–212). Expressions implying deterministic interpretation, such as "optimal elevation range" and "forecasts," were reformulated using statistically precise language (e.g., "highest predicted probability within the model"), and statements describing sinkholes as "very unlikely" were revised to indicate "lower modelled susceptibility" (e.g., Page 21, Line 586; Line 671).

In addition, a clarification has been added in the Discussion (Page 12, Lines 345–350) explicitly acknowledging that elevation may partly function as a spatial proxy variable given the limited elevation range and systematic gradient across the study area. This limitation emphasises that model outputs reflect relative spatial patterns rather than direct mechanistic causation.

Comment: Lines 609-615 (Groundwater response curve interpretation): The authors state that sinkhole probability „swiftly declines“ beyond 15.5 m groundwater level. This interpretation requires more critical discussion. From a hydrogeological perspective, there is no obvious mechanism that would cause erosion and dissolution processes to cease at higher groundwater levels. If anything, higher hydraulic gradients associated with elevated groundwater might be expected to increase erosion potential. Could GWL serve as a proxy for another factor not explicitly included in the model (e.g., aquifer properties, subsurface geology, or proximity to specific hydrogeological features)? I encourage the authors to discuss this possibility and avoid over-interpreting the response curve as reflecting a direct causal relationship.

Response: We acknowledge that a similar concern could arise for the groundwater variable and clarify that depth to water table in this study is not interpreted as a direct causal trigger of sinkhole collapse. In the Val d'Orléans floodplain, depth to water table integrates multiple hydrogeological characteristics - such as aquifer connectivity, conduit development, and interaction with the Loire River - that are difficult to represent explicitly at the regional scale. Its importance in the model therefore reflects its role as an indicator of favourable subsurface conditions for flood-related instability rather than a mechanistic driver. We have revised the manuscript to make this interpretation explicit and to emphasise that the results describe spatial associations within a susceptibility framework, not causal relationships.

Revision in manuscript: The interpretation of the depth to water table response curve (Lines 609–615) has been revised to remove causal language. The text now clarifies that depth to water table is treated as an integrative hydrogeological indicator within the susceptibility framework and may reflect broader subsurface conditions rather than a direct physical trigger of sinkhole collapse. Additional clarification has been added in the Discussion to emphasise that the results describe spatial associations rather than mechanistic causation.

Comment: Lines 603-609 (Alluvial thickness response curve): The response curve shows very low sinkhole probability where alluvial thickness is minimal (close to 0). The authors interpret this as consistent with the suffosion mechanism, which requires cover material. While this interpretation is reasonable for cover-collapse sinkholes, I would appreciate a brief discussion of whether solution sinkholes (without cover) could still occur in areas with minimal alluvium. Are there exposed karst features in the study area that might represent a different sinkhole type not captured by this model?

Response: The interpretation of low sinkhole probability at minimal alluvial thickness is indeed based on the dominant suffosion and cover-collapse mechanisms documented in the Val d'Orléans, where sinkhole formation primarily occurs within alluvial deposits overlying the karstified "Beauce" limestone (e.g., Perrin et al., 2015; Noury et al., 2018; Luu et al., 2019).

While solution sinkholes may locally exist in the study area, they are not the predominant sinkhole type captured in the inventory used for model calibration. In the revised manuscript, we have clarified this point and stated that the model reflects the sinkhole types most commonly observed in the study area.

Revision in manuscript: we have now clarified in the Discussion that the inventory is dominated by cover-collapse sinkholes in mantled karst settings and that the model primarily reflects these processes. We also state that solution sinkholes in areas of minimal alluvial cover are not the predominant type represented in the calibration dataset.

Comment: Lines 658-667 & Figure 9 (Jackknife test results): The jackknife test reveals that "Flood hazard zones" is among the weakest predictors when used in isolation (Figure 9). This finding appears to contradict the central premise of the paper that flooding is a key trigger for sinkhole formation. If flood hazard contributes minimally to the model's predictive power compared to elevation, groundwater, and alluvium thickness, what does this imply for the "multi-hazard" framing? (Be careful with causality here, as elaborated above). The authors acknowledge this partially (Lines 725-731), attributing it to spatial overlap with other variables. However, I encourage a more explicit discussion of whether this undermines the "multi-hazard" claim or whether it simply reflects that flood hazard is already captured by correlated variables.

Response: When considered on its own, the flood hazard variable has limited discriminatory power in a correlative MaxEnt framework, which explains its low jackknife contribution. This does not imply that flooding is irrelevant to sinkhole formation, but rather that its influence is expressed in combination with other conditioning factors. This issue again goes back to the fact that in the Val d'Orléans, flood hazard zones overlap spatially with other contributing variables (topography, depth to water table, geology and alluvium thickness). As a result, much of the spatial signal associated with flood-prone conditions is already captured by these conditioning variables.

This outcome does not question the role of flooding in triggering sinkhole collapses. Flooding acts primarily as a trigger, while the spatial occurrence of sinkholes is largely controlled by predisposing geological and hydrogeological conditions. For this reason, flood hazard alone has limited ability to distinguish sinkhole locations in a static susceptibility model, even though its influence becomes apparent when combined with other controlling factors. We believe that rather than undermining the multi-hazard framing of the study, this result highlights the distinction between triggering factors (flooding) and spatial susceptibility conditions (geo-environmental factors). We have revised the discussion to clarify this point and to avoid causal interpretation of variable importance.

Revision in manuscript: We have revised the interpretation of the jackknife results (Lines 848- 855) to clarify that the comparatively low standalone contribution of flood hazard zones reflects their limited discriminatory power within a static susceptibility framework. The Discussion now explicitly distinguishes between triggering processes (flooding) and predisposing spatial conditions (geology, topography, depth to water table), emphasising that spatial overlap among variables may reduce the isolated importance of flood hazard without undermining the multi-hazard framing of the study. Causal language regarding variable importance has been removed.

Comment: Lines 531-533 & 711-713 (Model performance vs. conclusions): The authors correctly characterize the AUC of 0.702 as "satisfactory" (Line 531), which indicates moderate discrimination ability. However, some statements in the manuscript appear to overstate the model's reliability (e.g., "the model forecasts the highest relative probability," Line 586-587; "Elevation appears to be the main variable," Line 714). I recommend that the authors temper these statements to align with the moderate predictive performance. Phrases such as "the model suggests" or "results indicate a tendency" would be more appropriate than definitive attributions.

Revision in manuscript: We have revised the relevant sections to temper deterministic language and ensure consistency with the model's predictive performance. Phrases such as "the model forecasts" and "optimal conditions" have been replaced with more cautious formulations (e.g., "the model suggests", "peak in predicted susceptibility", or "highest contribution to model performance"). We have also clarified that the results indicate relative susceptibility patterns rather than definitive causal relationships.

Minor Comments:

Comment: Figure 3: add "Arc" to GIS Pro

Revision in manuscript: "GIS Pro" has been corrected to "ArcGIS Pro" in Figure 3.

Comment: Figure 5: (e) is it relative or average depth? I assume the latter. Please correct.

Response: The variable represents the annual average depth to the water table. We have clarified this in the manuscript.

Revision in manuscript: The figure caption and the title within Figure 5e have been revised to read "Average depth to water table (m below ground surface)".

Comment: Figure 9 (p. 24): The color scheme (red, green, blue) makes the bars difficult to distinguish, particularly for readers with color vision deficiencies, but in its current form also for those without deficiencies. Please consider recoloring the bars.

Revision in manuscript: We have revised the colour scheme in Figure 9 by replacing the previously green “without variable” bar with grey to improve visual distinction and enhance readability.

Comment: Figure 10 (p. 26): The colors in the figure do not match the description in the text. Please verify and correct the color assignments in either the figure or the caption/text.

Revision in manuscript: We have corrected Figure 10’s caption to align the colour assignments with the bars shown in the figure (Blue for flood and red for flood-triggered sinkholes).

Comment: Line 64: “(Category 1)unleased“ → missing space before „unleashed“

Response/ Revision: Corrected.

Comment: Line 75: “800 houses(de Ruiter“ → missing space before citation

Response/ Revision: Corrected.

Comment: Line 111: Missing end of sentence: „…sciences. Bianchin …”

Response/ Revision: Corrected.

Remarks from the preceding review file validation: Figure 5:With the next file upload request, please improve the readability of the copyright statement.

Response/ Revision: The copyright statement in Figure 5 has been reformatted to improve its readability.

We hope that the revisions and clarifications provided satisfactorily address the reviewers' comments and improve the manuscript. We remain grateful for the constructive feedback and are happy to provide any further clarification if required.