Reply to Reviewer comments on Schulten et al., 2025, Scaling and exceedance probability of sinkhole collapse in evaporite karst (Dead Sea, Jordan)

Comments of Reviewer 1

The paper by Schulten et al. entitled “Scaling and exceedance probability of sinkhole collapse in evaporite karst (Dead Sea, Jordan)” presents new data on the scaling and probability of sinkhole formation in evaporite karst on the eastern Dead Sea shore (Jordan). Utilising high-resolution satellite imagery from 2017–2024, integrated with prior datasets, the research presents an updated inventory of sinkhole occurrence, extent, and dimensions through time over the period 1992-2024 for local geohazard and risk management at the eastern Dead Sea.

Overall, this is an appropriate subject area for the NHESS journal, and the developed research could help to improve the local geohazard and risk management at the eastern Dead Sea. The manuscript is well-written and developed even for inexperienced readers. However, I believe that minor improvements can be made, for example, by adding some aspects related to data processing, error assessment of outputs, and comparison with direct field surveys. Moreover, it is also required to add some details of previous works to have a complete understanding of the whole research. This work can be interesting and useful for the scientific community with some improvements, by better explaining the innovative aspects of the research and the insights it provides for improving the monitoring of sinkholes.

We are grateful for the constructive general comments offered by the reviewer and have addressed the specific comments below.

Line 49: Is “population” the most appropriate term to use? Maybe ‘site’ or “collection” would be better, or maybe I shouldn't put anything at all.

Response: The term ‘population’ as used here is appropriate; in statistics it simply refers to the entire set of observations. This term is also used in that sense when size-frequency analysis is applied to other geological domains, such as in fault studies (e.g. Walsh et al., 2002) or earthquake events (e.g. Farrell et al. 2009).

Farrell, J., Husen, S., & Smith, R. B. (2009). Earthquake swarm and b-value characterization of the Yellowstone volcano-tectonic system. Journal of Volcanology and Geothermal Research, 188(1-3), 260-276.

Walsh, J. J., A. Nicol, and C. Childs. "An alternative model for the growth of faults." *Journal of Structural Geology* 24.11 (2002): 1669-1675.

Line 266: Please define all acronyms (e.g., DSMs) in the text that appear for the first time. In Table S2, the term “DEM” is used, while the text refers to “DSM,” but they are not the same thing. Either the difference should be explained, or the text should be consistent.

Response: The reviewer is correct that DEM (Digital Elevation Model) and DSM (Digital Surface Model) are not synonymous. In our manuscript they reference two different datasets:

the DEMs listed in Table 2 are ~30 m resolution global elevation datasets used as an elevation reference for the orthorectification of the satellite scenes. The DSMs referred to in the main text are high-resolution (0.1m) surface models we generated from the photogrammetric surveys. Unlike the DEMs, these DSMs are not 'bare earth' elevation models, as they contain elevation data linked to areas of tall vegetation and to buildings. Therefore, we have defined both terms in Section 3.1, lines 267-270 of the revised manuscript and in the supplementary material in the caption of Table 2.

Line 268: Please briefly explain the workflow. The citation alone is not sufficient to ensure an understanding of the work done independently of previous work. In this way, the reader can better understand the workflow.
Response: We have added some additional context and citations (lines 281 - 284 of the revised manuscript) to allow the reader to more fully understand the workflow. The processing of the photogrammetric data to create digital surface models is explained in detail in an earlier publication, which is cited accordingly in the revised text.

Lines 274-278: It would be best to include this part of the explanation in the introduction to clarify things for the reader.

Response: We agree that this part of the text fits better earlier in the manuscript: indeed, it was somewhat superfluous as much of it was covered already in Section 2.2 (lines 175 - 187 of the original manuscript). Therefore, we have removed the entire first paragraph of section 3.2 in the original manuscript (previously lines 271 - 278) and partly incorporated it within Section 2.2 of the revised manuscript (lines 174 - 180).

Figure 13: The caption for Figure 13 is a bit long. It would be better to move some of the text to the discussion section.

Response: We agree with the reviewer and have therefore restructured the text such that the majority of the caption (lines 676 - 685 in the original manuscript) is now included within the manuscript body text (lines 682 – 694 of the revised manuscript).

Conclusions: should better highlight the innovative aspect of the research and its contribution to our understanding of the phenomenon.

Response: We have adopted the reviewer's suggestion and rewritten the conclusions section to be shorten ad more direct (lines 733 - 757). The revised text condenses the site-specific summary of results and is now ordered from the broadest implications to the local, comprising four main contributions:

1. Cover collapse sinkholes at the Dead Sea are a scale-bound rather than scale-free phenomenon, which resolves the outstanding power-law vs. log-normal disagreement in the literature and identifies the geometric and mechanical basis of that behaviour
2. This scaling enables the first exceedance-probability hazard assessment for the eastern Dead Sea shore.

3. Sinkholes at Ghor Al-Haditha present an evolving hazard to people and infrastructure, with formation continuing to migrate into new areas and to reactivate in areas previously thought dormant.
4. The study contributes to a wider reappraisal of the scaling of gravity-driven collapse phenomena.

Line 31 and line 41: Please avoid using double brackets and replace them with a semicolon in the manuscript.

Response: now fixed

Line 253: Ground Sampling Distance (GSD)

Response: now fixed

Line 601: “however.”??

Response: We have reordered the wording of this sentence to make it easier to understand.

Comments of Reviewer 2

The manuscript focuses on the phenomenon of collapse sinkholes at Ghor Al-Haditha, on the eastern shore of the Dead Sea, a process driven by evaporite karstification and accelerated by the anthropogenic lowering of the sea level. The work fills an important gap in the literature, until now concentrated on the western shore and divided in the statistical debate between scale-free (power law) and scale-bound (log-normal) size models. The paper also introduces as a novelty an updated inventory and a probabilistic hazard assessment.

To do this, the authors extend the existing catalogue to the period 1992-2024 through very high-resolution satellite imagery, mapping over 2000 sinkholes in over 30 years. The work analyzes the size distribution of the sinkholes both globally and by lithology (alluvium, mud, salt) and calculates the annualized exceedance probabilities for risk management along Highway 65 (a major road there). The main results demonstrate that the sinkhole population follows a log-normal distribution (scale-bound) and reveal a migration of the activity towards the north and towards the sea.

Overall, it is a solid, well-structured study supported by a dataset of great temporal and quantitative value. I compliment the authors. I have, however, noticed some margins for improvement, which is why I propose "accepted subject to technical corrections."

We are pleased that the reviewer considers our work to be of value. We thank them for their constructive comments, which we address below.

General comments

The images used cover GSDs ranging from about 2 m for "historical" images up to 0.3 m for the more modern ones. Observing Figure 7 (which has its start in 1992) I notice from about 2002 an almost linear increase in the Sinkhole count. I wonder if the effect of the resolution could in some way influence the reported growth trend and the different growth prior to 2004. Nothing wrong with all this, also because then you concentrate on the period 2004-2024, but I believe that a better discussion of this possible effect could benefit the manuscript.

Response: The initially non-linear increase in sinkhole count between 1992 and 2004 seen in Figure 7 is unlikely to be related to the lower GSDs of imagery from that period for two reasons. Firstly, the image resolution from 1992 to 2007 is consistently ~0.6 m (see Table S1). There is hence not a decline in image resolution over the relevant time interval. Secondly, equivalent data from western shore of the Dead Sea also show a non-linear sinkhole formation rate from the 1980s to the mid 2000s followed by a near-linear growth rate thereafter (cf. Fig. 4 of Yecheili et al., 2006; Fig. 5 of Avni et al., 2016, Figure S5 of Yizhaq et al., 2017). Mapping on the western shore has been primarily through repeated geological field surveys (Yizhaq et al., 2017), supplemented by manual assessment of aerial photographs and airborne LiDAR datasets (Yecheili et al., 2006; Filin et al., 2011; Avni et al. 2017; Abelson et al., 2017). Image resolution is therefore also unlikely to be a factor in accuracy of mapping on the western shore.

An alternative cause of the non-linear increase in sinkhole count from 1992-2004 is that we have undercounted sinkholes in this time period because evidence of their formation was erased. Such undercounting is also unlikely to explain the non-linear trend prior to 2004, however, for two reasons. Firstly, if we assume that the post-2004 linear growth rate of ~100 new holes per year was in fact present from 1992-2004, then we would expect a population size of over 1300 holes by 2004. This would require the erasure or mismapping of about 1200 holes. However, independent mapping of the Ghor Al-Haditha sinkholes from the late 1980s through the 1990s from aerial photos and fieldwork, as compiled in the report by Sawarieh et al., (2000), gives cumulative counts of 22 sinkholes by 1992 and 88 sinkholes by 1999. This includes several sinkholes that were infilled by local farmers. Our independent mapping via remote sensing data yields 18 sinkholes formed by 1992 and 58 sinkholes formed by 2000, which indicates some infilling-related undercounting in this period, but perhaps only on the order of a few 10's of holes rather than > 1000 holes. Secondly, as noted above, the same non-linear trend from 1980s to mid 2000s is seen on the western shore, including in areas of no agricultural development (e.g. the Ze'elim alluvial fan) where there is no incentive for erasure or infilling of sinkholes. Thus undercounting through erasure is also insufficient to cause the non-linearity of cumulative sinkhole count with time in Figure 7.

We have edited the related paragraph in the Results section to make the above points (Lines 473 - 479). Also, to be consistent in our focus on the post-2004, we have recomputed the distributions in Figs 8-10 to exclude the pre-2004 sinkholes. Since the excluded number of holes is < 4% of the total dataset, however, the impact of their exclusion on the reported distribution attributes is negligible.

Despite the monstrous work you have done for the manual mapping over time (compliments), I see in the dependence on manual work a possible criticality of the proposed method. Certainly a manual mapping is often more accurate than automatic or semi-automatic techniques, but it is obviously also subject to error and limits the reproducibility of the mapping. In the paper, a quantitative assessment of the reproducibility or the uncertainty is not reported (for example, a second independent operator who digitally remaps a test subset). A quantification of the mapping uncertainty, even if only qualitative and based on a selection of sinkholes per image in the most complex cases to map (for example presence of water/vegetation) would strengthen the analysis. Or, I suggest at least to discuss this point in the limitations.

Response: We agree with the reviewer that it is important to consider the variation in mapping between operators on the reproducibility of the mapping. The re-mapping experiment described in our response to the first general comment also addresses this limitation. Two operators independently mapped the same three zones, for two years, at three image resolutions, giving eighteen directly comparable cases per metric (Section S3, Figs S3 – S6).

The two operators agreed most closely on the number of sinkholes present, differing by an average of 11% in point count, and least closely on how those sinkholes were subdivided into discrete polygons, differing by an average of 20%. Differences in total mapped area averaged 16%. One operator delineated consistently more generous outlines than the other in 16 of the 18 cases, with a mean polygon area of 457 m² against 384 m². This same tendency explains why total mapped area does not decline with coarsening GSD as the counts do (Fig. S6). We suspect two effects to act in opposition here, where coarser imagery causes small sinkholes to be missed entirely, but causes those that remain visible to be delineated more generously, as edges are generalised and neighbouring depressions merge into single outlines.

Comparing the magnitude of the two sources of uncertainty is instructive. For sinkhole counts, varying the image resolution produced a median spread of a factor of 2.25, against only 1.10 between operators. For mapped area, the two were essentially equivalent (1.17 and 1.14 respectively). Uncertainty in mapped sinkhole area is therefore governed as much by operator delineation as by image resolution, whereas sinkhole counts are governed principally by resolution and are comparatively reproducible between operators.

We have attempted to mitigate the impact of operator variability in three ways. Firstly, the full time series was mapped by a single operator following a consistent protocol (as outlined in section 3.2), so that inter-operator differences cannot introduce artificial steps or trends into the temporal record. Secondly, our principal conclusions rest on sinkhole counts and size distributions rather than on absolute mapped areas, and counts are the more reproducible quantity. Thirdly, the systematic character of the area difference means it acts as a broadly uniform offset rather than as scatter, and so does not materially affect the shape of the size-frequency distribution on which our scale-bound interpretation depends. We have added a brief note on these limitations to the revised manuscript (Sect. 3.3, lines 382 – 393).

Specific comments

Lines 280-282: Here the theme related to manual mapping expanded in the general comments.

Response: see responses to general comments above.

Lines 403-407: the explanation of R is a bit too synthetic.

Response: we have added some text (lines 418 - 421 in the revised manuscript) to better explain how loglikelihood ratios (*R* values) are derived.

Line 733: Replace “XX”

Response: now fixed

Please re-check if all references are well cited and complete

Response: now fixed