Reviewer #1

Wilke et al. "Hail events in Germany; rare of frequent natural hazards" manuscripts presents the first study attempting to provide an in-depth comparison of MESH and VII for hail events in Germany and the first paper globally to look at recalibrating MESH from reports for C band radar. These results are of significant interest for the natural hazard and weather radar community; however, I wasn't able to complete a review due to many (minor) copyediting issues throughout the text, which made it difficult to focus on the science content. I'd strongly encourage the authors revise the manuscript (e.g., after a copyediting and structure review) and resubmit asap.

Here are some suggestions for the abstract

Abstract

line 2: "crowd sourcing" should be "crowd sourced"

line 4: "out of 6 years radar data" should be "using 6 years of radar data"

line 4-5: "over whole Germany" should be "over the whole of Germany"

line 5: The sentence "The size of hail depends heavily on the storm..." isn't clear to me? What properties of the storm are in question

abstract in general: The abstract at present doesn't read well, it feels more like a collection of short dot points rather than a summary of the paper. Please look at rewording it to improve the readability. It should provide an engaging summary of the motivations and findings of the paper, written to tie everything together.

Answer before revisions:

“Thank you very much for your thoughtful review of our manuscript. We sincerely apologize for the minor copyediting issues that hindered your review process. Your feedback is very helpful, and we appreciate your recognition of the importance of our study to the natural hazard and weather radar community. We acknowledge that the abstract currently lacks clarity, and we are committed to improving it by incorporating your insightful suggestions. Additionally, we will carefully revisit the overall structure of the manuscript to improve its readability and coherence.
Thank you again for your constructive feedback. We look forward to implementing these revisions and resubmitting our manuscript for your consideration.”

Revisions made based on authors response:

We have completely rewritten the abstract to provide a more concise and comprehensive overview of our study, its methodology, and key findings. We have revised the overall structure of the manuscript to improve clarity and flow of information. We have now incorporated both the case study and the survey as integral parts of the results section. This integration aims to present a more cohesive narrative of our research findings.The language throughout the entire manuscript has been carefully reviewed and refined to enhance readability and precision. By restructuring the content and improving the language, we believe the manuscript now reads better and presents our research more effectively.

We hope these changes address your concerns and improve the overall quality of our manuscript. We appreciate your time and expertise in reviewing our work and are open to any further suggestions you may have.

Reviewer #2

The manuscript “Hail events in Germany, rare or frequent natural hazards?” presents a strong and innovative analysis of hail events in Germany, utilizing advanced radar techniques and a diverse set of data sources. However, its impact is limited by a relatively short timeframe, over-reliance on radar data without sufficient validation or correction mechanisms, and a somewhat superficial analysis of auxiliary data sources such as crowd-sourced observations and insurance claims. There is significant potential for improvement.

Strengths:

1. Comprehensive Data Integration: The study stands out for its integration of diverse data sources, including radar, crowd-sourced reports, and insurance claims. This multifaceted approach provides a well-rounded perspective on hail events in Germany, ensuring a more complete understanding than relying on any single data type.
2. Innovative Use of Radar Technology: The application of advanced radar methods like MESH and VII demonstrates the authors’ technical proficiency. By using modern radar data to estimate hail sizes and occurrences, the study pushes the boundaries of traditional meteorological research.
3. Detailed Case Study and Real-World Application: The case study of the August 2021 hail event effectively highlights the strengths and limitations of crowd-sourced data, providing practical insights into how well lay observations compare with radar measurements. This adds a valuable real-world dimension to the analysis.

Directions for improvements:

1. Short Timeframe and Lack of Trend Analysis: The six-year period (2018–2023) used in the radar analysis is too brief to establish meaningful long-term trends. As hail events vary significantly year-to-year, a longer dataset or a more in-depth discussion of the limitations imposed by the short timeframe would enhance the study’s credibility.
2. Over-Reliance on Radar Data with Limited Corrections: While radar data is central to the study, its known issue of overestimating hail sizes is acknowledged but not adequately corrected. This over-reliance, without stronger validation or adjustment methods, weakens the conclusions and leaves room for potential inaccuracies.
3. Superficial Treatment of Crowd-Sourced Data and Insurance Claims: Though crowd-sourced data and insurance claims are included, the analysis does not fully explore their potential biases (e.g., urban reporting bias) or offer solutions to mitigate them. The insurance data, in particular, is not sufficiently explored for regional or structural factors, making this section feel underdeveloped relative to the overall scope of the study.

Answer before revisions:

“Thank you for your thoughtful review of our manuscript on hail events in Germany. We appreciate your insights and the opportunity to clarify and enhance our work based on your feedback.
We acknowledge that long-term trend analysis is a valuable aspect of meteorological studies, but we agree with the reviewer that six years of data is not sufficient for a trend analysis. Therefore, a trend analysis was not the primary goal of our research. Our focus was on utilizing advanced radar techniques to analyze recent hail events. However, we recognize the importance of understanding trends over time. In the revised manuscript, we will include an outlook discussing potential data sources that could be leveraged for future trend estimation, such as historical weather records and climate models. Additionally, we are committed to reanalyzing the data continuously as new data will be made available.
We are aware that radar data has its limitations, particularly regarding overestimation of hail

sizes. To address this, we undertook our own calibration of the MESH (Maximum Estimated Size of Hail) methodology as part of our study. This calibration effort aims to improve the accuracy of our hail size estimations. We will add a paragraph about how other studies deal with these uncertainties to the introduction e.g. the empirical correction done by Brook et al. (2024) and threshold-based optimization derived by CNNs by Forcadell et al. (2024).
Brook, Jordan P., et al. "A Radar-Based Hail Climatology of Australia." Monthly Weather Review 152.2 (2024): 607-628.
Forcadell, Vincent, et al. "Severe hail detection with C-band dual-polarisation radars using convolutional neural networks." EGUsphere 2024 (2024): 1-43.
We truly appreciate your insightful suggestion to explore the implications of bias in crowd-sourced data within our analysis. To enhance this aspect, we plan to provide a clearer comparison between population density and crowd-reported observations. We recognize that our reliance on insurance data, which is derived solely from postal code areas, may not provide a complete picture of hail events (see Figure 1). This limitation is why we chose to leave spatial analysis out of our examination of insurance data. By focusing primarily on larger hail events, we may inadvertently overlook occurrences of smaller hail, which are equally significant."

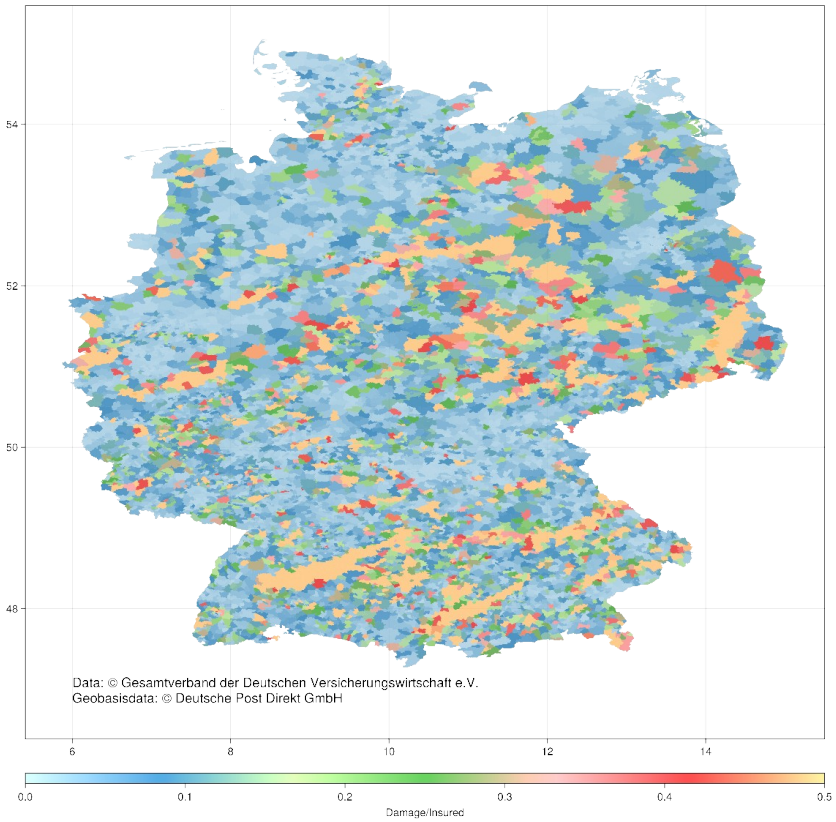

Figure 1: Loss [€] / Insured Value [€]

Revisions made based on authors response:

1. Short Timeframe and Lack of Trend Analysis

We have addressed the concern regarding trend analysis by explicitly clarifying that such analysis falls outside the current study's scope. Additionally, we have introduced a discussion highlighting the

potential of modelled data for trend analysis, noting that these time series are significantly longer compared to our radar data time series.

„Overall, the analyzed period of 6 years is not long enough to define a trend and was out of scope of this study. For trend analysis modelled data (Battaglioli et al., 2023; Wilhelm et al., 2024) might be a better fit as long as other consistent time series are too short.“ (p. 23, l. 342 ff.)

2. Over-Reliance on Radar Data with Limited Corrections

We added a new paragraph to the introduction specifically addressing hail size corrections for C-band radars. This addition provides context on existing methods for correcting hail sizes derived from radar data. We have explicitly stated that our primary reason for developing our own power law fit is to overcome the need of corrections. This clarification helps to justify our methodological choices and demonstrates the thoughtful consideration behind our approach.

„Originally developed for S-band radars in the US, Brook et al. (2024) showed that MESH used with C-band radars tends to overestimate hail sizes. To overcome this issue, Brook et al. (2024) introduced an empirical correction based on the matching of S-band and C-band radars in overlapping regions. Forcadell et al. (2024) utilize convolutional neural networks (CNNs) to obtain MESH values due to threshold-based optimization. In this study, we undertook our own calibration of the MESH formula with hail reports compared to the SHI derived from German C-band radar values (Section 4.5).“ (p. 2, l. 47 ff.)

3. Superficial Treatment of Crowd-Sourced Data and Insurance Claims

We have carefully enhanced Figure 11 (here Figure 2) to highlight the aspect of bias in crowdsourced data. The revised figure now clearly demonstrates that hail observations are more concentrated in areas with high population density. To explicitly show this connection, we normalized the number of observations against population data. Through this normalization process, the previously observed hotspots were effectively erased, allowing us to conclusively state that an urban reporting bias exists in hail observations. This methodological approach provides a more nuanced understanding of the spatial distribution of reported hail events.

Regarding the insurance data, we have included a clarification explaining why we did not conduct a spatial analysis. This explanation provides transparency about our methodological choices and helps readers understand the constraints and considerations in our research approach.

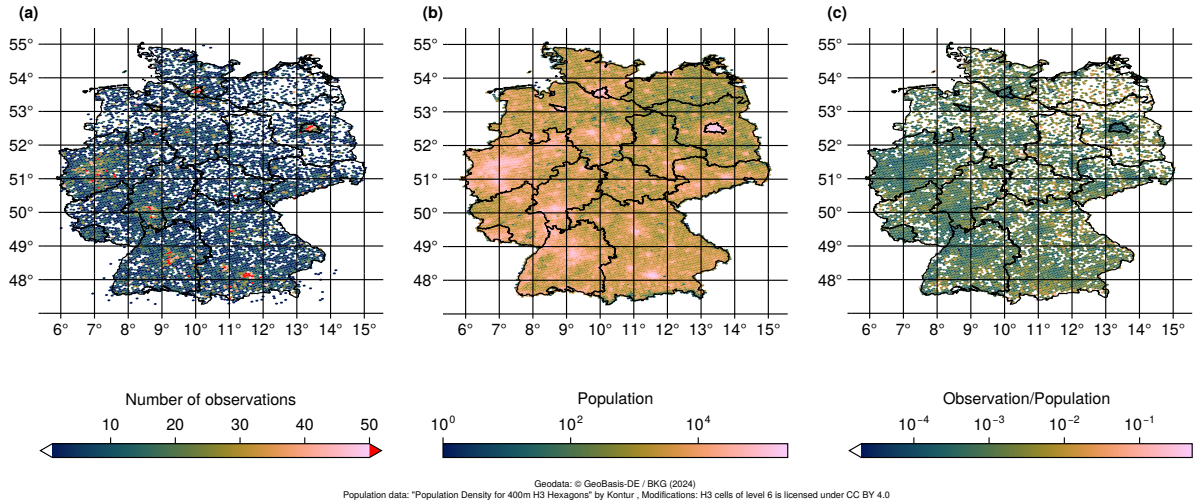

Figure 2: (a) Number of crowd observations in Germany 2001–2023 based on data from station observations, ESWD and WarnWetter-App. (b) Population in Germany (c) Normalized number of observations to the population

„The distribution of hail reports in Germany is displayed in Fig. 10(a) in a h3 grid of level 6 (Uber Technologies Inc., 2018). There are some hotspots visible in the south of Bavaria (close to Munich), in the middle of Baden–Wuerttemberg (close to Stuttgart), North Rhine–Westphalia in the Ruhr valley, Hessia in the Rhine–Main region and Berlin. In the north of Germany there are only a few hail reports. All areas with hail hotspots have a high population (see Fig. 10(b)). To examine whether the observations underlay an urban reporting bias, Fig. 10(c) shows the number of observations in a cell, normalized to the population. No such hotspots are visible any longer, only the Pre–Alps have a slightly higher number of observations than the rest of Germany. Some cells are more noticeable because of station observations that have been in one place for a long time and therefore reported hail for each day if apparent. In conclusion, the absence of hotspots in the normalized Fig. 10(c) clearly emphasizes the existence of an urban reporting bias, as there is no longer a visible difference in hail occurrence over Germany." (p. 13, l. 240 ff.)

The data is derived solely from postal code areas and may therefore not provide a complete picture of hail events. This limitation is why we chose to leave spatial analysis out of our examination of insurance data. By focusing primarily on larger hail events, we may inadvertently overlook occurrences of smaller hail, which are equally significant. So we focus on the temporal analysis of hail occurrences of insurance data. (p.4, l. 125 ff.)