Review Climatology of Large Hail in Europe: Characteristics of the European Severe Weather Database

The paper presents a statistical analysis of large hail reports from the ESWD for Europe. Analyses performed for 120- and 20-year periods include time series of reports and hail days, diurnal and seasonal cycles, annual distributions of hail sizes, and trends in temporal accuracy. Additional emphasis is given to reports from Poland, which has very high numbers of reports in some 10-year periods since 1930.

In general, the paper is well written and clearly structured.

Thank you for your kind words on our manuscript. We appreciate the time that you spent to provide these comments to improve our manuscript.

However, I have some **major concerns**, mainly about the quality and reliability of the data the analyses are based on, and the scientific content.

1. It is difficult for me to see new scientific results and profound conclusions that provide new insights into hail statistics. The paper is of course nice to read, but the scientific value seems to be low.

   We disagree with this assessment. This manuscript analyses more European hail reports than any other study. The closest comparator contains 39,537 reports for a 13-year period. We show the value of QC0+ data in the ESWD. We also examine how these reports have changed over time. The manuscript also documents the large addition of Polish data in the last century. All of these contributions justify the publication of this work.

   Also, the conclusion section is more or less a summary rather than a presentation of conclusions and interpretations.

   Indeed, we wrote the conclusion to summarize the manuscript. This approach is an entirely acceptable method for concluding a scientific manuscript. According to the instructions of the journal (https://www.natural-hazards-and-earth-system-sciences.net/submission.html), all that is specified is that each submission contain a conclusions section. At least some scholars of the scientific communication process argue that a conclusions section should contain no new information (e.g., Geerts 1999; Schultz 2009, p. 44) or allow for the possibility of the type of conclusion written here (e.g., Glasman-Deal 2021, p. 245). No change to the manuscript.

   Geerts, B., 1999: Trends in atmospheric science journals: A reader's perspective. *Bull. Amer. Meteor. Soc.,* **80,** 639–651.

   Glasman-Deal, H., 2021: *Science Research Writing for Native and Non-native Speakers of English.* World Scientific, 356 pp.

   Schultz, D. M., 2009: *Eloquent Science: A Practical Guide to Becoming a Better Writer, Speaker, and Atmospheric Scientist.* American Meteorological Society, 412 pp., https://doi.org/10.1007/978-1-935704-03-4.

2. As far as I understood from the manuscript, the authors considered all ESWD reports in their analyses, irrespective of multiple reports from a single storm, or the country or region affected. Given the large differences of prevailing reports among European countries (as shown in the Table), it can be assumed that the results are dominated by individual countries (e.g. Germany, Russia, Poland), leading to large uncertainties in all estimated quantities.

   We are not exactly sure what the reviewer's point is here, so if we are mistaken, we apologize. Indeed, some countries will have a greater number of reports per storm than others, irrespective of urban-reporting and daytime-reporting biases. This has previously been identified by several authors (e.g., Groenemeijer and Kühne 2014; Punge and Kunz 2016; Antonescu et al. 2017; Púčik et al. 2019) and has been alluded to in the manuscript. We will, however, make this point clearer in section 2 to fully communicate the biases present in the dataset.

3. In the same sense, hail reports are biased towards daytime and towards larger cities. This effect is difficult to estimate, but at least a profound statement is required (although a spatial analysis with respect to the distance of reports to larger areas may help to get an estimate of the latter effect).

   See our response to the previous comment.

4. Is a hail day one with at least one report across Europe (that would make no sense), or have you considered some threshold? For example, is a day with only one 2 cm report considered the same as a day with thousands of reports and hailstones larger than 10 cm? That would be strange.

   This study aims to look at the ESWD as a whole. Therefore, yes, the entirety of Europe has been considered. Indeed, one hail day is one where at least one hail observation of 2+ cm has been reported. The concept of a hail day is similar to that of a lightning day or tornado day, concepts that are well accepted in the severe weather community. In principle, hail days should be more robust to these fluctuations in reporting individual hail reports, which is why we have showed this quantity. No change to the manuscript.

   Furthermore, it makes no sense to define a hail day for the whole of Europe, with its wide variety of local climates. I would rather suggest limiting it to countries with a high number of reports, for example. I would also suggest considering different thresholds for both hail size and number of reports.

   The point here is not to look at individual days, but to put this in perspective on the continent over a longer time period. In principle, hail days should be more robust to these fluctuations in reporting individual hail reports, which is why we have showed this quantity. Thus, the hail-day concept is exactly intending to solve the problem the reviewer identified. The reviewer's proposed solution makes a rather simple concept of a hail day into a much more complicated matter. Furthermore, the reviewer's solution of concentrating on countries with higher reports would not remove any variability within climates. No change to the manuscript.

5. Point 4 also refers to the other analyses, such as the annual and diurnal cycles. It is mentioned that Púcik et al. (2019) divided the study area into at least two parts due to the different climates. Why did you not follow this?

   Although we agree that Europe encompasses many climates, how to divide these up can occur in numerous ways. We chose not to classify different climatological zones,

in part because of this ambiguity and in part because this was beyond the scope of the research. For example, one may choose to differentiate between a more maritime or more continental climate, but these may then contain other factors such as mountain ranges or plains. Hence, we decided to stick to a general overview of the reported distribution of large hail in Europe. No change to the manuscript.

6. A climatological period is usually defined as 30 years or more. It also includes spatial analysis. Neither is the case in this paper. Therefore, I suggest changing both the title and the wording in the manuscript.

   From the *Glossary of Meteorology*, *climatology* is defined as "The description and scientific study of climate. Descriptive climatology deals with the observed geographic or temporal distribution of meteorological observations over a specified period of time. Those climatological data can be averaged over 30 years to produce climatological standard normals." (https://glossary.ametsoc.org/wiki/Climatology). Thus, a 30-year period is only relevant for defining climate normals and is not a factor with climatologies of weather events. Also, the geographic distribution is not required for a climatology. Thus, our study fits perfectly with the accepted definition of a climatology.

   Furthermore, *NHESS* commonly publishes climatologies of weather events that are not 30-year periods.

   18 years: Gatzen, C. P., Fink, A. H., Schultz, D. M., and Pinto, J. G.: An 18-year climatology of derechos in Germany, Nat. Hazards Earth Syst. Sci., 20, 1335–1351, https://doi.org/10.5194/nhess-20-1335-2020, 2020.

   15 years: Burcea, S., Cică, R., and Bojariu, R.: Radar-derived convective storms' climatology for the Prut River basin: 2003–2017, Nat. Hazards Earth Syst. Sci., 19, 1305–1318, https://doi.org/10.5194/nhess-19-1305-2019, 2019.

   10 years: Pacey, G., Pfahl, S., Schielicke, L., and Wapler, K.: The climatology and nature of warm-season convective cells in cold-frontal environments over Germany, Nat. Hazards Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/nhess-2023-39, in review, 2023.

   10 years: Akkoyunlu, B. O., Baltaci, H., and Tayanc, M.: Atmospheric conditions of extreme precipitation events in western Turkey for the period 2006–2015, Nat. Hazards Earth Syst. Sci., 19, 107–119, https://doi.org/10.5194/nhess-19-107-2019, 2019.

   For these reasons, we disagree with the premise of the comment. No change to the manuscript.

Additional **minor review points** are those:

1. L21: 20-years is not a climatological period (major comment 5)

   We disagree. See our response to the previous comment. No change to the manuscript.

2. L30: "Large hail" for a diameter of > 2 cm is not a European definition, rather used by ESWD.

Thank you. Deleted “in Europe”.

3. L44-45: You may add that most of the hail climatologies / statistics (e.g., those cited in Touvinen et al., 2009) are outdated

   What the reviewer means by “outdated” is unclear and unfair to these studies. Indeed, some of the studies mentioned were published a number of years ago as implied by our statement of ““A summary of *past* European hail climatologies”. However, this does not mean their results are necessarily outdated. Moreover, readers would understand that a study published in 2009 is representative of the time in which it was published and of the dataset from which it was derived. Therefore, we disagree with the premise of this comment. No change to the manuscript.

4. L50: It should be noted here that some pan-European hail hazard assessments are available, e.g. from Punge et al. 2014 or Punge et al. 2017 based on overshooting top detections, from Rädler et al. 2018 using reanalysis, or from Taszarek et al. 2018 using multiple data sources. In this sense, the statement in L60 “…their work shed the first light on” is not true.

   Thank you for this clarification. Indeed, previous climatologies do exist, but are not based on the ESWD data as a singular entity. We have clarified this by putting “from surface reports” at the end of the sentence.

   Moreover, comparing these other climatologies with our results will be discussed in the results sections, per a comment by Reviewer 3. We believe that these revisions will also address this present comment.

5. I miss a better motivation and scientific objectives of the paper. “Increasing the size of the dataset through…extending the period of analysis” is too weak when only 2 additional years are considered.

   This comment is unfair. The reviewer has selectively edited this sentence to misrepresent what we actually wrote in the original submission.

   “In the present article, we explore whether increasing the size of the dataset through *lowering the quality-control levels of the reports* and extending the period of analysis yields comparable results, increasing the generality of Púčik et al.’s (2019) results.”

   So, our analysis was also about adding cases through lowering the quality-control levels of the reports, not only extending the time period. These two changes resulted in an increase in the number of reports from 39,537 (Púčik et al. 2019) to 62,053 (present study), a 57% increase in the size of the dataset.

   But, our study is about more than just increasing the size of the dataset. We also had different purposes to Púčik et al. (2019), which again were not mentioned by the reviewer.

   “In doing so, we also document the reporting characteristics of the database as a function of time both throughout the 20th century and within the last 20 years. In particular, we seek the possible existence of a relatively homogeneous period of time in the database that could be used as a baseline for climatologies and climate-change studies.”

Thus, we feel that we have clearly stated our motivation and scientific objectives, despite the manipulated and truncated quotation provided by the reviewer. No change to the manuscript.

6. P3, 2nd paragraph: Why did you not use the most recent data until 2022? The analyses seem to be easily reproducible.

   This manuscript was a result of an undergraduate dissertation (see the acknowledgements). The study commenced in late 2020. The data was sent from the ESWD, so the dataset was set as of late September 2020. Further analysis was not necessary. Nevertheless, we have added "at the time this study commenced" in section 2 to make it clear to other readers that the scope of the dataset was determined at this time.

   L98-105: This is the correct designation of the quality levels; in the later text they are incorrectly quoted.; L106: "…plausibly checked QC1…", but this is "report confirmed"

   Deleted "plausibly checked". Thank you.

7. Also in L225-226 it should read "report confirmed"

   Added "confirmed". Thank you.

8. L157: "…ability to detect reports linked to the same event, and hence have removed duplicate events from the dataset". This would make no sense at all and is not the case. In the papers cited (e.g. Wilhelm et al., 2020) it is clear that a single streak is covered by several reports.

   The point made here is that fewer reports have been needed for the same quantity of hail days over recent years than previously. Therefore, we are just speculating a few reasons for this. No change to the manuscript.

   Just a small correction to note in this comment: The citation should be Wilhelm et al. (2021), not (2020).

9. L77-79: Kunz et al. (2020) estimated annual and diurnal cycles not from ESWD data, but from radar-derived potential hail streaks (Z > 55 dBZ). These streaks were also combined with ESWD reports. The main difference is not the quality level of the ESWD reports considered because as written in Sect. 2, 70.4% were QC1 and 29% were QC+, leaving only 0.6% at Q0 level.

   We presume this refers to lines 177–179 where we cite Kunz et al. (2020), not lines 77–79.

   Thank you for this clarification. We have revised the sentence to the following:

   "These distributions are also similar to those from Kunz et al. (2020, their Fig. 2a) for hailstorms in central Europe using radar-derived hail streaks combined with all quality levels from the ESWD, indicating that this dataset derived using different methods is a reliable source of large-hail data."

10. L188: Can you briefly describe how you converted UTC to LT?

All reports have the country of origin listed and all times are in UTC. By looking at each country on an individual basis, these were converted to LT taking daylight savings into account. No change to the manuscript.

11. L191-192: see comment (9); Although the diurnal cycles of Kunz et al. (2020) have a resolution of only 3 hours, there are some differences, which may be due to different study areas?

    In fact, Figure 4 (local time) in the present manuscript if converted to a bar chart and Fig. 2b in Kunz et al. (2020) are quite similar. Sure, small differences will be due to different study areas and different years, but we don't see that. No change to the manuscript.

12. Fig 7: This figure is very interesting, but again not very valuable for the whole of Europe (and the under-reporting in most countries). I suggest that this type of figure be reproduced for countries where the number of reports is highest according to the Table.

    We find that this figure remains interesting by showing that there is not that much variation in the peak hail time across Europe, even between different climatic zones and countries. However, we do see the value in adding a table showing the proportion of hail days per year by country. We also believe that a figure showing the annual distribution of hail reports per country could be interesting, as a more even spread would suggest more consistent reporting over the years.

    L248 and L300: Did you use the Pearson product-moment or Spearman rank correlation coefficient? The latter would be more appropriate due to the obvious deviation from a normal distribution.

    We used the Pearson product-moment. However, the reviewer is correct and the Spearman correlation would make more sense. We will amend this.

13. L287-289: The main reason for the high number of reports in Germany is obviously that ESSL was founded here.

    Yes and no. The ESWD grew out of other data-collecting efforts such as TorDACH (tornado dataset from Germany, Austria, and Switzerland). So, although there was a focus on Germany, it was not strictly limited to the founding of ESSL. No change.

    It should be mentioned that in some countries severe weather reports are collected by other institutions, e.g. KERAUNOS in France. Moreover, crowd-sourcing via meteo apps is well known and emerging in some countries, such as the MeteoSwiss app, which has collected >100,000 reports in recent years (compared to only 266 ESWD reports). So we should not blame spotters for being less enthusiastic.

    The wording as written is precise. There are two factors in play here, and our text is clear in both of those factors. Storm-spotter networks may be more or less enthusiastic about collecting reports within their own countries ("existence, size, and enthusiasm of spotter networks within each country"), and such networks may vary in how effective they are at contributing those reports to the ESWD ("variations in the ability or enthusiasm of citizens to input into the ESWD"). No change to the manuscript.

14. L315 and others: I'm not sure about the comparability with the study by Suwala (2011), as they used station data over a period of 8 years. Station data often do not distinguish

between hail diameters, but rather consider ice pellets or graupel in the same class as hail (see also the review by Punge and Kunz, 2018). This could at least explain the discrepancies with the large number of hail events in the cold season. However, this is frankly speculative, as there is no information available for the Polish station.

This is a fair point. We have now added some text to clarify this point. “Although these results may indicate a cool-season preference for hail, there is the possibility that ice pellets or graupel might have been classified as hail (e.g., Punge and Kunz 2018).” Thank you for this information.

15. L324: Again, a separation by region would be desirable.

    Please see our response to point 5.

16. L374-377: The trend directions are not that clear. Eccel et al. (2012) or Manzato et al. (2023) found no positive trends in hailpad data in northern Italy, but fewer and larger hailstones. Dessens et al. (2015) found almost the same in their hailpad data in France. You may cite here the review paper by Raupach et al. (2021).

    Thank you. We have added a sentence to the manuscript. “Other studies have also concluded that there were no positive trends in the frequency of hail in hailpad data in northern Italy (e.g., Eccel et al. 2012; Dessens et al. 2015; Raupach et al. 2021; Manzato et al. 2023).”

Edits/Typos:

L18: “…dataset for severe convective storms **reports**.” Otherwise it’s not true, as there are SCS statistics available from model data or overshooting top reports (see minor point X)

Fixed. Thank you.

L20: “…to evaluate **hail reports** from… “ (you did not evaluate the database)

Fixed. Thank you.

L38: “on 10 June”; you refer here to the supercell that hit the city of Munich on that day.

Now reworded to: “Other similar events occurred over southern Germany on 10–12 June 2019, with one storm producing 6-cm hailstones and causing EUR 1 billion in damages”.

L43: “…intensity, and hailstone size.” On L23, you wrote “intensity, as measured by maximum hail size..”, but here you both intensity and size.

“Intensity” has been deleted.

L54: “..which help**s**..”

Fixed. Thank you.

L83: delete “insurance data information”: in the ESWD data, I see only 3 entries in 20 years that are from an insurance company; this is not worth mentioned here

So, it is not a lot of information, but it is present in the ESWD. So, our statement is correct. No change to the manuscript.

L84: delete “organizations”; a large number of reports are not from organizations rather than from trained (and well-known) spotters

Fixed. Thank you.

L93 “…also examin**ed**…” above and below you used past tense

Fixed. Thank you.

L215: “…frequency of **events**…” for one event, the frequency is = 1;

Fixed. Thank you.

L218: “…**is** more spherical…”

We disagree. “Were” is being used in the past subjunctive tense and is being used correctly. No change to the manuscript.

L224: decreased --> decreases

We disagree. This verb is best as past tense, consistent with our interpretation of the data in the figures, which occurred in the past. No change to the manuscript.

L273: delete from

Fixed. Thank you.

L298: include a comma after Fig. 2

We do not agree that a comma is needed here. No change to the manuscript.

L302: implies --> imply

Fixed. Thank you.

L330: mentions --> mentioned

Fixed. Thank you.