

This study explores some characteristics of hail in Europe based on hail reports from ESWD. A distinction is made between hail days and hail reports. The data are summarized in graphics. The seasonality and diurnal cycle are described and data with two different quality flags compared. The time evolution of the reported hail stone sizes is illustrated.

Thank you for these comments. They will help us improve the manuscript.

One major point that I would like to raise is that the paper does provide only very limited information about the data in the data base, the quality check procedures, and the methodology that is used to analyse the data. This information is crucial for the interpretation of the data and the discussion of the limitation and uncertainties of the data set (more detailed information on that point is listed below).

We agree that some of that information is needed to understand this study. Where we felt that information was needed or relevant to our arguments or results, we have included that information. However, there are many other documents that contain the procedures and dataset information in much more detail than we can provide. For more information on the functioning of the ESWD, please refer to any of these documents: Groenemeijer et al. (2009, 2017), Dotzek et al. (2009), and Groenemeijer and Liang (2020).

The second point is that the analyses are qualitative. This is ok but there should be no statements about changes over time in the abstract without underlying statistical analyses.

We certainly appreciate the importance of quantitative analyses where relevant and possible. But, we disagree with the rigid statement “no statements about changes over time...without underlying statistical analyses”. Certainly, some signals can be so strong or so weak that statistical analyses are not needed to confirm what is visually apparent, especially with over 60,000 hail reports. We address each statement individually in the comments below.

Major:

- Please provide (a lot) more information about the data sources of ESWD. What are all the data sources of ESWD (e.g. does it also contain insurance data)? How big is the fraction of each data source (e.g. crowd-sourced vs. observers from weather services)? How do the different data sources change over time? This information is important as the uncertainty crowd-sourced data and insurance data is quite high

How large is the fraction of each data source in each quality class? How exactly are the quality classes assigned? What does plausibility checked mean exactly? A cross-check against radar information? A cross check against newspaper reports? Which data source provides typically, which types of variables (e.g. mean and max. size of the hail stones).

Please include all of this information in the methods section.

This level of detail is tedious and unnecessary for most of what we are doing. Other published articles that have used the ESWD have not had to answer these types of questions to the same level of detail. These published articles do not contain this information, even if such answers were available.

We agree that some of that information is needed to understand this study. Where we felt that information was needed or relevant to our arguments or results, we have included that information. However, there are many other documents that contain the procedures and dataset information in much more detail than we can provide. For more information on the

functioning of the ESWD, please refer to any of these documents: Groenemeijer et al. (2009, 2017), Dotzek et al. (2009), and Groenemeijer and Liang (2020).

Hail days: Please provide more information in the methods section how you identify hail reports and hail days?

The reports come directly from the ESWD as individual records in a spreadsheet. Therefore, we do not 'identify' these ourselves. How hail days are computed is already described in the text: "The annual number of large-hail days was derived from the annual number of large-hail reports by removing duplicate dates." We just counted up the number of unique dates in the dataset to obtain the number of hail days. No change to the manuscript.

How can you have more hail days than hail reports (Figure 1)?

There are never more hail days than hail reports per year. The scales are on different axes: left y axis for reports and the right y axis for days. No change to the manuscript.

Hail events: Please explain how you remove duplicate dates. How do you define an event? What counts as a duplicate date? If the report is exactly at the same location? In the same country? How much time difference do you allow for? How accurate are the report locations?

Duplicate dates are simply when more than one hail event is recorded across Europe on the same day, regardless of location. When we considered hail days per country, then the same procedure applies.

Time accuracy: If I understand it right, this information is self-declared? Has it every been verified against independent data (radar, satellite information)? How is this information obtained for historical data? Please expand the discussion of this variable to include these aspects in the methods section. How reliable is time information generally in crowd-sourced data?

This information is added by the ESWD data manager at ESSL. It is not self-reported. It is unclear how to verify this value as we never have the true time of the event down to the second. The time accuracy quantity is just a statement of how precise the reporting time of the event is. For example, if the newspaper story and photo that confirms the hail event says 15 local time, then the time accuracy is one hour. That is all. No changes to the manuscript.

Location accuracy: how is this parameter estimated? How is it verified?

Same as the previous comment. No changes to the manuscript.

2) How do your findings compare to hail climatologies based on radar/satellite and proxy indicators?

This is a great suggestion. Comparing our results to those of previous studies are further discussed in section 8, among other places in the text where they are most appropriate.

We believe that the following papers would be relevant to our findings.

Sanchez, J.L., Merino, A., Melcón, P., García-Ortega, E., Fernández-González, S., Berthet, C. and Dessens, J., 2017. Are meteorological conditions favoring hail precipitation change in Southern Europe? Analysis of the period 1948–2015. *Atmospheric Research*, 198, pp.1-10.

This paper discusses the variability in hail days from different climates in France and Spain over the time period 1948–2015. They highlight that only small spatial variations have a large impact on the number of hail days recorded over the time period. Furthermore, the paper emphasises that different climates have different peak hail months within the year. Therefore, we believe that an annual distribution chart per country could also be an interesting addition to the manuscript.

Rädler, A.T., Groenemeijer, P., Faust, E. and Sausen, R., 2018. Detecting severe weather trends using an additive regressive convective hazard model (AR-CHaMo). *Journal of Applied Meteorology and Climatology*, 57(3), pp.569-587.

This study uses reanalysis data from 1979 onwards, but also compared them to ESWD reports. This highlights that observed events play an important role even for other methods of researching severe events to ensure that results are plausible.

Taszarek, M., Allen, J., Púčik, T., Groenemeijer, P., Czernecki, B., Kolendowicz, L., Lagouvardos, K., Kotroni, V. and Schulz, W., 2019. A climatology of thunderstorms across Europe from a synthesis of multiple data sources. *Journal of Climate*, 32(6), pp.1813-1837.

This study looks at different datasets in order to investigate severe storms in Europe. They argue that by using different datasets, and therefore data collection methods, you can compare the spatial and temporal resolutions of these. This highlights the importance of understanding the strengths and weakness of different datasets, as well as understanding their composition, which our study provides from the ESWD at a lower quality-control level than previously published.

Taszarek, M., Brooks, H.E., Czernecki, B., Szuster, P. and Fortuniak, K., 2018. Climatological aspects of convective parameters over Europe: A comparison of ERA-Interim and sounding data. *Journal of Climate*, 31(11), pp.4281-4308.

We also believe that this study could make an interesting comparison, as it investigates annual distribution by region, with respect to the underlying meteorological factors that provide ideal environments for severe storm production.

Minor points:

Abstract: The intensity as measured by \diamond larger hailstones are rarer than smaller hailstones.

Of course they are. Nevertheless, the maximum hail size is often the measure of the intensity of a hailstorm, as commonly accepted by the meteorological community. No change to the manuscript.

Introduction: there is a body of literature that discusses hail climatologies in Europe based on indirect observations (radar, satellite data) and proxy indicators (soundings, reanalysis). I recommend a qualitative comparison of the results with this body of literature and hence a brief discussion of this research branch in the introduction.

Good suggestion. The third paragraph of the introduction discusses those past climatologies. The comparison to the results of the present study are further discussed throughout the text where they are most appropriate and in section 8. See suggested studies in point 2.

L86 please add: at the time of this study ...

Yes, fixed.

L106 please mention if these levels are inclusive, i.e. are all QC1 also included in Q0+?

No, these are separate categories. Each report only has one category. We have revised the manuscript to be more clear, referring to a “single” quality-control level and being explicit about which categories were included: “Půčik et al. (2019) used only plausibly checked QC1 and QC2 events” and “this present study uses QC0+, QC1, and QC2.”

L106 please mention that the quality control is strictly against reports from ESWD with a higher quality flag not with other data sources.

Indeed. Fixed. Please see our response to the above comment.

L179 “This data set” is referring to QC0+ correct? Please state so explicitly.

Correct. We have added “this larger dataset including QC0+ events...” to clarify.

L205

No comment was provided here, just a line number. It might be that there was a missing “and” in this sentence. We have fixed that. Thank you.

L225 pausibly \diamond plausibility?

The wrong QC description was given here. In fact, this report was “confirmed”, and we have changed the wording accordingly to “confirmed”. Thank you.

L234 does not change dramatically \diamond please compare distributions and check for stat. significance

Figure 7 has very little variability in the data year to year. This fact can be seen visually by readers. If there is no visual indication of a reasonable amount of variability, then there will be no reason to test for statistical significance. No change to the manuscript.

L237 a period of stability \diamond of what?

Stability in reporting. Specifically, on average over this time period, roughly the same percentage of each bin size were reported. We have revised the text to include “in reporting”.

Figure 7: Can there be a bias towards larger hail stones in crowd-sourced data? People being eager to report large hail stones? How does it compare to ground observations from e.g. hail pads?

The biases in the hail-size reports have been documented in other publications. Indeed, as likely as it is that maximum hail sizes may be overestimated, maximum hail sizes may also be undersampled. All the data in this manuscript (as with all other hail studies using the ESWD) will have the same issues (we prefer the word “issue” rather than “bias”). There are few studies that use direct measurement from hail pads. Even so, it is unclear how such studies could be applied to our study. So, while we do not disagree with the reviewer, we have no way of rectifying these issues. No change to the manuscript.

L276 This statement is a bit patronizing, I recommend removing it.

We have rephrased this sentence to “Figure 10 also indicates the countries for which there is opportunity to improve engagement in severe-weather reporting.”

L281 suggest to add “People in countries ...”

Revision is unnecessary as it refers to countries as a combined effort by spotters and other organizations (including the national hydrometeorological services) in each country, not an evaluation of personal contributions by any one person. No change to the manuscript.

L287 Note that some countries such as Switzerland and Germany have national hail crowd-sourcing programs organized through the National weather services that might explain why there are fewer entries to ESWD

Germany has the most large-hail reports (4956) and Switzerland has among the least. So, it would seem difficult to argue that such programs are influencing the number of reports. We would feel uncomfortable making such an assertion, given our results. No change to the manuscript.

Section 7: It is not entirely clear why you dedicate such a detailed analysis to the data from Poland.

Figure 1 shows some unusual behaviour in the 1930s, 1940s, and 1950s. Further analysis of those unusual maxima in hail reporting results in Figure 12, which shows that these maxima are entirely a result of Polish data. This is the premise of our argument in the first paragraph in section 7. Not mentioning these unusual maxima would be inappropriate.

However, recognizing that we have a large number of reports from a different century, it would be interesting to compare their characteristics to the characteristics of the modern dataset from the ESWD. This comparison is the reason for Figure 13, as well as the rest of the text in section 7. We believe it is a worthwhile inclusion to the manuscript. No change to the manuscript.

L353 the reported location accuracy

Included “reported”.

L359 I do not yet understand how you come to this conclusion about hail trends.

“Hail trends” could have been better worded for clarity. We have revised these words to “distributions of hail size, frequency, and location”.

L377 consistent = homogeneous? If not, what do you mean exactly by consistent?

“consistent” has been deleted.

L396ff These statements need to be supported by statistical analyses

See our previous argument. No change to the manuscript.

L399 the reported time accuracy

Adding “reported” is unnecessary. “Reports” or “reported” is already used three times in this sentence. The implication is clear. No change to the manuscript.