Review for NHESS-2020-138, Revision 1

General Comments

The authors have made significant revisions to the paper based on the reviewers' comments, including provision of technical details on the French and German radar datasets and an assessment of the Froude number on hail days over the Massif Central using reanalysis data. I would like to thank them for these efforts. However, some of the new material was very difficult to understand (particularly the details of the radar tracking algorithm), while other parts have raised further questions that need to be addressed. I have a large number of additional comments; however, I expect most of these to be fairly simple to address. As such I am recommending further minor revisions.

Specific Comments

L19-20: Hail can also be produced by single-cell "pulse" storms; however, large hail is almost always associated with organised convection, particularly supercells (Smith et al. 2012; Wapler et al. 2016)

L23-24: A decrease in θ_e with height indicates that the atmosphere is potentially unstable. According to Markowski and Richardson (2010), potential instability is not generally considered to play a role in the destabilisation of the atmosphere that precede convective initiation. The key instability for deep, moist convection is conditional instability, which exists when θ_e* (the equivalent potential temperature that the environment would have if it were saturated) decreases with height. However, I don't think you need to get into these definitions; just note that conditional instability is needed for lifted parcels of air to become positively buoyant.

L28-30: Other studies you might mention here include Brooks et al. (2003), Johnson and Sugden (2014), and Taszarek et al. (2017).

L31-39: Since the Spanish Plume isn't mentioned again I don't think you need to discuss it in detail here.

L67-69: I know I requested you add the definition of overshooting tops (OTs), but it rather breaks up the flow of this sentence. Also, this definition is relevant to the visual identification of an OT, but in satellite imagery these features are typically identified based on infrared brightness temperature signatures. As such I would suggest getting rid of the definition and instead including something like the following "...such as overshooting tops (an indicator of strong convective updrafts) in satellite imagery…"

L70-71: The statement "the link between the observed quantities and hail occurrence at the surface is less reliable than using radar measurements" doesn't really make sense and in any case only really applies to satellite measurements, not model data. I would recommend making a more general statement in this sentence about these approaches, then noting the specific limitations of satellite- and model-based hail proxies in the subsequent sentences.

The key benefit of radar proxies over overshooting tops is that the former is a more direct measure of hail within a storm (presence of large reflectivities), whereas the latter just indicates the presence of a strong updraft that might produce large hail.

L73-75: I think there is an opportunity here to emphasise the unique nature of the hail climatology you present and, in doing so, better link these two paragraphs. You could note that the advantage of satellite- and model-based hail proxies is that they can cover a wide geographic area, whereas radar-based climatologies are typically limited to a single country or region. However, yours is, I believe, the first study to combine radar observations from multiple countries.

L96-98: Does this mean that, for a given time, a CAPPI was taken from each radar and these were composited onto a common grid by taking the maximum value at every grid point? If so, at what altitude are the CAPPIs?

L101-113: I don't think you need to mention specific projects such as Panthère or RHyTMME, though you could add the relevant references (Tabary, 2007; Beck and Bousquet, 2013) to the end of the opening sentence of this paragraph. All you really need to note here is how many radars were present in the network and when these were replaced or upgraded.

L117-119: It looks like complete coverage of Luxembourg is only provided by one of the German radars. Perhaps this discussion of the inclusion of Belgium and Luxembourg should be moved to the opening paragraph of section 2.1, since it relies on both French and German data.

L122-123: The dual-pol variables aren't used so no need to mention them here.

L129-130: It's not obvious to me how a VPR (vertical profile of reflectivity) could be used to account for beam broadening with range. My understanding is that VPR corrections are used to extrapolate measurements taken aloft to the surface. This issue is worse at long range because the beam height increases (due to both the non-zero elevation angle and the curvature of the Earth); however, this has nothing to do with beam broadening, which tends to lead to an increase in partial beam filling with range.

L135-141: Is this discussion of QPE and associated quality control relevant since you are using reflectivity data only? Or do you take the corrected rain rates and convert them back to reflectivities? If not then all this detail can probably go.

L154-155: As noted in my original review, these regions are covered, according to Fig. 2, but only by a single radar. In your reply, you say "Looking at long-term radar composites, the far north place near the Danish border and the southeastern part of Bavaria have no reflectivity values; we rather see some values near the location of the radar stations." This leaves two possibilities. Either the German radar composite excludes pixels that are covered by only a single site or both these regions happen to be affected by beam blockage. It would be good to know what the reason is.

L160: Is orographic shading actually corrected for (e.g. by interpolating from higher tilts) or are pixels that are affected by it simply masked or labelled as low quality?

L160-161: Again is the conversion from Z to R relevant, since you're only using reflectivity?

L169-170: How do you get a "terrain-following near-ground reflectivity"? Are VPRs used to extrapolate values to the surface? Also it doesn't make sense to say CAPPI here since CAPPI stands for "constant-altitude plan position indicator". Here the "constant altitude" typically refers to above radar level (ARL) or above sea level (ASL), rather than above ground level (AGL), which is what you'd have if the data really are "terrain-following".

L204: You should say 0.25° resolution rather than 31 km as that is what you used (as I can see from Fig. 4). I think 31 km refers to the underlying Gaussian grid from which the 0.25° products are derived.

L216-217: I'm not sure what you mean here; please rephrase.

L224: How are you defining "high reflectivity" here?

L236-239: I would suggest using Z_{MAX} and Z_{RC} to indicate the maximum reflectivity in an ROIP and the threshold for RCs, respectively.

L239-240: Presumably this means that if an RC's maximum reflectivity briefly (say for one or two time steps) drops below 55 dBZ it will be treated as two separate convective cells, is that correct? If so, this is an important limitation and should be noted.

L241-242: Change "surface area" to just "area" as the former implies a 3D object. Also, rather than "radar bins" I presume you mean pixels? In which case, this quantity should have units of km^2, both here and in Table 1. I would suggest adding symbols to represent these variables; maybe A_{RC} for the minimum RC area and A_{hail} for the minimum area with reflectivities ≥ 55 dBZ. Finally, can you provide any justification for the inclusion of these two criteria? It makes sense to filter out very small cells as these could be spurious, but is there much sensitivity to the specific choice of thresholds?

L251: FAR = b / (a + b) is the false alarm *ratio*, whereas the false alarm *rate* (also known as the probability of false detection) is defined as F = b / (b + d). It seems this mistake was present in Puskeiler et al. (2016) but must have been missed by that paper's reviewers. You could maybe add a note to this effect: "...the false alarm ratio (FAR; incorrectly labelled the false alarm *rate* in Puskeiler et al. 2016) was 0.4..."

L253-254: Change "The algorithm assigns the RC of the previous radar composite to the actual composite" to "The algorithm associates RCs between consecutive radar composites".

L255-256: What specifically are the similarity criteria?

L257: I'm guessing that the "velocity factor" is intended to account for uncertainties in the motion estimates. Is there any reason for using 0.6?

L258-260: The right-moving storm tends to be favoured in the northern hemisphere, whereas the left-moving storm is favoured in the southern hemisphere (due to mirrored shear profiles). However, I'm not sure that it's fair to say that "in most cases, the left-moving cell weakens very quickly". It may sometimes evolve into more of a multicell structure, while in unidirectional shear profiles both storms may persist as supercells. Furthermore, it's important to note that cell splitting may also occur due to changes in storm intensity that cause a single RC to break up (or vice versa in the case of cell merging).

L262-268: I'm afraid I really don't follow this explanation at all. It seems to imply that single cell associations where there is a significant change in cell area can be labeled as splits or mergers, but that doesn't make sense. A split would normally be defined where a cell at time t can be associated with two or more cells at time $t + \Delta t$. In this case the choice has to be made as to whether all or only one (e.g. the largest or most intense) of the "child" cells inherit the history of the "parent" cell. Similarly, a merger is defined where multiple cells at time t are associated with a single cell at time $t + \Delta t$. In this case, a choice has to be made how to assign a history to the child cell. I suggest you completely rework this description (possibly adding a schematic) to make it easier to understand.

L276-277: Motion vectors include both speed and direction, so I don't think you need to say "as well as the track direction of the convective cells".

L279: If I am understanding this correctly, the shift vector is a spatial increment rather than a velocity. In this case you shouldn't use du and dv, but rather dx and dy. If it *is* a velocity vector then the components should be u and v.

L281-287: Again, I found this explanation really hard to follow. I tried reading the equivalent explanation in Puskeiler et al. (2016), but that is equally perplexing. In my head, the way this type of procedure would work would be to take a cell at two consecutive times and then shift the early cell forward in time and the later cell backward in time and average the two. This would be done for multiple intermediate time steps in order to create a smooth track. However, it sounds like the procedure used here is considerably more complicated. Again, I think this needs a complete rewrite in order to make it comprehensible.

L318: Presumably by "hail days" you mean hail days within the subdomain shown in Fig. 4 as opposed to anywhere in your analysis domain. Please clarify in the text.

L320: Again, you should say 0.25° rather than 31 km.

L322: Here and elsewhere in this section you should change "hail days" to "hail days per year"

L326-327: Is this flow convergence perhaps the signature of a surface cold front or pressure trough? The circulation associated with a front/trough might favour the development of severe storms (through the associated generation of mid-level instability and advection of low-level moisture), but flow interactions with the Massif Central could still act as a focussing/initiation mechanism.

L337: Surely the winds should be averaged over the lowest 1200 m to be consistent with the definition of H?

L338-341: The definition of the Brunt–Väisälä frequency is pretty standard so I don't think you need to cite Huschke (1959). However, I would recommend that you modify the equation so that it is an expression for N rather than N^2. Also, you can just use θ_v for the virtual temperature (you don't need to refer to it as "ambient").

L345-346: You might want to note that this deflection of the flow is unlikely to show up in Fig. 4 due to the fairly coarse resolution of the ERA5 reanalysis (the Massif Central will be much lower and smoother in the IFS than is shown in the figure).

L350-353: This sentence is overly long and should be rephrased. You could potentially get rid of the bit starting "by referring to"

L390-395: The Froude number that you obtain is actually a bit larger than that estimated by Kunz and Puskeiler (2010), even though they used a smaller value of H (1000 m). Presumably then, your values of U are larger and/or your values of N are smaller. Can you comment on these differences?

L411: I would suggest changing "is partly caused by" to "show an association with" as causality has not been firmly established.

L415: I've checked an Puskeiler et al. (2016) don't actually examine interannual variability in hail frequency. Nisi et al. (2018) do, but they find much higher interannual variability compared to this study. Perhaps this reflects the much larger study domain considered here.

L430-431: You don't show maps for 2012 so this sentence can probably be deleted.

L438-440: I'm not sure this last sentence is needed. Given the size of your study domain it is hardly surprising that years with a below-average number of hail days overall could still have a few localised high-impact events. Perhaps you could simply rephrase what you have to make this point. Also 2013 is only slightly below average and the average is arguably dragged up by the anomalously high number of hail days in 2006.

L451: What you show in Fig. 7 isn't a 10-day moving average as the averaging windows don't overlap. Instead you could say that you calculate the average number of hail days for consecutive 10-day periods. Alternatively, given the short length of your climatology, you might consider just plotting the relative frequency of hail in each month. This might also make for an easier comparison with previous studies.

L468-469: Can it really be argued that these are distinct maxima? This could just be an artefact of the relatively short length of your climatology.

L473-474: It doesn't make sense to describe this as "a right skewed distribution" since you're not really talking about a distribution but a time series.

L479: Which of the aforementioned studies used hail pad data? What data sources did the other studies consider? And why would the hail pads being clustered “near the subdomains influenced by maritime air mass” lead to an earlier seasonal peak?

L484: It’s not really right to say “confirmed” here as the Lukach and Delobbe (2013) study obviously came before yours and showed an earlier maximum. To me the seasonal cycle for subdomain BEL looks pretty flat, but this is because you have plotted the number of hail days rather than the relative frequency.

L486: Since you only consider the first time that CCTA2D detects the cell this analysis pertains more to the development of hail storms, rather than their overall diurnal cycle. As stated in my original review, it would make more sense to consider all times when a storm exceeded 55 dBZ as this would account for storms that persist for multiple hours. This would also make comparisons with previous studies easier, since I imagine most of these considered all hours with hailstorms, rather than just the hour in which storms developed.

L492: There’s no need to keep reminding the reader where each region is located as this has already been stated and is shown in Fig. 3. The same comment applies to L495-496 and L498.

L495-496: The peak at 16 LT is only slightly above the values for the adjacent hours. Given the relatively short length of your climatology I’m not sure you can read too much into this difference.

L510-511: The peak at 18 LT is much later than what you and most other studies find. Can this difference be explained?

L512-531: I’m really not convinced that this section adds much, if anything, to the manuscript. The distribution of track onset locations seems pretty consistent with the overall distribution of hailstorms shown in Fig. 3, with fewer points overnight and during the morning and more points during the afternoon and evening (as one would expect from Fig. 8). Unless you can quantitatively show that some regions show a *disproportionately* high/low onset frequency for a given time (i.e. many more/less onsets than one would expect based on overall hailstorm frequency) I would suggest getting rid of section 4.4 altogether.

L513-515: Get rid of “and to distinguish between mechanisms triggering nighttime events and convection being triggered within the boundary layer occurring preferably in the afternoon and early evening” - it makes the sentence overly long and isn’t needed.

L518: Hours less than 10 should be written as 00, 03, etc. The same comment applies to L519, L522, and 530.

L540: I would say “a length less than 10 km”. There’s no need to include the symbol “L” if you’re not going to include it in an equation.

L541-542: Figure 10 only shows track lengths up to 100 km so perhaps the values you quote here should correspond to track lengths of 20–100 km and > 100 km, rather than 20–150 km and > 150 km.

L559-561: In my original review I suggested that it might make more sense for orientation to be computed either as the angle of a line connecting the first and last points in the track or by fitting a line of best fit to all points in the track. In your response you argued that these methods “fit for straight (e.g. undeviated) swaths only”. However, your method still does not account for curved storm tracks since it only considers a single pair of points in the track (before and after the centre point). Furthermore, it is likely to be more sensitive to sudden changes in cell direction associated with splits/mergers or changes in cell area. Assuming you have the start and end positions of the cell track I would suggest using these to compute the orientation as it is simply but consistent with how track length is defined.

L563: Technically a west-to-southwest direction would be from 225 to 270°. As such I would just say “from between 200 and 260°.”

L576-578: This sentence needs reworking. First, you should change “none or only several” to “only a few”. Second, “along the European coastlines” isn’t very specific and is arguably repetition since Brittany and north Germany could be classed as “along the European coastlines”. Third, I would say “farther inland” rather than “far off the coasts” as the latter implies offshore. Finally, you should quote hail frequencies for both coastal and inland regions in days per year.

L580-581: “The high spatial variability in the number of radar-derived hail days and the increasing number around orographic structures…” - I’m not sure what you mean by this; please rephrase.

L583: Is the diurnal cycle of hail that different between different regions? It’s hard to say from Fig. 8 because it plots the absolute number of hail days rather than the relative frequency.

L603-604: X-band radars are actually less suitable for hail detection because their signal is strongly attenuated by large precipitation particles. You might instead mention the use of dual-polarisation measurements, which can provide more accurate detection of hail (e.g. Heinselman and Ryzhkov, 2006).

Figure 1: In my previous review I requested that you use different line thicknesses or colours for country and region/state borders, so that readers less familiar with European geography can distinguish between the two; however, this change does not appear to have been made.

Figure 2: Since the X-band radars aren’t included in your analysis they probably should be removed from this figure. You might also consider using different symbols for those radars that were replaced or upgraded to dual-polarisation during the study period.

Figure 3: As noted in my original review, a colorbar would be more appropriate than the individual blocks with value ranges, since these imply gaps (e.g. between 0.6 and 0.7 day per year). I would also suggest using an increment of 0.5 days per year as this is much more intuitive (0.5 days per year = once every two years). One more thing. In transforming the projection of this plot, the boxes defining the different regions appear to have become distorted, such that their edges don’t properly line up. Can this be rectified?

Figure 4: It looks like you have two coastlines in this map - can you get rid of the coarse one? Again, I would suggest using an increment of 0.5 days per year for the hail frequency contours (values of 2.5, 3.0, and 3.5 days per year). Finally, could you add a box showing the area for which the Froude number was evaluated?

Figure 6: Again, use a colorbar rather than the individual blocks for each colour range (with increments of 1 day), if possible.

Figure 7: The shading under the curves isn't needed so I suggest getting rid of it. The same comment applies to Fig. 8. I would also shift the curves so that points are centred on each 10-day window, rather than at the end of the window.

Figure 8: The results in this and the previous figure might be better presented as relative frequencies for each month/hour rather than the absolute number of hail days, as this will allow you to better infer differences in the seasonal and diurnal cycles of hail between the five regions. Just a suggestion.

Figure 9: Hours less than 10 should be written as 00, 03, etc.

Figure 11: My first impression was that this figure, and the ones before and after, show relative frequency of the y axis. However, for this one the values for the first four bars alone add up to more than 100 %. Is this an error?

Technical Corrections

L4: Change "reflectivity radar data and lightning data" to "radar reflectivity and lightning data".

L8: Change "or" to "and".

L20: It should be "mesoscale convective *systems*" not "mesoscale convective *storms*".

L22: "subtle" not "subtile".

L30: It should be "aside from" rather than "aside of".

L94-95: "that requested some computation adjustment into the national radar composite" - I'm not sure what you mean here; please rephrase.

L95: Change "up to nowadays" to "onward". Also, should it be "2015 onward" rather than "2014 onward", since data were available for 2014?

L219: Change "Despite" to "Although" and "has included" to "includes".

L240: When referring to RCs you should use the pronoun "an" rather than "a" since the initialism is read as *arr-see*. There are many other occurrences in section 3.3 where this needs to be corrected.

L245: Change “BZ” to “dBZ”.

L260: Change “after they have been splitted” to “after they have split” or “following the split”.

L290: It should be “Kunz et al. (2020)” not “(Kunz et al., 2020)”.

L293-294: I know what you’re trying to say here but “prevail” isn’t the right word. The key point is that there are far fewer hail reports in France, Belgium, and Luxembourg than in Germany.

L318: Get rid of “(in terms of speed and direction)”.

L323-324: Get rid of “named” in both sets of parentheses.

L325: Change “northern” to “southerly” (wind directions refer to the direction from which the wind is blowing).

L355: “requires”

L357: Get rid of “referred to as”. The same change should be applied on L377, 387, L401, and L404.

L363: “1424 m” (get rid of the period).

L364: Changed “mentioned as” to “labelled”. The same change should be applied on L388.

L365-366: You should put “e.g.” at the start of the list of references rather than having “among others” at the end.

L384: This shouldn’t be a new paragraph.

L427: This shouldn’t be a new paragraph.

L437: Change “entire” to “all of”.

L458: Change “Pyrenean” to “Pyrenees”.

L472: Get rid of “located in the very southwest of France” (this is obvious from the name).

L483: Change “upper western part of Belgium” to “much of Belgium”.

L499: Change “reminds” to “remains”.

L542: Get rid of the comma after “MCSs”.

L548: This shouldn’t be a new paragraph.

References

Brooks, H.E., Lee, J.W. and Craven, J.P. (2003) The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. Atmospheric Research, 67, 73–94.

Heinselman, P.L. and Ryzhkov, A.V. (2006) Validation of polarimetric hail detection. Weather and Forecasting, 21, 839–850.

Johnson, A.W. and Sugden, K.E. (2014) Evaluation of sounding-derived thermodynamic and wind-related parameters associated with large hail events. Electronic Journal of Severe Storms Meteorology, 9, 1–42.

Smith, B.T., Thompson, R.L., Grams, J.S., Broyles, C., and Brooks, H.E. (2012) Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Storm classification and climatology. Weather and Forecasting, 27, 1114–1135.

Taszarek, M., Brooks, H.E. and Czernecki, B. (2017) Sounding-derived parameters associated with convective hazards in Europe. Monthly Weather Review, 145, 1511–1528.

Wapler, K., Hengstebeck, T., and Groenemeijer, P. (2016) Mesocyclones in Central Europe as seen by radar. Atmospheric Research, 168, 112–120.