Review on "The climatology and nature of warm-season convective cells in cold-frontal environments over Germany" by Pacey et al.

This study presents a climatology of convective cells associated with cold fronts in a front relative coordinate frame and compares these to convective cells occurring in non-cold frontal environments. Convective cells are shown to be much more frequent on cold front days than non-cold front days and the most likely location for convective cells to develop on cold front days is found to be 350 – 400 km ahead of the front. Overall, the manuscript is clear, well written and the results are supported by evidence. However, I have three major concerns regarding this manuscript.

The first major comment concerns how well this study fits the scope of this journal and the broader context of the results. In the manuscript the link to actual hazards is weak and little emphasis is given to this aspect. Lightning is considered but relatively briefly. It should be clearer how the results of this study inform about meteorological hazards. This study also only focuses on Germany, a choice which is motivated by the availability of radar data. While the authors do state that the study should be expanded to all of Europe, the current manuscript may be of limited interest to readers from other places than Germany. At a minimum the authors should attempt to address the question of how do these results apply to elsewhere in the world? Do they only apply over continental areas for example?

The second major comment regards the assumption that the surface front is 300 km ahead of the 700-hPa front. This is an oversimplification and likely is not accurate in many cases. Specific points:

1. It is stated that this assumption is based on ERA5 data, but this is not presented – it should certainly be shown even if only as supporting material.
2. It is stated that the surface convergence zone is 300 km ahead of the 700-hPa front. In some cold frontal cases the wind shift (i.e. convergence zone) is not co-located with the temperature gradient so it may not be the thermal gradient which is 300 km ahead of the 700-hPa front. It is also inconsistent to use convergence to locate the surface front but a thermal gradient for the 700-hPa front.
3. This simple approach does not consider kata-cold fronts in which the front appears to slope forward with height as the cold air aloft has overrun the surface front. Kata cold fronts can certainly trigger elevated convection and these fronts should be considered separately.

The third major comment concerns the clustering presented in section 3.3.3. Specific issues here are:

1. The manuscript lacks details on exactly how the clustering was done (e.g. was any normalisation on the input features performed?).
2. The choice of the number of clusters appears subjective whereas the silhouette score and elbow plots could be used to better justify the final number of clusters.
3. The number of clusters (30) is too large to be of practical use to e.g., forecasters
4. The justification for removing 6 of these 30 clusters is not clear and it appears that the clustering has identified 6 clusters which are not physically consistent – strongly suggesting that the clustering has not been performed in an optimal manner.
5. The only outcome of the clustering that is presented is the number of cells and the location of the front. It would be helpful for forecasters to also see additional meteorological variables associated with each of these clusters, for example, the MSLP and equivalent potential temperature.

Below I also list some minor comments which would certainly improve the manuscript:

Minor comments

1. Line 41 – 42. Please expand this sentence to make it clearer. It needs to be stated that this is due to the frontal surface sloping rearwards with height.
2. Line 62, Question 3. This could be written in a manner so it can stand alone and does not need a reader to refer back to Q1 and Q2. This would likely make it clearer and easier to understand.
3. Line 125. Figure 1. Could the domain where fronts are identified in be marked on this figure?
4. Section 2.1.1. The criteria used to identify the fronts are quite large so will only identify quite strong fronts in terms of the thermal gradient. Do the results depend on these thresholds, and in particular, do the fronts still hold if weaker fronts are also considered? If not, it should be stressed more clearly that these results only apply to strong cold fronts.
5. Section 2.3. How were these four examples selected and how representative of the whole data set are they? They look like quite standard fronts, so I am wondering if the method works well with more complicated or less uniform fronts.
6. Section 2.3 / method. “Timesteps containing two or more cold front lines in the domain were omitted”. Since timesteps with two fronts present are omitted, this means that this method only works for a small area and could not be expanded to e.g., European scale (even if the radar data was available). This is a notable limitation of this method which should be clearly highlighted, or the method improved to allow two or more fronts to be present at the same timestep.
7. Related to the point above, neglecting timesteps with two or more fronts present means that double fronts will be automatically ignored which may add a systematic bias to the results.
8. Line 176. There is a typo here “in In Figure 2a…”
9. Section 3.1. There are a lot of numbers, percentages especially at the start of this section and it is hard to read. Many of these numbers etc, are in Table 1, but Table 1 is not referred to much here. It would help a reader to refer to Table 1 more. Furthermore, this section may be clearer if the number of cold front days was discussed first, and this information was added to table 1.
10. Line 199, should this be a comma before All rather than a period?
11. Line 210. Is the surface convergence influenced by land use, coastlines, topography etc.?
12. Line 218 and elsewhere after this the phrase “pre-700-frontal environment…” is used. Should hPa be added after 700 here?
13. Line 272 – 274. How does this spatial climatology of convective cells relate to the spatial climatology of fronts as shown in Figure 1? Adding a sentence to relate these aspects would be helpful for a reader.