This paper has been much improved by the revisions. I applaud the authors for including an analysis using SED reports, and for very helpful figures like Fig. 6. It is also nice to see the change in criteria to match better with current definitions of derechos.

Despite the many improvements, I do still have several major concerns that require revisions to the paper. The first one below is particularly serious, but understandable since this happens often when one paper is undergoing revision while another that is very similar by different authors suddenly is published. In the present situation, that paper is so closely related to your own that you will want to not only refer to it, but also change a comparison that you currently make to a rather different dataset (more of an apples to oranges comparison) to instead compare to the new data in the new paper (much more of an apples to apples comparison). The major revisions I request are listed below:

1. One big problem is that although you did include some references to the Part 1 paper by Squitieri and coauthors in 2023 based on my earlier review, you have failed to recognize that Part 2 has already been published (2025, BAMS). That paper has some of the same type of analyses that you do (spatial analyses) using the new definition of a derecho hinted at in the 2023 paper and following Corfidi (2016) as you have done. Clearly, it is very important that you add comparison to this other study now that it has been published. Your paper makes an excellent companion piece to that one, since there is now a lot of similarity, which allows you to truly do a precise comparison and establish how well AI may be working. My own quick glance shows they find very few derechos per year, ranging from 1 to 8 per year. It seems to me that you should replace your comparison with an old 2 years of SPC data (50 cases) with a comparison to the new Squitieri et al. (2024) paper, which is a far better apples to apples comparison with your definition of a derecho. In addition, the really small number of derechos they now find in the US per year raises some questions about some of your tunable parameters. In my third major comment below, I talk about how you should consider your overestimate of ISD-only derechos as evidence you have dropped your threshold too low. In light of the Squitieri et al. (2025) paper, an increase in that threshold would not only result in better agreement between your SED- and ISD-determined number of derechos, but would also likely lower your totals, matching a bit better the small numbers found in the 2025 paper.

2. At line 414 when you are describing your criteria, how do you determine a fraction of SED sites above 20%? I can understand with the other dataset of surface observations that you must look at the total number of stations falling within your derecho swath, and you can figure out what 20% of those stations are. However, for the SED reports, there is no grand total to use. Reports come in from whoever bothers to make the report. Near a city, many reports can come in close proximity to each other, while in wide open areas with low population, reports can be few and far between. There is no total number of stations, so it is impossible to compute a percentage or fraction. More explanation is needed here because I have no idea how you can even attempt to apply this criterion to SED reports.

3. In the discussion from lines 633-645, to me, that agreement does not seem very good. I agree it is not terrible, but the union of all 3 possibilities is 322 cases. 172, or about 55%, are matched

with the use of both datasets. You admit that you used an arbitrary adjustment downward to set a threshold for your ISD cases. Wouldn't this discussion here have been a really good starting point to explore what happens with different ISD thresholds? I would think a good scientific approach would have been to use the threshold in an iterative manner to get the best match. If you are truly correcting for the lack of ISD stations by using a lower threshold, wouldn't it be best to have found a threshold that results in the largest intersection with the most cases showing up in both datasets and the least number of cases only showing up in one or the other? This seems like relatively low-hanging fruit. Instead, you did not do any sensitivities to the threshold (I do realize you said your general behavior between the cases is relatively similar) but since you seem to be arguing that this is a product the community can already go use, I would have thought it would be very important to at least try to get the best match possible. For some of the reasons you mentioned regarding errors in SED, you will probably never be able to eliminate those cases that show up there but do not show up in ISD. However, this was a substantially smaller number than what you found showing up in ISD but not SED. That overestimate in ISD cases is surely linked to having the bar set too low for your wind threshold. What would happen if you raised your arbitrary threshold by 2.5 m/s? Your discussion of sensitivities later on in the paper is useful, but you do not mention this particular sensitivity, which seems the most correctable of all of them to me. You know that you are arbitrarily pulling a number out of the air, smaller than the usually used severe threshold, because you correctly point out that a smaller threshold should be justified due to a lack of stations. But there is nothing we can point to in science to know exactly how much below the original threshold you should be. Wouldn't a fine tuning of the numbers right in this discussion have been a good way to determine the best threshold to have adjusted downward to?

The following are other minor revisions that are needed:

1. Change line 114 from "this study" to "The present study". We cannot tell from "this" if you mean your study or one of the ones you just mentioned.

2. Lines 233-234, what do you mean by a negative sample? Is this a bow echo that does not have damaging wind associated with it during the hour? Please provide more details. In this paragraph, it sounds like you are only looking at 54 known derechos, and hourly images. It seems like 566 positive and 4000 negative are far too many if you are just using hourly images for 54 events.

3. In Figure 12 and the discussion around it, I am confused. What is the difference between wind reports in an DMCS versus a derecho? I thought a DMCS is an MCS that produces a derecho. Is the information in panel c simply wind reports that happen in an MCS that did have a derecho, but falling outside the location or time window that you define to be a derecho? I don't feel like this is very clear in the paper, and believe when you start showing statistics relating to these different types of systems, you need to remind the reader what the difference is, or in this case, how you can have differences in reports from DMCSs compared to derechos.