

Dr. Graciela Raga
Editor
ESSD
April 8, 2025

Subject: Revision of manuscript # essd-2024-112

Dear Dr. Raga,

Thank you and the reviewers for your thoughtful and thorough review of our manuscript. In addition to comments from Reviewer #1, we also received feedback from Reviewer #3 via email. Reviewer #3 was unable to meet the submission deadline and upload comments to the ESSD system. However, we have nonetheless carefully considered their input alongside the other reviewer's remarks. Additionally, we have made further minor corrections to the manuscript to ensure it meets the publication standards.

Please direct any correspondence regarding this manuscript to me (jianfeng.li@pnnl.gov) or Dr. Zhe Feng (zhe.feng@pnnl.gov). We sincerely appreciate your time and effort throughout this process.

Sincerely,
Jianfeng Li
Atmospheric, Climate, and Earth Sciences Division
Pacific Northwest National Laboratory
Richland, Washington, US, 99354

Response to Reviewer #1

Thank you for your careful and thorough reading of this manuscript and your thoughtful comments and suggestions. Our responses follow the reviewer's comments (*in Italics*).

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:

Reply:

Thank you for reviewing the manuscript carefully and thoroughly again. We sincerely appreciate your efforts to help improve the manuscript. We have addressed all your comments 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.

Reply:

Thank you so much for pointing out to us about the publication of the new papers (Squitieri et al., 2025a, b), which are directly related to our study. We have added a comparison between our study and Squitieri et al. (2025b) (Lines 519-526 in the revision-tracked main manuscript) and cited the paper a few more times when describing the derecho climatological characteristics (Lines 589-592, 618-619, 622, and 633-637). However, we would like to keep the comparison

against the NOAA SPC data, which provides the exclusive event-scale information. The total derecho number comparison cannot guarantee that our dataset captures the correct “derecho” events even if the total counts are similar. The derecho dataset developed in Squitieri et al. (2025b) was not made available publicly, preventing an event-based comparison as what we have done with the NOAA SPC data. We acknowledge the limitation of comparing our dataset with the NOAA SPC data, but it is critical to show that our detection algorithm identifies the “correct” events. As noted in the response below, differences between Squitieri et al. (2025b) and our dataset may be attributed to many reasons, suggesting a detailed event-based comparison combined with sensitivity analysis would be useful in the future. For questions related to ISD, please refer to our responses to the third major comment.

Lines 519-526: ‘Our derecho numbers are also higher than those from Squitieri et al. (2025b), who identified 70 SED-based derechos during 2000-2022 based on the physically-based definition criteria from Corfidi et al. (2016) but with much stricter gust requirements (e.g., at least five reports of very damaging gusts ($\geq 33.53 \text{ m s}^{-1}$)) for a 400-km-long gust swath (Squitieri et al., 2025a, b). The discrepancies among the present study, Corfidi et al. (2016), and Squitieri et al. (2025b) could be attributed to the different gust criteria used in the derecho definitions but also likely stem from differences in the methods used to calculate gust swath length and width, the criteria for forward propagation, and the diverse observational source datasets used in the derecho detection.’

Lines 589-592: ‘While if we require at least five very damaging gust reports when using SED, the derecho count decreases substantially from 220 to 125, which is still larger than but much closer to the estimates by Squitieri et al. (2025b) (70 derechos between 2000 and 2022).’

Lines 634-636: ‘However, our dataset has almost no derechos in the cold seasons, which is generally not the case in previous studies except for Squitieri et al. (2025b), which also used physically-based criteria to detect derechos.’

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.

Reply:

Thank you for your comments. Yes, the 20% criterion is primarily used for ISD gust speeds, and SED seldom reports gust speeds weaker than strong gusts. Implementing the 20% criterion for SED data excludes only one MCS from being considered a potential DMCS, which we can also remove by manually correcting the falsely identified bow echoes, although the latter is more time-consuming. However, to keep the criteria consistent between ISD and SED as much as possible, we still apply the 20% criterion to SED data, which does not produce any adverse impact on the derecho detection according to our manual check of all identified derechos and high-wind-producing convective systems.

The implementation of the 20% criterion for SED data is rather straightforward. Each SED gust report has a location (latitude and longitude), as mentioned in Lines 194-198, and we consider the location a “site”. Then we can count the number of unique “sites” with damaging gusts or gust speeds weaker than damaging gusts.

We have added a sentence in Lines 445-447 to make the 20% criterion more understandable.

“It is noteworthy that this criterion is primarily applicable to ISD data, and its implementation for SED data excludes only one MCS from being considered a potential DMCS.”

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?

Reply:

Thank you very much for your comments and suggestions. Firstly, we want to emphasize that both ISD and SED datasets have limitations, and the derecho datasets developed based on the two gust speed datasets inherit those limitations. Neither of the derecho datasets nor their combination can be considered the ground truth. An overlap of 172 derechos between the ISD-based (274 derechos) and SED-based (220 derechos) datasets is imperfect but may be acceptable considering various uncertainties involved in developing the derecho datasets. Secondly, we acknowledge that we do not have a robust scientific reason to select 17.43 m s^{-1} as the ISD gust speed threshold to detect derechos. There are a few named gusts from the National Weather Service, including violet gust ($> 41.13 \text{ m s}^{-1}$), very damaging gust ($\geq 33.53 \text{ m s}^{-1}$), damaging gust ($\geq 25.93 \text{ m s}^{-1}$), and strong gust ($\geq 17.43 \text{ m s}^{-1}$). We selected the strong gust speed as the

threshold for ISD because it ranks right below damaging gust. The threshold is undoubtedly adjustable. Following your suggestions, we have performed several sensitivity tests with different ISD gust speed thresholds. The derecho number is sequentially reduced from 274 when using a threshold of 17.43 m s^{-1} to 255, 229, 210, and 157 for a threshold of 18, 18.5, 19, and 20 m s^{-1} , respectively. Even if the total derecho numbers are very close between the ISD-based and SED-based datasets when we use a threshold of 18.5 m s^{-1} for ISD, their overlap reduces to 152 (Figure R1), suggesting the challenge of simultaneously matching the total derecho numbers and increasing the number of overlapping derechos derived from the two datasets. This is within our expectation. The difference between the ISD-based and SED-based derecho datasets stems from the inherent limitations of the ISD and SED gust speed datasets, which cannot be eliminated by simply tuning the gust speed threshold. Per your request, we have added the sensitivity test results in Lines 586-589 in the revised main manuscript, as follows.

‘For example, to reduce the ISD-based derecho count to the SED-based level, we must increase the ISD gust speed threshold in Criterion 4 in Section 4.1 from 17.43 m s^{-1} to 18.5 m s^{-1} ; using the latter threshold produces a derecho number of 229, 152 of which overlapped with the SED-based derecho dataset.’

However, we do not plan update our ISD derecho dataset using the new threshold (18.5 m s^{-1}) due to two reasons: 1) 18.5 m s^{-1} does not correspond to any named gusts; 2) the lack of a ground truth to support any fine tuning. For users who are interested in derechos that produce stronger gusts, we encourage them to adjust those thresholds based on their specific scientific needs. As we highlighted in Lines 596-599, to make our dataset useful to the broad users, we store all the key parameters in the dataset to make it flexible; it is straightforward to further tune those thresholds per the users’ specific needs.

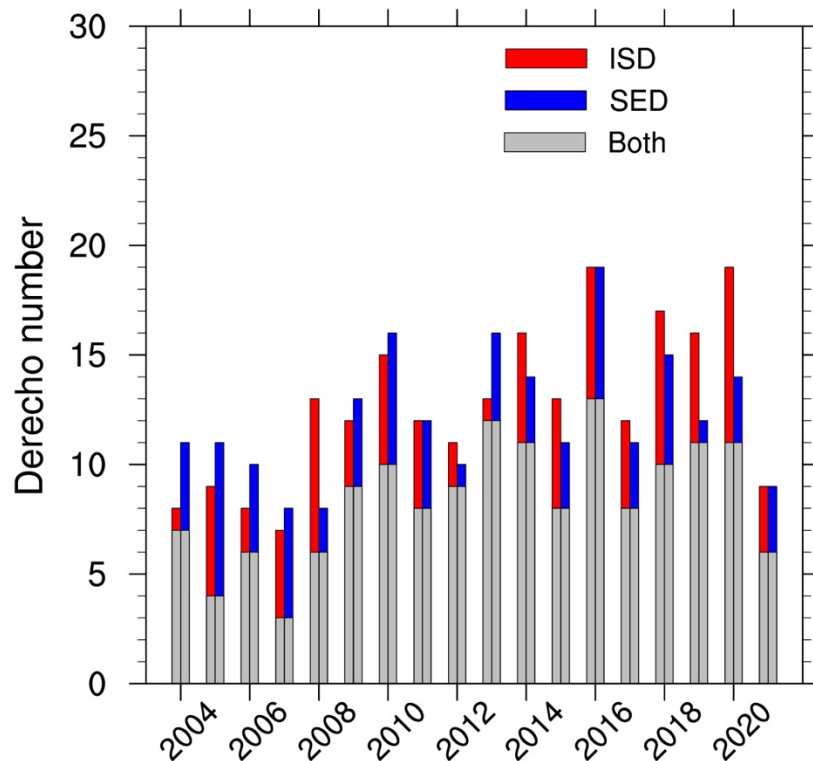


Figure R1. Same as Figure 8 but using a gust speed threshold of 18.5 instead of 17.43 m s⁻¹ for ISD.

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.

Reply:

Changed. Please see Line 103 in the revised manuscript. Thank you!

“The present study applies a semantic segmentation convolutional neural network (CNN) to detect bow echoes automatically from two-dimensional composite (column-maximum) reflectivity (Z_{Hmax}) data in the United States”

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.

Reply:

Since our machine learning approach is used to identify bow echoes, positive and negative samples only differ in the existence of bow echoes. However, since the negative samples are randomly selected from the entire radar reflectivity dataset, few negative samples may contain bow echoes. We must emphasize that the few “bad” negative samples have minimal impact on the training of a Dense Net (Section 3.1.2), and moreover, this is just the first step of our machine learning part, and we only used Dense Net to collect more new positive samples due to its high false positive rate, as mentioned in Lines 233-237.

It is noteworthy that the initial 566 positive samples are from the 54 named DMCSs, while the negative samples are randomly selected from the entire radar reflectivity dataset embedded in the MCS dataset (18 years of hourly data). As we mentioned in the first round of revision, the mean derecho lifetime is ~11 hours, it is not hard to find 566 positive samples from 54 named DMCSs. The initial 566 positive samples had been validated independently by two of the coauthors and have high qualities. However, we understand the limitation of the initial positive samples, which are from a limited number of derechos and may not be representative. Therefore, we generated a much more diverse positive and negative samples through Dense Net and pseudo-labeling, and the final samples include 1699 positive and 1978 negative samples, as explained in Sections 3.1.2 and 3.1.3.

We have added more information in Lines 219-222 and hope it is more understandable.

“The number of bow echo samples varies among different DMCSs, and 566 positive samples (with bow echoes) are obtained in total. 5400 negative samples (generally without bow echoes) are also randomly selected from the entire 18 years of radar reflectivity data embedded in the MCS dataset.”

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.

Reply:

Thank you for your comments. We have clarified the discussion around Figure 12 in Lines 659-661, as follows. The DMCS damaging gust reports contains both the derecho associated damaging gust reports and those falling outside. Therefore, Figure 12d is a subset of Figure 12c.

“Notably, damaging gust reports associated with a DMCS include those from the corresponding derecho as well as those falling outside the derecho location or time window.”

Reference:

Squitieri, B. J., Wade, A. R. and Jirak, I. L.: On a modified definition of a derecho. Part I: Construction of the definition and quantitative criteria for identifying future derechos over the

contiguous United States, Bulletin of the American Meteorological Society, 106(1), E84-E110, <https://doi.org/10.1175/BAMS-D-24-0015.1>, 2025a.

Squitieri, B. J., Wade, A. R. and Jirak, I. L.: On a modified definition of a derecho. Part II: An updated spatial climatology of derechos across the contiguous United States, Bulletin of the American Meteorological Society, 106(1), E111-E124, <https://doi.org/10.1175/BAMS-D-24-0140.1>, 2025b.

Response to Reviewer #2

Thank you for your careful and thorough reading of this manuscript and your thoughtful comments and suggestions. Our responses follow the reviewer's comments (*in Italics*).

Review of the most revised (dated 7 February 2025) version of ESSD manuscript 2024-112 (numbers refer to line numbers in the 7 February 2025 manuscript).

92. *Consider adding Corfidi et al. 2016 as a reference at the end of line.*

Reply:

Added. Thank you. Please see Lines 93-94 in the revision-tracked main manuscript.

221. *"Timings" is vague; is the indicated time that of the "bow echo" output?*

Reply:

It refers to the occurrence time of the bow echo. We have rewritten the sentence as follows (Line 225).

“The subplot titles indicate the bow echo occurrence times.”

239. *Likewise, "distinct" is vague; what is meant by "distinct"? Easily recognizable?*

Reply:

Here, “distinct” refers to “diverse” or “a broader range of”. We meant that the updated 500 positive samples are from a broader range of “sources” to improve their representation. We have changed the sentence as follows (Lines 242-243).

“these samples have higher diversity than the initial bow echoes generated from the named derechos on Wikipedia because they are drawn from a broader range of events”

256. *I'm not sure how "probability of 0.1" applies here; a brief phrase to explain would be helpful.*

Reply:

Thank you for your comment. A probability of 0.1 corresponds to 10% of the pixels. For each pixel, the probability to add random noise to it is 0.1. We have clarified it in Line 260, as follows.

“First, random salt and pepper noise is added to 10% of the pixels in each sample with a probability of 0.1 (i.e., it has a probability of 0.1 to add random noise to a pixel).”

370-400. Great to see the modifications that were made here compared with what appeared in the first version of the manuscript --- especially the method used to identify forward propagation (i.e., lines 381-388).

Reply:

Thank you for your positive comment!

402. Good update to the figure (cf. the figure in the first version of the manuscript)!

Reply:

Thank you!

608. Does figure S6 really show this? "SED" is not mentioned in the caption for Figure S6 in my version of the Supplement.

Reply:

Please find below the updated Figure S6 in the first round of revision. Compared to Figure 9 in the main manuscript, the northwest-southeast axis is more noticeable.

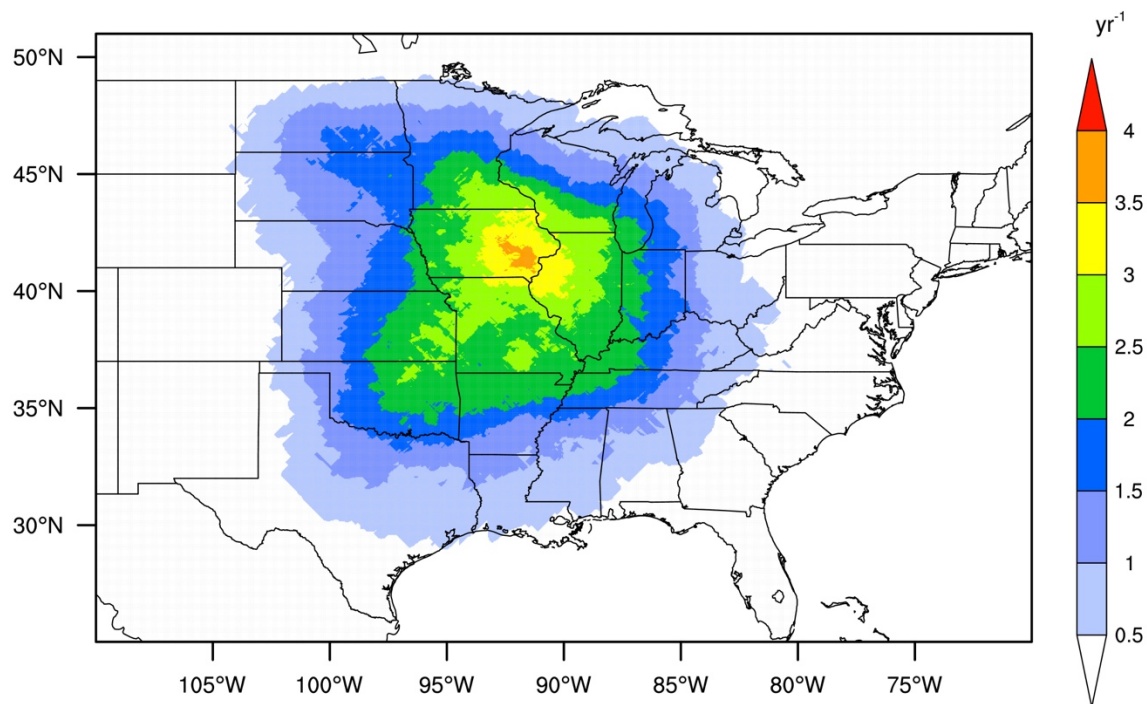


Figure S6. Same as Figure 9 but for the SED-based dataset.

616. "Code" should be "cool."

Reply:

Thank you very much for pointing out the error. We have changed “code” to “cold”. Please see Lines 634-635 in the revision-tracked main manuscript.

626. *The axes also may represent two distinct subsets or populations of progressive (forward-propagating) derechos that move (1) primarily west to east vs. (2) those that move more north to south.*

Reply:

Thank you so much for pointing out the issue. Based on our most recent self-organization map analysis, the two axes indeed correspond to different types of derechos associated with different large-scale meteorological patterns. We have corrected the sentences as follows (Lines 646-650).

“The axes may represent two distinct types of progressive derechos associated with different large-scale meteorological patterns.”

630. *Ditto first sentence in comment made for line 608.*

Reply:

Please find below the updated Figure S7 in the first round of revision.

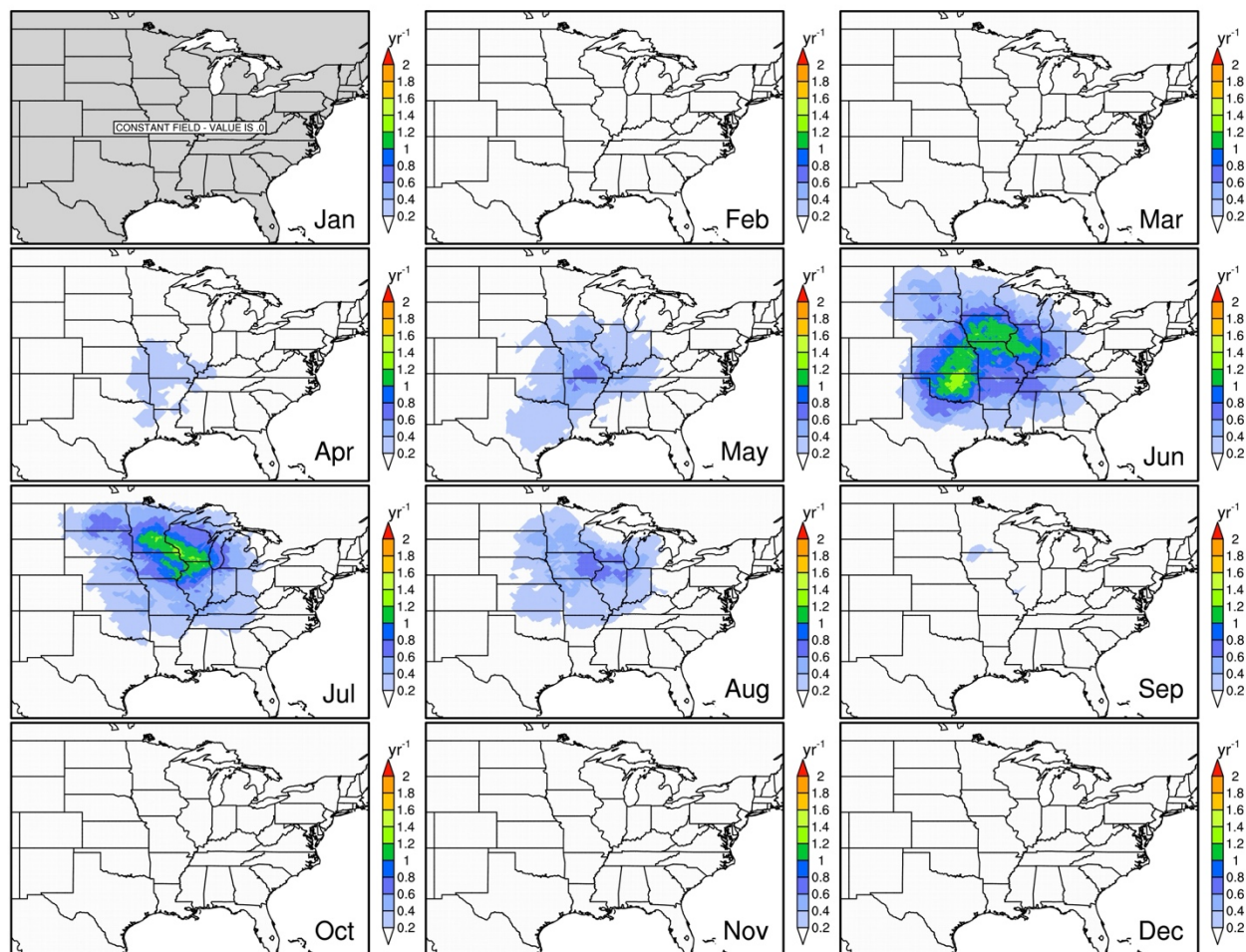


Figure S7. Same as Figure 11 but for SED-based derechos.

643. Is the meaning of "PDF" introduced prior to this point? If not, I suggest spelling out "probability distribution function."

Reply:

Corrected on Line 666. Thank you!

664. My copy of the Supplement references a Figure 14 which does not exist in the revised version of the main manuscript.

Reply:

Yes, the supplement has been updated in the first round of revision.

Your derecho dataset, based on a method that combines machine learning identification of relevant radar features with a physically-based definition of a derecho, should be a valuable resource for the meteorological community.

Reply:

Thank you. We greatly appreciate your kind acknowledgment.