Response to Reviewer #1

Thank you for your careful and thorough reading of the manuscript and your thoughtful comments and suggestions. We apologize for the delay in revising the manuscript, as we spent a significant amount of time on the manual validation and inspection of the identified derechos. According to all three reviewers' comments and suggestions, we have made several significant improvements in our manuscript, which we want to highlight before the point-by-point response.

Firstly, we have improved the language of many sentences so that the manuscript is more understandable, and we have also changed the wording of some terms so that they are more distinguishable and accurate. A derecho is a windstorm, while an MCS is a convective system. They are different concepts. Therefore, we change "derecho" to "derecho-producing MCS" when we refer to the MCS associated with a derecho, and "derecho feature" has been changed to "derecho." The dataset developed in this study includes tracking of both derechos and corresponding derecho-producing MCSs.

Secondly, we have incorporated a "forward propagating" criterion in our derecho detection algorithm. Our initial understanding of the term "forward propagating" was incomplete, and we failed to recognize its critical role in defining a derecho proposed by Corfidi et al. (2016). After careful consideration and evaluation, we have adopted and modified two criteria from Corfidi et al. (2016) to establish the definition of "forward propagating." One is that the acute angle between the averaged bow echo orientation and the bow echo series' propagation direction is larger than 45°, and the other is that the bow echo series' propagation speed is at least 30% faster than the 500-hPa background wind speed. Implementing the "forward propagating" criterion removes many windstorms externally forced by extratropical cyclones, aligning with the purpose of Corfidi et al. (2016), which intends to define derechos as internally driven windstorms. "Externally forced" and "internally driven" reflect distinct physical formation mechanisms of those windstorms, which is why Corfidi et al. (2016) proposed a physically based derecho definition. With the updated detection algorithm, the derecho number between 2004 and 2021 has been substantially reduced from 556 to 274 (for ISD) and 220 (for SED). In addition, due to the inclusion of the "forward propagating" constraint in our derecho algorithm, we have decided not to change the name of our dataset to "high wind-producing bow echo." The derecho detection algorithm description has been updated accordingly in Section 4, and we have also updated all the results in Section 6 based on the improved dataset.

Thirdly, we have developed another parallel dataset using gust speeds from the Storm Events Database (SED). Now, our derecho dataset consists of two subsets: one based on gust measurements from the global hourly Integrated Surface Database (ISD) and the other based on SED gust speeds. Although there are some discrepancies between the two subsets, their agreement is much larger than their difference (Figures 8-11 in the revised main manuscript; or Figure R1 below). Since both ISD and SED gust speeds have limitations and uncertainties, minor differences between the ISD-based and the SED-based datasets are expected and understandable. In addition, the general agreement between the two datasets indicates that our usage of lower gust speed criteria for ISD measurements is reasonable and does not change the derecho number much. We must emphasize that using lower gust speed criteria for ISD measurements than SED reports does not mean that the ISD-identified derechos are weaker than the SED-identified ones

(or even not derechos). This is a compromise, considering that ISD stations are limited and may miss many damaging gusts, as we highlighted in Lines 522-532 in the revision-tracked main manuscript (as below).

"We emphasize that, in Criterion 4, our ISD gust speed criteria are weaker than the SED gust speed criteria as well as those of previous studies (Squitieri et al., 2023; Bentley and Mote, 1998; Johns and Hirt, 1987), which also estimated the gust swath based on SED damaging gusts. As mentioned in Section 2.2.2, most SED gust reports are estimates, while ISD provides gust measurements from weather stations. SED estimates can capture potential damaging gust occurrences over a much larger area, although with large uncertainties. In contrast, due to the limited coverage of observational sites, real-time ISD measurements may miss substantial damaging gust occurrences in nearby regions. Therefore, we lower the gust speed criteria to capture potential derechos when using ISD measurements. It does not mean that the ISD-based derechos are weaker than the SED-based ones or even not derechos, as elaborated in Section 5."

Fourthly, due to the incorporation of "forward propagating" in the detection algorithm and the development of the SED-based derecho dataset, we have updated some sensitivity tests, evaluations, and the comparisons of our datasets with the NOAA SPC data in 2004 and 2005 and previous studies. Please see Section 5 for further details.

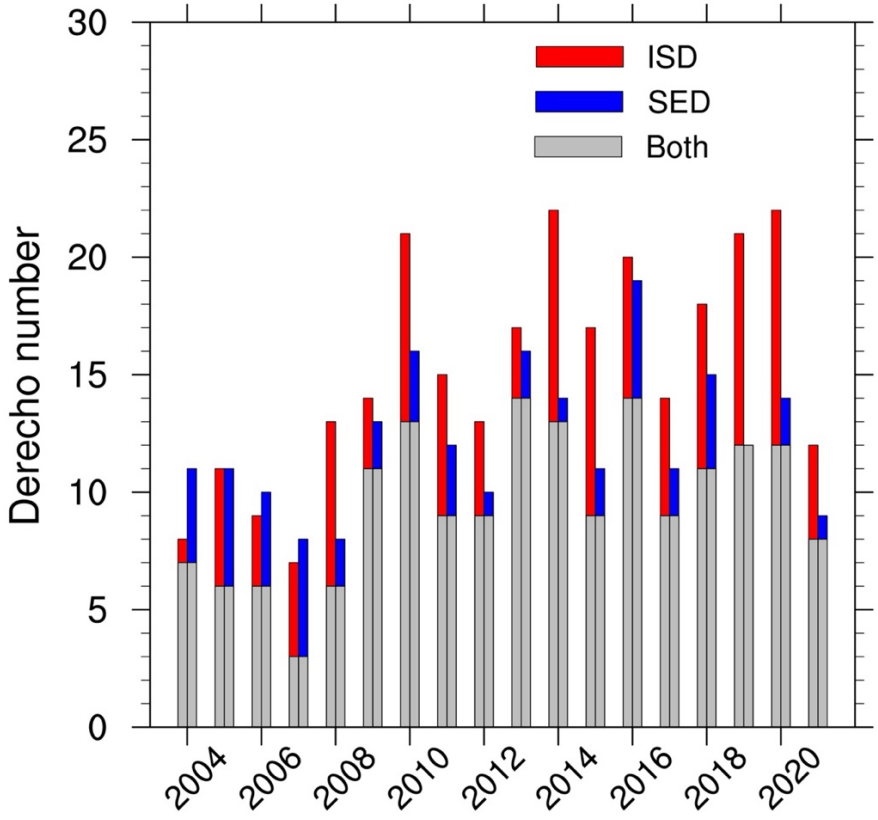

Figure R1. Bar chart of the annual derecho numbers from the ISD-based and the SED-based datasets from 2004 to 2021. Gray shading denotes derechos captured by both datasets, red shading refers to derechos only identified when using ISD gust observations, and blue shading represents SED-only derechos. The figure is the same as Figure 9 in the revised main manuscript.

The general idea of an AI-based objective tool to identify derechos is fascinating, and such a tool could be very useful. The approach taken in this paper has promise and it is exciting that some reasonably good results are obtained. However, I feel like there were some poor choices made in the design of the tool that harm its performance, and I list those below. In addition, I wonder if the timing of a paper like this is not good, in light of the fact that a very new physics-alone-based definition of a derecho is being proposed by some severe weather scientists (e.g. Squitieri 2024, abstract 16A.3 for AMS 28th Conf. on Sev. Local. Storms). In my own work with derechos and severe thunderstorm winds in recent years, I was aware that many in the community feel the definition much change from its current unusual focus on after-the-fact wind measurements. In meteorology, we normally define things based on their physics, and it sounds like that is what is being proposed by the severe weather community within SPC, which typically has been tasked with identifying derechos. It seems it would be much better for the AI work in this present study to be better coordinated with the group proposing the new definition, so that the AI tool is designed to work with the new definition. As it stands, I fear the AI tool will be obsolete from day 1, particularly in light of my concerns listed below. The Squitieri 2024 abstract indicates that the new definition reduces the number of derechos per year substantially, which would make the present study's overestimate of derechos even worse. In summary, I believe the problems associated with the design of this tool are so severe that the paper cannot be published in its current form. The AI tool is really just identifying bow echoes that produce strong winds, not derechos. This presents numerous fatal problems in the discussion of results.

Reply:

Thank you for your comments and interest in the AI tool. As discussed in the manuscript, the segmentation CNN is used to identify bow echoes, which is only part of the derecho detection. We shared the bow echo dataset as well as the segmentation CNN code with the public on Zenodo (Lines 865 – 870 in the revision-tracked main manuscript, or as below). Any group can use the AI tool or the CNN-identified bow echoes in their studies, regardless of whether they are related to derecho detection or not.

"For researchers interested in applying the segmentation CNN for bow echo detection in different regions or time periods, or in leveraging the CNN-identified bow echoes for other studies, we provide access to the bow echo segmentation code and datasets at https://doi.org/10.5281/zenodo.10822721 (Geiss et al., 2024). This repository includes the trained CNN weights and detailed usage instructions. Additionally, a video supplement demonstrating the bow echo segmentation scheme is available at https://youtu.be/iHWY_OhaVUo and is permanently archived in the above Zenodo repository."

Before submitting the manuscript, we discussed the derecho definition issue with the group that submitted the AMS abstract you mentioned. Unfortunately, we did not reach an agreement on the derecho definition. According to our understanding, except for distinct approaches to identifying bow echoes, the primary difference between our algorithm and their derecho definition is the gust speed criteria. The AMS abstract used gust speeds from SED and required a derecho to have at least five reports with gust speed ≥ 33 m s^{-1} in the 400-km long wind swath, with at least three of the ≥ 33 m s^{-1} gust reports measured; otherwise, there must be at least three damaged-estimated gust reports with a speed ≥ 44 m s^{-1}. We understand their motivation to use much stricter gust speed criteria. Those derechos have more significant socioeconomic impacts than relatively weaker ones. But it's arguable that their derecho definition is physics-based, while ours

is not. Both definitions are based on Corfidi et al. (2016), where the physical parts are more reflected in the usage of bow echoes with mesovortex and/or rear-inflow jets, cold-pool driven, and forward propagating. We do not mean that different gust speeds are all generated by the same physical mechanism. They may be associated with different physical processes (which requires additional analysis after we develop a derecho dataset), but in the definition, the gust speed criteria are more related to socioeconomic impacts. As we mentioned above, even if we use SED gust reports and the same gust speed criterion from Corfidi et al. (2016) in our derecho detection algorithm, the derecho number is still similar to that based on ISD with a weaker gust speed criterion. We used a lower gust speed criterion for ISD because of the limitations and uncertainties of the gust speed dataset, and it does not mean that the ISD-based derechos are weaker or that they are not derechos. Moreover, even if NOAA plans to use stricter gust speed criteria in the future, it does not undermine our study. Our dataset is flexible, and we store all the key parameters for each derecho in the dataset, including gust speed reports from SED or ISD. Users can easily filter out stronger derechos using existing variables in our dataset if they are interested in the strong derechos, which would reduce the number of events in the full derecho dataset.

We also want to clarify that we do not intend to use the algorithm for operational purposes by developing the detection algorithm and derecho dataset. The segmentation CNN bow echo identification is imperfect and could suffer from uncertainties, which is why we invested substantial time on manual inspection of the dataset and those MCSs that produced strong gusts. For meteorological operations in the United States, instead of manually correcting the AI detection errors, it may be easier for an experienced meteorologist to label bow echoes and identify derechos manually when the data volume is small. When the data volume is large, as in this application with 18-years of hourly data, the AI classifier is much more efficient. Besides developing the 18-year derecho climatology over the United States, we expect that the CNN bow echo identification tool and the automatic derecho detection algorithm can be used in other areas including application to model results with large volumes of data. This is another reason why we prefer to use the ISD-based derecho subset instead of the SED-based subset: other countries often do not have as much data as the United States, especially the SED data. However, many countries have some measurements from weather stations, and they can easily adapt our detection algorithm to fit their specific conditions.

I am curious since prior definitions of derechos have included the length requirement (roughly 400 miles of damaging winds, or 650 km, in the more recent stricter definition of Corfidi et al. 2016 that you cite), why you would shorten the MCS longevity requirement to just 6 hours? If one assumes a typical MCS translation speed of 50 km/h, then an MCS lasting only 6 hours would only move over an area 300 km long, or less than half the pathlength needed in this prior definition of derechos. Even with the older 400 km pathlength requirement that you mention, many events would not meet it if moving at 50 km/h. Also, even if one acknowledges that derechos often move rather quickly (I believe the 2020 Midwestern one moved at around 90-100 km/h, which is about as fast as they can move), a 6-hour lifetime would still not result in a damage path quite long enough to match the 650 km definition. It is true that MCSs can produce significant damage with wind swaths shorter than 650 km, but if that is the focus of your work, you should not be referring to it as a derecho climatology. A high wind-producing bow echo climatology would be more appropriate for the title. In fact, in light of my opening paragraph

and the ongoing efforts to change the definition of a derecho substantially, this change in title might avoid even more serious problems. It much better defines what your tool does since it is broader in its wording. I would strongly suggest that change.

Reply:

Thank you for your comments. The wording in the last version of our manuscript might cause some confusion. We have improved the wording of some terms, as explained at the beginning of the response, which we hope can address your concern. In detail, the study aims to identify derechos (and derecho-producing MCSs) from the MCS dataset, which comprises derecho-producing MCSs and non-derecho-producing MCSs. Using a lifetime threshold of 6 hours in our MCS detection algorithm does not affect our derecho (and derecho-producing MCS) detection, which explicitly requires a gust swath of at least 650 × 100 km^2 associated with an MCS (Line 418 in the revision-tracked main manuscript, or as below).

"The gust swath must be at least 650 km in length and 100 km in width."

According to our current dataset, the average derecho and derecho-producing MCS lifetimes are 11.4 and 27.8 hours respectively when we use ISD in our derecho detection. The numbers change to 10.5 and 29.4 hours when SED is used in the derecho algorithm.

According to our dataset, the typical derecho propagation speed is about 25 m s^{-1}, hence 650 km is roughly equivalent to 7 hours. Considering the potential CNN bow echo identification uncertainties, we change the lowest derecho duration from 2 hours to 5 hours in Line 397 in the revision-tracked main manuscript. In fact, the lifetime threshold is not so important since we have set a swath threshold of 650 × 100 km^2, which already implies a duration threshold.

"The derecho must persist for at least 5 hours, with a bow echo present for at least 80% of its lifetime."

I am troubled by your use of the surface wind station database for the winds used in your system. This is NOT how derechos are classified operationally now. Instead, classification is based on the Storm Events Database (SED), which I would assume includes more actual measurements (such as home weather stations) and estimates of wind speed based on damage. In fact, roughly 90% of all the severe wind reports in that database do not involve a measurement and are instead estimates based on damage. The definition of a derecho that I found on SPC's website does not restrict derecho classification to just measured winds. The Storm Events Database is rather robust for the period 2004-2021 that you are using, and thus it is puzzling why you would not have used that for your training? Your use of just the surface station database is even more puzzling considering my comment earlier that your reduction in the longevity requirement for MCSs would lead your system to call many events derechos that would not meet the past requirements of a swath of severe winds at least 650 km long (or even 400 km long in many situations). Thus, you made one choice that really makes it easier to call a system a derecho, but then this choice of where you obtain wind information would do the opposite, making it much more difficult for systems to meet prior thresholds to be called a derecho. Tirone et al. (2024) used the SED in their training of a ML tool, so there is no obstacle to using that information as a source of thunderstorm wind information. I believe you need to

test the sensitivity of your results to your use of a very limited database of thunderstorm winds, by examining changes that happen if you switch to using the SED. I see that later starting at about line 350, you provide good explanation which I think deals with my concern. It seems you acknowledge the deficiencies in your choice of this surface data network for wind and thus make numerous modifications to try to account for the deficiencies (like lower wind thresholds, use of broader ellipse containing reports, etc). You talk some about how damage estimates take some time to be performed, but it is unclear if this is the primary motivation for your use of the problematic surface wind network dataset. I believe you need to provide some of this justification earlier when you first mention that surface network, since most readers familiar with severe weather reports will question why you are not using the SED. Derechos are often classified now within 24 hours of their occurrence as the process of gathering both the measurements and having some estimates is quick. From what I have read so far, I do not see a reason why 24 hours is too long for what you are doing. You need to make a stronger case for your use of the surface wind measurements. Are you planning for your tool to be used operationally in a setting where it must alert forecasters the moment that a system has reached the requirement to be called a derecho? I guess I do not see why this would be so urgent. With all ongoing derecho research that I am aware of, your tool could easily be applied one day after an event, or even a week or month afterward, so there does not seem to be a valid reason to avoid using the SED.

Reply:

Thank you for your comments. Please see our above answers and clarifications. Besides, we also want to clarify that our AI tool is only used to detect bow echoes from composite radar reflectivity. The training of our bow echo detection tool does not use gust speed.

When you refer to bow echo samples on line 180, it would be helpful to know the time resolution of the radar images you use. In recent years, radar reflectivity images often update nearly every minute or two. For readers to be able to put into context your 566 positive identifications, they need to be able to figure out how many images in total might be getting evaluated in the normal lifetime of a derecho. It is possible the citation about Gridrad would mention the time frequency of the Gridrad products, but this is simple information to supply in your own paper and is absolutely necessary. For a derecho lasting 8 hours, if radar images are available every 5 minutes (the traditional frequency of NEXRAD scans), there would be 96 for one case, and thus your 566 total bow echo scans from 54 events would be a tiny fraction of the lifetime of the derechos. If the images are hourly, then 566 implies every hour of every derecho must include a bow echo. Thus, this information is critically important.

Reply:

Thank you for your comment and suggestion. Gridrad has a temporal resolution of 1 hour. We added "hourly" before "Z_{Hmax}" in Line 230 in the revision-tracked main manuscript, as well as Lines 143 and 144.

'We manually label times with apparent bow echoes through visual inspection of hourly Z_{Hmax} associated with the tracked DMCSs.'

I believe the results you state starting at about line 422 reflect some of the harm done by choices you made that I take issue with in my earlier comments. You do mention your reduction in wind threshold, but this could likely have been avoided had you used the SED instead of the surface wind network. Likewise, your overestimate compared to other studies is probably influenced by your unusual choice to reduce the MCS longevity criterion. You should try using a longer threshold than 6 hours to see how the numbers change. I do not believe you have made a case as to why you needed to reduce it to 6 hours. If your line 433 is implying that there may be derechos in the SPC database that are not really derechos since they did not come from organized convection, I highly doubt that. SPC usually avoids even showing wind damage reports unless they are clearly due to convection. There is no way I can conceive that a system would get listed as a derecho, because of that path length requirement), if it did NOT come from an MCS. Perhaps I have misinterpreted what you say in line 433, but if so, you need to be more direct in explaining why your MCS criteria is so important. From what I know about the SPC database, all events being called derechos had to be MCSs. For your AI work, obviously you need to have some criteria to ensure a system is an MCS, but when humans classify derechos, they are automatically ensuring this. Thus, I think the only impact of your MCS criteria would be if you are comparing your number of cases to numbers in other studies that used AI to classify derechos (and presumably to explain why you should end up with a smaller number than other studies that may not have bothered to ensure the winds were happening due to an MCS). In the context here, in your paper, however, I do not believe it makes sense.

Reply:

Thank you for your comment and suggestion. Since we now include the SED-based dataset in the study, we have rewritten Section 5 about the evaluation and uncertainties of our dataset. Please also see above for our answers and clarifications.

Your discussion around line 450 again seems like it is needed there only because you did not use the SED database. If you had, you could avoid having to explain so many possible caveats.

Once again, starting around line 552, this discussion is reflecting the serious flaw in your design of your experiment. You chose to define derechos in a way inconsistent with the already problematic fact that at least two standard definitions exist. It is like you chose to study cats, but are calling them dogs, and now you are having to explain numerous differences between your study and prior ones. Because you used a different definition, it is impossible to know how serious the differences are. As I stated earlier, you are really just identifying bow echoes that were associated with strong winds. Not derechos. The question here becomes, is there a physical reason why strong wind bow echoes do not show the NW-SE swath of enhanced occurrence that is present for derechos? Or is there some fundamental issue with AI that is resulting in the difference. We cannot know because you chose to compare apples to oranges. The sensitivity test you do on line 554 is interesting and may offer some insight into my question above. But I believe your neglect of using the SED still complicates the interpretation you are providing here.

Reply:

Thank you for your comment. In this study, we do not propose a new derecho definition but follow the definition from Corfidi et al. (2016). However, due to the limitation of some source datasets, we modify some thresholds, especially for those related to ISD gust speeds. This is a compromise. The derecho definition from Corfidi et al. (2016) is different from conventional ones. Since there are no other derecho climatologies based on Corfidi et al.'s definition, we can only compare our dataset with results based on conventional definitions. We disagree that this is a comparison between apples and oranges. The derecho definition is continuously evolving. Even the one in the AMS abstract you mentioned is different from previous definitions. On the other hand, the NW-SE swath of enhanced occurrence is observable in our updated yearly and monthly climatology (Figures 9, 11, S6, and S7), especially in July. For SED, please see our answers and clarifications in the above.

In lines 593-594, it would have been good to try to compare your rather large percentage of wind reports being due to derechos and DFs to an estimate of what prior studies showed. I have a feeling your number is much higher, which is consistent with the fact that you are actually studying all bow echoes that produce strong wind, and not true derechos.

Reply:

Thank you so much for pointing out the issue. We made a mistake with the calculation. We averaged the fractions of all sites. Since many sites have a few damaging gusts but the fractions are ~100%, averaging the fractions would overestimate the weight of sites with a few damaging gusts. We now just count the number of damaging gusts regardless of the sites, and the fraction reduces to ~3.1% based on our new dataset. Please see Lines 35 and 821 in the revised main manuscript.

"Additionally, during the study period, derechos account for approximately 3.1% of measured damaging gusts ($\geq$ 25.93 m s^{-1}) over the eastern United States."

"On average, DMCSs contribute 4.0%, and derechos contribute 3.1% of all damaging gust occurrences."

References

Corfidi, S. F., Coniglio, M. C., Cohen, A. E., and Mead, C. M.: A proposed revision to the definition of "derecho", Bulletin of the American Meteorological Society, 97, 935-949, https://doi.org/10.1175/BAMS-D-14-00254.1, 2016.

Response to Reviewer #2

Thank you for your careful and thorough reading of the manuscript and your thoughtful comments and suggestions. We apologize for the delay in revising the manuscript, as we spent a significant amount of time on the manual validation and inspection of the identified derechos. According to all three reviewers' comments and suggestions, we have made several significant improvements in our manuscript, which we want to highlight before the point-by-point response.

Firstly, we have improved the language of many sentences so that the manuscript is more understandable, and we have also changed the wording of some terms so that they are more distinguishable and accurate. A derecho is a windstorm, while an MCS is a convective system. They are different concepts. Therefore, we change "derecho" to "derecho-producing MCS" when we refer to the MCS associated with a derecho, and "derecho feature" has been changed to "derecho." The dataset developed in this study includes tracking of both derechos and corresponding derecho-producing MCSs.

Secondly, we have incorporated a "forward propagating" criterion in our derecho detection algorithm. Our initial understanding of the term "forward propagating" was incomplete, and we failed to recognize its critical role in defining a derecho proposed by Corfidi et al. (2016). After careful consideration and evaluation, we have adopted and modified two criteria from Corfidi et al. (2016) to establish the definition of "forward propagating." One is that the acute angle between the averaged bow echo orientation and the bow echo series' propagation direction is larger than 45°, and the other is that the bow echo series' propagation speed is at least 30% faster than the 500-hPa background wind speed. Implementing the "forward propagating" criterion removes many windstorms externally forced by extratropical cyclones, aligning with the purpose of Corfidi et al. (2016), which intends to define derechos as internally driven windstorms. "Externally forced" and "internally driven" reflect distinct physical formation mechanisms of those windstorms, which is why Corfidi et al. (2016) proposed a physically based derecho definition. With the updated detection algorithm, the derecho number between 2004 and 2021 has been substantially reduced from 556 to 274 (for ISD) and 220 (for SED). In addition, due to the inclusion of the "forward propagating" constraint in our derecho algorithm, we have decided not to change the name of our dataset to "high wind-producing bow echo." The derecho detection algorithm description has been updated accordingly in Section 4, and we have also updated all the results in Section 6 based on the improved dataset.

Thirdly, we have developed another parallel dataset using gust speeds from the Storm Events Database (SED). Now, our derecho dataset consists of two subsets: one based on gust measurements from the global hourly Integrated Surface Database (ISD) and the other based on SED gust speeds. Although there are some discrepancies between the two subsets, their agreement is much larger than their difference (Figures 8-11 in the revised main manuscript; or Figure R1 below). Since both ISD and SED gust speeds have limitations and uncertainties, minor differences between the ISD-based and the SED-based datasets are expected and understandable. In addition, the general agreement between the two datasets indicates that our usage of lower gust speed criteria for ISD measurements is reasonable and does not change the derecho number much. We must emphasize that using lower gust speed criteria for ISD measurements than SED reports does not mean that the ISD-identified derechos are weaker than the SED-identified ones

(or even not derechos). This is a compromise, considering that ISD stations are limited and may miss many damaging gusts, as we highlighted in Lines 522-532 in the revision-tracked main manuscript (as below).

"We emphasize that, in Criterion 4, our ISD gust speed criteria are weaker than the SED gust speed criteria as well as those of previous studies (Squitieri et al., 2023; Bentley and Mote, 1998; Johns and Hirt, 1987), which also estimated the gust swath based on SED damaging gusts. As mentioned in Section 2.2.2, most SED gust reports are estimates, while ISD provides gust measurements from weather stations. SED estimates can capture potential damaging gust occurrences over a much larger area, although with large uncertainties. In contrast, due to the limited coverage of observational sites, real-time ISD measurements may miss substantial damaging gust occurrences in nearby regions. Therefore, we lower the gust speed criteria to capture potential derechos when using ISD measurements. It does not mean that the ISD-based derechos are weaker than the SED-based ones or even not derechos, as elaborated in Section 5."

Fourthly, due to the incorporation of "forward propagating" in the detection algorithm and the development of the SED-based derecho dataset, we have updated some sensitivity tests, evaluations, and the comparisons of our datasets with the NOAA SPC data in 2004 and 2005 and previous studies. Please see Section 5 for further details.

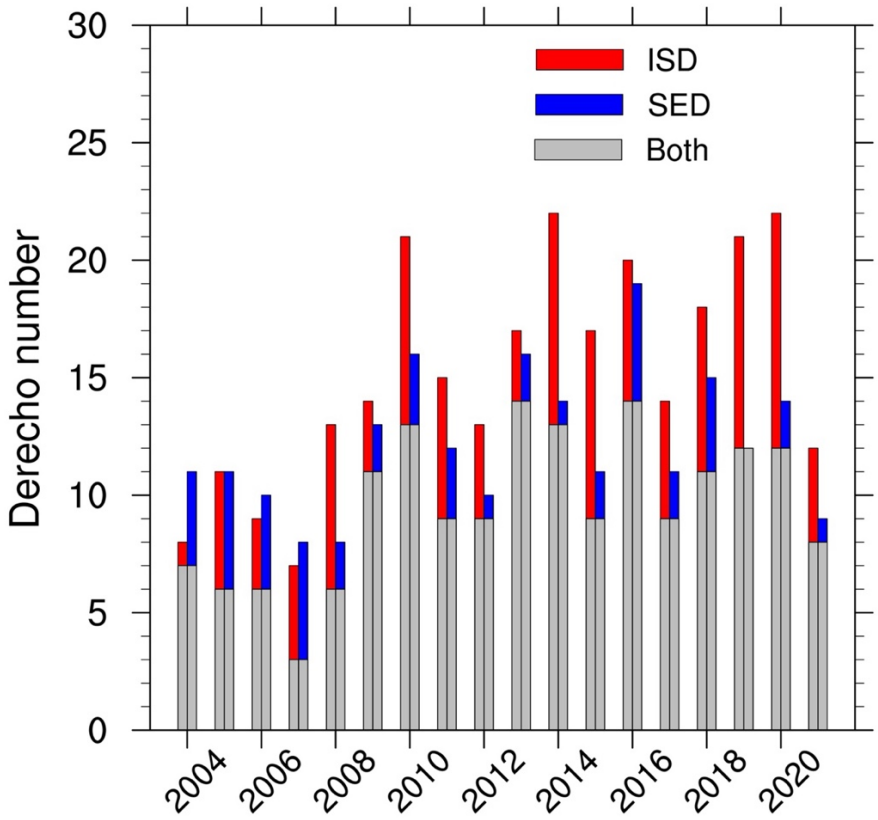

Figure R1. Bar chart of the annual derecho numbers from the ISD-based and the SED-based datasets from 2004 to 2021. Gray shading denotes derechos captured by both datasets, red shading refers to derechos only identified when using ISD gust observations, and blue shading represents SED-only derechos. The figure is the same as Figure 9 in the revised main manuscript.

This article describes the development of a machine learning approach to create a derecho climatology across the United States. The novelty and originality of the work should be praised. The authors for the the most part have a well reasoned approach and methodology to creating this dataset, however there are a few major items of concern that stood out during this review:

I struggle with understanding which definition of a derecho the authors are using and also relying on to classify a feature as a derecho. In the background/introduction the authors present a history on the evolution of the definition of a derecho. I encourage the authors to keep this in the introduction, but I also encourage to authors to present the definition of a derecho they chose for their methodology clearly and provide additional reasoning on why the this specific definition was chose. The authors should really try to use a definition that most closely represents the official definition used by the National Weather Service and/or Storm Prediction Center. Using a definition that either has a shorter length requirement (or longer one) would impact the number of derechos that are classified in your results.

Reply:

Thank you for your comments and suggestions. As explained above, we change the wording of "derecho features" and hope the manuscript is more understandable now. In the introduction, we highlight that this study aims to develop a derecho dataset following the definition proposed by Corfidi et al. (2016) (Lines 95-101 and 114-118 in the revised main manuscript, or as below). We did not intend to change anything in Corfidi et al.'s definition, but because of the limitations of the wind gust datasets and uncertainties of bow echo identification, we modified some thresholds and introduced the details of how we apply Corfidi et al.'s definition to our available datasets in Section 4. This is a compromise but not a change of definition. Follow your suggestions, we have rewritten Section 4: we first provide a relatively simple derecho definition (Lines 394-419 in the revision-tracked main manuscript, or as below) and then explain the details separately.

"Considering the inconsistent thresholds used in the above studies and the lack of physical mechanisms in their derecho definitions, Corfidi et al. (2016) proposed a stricter and more physically based derecho definition, which required the existence of sustained bow echoes with mesoscale vortices or rear-inflow jets and a nearly continuous wind damage swath of at least 100 km wide along most of its extent and 650 km long. In addition, the wind damage must occur after the convective system was organized into a cold-pool-driven forward-propagating MCS."

"This 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, which are then combined with an MCS tracking dataset and surface gust speeds to identify derechos using criteria adjusted from Corfidi et al. (2016)."

"Our final criteria are summarized below, with detailed explanations provided afterward (Figure 6).

1) A derecho must be attached to an MCS from the MCS dataset.

2) The derecho must persist for at least 5 hours, with a bow echo present for at least 80% of its lifetime. In addition, gaps between successive bow echo occurrences cannot exceed two hours. All bow echoes must belong to the same bow echo series, as defined in the subsequent explanation.

3) The derecho bow echo series must exhibit forward propagation, based on two modified criteria from Corfidi et al. (2016):

    - The acute angle between the averaged bow echo orientation and the bow echo series' propagation direction must exceed 45° (Figure 6).
    - The propagation speed of the bow echo series must be at least 30% greater than the background mean wind speed at 500 hPa, derived from ERA5 data. The methodology for calculating the averaged bow echo orientation, bow echo series' propagation direction and speed, and the background mean wind speed is detailed in Appendix A.

4) Derecho-associated gust speed criteria vary based on the gust speed source dataset:

    - For ISD data: Within 100 km of the derecho-accompanied bow echoes (termed the "derecho area"), there must be at least 10 sites with strong gusts (≥ 17.43 m s^{-1}) and at least 1 site with damaging gusts (≥ 25.93 m s^{-1}).
    - For SED data: At least 10 locations must report damaging gusts.
    - The fraction of sites with strong/damaging gusts (ISD) or damaging gusts (SED) must be ≥ 20%.
    - Gaps between successive strong (ISD) or damaging (SED) gust reports cannot exceed two hours.
    - The gust swath must be at least 650 km in length and 100 km in width. Swath length and width calculations are explained below."

I do not understand the inclusion of surface wind speed observations in this manuscript. Derechos are classified operationally through the Storm Events Database (i.e. local storm reports), not through surface wind observations.

Reply:

Please see above our explanation and clarification.

The organization of introduction needs quite a bit of improvement as well. It was very difficult to follow in terms of readability, partly compounded by the presentation of all the definitions of derechos. The introduction also presents Figure 1 which is a very very busy figure and in its current form, takes away from the paper. I recommend the authors overhaul the section to provide clarity on previous research, the definition of a derecho and motivation for their great ideas as far as developing this database.

Reply:

Thank you for your comments and suggestions. Considering many researchers are not familiar with derechos, Figure 1 can help them understand the basic characteristics of the phenomenon quickly. We have simplified Figure 1 per your request and reorganized the sentences for the first derecho definition in Lines 69-82 in the revision-tracked main manuscript (as below). Also, as mentioned at the beginning of this document highlighting the key changes in the revision, we have changed the wording of some terms and hope these changes can improve the reading of the manuscript.

'Specifically, they required a derecho to satisfy the following six criteria.

1) There must be a concentrated area of reports with wind damage or convective gusts > 25.7 m s^{-1}, and the major axis length of the area must be at least 400 km.
2) Those wind damage or convective gust reports must show a pattern of chronological progression, either as a singular swath or a series of swaths.
3) The concentrated area must have at least three reports of either F1 damage (32.7-50.3 m s^{-1}) (Fujita, 1971) or convective gust of at least 33.4 m s^{-1} separated by ≥ 64 km.
4) At most 3 hours can elapse between successive reports of wind damage or gust > 25.7 m s^{-1}.
5) The associated convective system must have temporal and spatial continuity in surface pressure and wind fields.

6) If multiple swaths of wind damage or gust reports > 25.7 m s^{-1} exist, they must be from the same MCS event.'

It is difficult to evaluate the results that are presented, especially with the current derecho definition that is used. The current definition that is used (and with sfc wind obs) makes the number of derechos classified by this current form of research difficult to believe. Hopefully an overhaul in the definition used will provide a more realistic number of derechos identified. I do like Figures 9, 10, and 11 in presenting the results. These are great and easy to interpret graphics. I encourage these graphics to stay but adjusted with potential adjustments from the reviews. I would like to see potentially see how the next iteration of these graphics compare to actual confirmed derechos from the same time period.

Reply:

Thank you for your comments. We want to clarify that we follow the derecho definition proposed by Corfidi et al. (2016), which is different from conventional definitions. Since a derecho climatology based on Corfidi et al.'s definition is unavailable, we can only compare our results with prior studies based on conventional definitions. For the details on how we improve the manuscript, please see our response at the very beginning.

References

Corfidi, S. F., Coniglio, M. C., Cohen, A. E., and Mead, C. M.: A proposed revision to the definition of "derecho", Bulletin of the American Meteorological Society, 97, 935-949, https://doi.org/10.1175/BAMS-D-14-00254.1, 2016.

Response to Reviewer #3

Thank you for your careful and thorough reading of the manuscript and your thoughtful comments and suggestions. We apologize for the delay in revising the manuscript, as we spent a significant amount of time on the manual validation and inspection of the identified derechos. According to all three reviewers' comments and suggestions, we have made several significant improvements in our manuscript, which we want to highlight before the point-by-point response.

Firstly, we have improved the language of many sentences so that the manuscript is more understandable, and we have also changed the wording of some terms so that they are more distinguishable and accurate. A derecho is a windstorm, while an MCS is a convective system. They are different concepts. Therefore, we change "derecho" to "derecho-producing MCS" when we refer to the MCS associated with a derecho, and "derecho feature" has been changed to "derecho." The dataset developed in this study includes tracking of both derechos and corresponding derecho-producing MCSs.

Secondly, we have incorporated a "forward propagating" criterion in our derecho detection algorithm. Our initial understanding of the term "forward propagating" was incomplete, and we failed to recognize its critical role in defining a derecho proposed by Corfidi et al. (2016). After careful consideration and evaluation, we have adopted and modified two criteria from Corfidi et al. (2016) to establish the definition of "forward propagating." One is that the acute angle between the averaged bow echo orientation and the bow echo series' propagation direction is larger than 45°, and the other is that the bow echo series' propagation speed is at least 30% faster than the 500-hPa background wind speed. Implementing the "forward propagating" criterion removes many windstorms externally forced by extratropical cyclones, aligning with the purpose of Corfidi et al. (2016), which intends to define derechos as internally driven windstorms. "Externally forced" and "internally driven" reflect distinct physical formation mechanisms of those windstorms, which is why Corfidi et al. (2016) proposed a physically based derecho definition. With the updated detection algorithm, the derecho number between 2004 and 2021 has been substantially reduced from 556 to 274 (for ISD) and 220 (for SED). In addition, due to the inclusion of the "forward propagating" constraint in our derecho algorithm, we have decided not to change the name of our dataset to "high wind-producing bow echo." The derecho detection algorithm description has been updated accordingly in Section 4, and we have also updated all the results in Section 6 based on the improved dataset.

Thirdly, we have developed another parallel dataset using gust speeds from the Storm Events Database (SED). Now, our derecho dataset consists of two subsets: one based on gust measurements from the global hourly Integrated Surface Database (ISD) and the other based on SED gust speeds. Although there are some discrepancies between the two subsets, their agreement is much larger than their difference (Figures 8-11 in the revised main manuscript; or

Figure R1 below). Since both ISD and SED gust speeds have limitations and uncertainties, minor differences between the ISD-based and the SED-based datasets are expected and understandable. In addition, the general agreement between the two datasets indicates that our usage of lower gust speed criteria for ISD measurements is reasonable and does not change the derecho number much. We must emphasize that using lower gust speed criteria for ISD measurements than SED reports does not mean that the ISD-identified derechos are weaker than the SED-identified ones (or even not derechos). This is a compromise, considering that ISD stations are limited and may miss many damaging gusts, as we highlighted in Lines 522-532 in the revision-tracked main manuscript (as below).

"We emphasize that, in Criterion 4, our ISD gust speed criteria are weaker than the SED gust speed criteria as well as those of previous studies (Squitieri et al., 2023; Bentley and Mote, 1998; Johns and Hirt, 1987), which also estimated the gust swath based on SED damaging gusts. As mentioned in Section 2.2.2, most SED gust reports are estimates, while ISD provides gust measurements from weather stations. SED estimates can capture potential damaging gust occurrences over a much larger area, although with large uncertainties. In contrast, due to the limited coverage of observational sites, real-time ISD measurements may miss substantial damaging gust occurrences in nearby regions. Therefore, we lower the gust speed criteria to capture potential derechos when using ISD measurements. It does not mean that the ISD-based derechos are weaker than the SED-based ones or even not derechos, as elaborated in Section 5."

Fourthly, due to the incorporation of "forward propagating" in the detection algorithm and the development of the SED-based derecho dataset, we have updated some sensitivity tests, evaluations, and the comparisons of our datasets with the NOAA SPC data in 2004 and 2005 and previous studies. Please see Section 5 for further details.

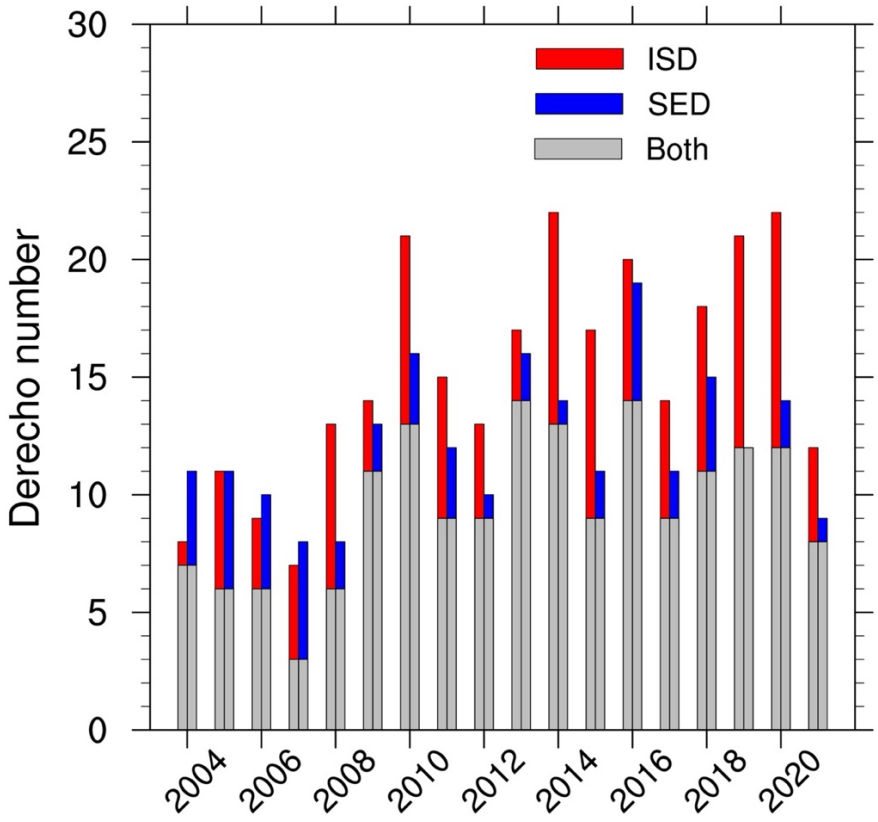

Figure R1. Bar chart of the annual derecho numbers from the ISD-based and the SED-based datasets from 2004 to 2021. Gray shading denotes derechos captured by both datasets, red shading refers to derechos only identified when using ISD gust observations, and blue shading represents SED-only derechos. The figure is the same as Figure 9 in the revised main manuscript.

Major Comments

This paper presents a novel, machine-learning scheme to objectively identify derecho-producing convective systems. The paper reads well and, not withstanding the minor comments indicated below, to this reviewer seems both well-organized and well-presented. I have only two main comments.

First, being familiar with the vagaries of the severe-weather report database, I support the authors' use of the Integrated Surface Database (ISD). While the ISD arguably is subject to its own limitations, the quality-control algorithms used offer a higher and more universal level of uniformity than that associated with the severe-weather report database. The large number of derecho-producing convective systems identified with the current approach of the authors compared with those of previous studies largely reflects the rather low wind threshold employed; use of a somewhat higher threshold (and/or duration threshold) would lower the number of events identified. Obviously the true frequency of derecho-producing events remains unknown; the lower frequencies suggested by previous studies may be somewhat low.

Reply:

We sincerely appreciate your support for our usage of the ISD. To make the study more reliable and robust, and in response to the other reviewers, we have developed an additional dataset based on SED. Comparison between the two datasets justifies the usage of lower wind thresholds for the ISD. The overestimation in our initial manuscript is mainly due to incomplete implementation of the "forward propagating" criterion from Corfidi et al. (2016). Revising the "forward propagating" in our detection algorithm removes many externally forced systems. For more details, please see the highlights of our responses above.

The "true frequency" point brings into mind the main purpose of the present study --- objective identification of derecho-producing MCSs. The number of systems identified is sensitive to the underlying definition used in the scheme. This is where difficulty has arisen in the past and to some extent continues with the present paper. The omission of "forward propagating" from the current definition of the authors (page 516 ff) is problematic. Sustained forward-propagation is a fundamental aspect of derecho-producing convective systems. In absence of such a criterion one could argue that the approach described in the present paper is closer to that of a bow-echo detection scheme. Derecho-producing convective systems could be described as arising from bow-echo producing processes --- including rapid, sustained forward propagation --- that remain active for extended periods of time. I suggest re-visiting the abortive attempts made (lines 517-519) to identify the presence of forward propagation and refine the ML approach used here.

Reply:

Thank you so much for the suggestions. As mentioned above, our initial understanding of "forward propagating" is incomplete. We thought it was just a direction metric. After careful reading of Corfidi et al. (2016), we realize the importance of this criterion and its relationship with physical processes. After implementing the "forward propagating" criterion (Lines 401-408 in the revision-tracked main manuscript), the dataset quality has been much improved.

"The derecho bow echo series must exhibit forward propagation, based on two modified criteria from Corfidi et al. (2016):

- The acute angle between the averaged bow echo orientation and the bow echo series' propagation direction must exceed 45° (Figure 6).
- The propagation speed of the bow echo series must be at least 30% greater than the background mean wind speed at 500 hPa, derived from ERA5 data. The methodology

for calculating the averaged bow echo orientation, bow echo series' propagation direction and speed, and the background mean wind speed is detailed in Appendix A."

Minor comments (numbers refer to line numbers in version of 24 June 2024)

40. Change "magnitude" to "impact"

Reply*:*

Corrected. Please see Line 49 in the revision-tracked main manuscript. Thank you!

"Ashley and Mote (2005) demonstrated that derechos could be as hazardous as and were comparable in impact to most hurricanes and tornadoes in the United States between 1986 and 2003."

190. Consider adding a parenthetical description of "skip connections"

Reply*:*

Added. Please see Line 243-244 in the revision-tracked main manuscript. Thank you!

"Dense Nets are notable for their large number of skip connections (which create multiple paths for data to flow through the network without passing through every layer), and they can achieve comparable performance to very large classifier CNNs with only a fraction of the trainable parameters."

201. Clarify what is meant by "more distinct"

Reply*:*

The initial 556 positive samples are from 54 named derechos. The new negated 500 positive samples are from > 100 convective systems. Here, "more distinct" just refers to the samples from diverse "sources" to improve its representation.

267. Define or reference "binary cross entropy loss"

Reply*:*

A reference added. Please see Line 323 in the revision-tracked main manuscript. Thanks.

"It is trained using binary cross entropy loss (Bishop, 2006) on masks generated from its 384, 192, 96, 48, 24, and 12-pixel resolution feature representations (Huang et al., 2020)."

307-310. Well-stated

Reply:

Thank you!

312. Not completely sure what is meant by the "upper" and "lower" parts of the table

Reply:

Corrected to “upper triangular” and “lower triangular”. Please see Lines 369-370 in the revision-tracked main manuscript. Thank you!

366. Add parenthetically, "Derecho feature" after "DF"

Reply:

Thank you for your suggestion. We change the wording of “derecho feature” to “derecho” throughout the manuscript.

371-372. Not sure that this criterion would always be helpful...

Reply:

It is not always helpful. But it does help remove some extratropical cyclones, especially for the ISD. We manually check every identified derecho and high-wind-producing convective system to ensure the criterion does not produce any adverse impact.

385. Makes sense!

Reply:

Thank you! By comparing our SED-based and ISD-based subsets, we find lowering the gust speed criterion is reasonable for the ISD.

433-434. Agree with this focus; consider also maintaining the forward-propagating aspect.

Reply:

Thank you for your suggestion. The sentence has been deleted since we rewrite the evaluation section (Section 5) due to the incorporation of forward-propagating and the inclusion of SED.

455. Should "as" be "than"?

Reply:

The sentence has been deleted. Thank you.

480. Consider parenthetically adding "cold-cloud shield" given that acronym has not been used since line 129.

Reply*:*

Added. Please see Line 673 in the revision-tracked main manuscript. Thank you!

"The 50% areal overlap threshold in PyFLEXTRKR, which links consecutive cold cloud shields (CCSs), may fail to capture very fast-moving convective systems using hourly satellite and NEXRAD data."

520. The addition of a simple schematic to illustrate the angles mentioned would be helpful

Reply*:*

Thank you for the suggestion. We have updated the algorithm schematic (as Figure R2 below) to illustrate the angle. Besides, in Appendix A, we provide more details on how we calculate it.

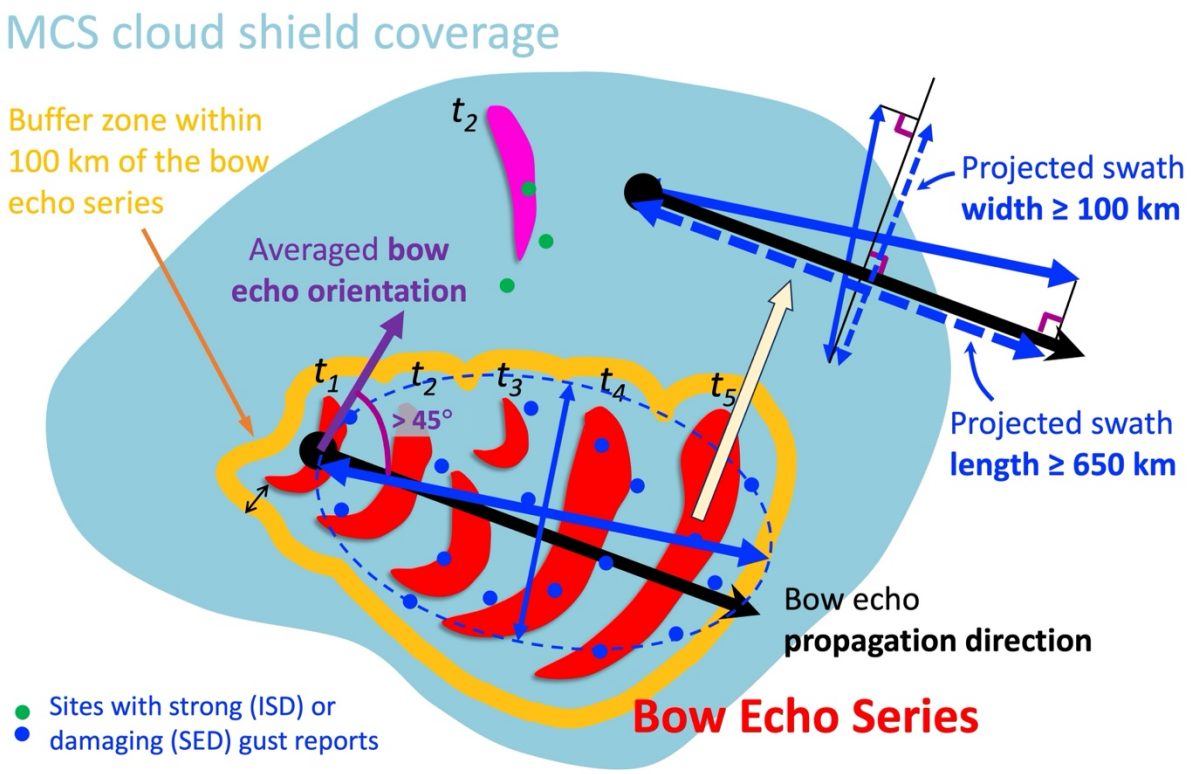

Figure R2. Schematic of the automated detection algorithm.

525-526. Good to see this explicitly stated

Reply*:*

Thank you for positive comment. Unfortunately, our understanding of “forward propagating” was incomplete. We remove the sentence and incorporate “forward propagating” in the derecho detection algorithm.

556. Should "west-east" be "northwest-southeast"?

Reply*:*

The sentence is deleted. We remove this sensitivity test, which is unnecessary after the development of the SED-based dataset.

588. Should "2014" in Figure 12 caption be "2004"?

Reply*:*

Corrected. Please see Line 815 in the revision-tracked main manuscript. Thank you for pointing out the error.

676. Capitalize "Weather" and "Review"

Reply*:*

Corrected. Please see Line 978 in the revision-tracked main manuscript. Thank you!

712. Add publication in which this manuscript appeared

Reply*:*

This is a report, and we add the website of the report. Please see Line 1025 in the revision-tracked main manuscript.

731. Add year of publication (2020 (?), per line 234)

Reply*:*

Added. Please see Line 1047 in the revision-tracked main manuscript. Thank you!

738. Capitalize "Python"

Reply*:*

Corrected. Please see Line 1055 in the revision-tracked main manuscript. Thank you!

791. Capitalize "Atmospheric Sciences"

Reply*:*

Corrected. Please see Line 1116 in the revision-tracked main manuscript. Thank you!

References

Corfidi, S. F., Coniglio, M. C., Cohen, A. E., and Mead, C. M.: A proposed revision to the definition of “derecho”, Bulletin of the American Meteorological Society, 97, 935-949, https://doi.org/10.1175/BAMS-D-14-00254.1, 2016.