1   Supplementary information for

2   A derecho climatology (2004-2021) in the United States

3   based on machine learning identification of bow echoes

4   Jianfeng Li[1, *], Andrew Geiss[1], Zhe Feng[1, *], L. Ruby Leung[1] Yun Qian[1], Wenjun Cui[2]

6 [1]Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory,
7 Richland, Washington, USA

8 [2]Cooperative Institute for Severe and High-Impact Weather Research and Operations, University
9 of Oklahoma, Norman, Oklahoma, USA

10 [3]National Severe Storms Laboratory, National Oceanic and Atmospheric Administration,
11 Norman, Oklahoma, USA

12 *Correspondence to Jianfeng Li (jianfeng.li@pnnl.gov) and Zhe Feng (zhe.feng@pnnl.gov)

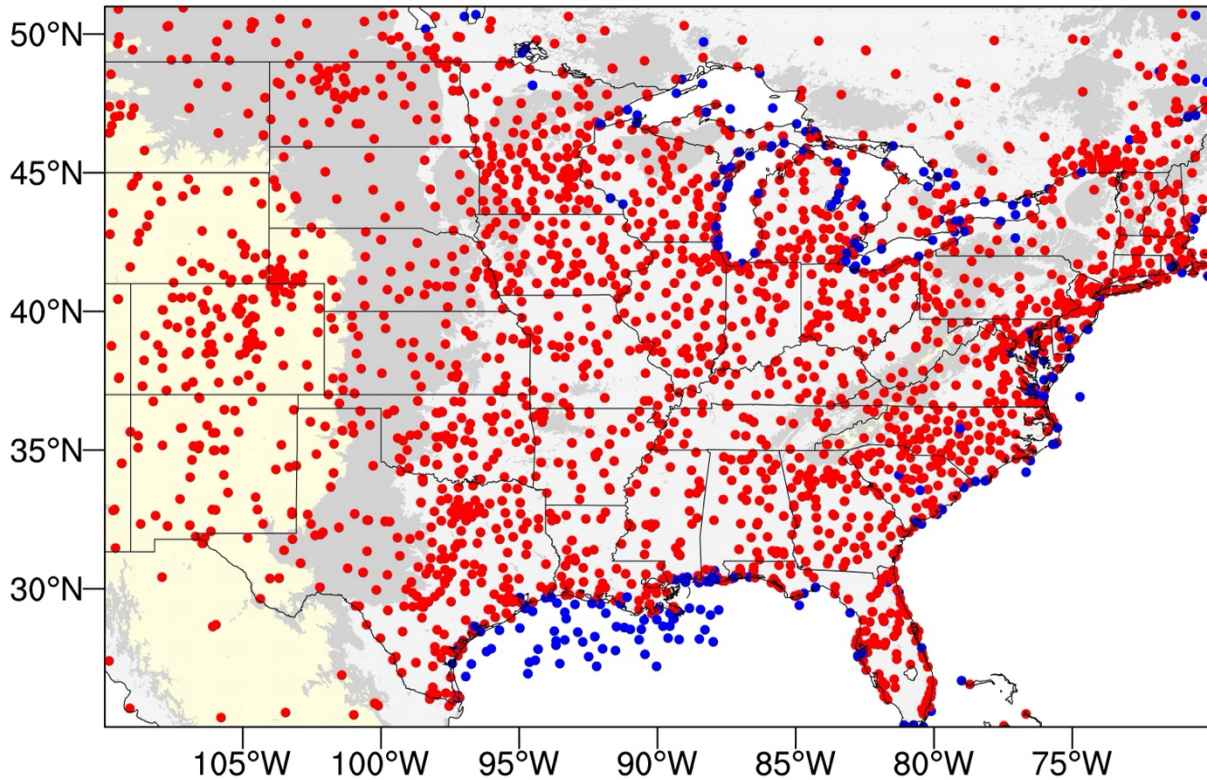



Figure S1. Locations of gust speed observational sites used in the study. There are 4,260 sites, 3,954 of which are over land (red points), while the rest are over the ocean or lakes (blue points). We use the Advanced Research Weather Research and Forecasting (WRF) Preprocess System (WPS) to generate a 4-km land cover map to determine the land type associated with each observational site. Light-yellow shading denotes an elevation greater than 1000 m; light-gray shading denotes an elevation between 400 m and 1000 m; and smoke-white shading denotes an elevation less than 400 m. Background white is for the ocean and lakes.

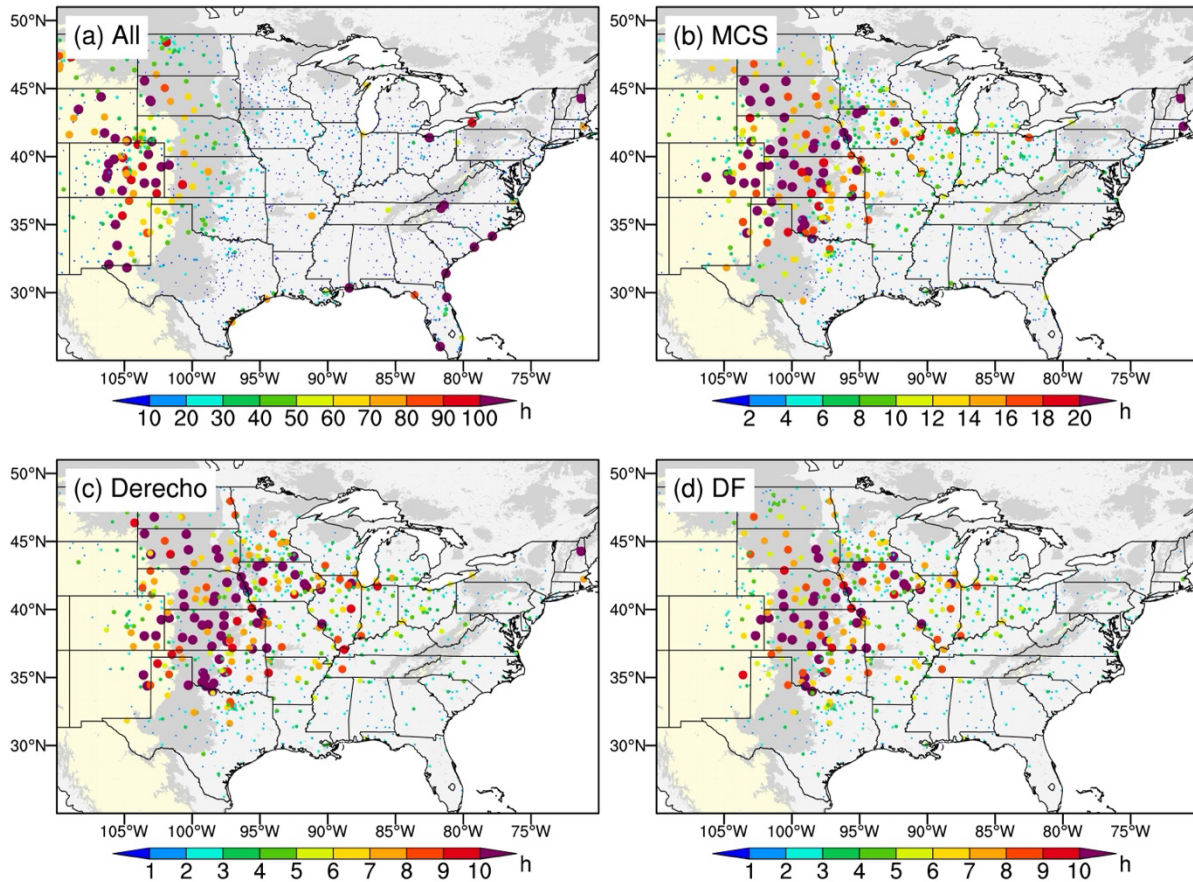



Figure S2. Same as Figure 13, but the frequencies of damaging gust occurrences from mesoscale convective system (MCS), derecho, and derecho feature (DF) are shown in (b), (c), and (d) instead of fractions. Non-derecho MCS events overlapping with tropical cyclones (TCs) are excluded in (b).

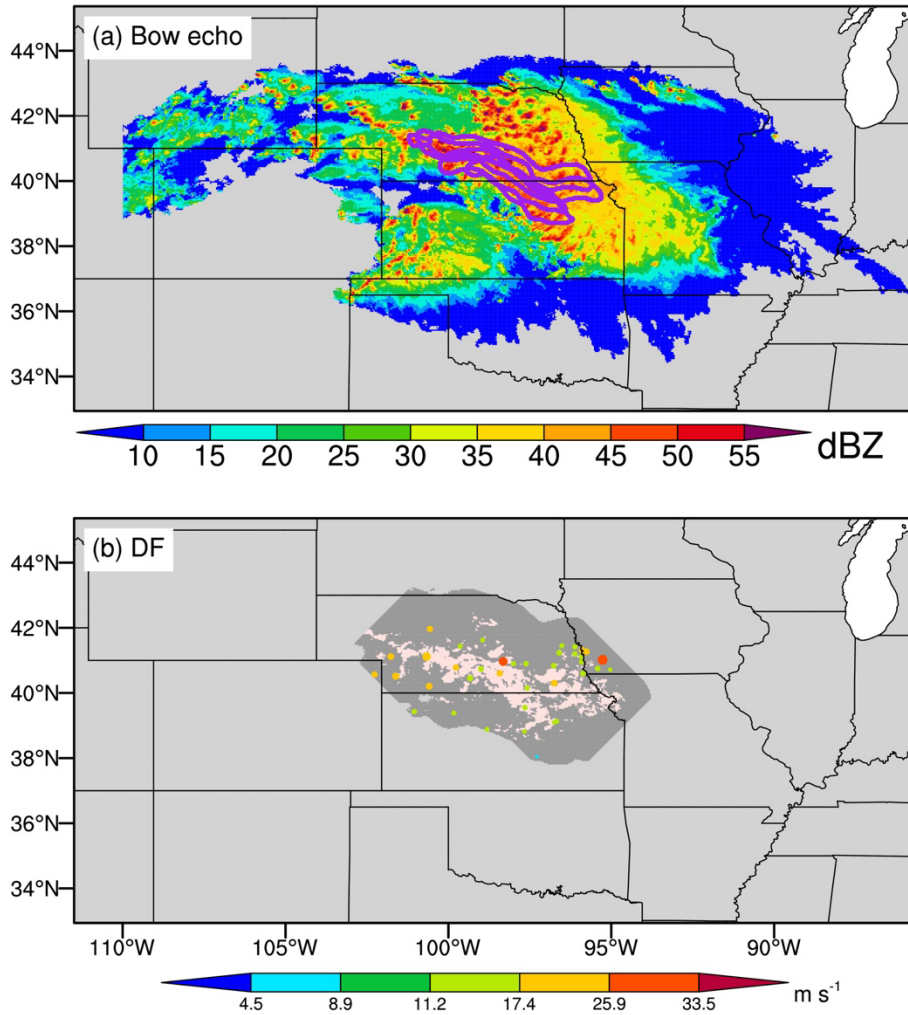



Figure S3. Same as Figure 7 but for an MCS that occurred on 1-2 August 2013. The automated detection algorithm falsely classifies the MCS as a derecho due to the false identification of bow echoes by the segmentation CNN. The figure title refers to the derecho timing range.

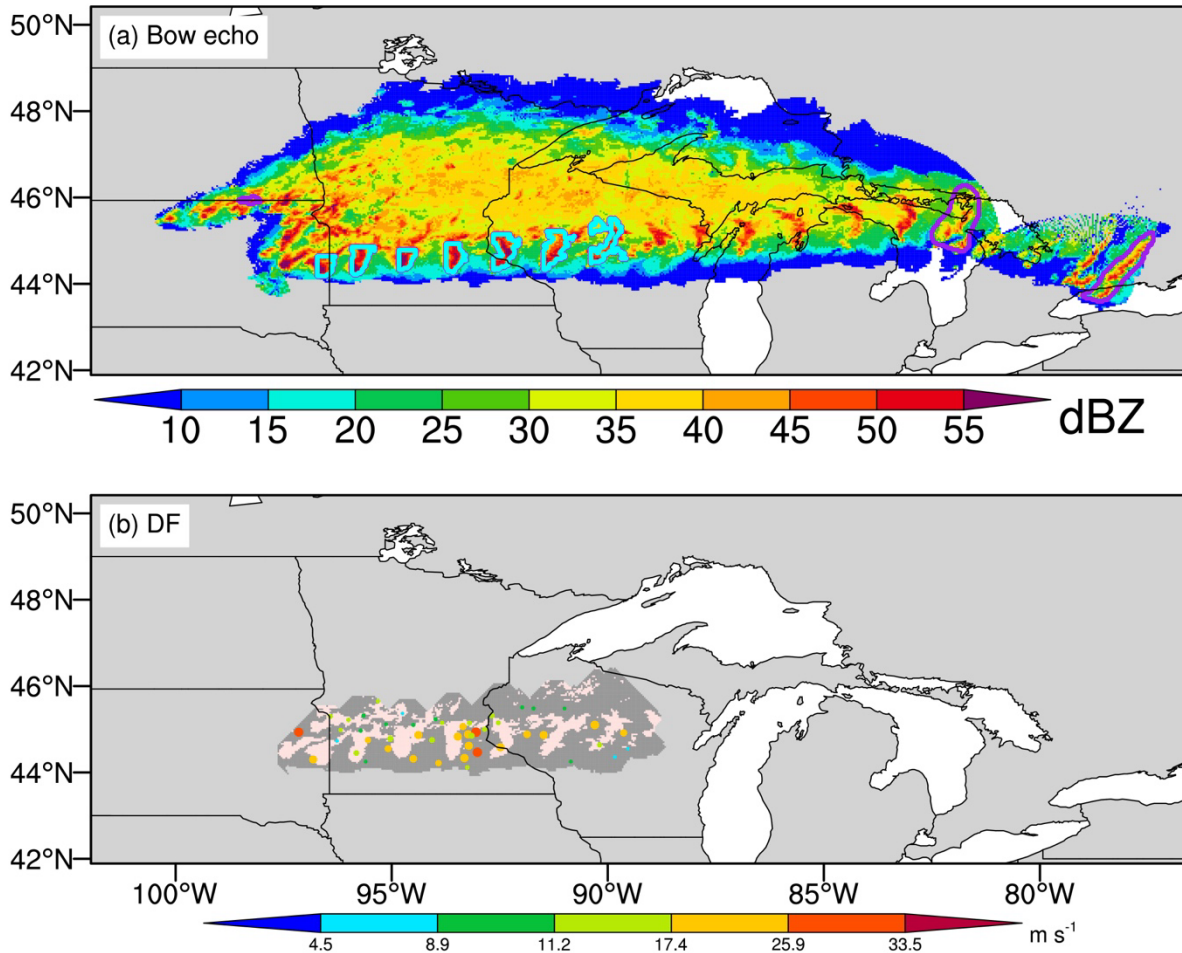
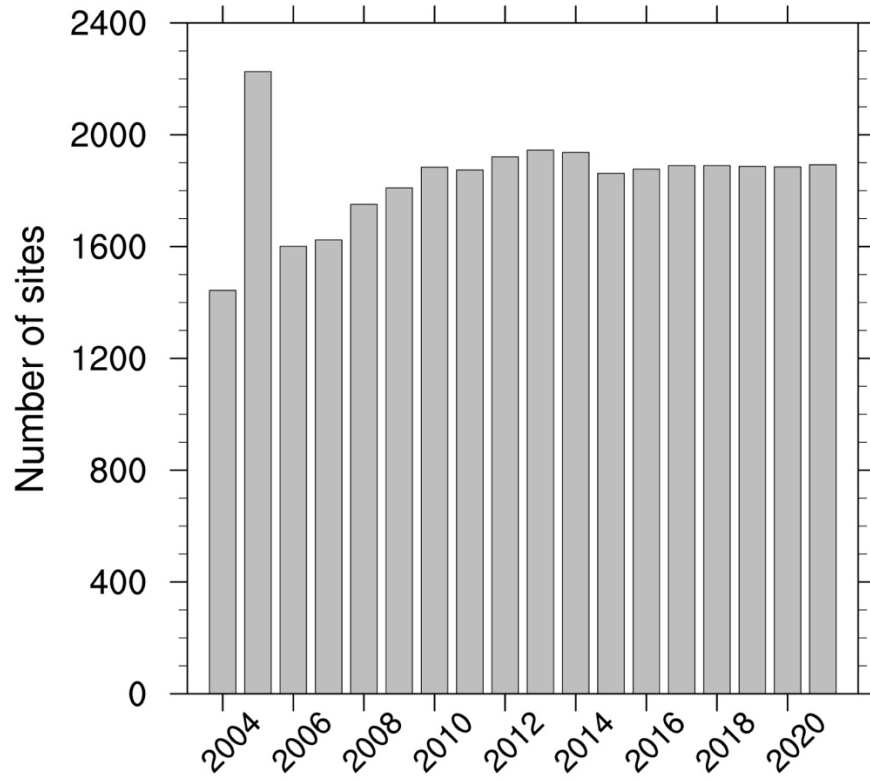



Figure S4. Same as Figure S3 but for a derecho that the automated detection algorithm misses due to a failure in bow echo identification. In (a), purple contours denote CNN-identified bow echoes, while cyan contours refer to bow echoes labeled manually.

36
37 Figure S5. The annual number of gust speed observational sites used in the derecho identification between
38 2004 and 2021.

39

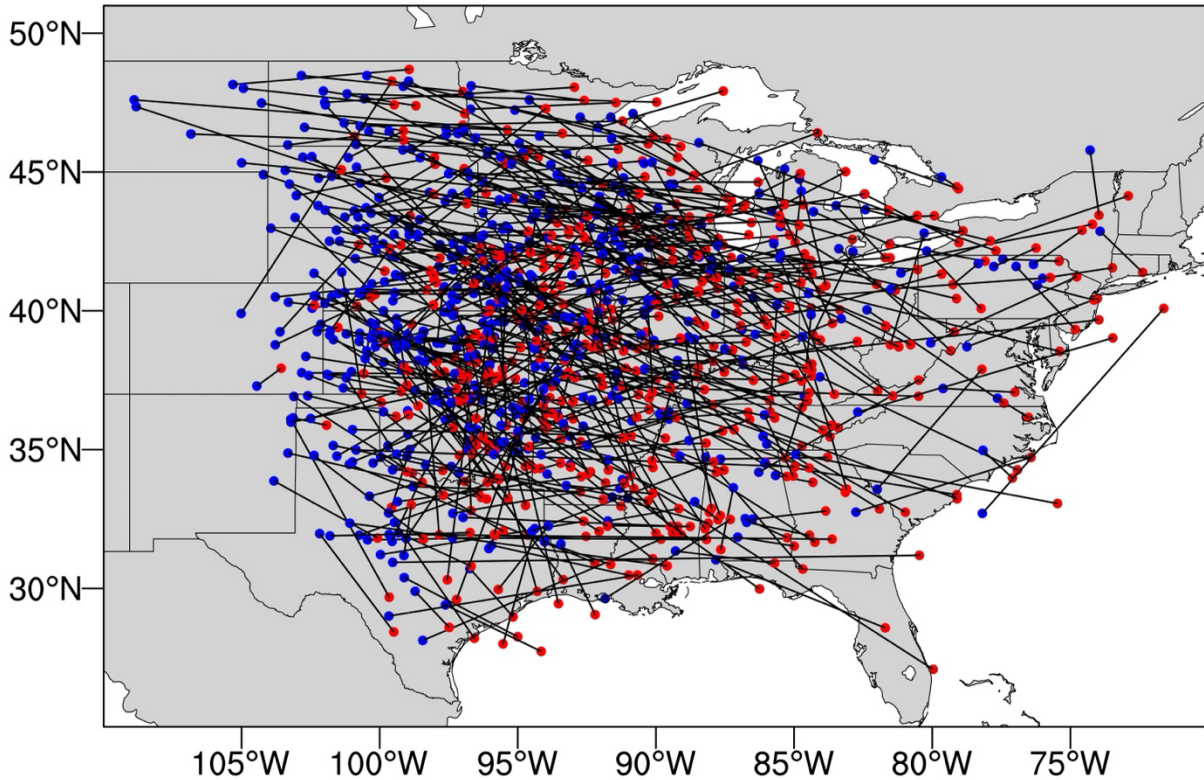
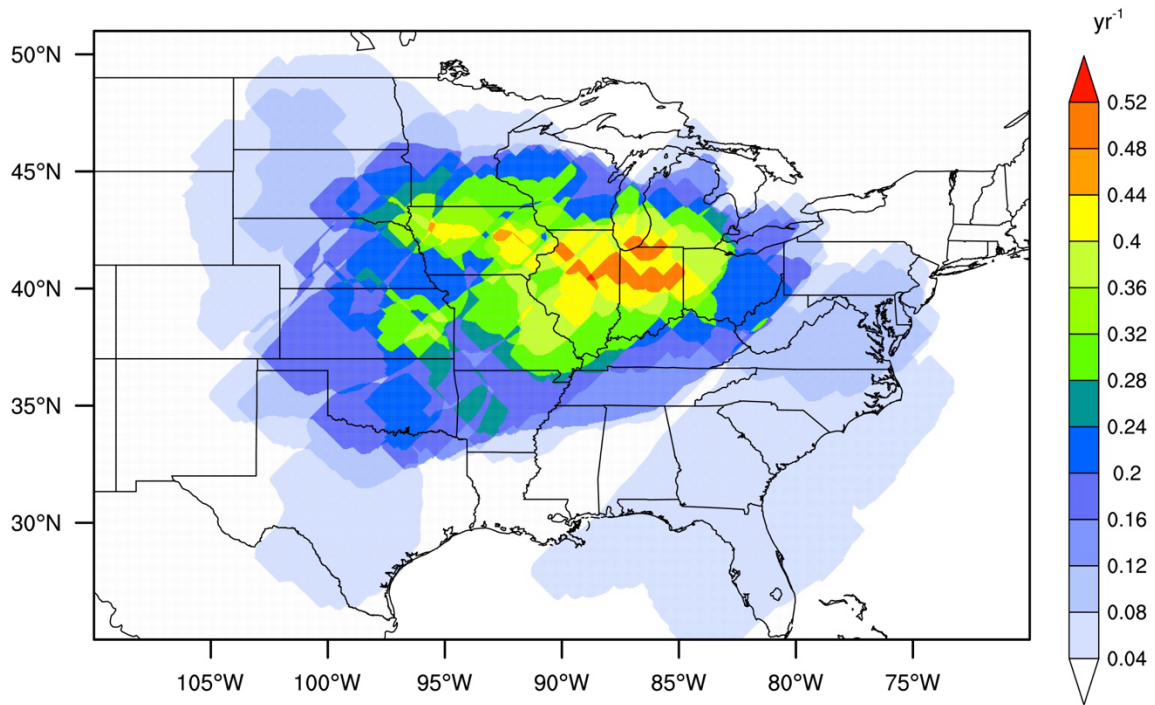
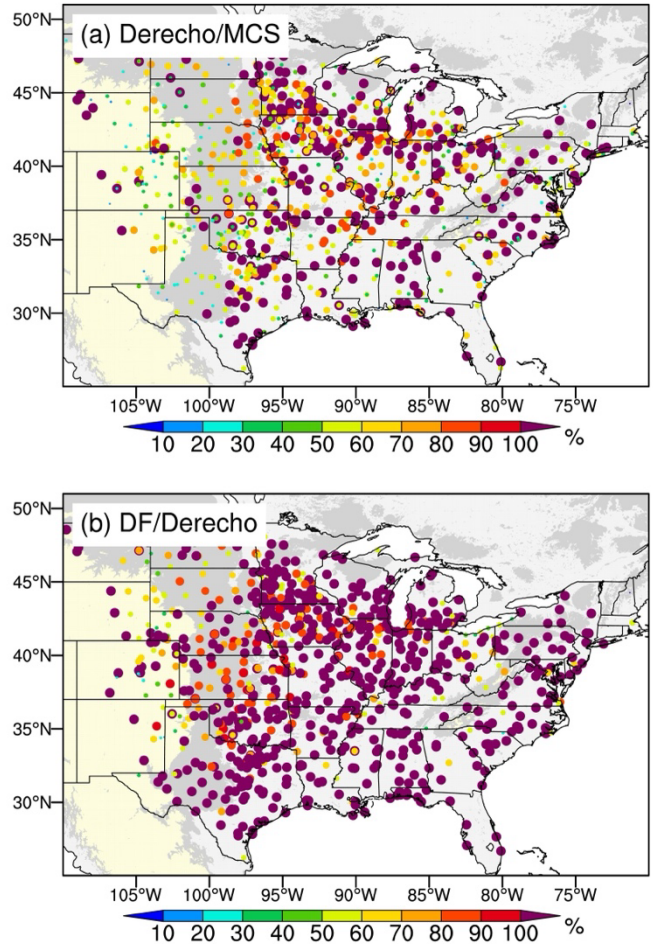
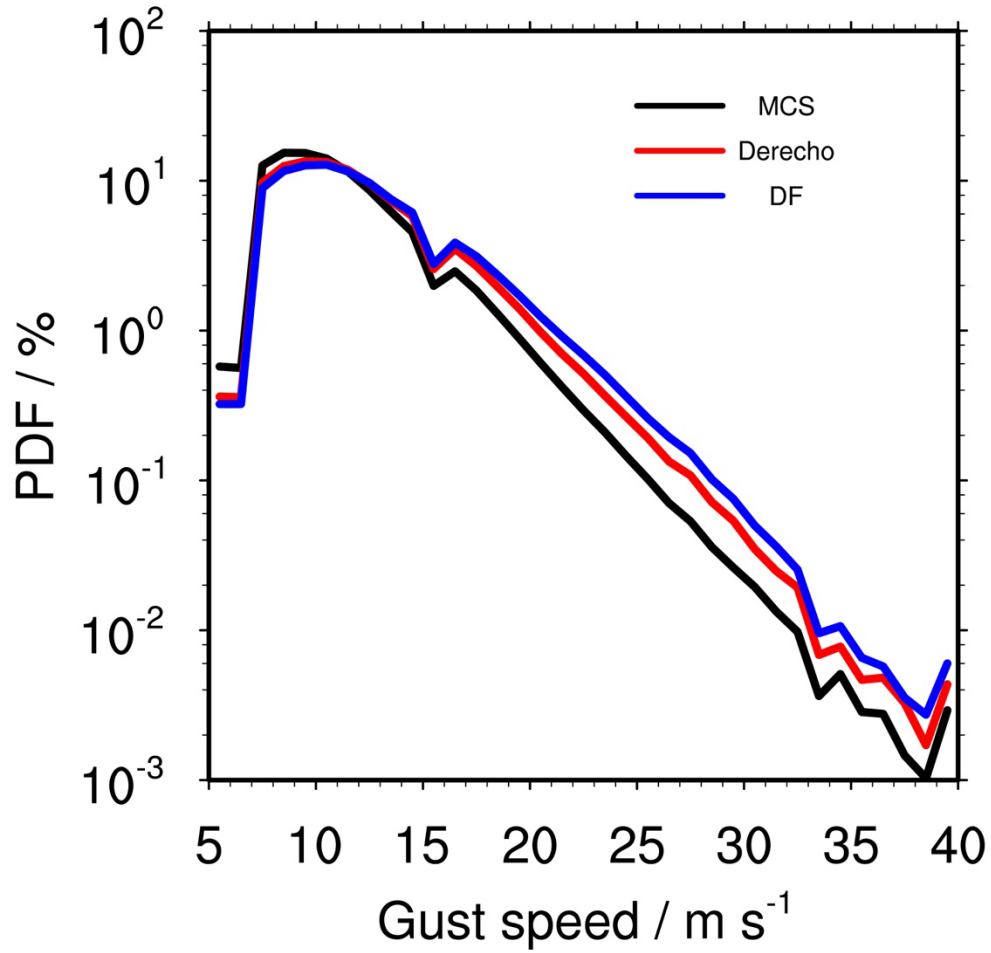
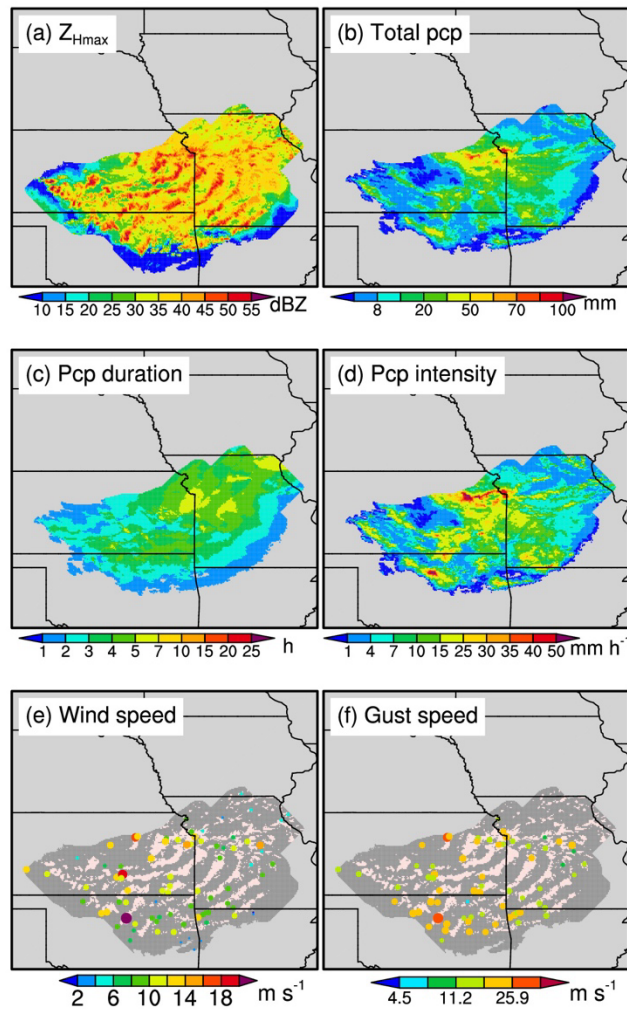



Figure S6. The derecho paths (black lines) during their DF periods. Each blue dot denotes the location of a derecho at the beginning of its DF period, while the corresponding red dot refers to the location at the end of the DF period. The location is calculated as the mean latitude and longitude of all grid cells within the DF area and covered by the derecho at the given time. Only grid cells with $Z_{Hmax} \geq 20$ dBZ are considered in the calculation. If there are no grid cells with $Z_{Hmax} \geq 20$ dBZ, we will move to the next hour (for blue dots) or the previous hour (for red dots) until we find grid cells with $Z_{Hmax} \geq 20$ dBZ.

48
49 Figure S7. Same as Figure 10 but for a sensitivity test with the gust swath calculation based on ≥ 10 sites with
50 damaging gusts. Using the updated criteria, 19 derechos are identified.

51

52

53 Figure S8. (a) Relative contributions of derecho-associated to MCS-associated damaging gust occurrences
54 between 2004 and 2021 at weather stations over the United States east of the Rocky Mountains. (b) is the same
55 as (a) but for relative contributions of DF-associated to derecho-associated damaging gust occurrences. We
56 exclude non-derecho MCS events overlapping with TCs in (a).

57

58
59 Figure S9. PDFs of land gust speeds associated with MCSs, derechos, and DFs in the United States east of the
60 Rocky Mountains. We exclude non-derecho MCS events overlapping with TCs.

61

62
63   Figure S10. Same as Figure 14 but for the spatial evolutions during the derecho DF period. The figure title
64   refers to the DF timing range.