

Title: A high-resolution unified observational data product of mesoscale convective systems and isolated deep convection in the United States for 2004 –2017

Authors: Jianfeng Li, Zhe Feng, Yun Qian, L. Ruby Leung

Journal: Earth System Science Data

General comments

The authors present a new methodology for identifying mesoscale convective systems based on the combination of three different sources: satellite imagery, weather radar volumetric mosaics and rainfall charts obtained from the merging of radar estimation and rain gauge values. The work results interesting but there are some points that should be solved before its accepting. One is the large number of typos associated with the table and figures references (what is S1, S2, ...?). Second one is conceptual, and more important to me: when the authors define isolated deep convection, they do not refer in any case to supercells. Besides, while the limitations of the methodology about the spatial and temporal scales are minimized in the case of MCS (because of their extent and duration), the part of isolated convection does not look like solved as clearly. I think that the authors should try to explain better the limitations (if exists) about this issue or, at least, explain why the results are not affected by this point. Finally, the number of results is excessive and, in my opinion, deviates the attention about the main objective of the research: the application of the new methodology. On the contrary, they do not compare their results with other methodologies, which are easy to find and can verify the lines provided by the current manuscript. In the next lines, the authors will find more detailed comments regarding some other points.

Abstract

Acronyms (MCS, IDC, FLEXTRKR), references (Li et al., 2020) and web pages (<http://dx.doi.org/10.25584/1632005>) are not frequent and preferable not included in abstracts. Do you consider strictly necessary for the understanding of the text to maintain them? In my opinion, at least the last sentence should be removed

Introduction

Although Doswell et al (1996) is still one of the reference paper in this field, there are many more recent research manuscripts that are noticeable to include in the L46-61 paragraph, for instance:

- Brooks, H. E., Doswell III, C. A., & Kay, M. P. (2003). Climatological estimates of local daily tornado probability for the United States. *Weather and Forecasting*, 18(4), 626-640.
- Taszarek, M., Allen, J. T., Púčik, T., Hoogewind, K. A., & Brooks, H. E. (2020). Severe Convective Storms across Europe and the United States. Part II: ERA5 Environments Associated with Lightning, Large Hail, Severe Wind, and Tornadoes. *Journal of Climate*, 33(23), 10263-10286.

L62-70: “deep convection” is repeated five times in the same paragraph. Please, modify the text using other options

L75-76: when you introduce IDC, are you including supercells? If the answer is yes, can you confirm that all the sentences that following this are true? In special, I disagree with the points about the higher rain rates, larger echo top heights, and greater ice masses.

Again, lines 80 and 82 depend on if you consider supercells or not in the IDC database

L103-104: “We produce the data product”?

Source datasets and algorithms

L120-121: “We only use the hourly Tb data in the FLEXTRKR algorithm discussed below, as all other datasets are only available at an hourly interval” Do you think that this time resolution could have any influence in the results?

Figure 1: Maybe you should include a small map of the whole American continent and a box for the zoomed area shown in the current caption

L202 (and many more): you cite “table S1” in the text, but I was not able finding this table in your manuscript.

Para 283-288: according to these lines, maybe you should change the label of “IDC” category.

In figure 2(b), it seems that there are more categories that the maximum number of the legend. Is this it?

Results

This section results too much extend and hard to follow (because its density and the large number of interesting results). However, I miss the comparison of your results with other works such:

Fritsch, J. M., R. J. Kane, and C. R. Chelius, 1986: The Contribution of Mesoscale Convective Weather Systems to the Warm-Season Precipitation in the United States. *J. Climate Appl. Meteor.*, 25, 1333–1345, [https://doi.org/10.1175/1520-0450\(1986\)025<1333:TCOMCW>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<1333:TCOMCW>2.0.CO;2).

Or the cited:

Haberlie, A. M., and W. S. Ashley, 2019: A Radar-Based Climatology of Mesoscale Convective Systems in the United States. *J. Climate*, 32, 1591–1606, <https://doi.org/10.1175/JCLI-D-18-0559.1>.

Geerts, B. (1998). Mesoscale convective systems in the southeast United States during 1994–95: A survey. *Weather and Forecasting*, 13(3), 860-869.

Then, my suggestion is reducing the results to the most interesting one (for instance, the percentage of contributing rainfall for each type) and comparing with the others works. This is also because the goal of the paper is to present the methodology, but not the “climatology”. Then, the authors could have the opportunity of publishing the climatological results in another manuscript.

Uncertainties of the data product

About lines 540-549: there are many more radar errors that can affect NEXRAD or other network, e.g. beam blockage, false echoes related with EM interferences, solar interferences, volumetric conus influence, among others. Have they considered or they can appear in the volumes?

L566-567: “we identify the most robust MCS/IDC events satisfying all the criteria based on the three datasets” Which percentage of data satisfies the whole set of criteria?

About Stage IV: do you think that geo-statistics contribute to the error, at the time of generating the final product?

L584-586: The sentence “Most grid cells in the US have less than 2% missing hours, which should have a negligible impact on the data product.” Is, at least, debatable. According to figure 5, it is difficult finding pixels with more than 170 hours of rainfall per year (combining both maps). This is less than 2% of the yearly hours (8760). If most of those missing hours coincide with a part of the rainfall period, the results changing notably. Please, explain better this point.

L647: What is for you “most of the important MCS”?