

Weather Clim. Dynam. Discuss., referee comment RC1
<https://doi.org/10.5194/wcd-2022-63-RC1>, 2023
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on wcd-2022-63

Anonymous Referee #1

Referee comment on "A composite approach to produce reference datasets for extratropical cyclone tracks: application to Mediterranean cyclones" by Emmanouil Flaounas et al., Weather Clim. Dynam. Discuss.,
<https://doi.org/10.5194/wcd-2022-63-RC1>, 2023

Review of "A composite approach to produce reference datasets for extratropical cyclone tracks: Application to Mediterranean cyclones"

I enjoyed reading this interesting and well-written paper. The authors use a novel approach to combining multiple cyclone detection and tracking schemes, which aims to identify those Mediterranean cyclones that are consistently detected between methods. They also evaluate this using a combined dataset from 5 subjective analysts, demonstrating the uncertainty inherent in subjective tracking datasets. I have a range of relatively minor comments and suggestions that I hope will improve an already very good paper.

Specific comments:

1. Colour schemes - many of the figures in your paper use colour schemes that include both red and green and are not colourblind-friendly. While I acknowledge that finding 10 unique colours for the 10 methods in e.g. Figure 4 is difficult, the authors could at least use viridis or a diverging red-blue colourbar for Figures like 5 and 8.

2. While the focus of the paper is on the objective tracking results, I would love to see a bit more analysis of the subjective tracking, as there may be some additional insights in your subjective dataset that will be useful for future research to draw on. What proportion of the 120 cyclones were identified by all 5 experts? What proportion were identified by none? If you designed your dataset based on historical case studies I assume it tends to

include medicanes with major impacts - does the fact that many of these are not identified by experts indicate that your duration criteria is too restrictive for studies that want to identify impactful events? How were the case studies distributed throughout the year, and was the matching between the experts better in winter?

Supplementary Material 1 mentioned at L247 does not seem to be currently available, but I hope it includes a summary table listing the dates of the 120 events and how many of the experts identified them.

3. At L310 you attribute the varying seasonal cycles to weaker cyclones being more common in the summer months. If that were the case I would expect to see a correlation between the total number of cyclones identified by a method and the proportion of lows in the summer months, but that does not seem to be the case - M07 has the second-highest frequency of lows and M08 has the second-lowest, but they both seem to have very uniform seasonal distributions. I think this needs further assessment, as there may be some more complex factors at play e.g. related to the spatial distributions of lows in different methods.

4. Section 3.2/Figure 7 - I find it interesting and surprising that the similarity scores in Figure 7 seem to be uncorrelated with the total number of lows each method generates in Figure 4 - I would have expected M07 to perform a lot better given its high track frequency. Do you have any explanation of this?

5. Figure 9 would be more useful I think if it showed the "hit rate" (the proportion of subjective lows detected), so that the denominator is the same for the whole plot.

6. L462 - Please share the correlation values - Figures 11c and d look very different to my eye, as Fig 11d has almost no lows in the Atlantic or the Black Sea but in Figure 11c the numbers in those regions are only slightly lower than in the Mediterranean sea. If the correlations are only calculated over a smaller subregion e.g. the Mediterranean sea, maybe show a contour around that area in Figure 11.

7. Comparing Figure 10a and Figure 4, it becomes obvious that because the numbers of cyclones vary significantly between datasets, lows identified using 2 or 3 methods will be dominated by agreement between just the subset of methods with high frequencies (M03, M07, M06 and M01). The paper would benefit from a figure that tries to quantify this, to understand which methods are most responsible for determining the seasonality/spatial patterns of the combined tracks that are then obvious in Figures 10b and 11.

I would imagine something like Figure 8, showing that e.g. for a confidence level of 2 (fake numbers), 60% of composite tracks include M03 but only 20% include M08, since M08 is less common. This increases with confidence level, so at level 10 each method is included 100% of the time, by definition. Similarly, a plot showing what proportion of all tracks for a method are included in the combined dataset - e.g. that a confidence level of 8 includes 50% of M08 tracks but only 5% of M03 tracks. Bonus if this could be shown for different months or seasons.

8. Conclusions - I think the conclusions would benefit from some discussion of potential extensions/applications. Which confidence level do you think would be most applicable to identifying medicanes with significant impacts e.g. rain, given the tendency to favour long-lived events meant that many of the case study events were not identified by your subjective analysis? How applicable do you think this approach would be for cyclones in other areas or globally?

Minor comments:

9. I don't really like the term "bogus tracks" - it implies that the tracks that are only detected by some methods are wrong/bad, when some of those may indeed be real cyclones that caused real impacts. I'm not sure of a better term to use, but it's something to consider when you discuss them.

10. Figure 2d/L207 - I assume that the track components that were rejected from this composite (i.e. the red dots) remain in the pool of data, so the red dots end up being a second track in the dataset with confidence ~ 5 ?

11. L373 - Do you have any explanation of why M01-M02 and M01-M08 have stronger similarities, e.g. linked to characteristics of the methods?

12. In most figures, the methods are ordered from M01 at the top to M10 at the bottom, but in Figure 6 M10 is at the top. It would be good for the order to be consistent.

13. The authors may want to cite Pepler et al. (2020), who combined two CDTM methods and showed that lows identified by both methods produced higher average rainfall totals than lows identified by only a single CDTM (their Figure 2).

14. I am surprised to see that the method of Kouroutzoglou (2011 etc) is not included in this paper, given that it was used for several key papers on medicane climatology and J Kouroutzoglou and H Flocas are both authors of this paper. Is there a reason for this?

15. The insight from lines 282-285 - that a similarity of 80% is as good as you can get even between subjective methods - is really interesting. I think it should be highlighted in

the conclusions.

Technical comments:

L169 etc - "Exemplary" is generally used to mean "very good". Unless you mean to say that the cases in Figure 2 are two of the best in the whole dataset maybe you should call them "cases" or "examples".

L628 - What is N for the ERA5 data?

References:

Kouroutzoglou, J., Flocas, H. A., Keay, K., Simmonds, I., and Hatzaki, M.: Climatological aspects of explosive cyclones in the Mediterranean, *Int. J. Climatol.*, 31, 1785–1802, <https://doi.org/10.1002/joc.2203>, 2011.

Pepler, A. S., Dowdy, A. J., van Rensch, P., Rudeva, I., Catto, J. L., and Hope, P.: The contributions of fronts, lows and thunderstorms to southern Australian rainfall, *Clim. Dyn.*, 55, 1489–1505, <https://doi.org/10.1007/s00382-020-05338-8>, 2020.