Review of “Quantifying forecast uncertainty of Mediterranean cyclone-related surface weather extremes in ECMWF ensemble forecasts. Part 1: Method and case studies” by Hartmuth et al.

Review by Michael Schutte

Recommendation: Accept subject to minor revisions

The study focuses on forecast uncertainty of Mediterranean cyclones with respect to the cyclones’ predictability, and their related wind and precipitation extremes. Based on ensemble forecasts from ECMWF, the authors outline a novel methodology to track and assess the representation of cyclone-related weather extremes. Three case studies illustrate the methodology and highlight case-to-case variability in forecast uncertainty of those extremes. Additionally, forecast uncertainty is greater if smaller regions are affected by extremes, during earlier lead times, and if the cyclone is not well captured by the ensembles.

Although my expertise lies not in Mediterranean storms, the methodology appears to yield valuable results. At the same time, the manuscript could be strengthened by considering the following aspects:

- Varying dependence between forecast uncertainty and cyclone lifetime
- A risk for confusing the goals stated in the introduction and those actually addressed by the paper
- Potential differences in bias between ENS and ERA5
- The fairness of comparing predictions of extremes for storms of different lifetimes

These points will be discussed further in the following.

Main comments

You note in several places (e.g., l. 11-12 and l. 413-417) that predictions are more uncertain during the earlier stages of the cyclones’ lifecycle. While this is consistent with Fig. 11 for storms Jan and Daniel, storm Denise appears to show the opposite behavior, with uncertainty being highest during the later stages of its lifetime. It may therefore be helpful to acknowledge this varying dependence between forecast uncertainty and cyclone lifetime, rather than presenting it as a general tendency.

Paragraph in l. 66-80: As a reader it is not immediately clear, which goals you aim to achieve in this study and which you investigate in the second part. It could be beneficial to highlight that these goals apply to both studies together in l. 67. Alternatively, the authors could also mention the goals relevant to this paper at first and outline goals for the second part separately.

l. 142-146: Considering the different IFS model cycles, how well do these thresholds align between ERA5 and ENS? Is the difference between ENS and ERA5 small enough to be negligible or might it have an effect on how well extreme objects are detected in the ensembles? It could be good to mention this, as it might limit the interpretability of the results otherwise.

You compare associated extremes and cyclones of different lifetimes. I think the three storms are well chosen, as they also reflect this aspect. However, one might have to be careful when comparing the results of the different storms with each other. For example, in Fig. 9, it is hardly possible to predict a later cyclogenesis for storm Denise, due to its very short lifetime. Thus, the algorithm might miss out any forecast that predicts a storm some days later, opposed to storm Daniel, with potential consequences for the predictability of P and G_{10}. The authors could at least point out this aspect, e.g., in Sec. 4.1.1.

Minor comments

l. 7-8: The authors could define in the abstract what they mean with 'objects of extreme surface weather' These two-dimensional objects are defined nicely in Sec. 2.4., but it might be helpful to provide a short explanation in the abstract, as well.

l. 12: The authors could add a sentence about implications at the end of the abstract to highlight the relevance of the study's results.

l. 34 'cyclone speed': As a reader it is not immediately clear what is meant with cyclone speed. Do the authors mean the propagation speed of cyclones, intensification speed/cyclogenesis, or 10-m wind speeds?

l. 106-107: Since you mention the region that the retrieved fields cover, you could also add the coordinates of the domain in parentheses.

l. 107: Why did you chose 0.5° grid spacing? As you also mention in the conclusions, a higher spatial resolution could be valuable in the case of investigating Mediterranean cyclones, e.g., with respect to cyclone detection and tracking (Aragão et Porcù, 2022).

Fig. 1: I really like this figure to visualize how the cyclone track matching algorithm works. Unfortunately, you don't refer to it in the text right now. For example, this could be done at the end of Sec. 2.3.

l. 136-138: How much does this methodological choice increase the number of 'random jumps' in identified cyclone tracks, e.g., in Figure 1, jumping from the first match to the second match and then to the third match? And does this have any potential drawbacks for the analysis of forecast uncertainty? For example, a jump in detected cyclone could result in a sudden change of P or G_{10} probabilities due to a different location of cyclone center.

l. 169-179: The first paragraph of Sec. 2.6 elaborates on the probability of extreme objects, in line with its title. However, the second paragraph justifies why the three storms have been selected as case studies. Thus, it could be beneficial to have a separate section for the second paragraph, e.g., 'Selection of case studies', or revising the title of Sec. 2.6.

Tab. 1: Are the SLP minimum values based on ERA5 data or on observations? Furthermore, you have mentioned socioeconomic impacts of Mediterranean cyclones, so I would suggest to add one more column for only land area/grid points affected by P, and in the same way one column for G_{10}.

Fig. 9: I really like how you visualize the probability of cyclogenesis in this figure. As especially subpanel (c) is rather long, the readability could be improved if you add thin horizontal lines, e.g., every 20%, in all subpanels instead of having only the 100% line.

l. 259-263: At lead time -120 h in Fig. 9a, you mention that half of the members are predicting cyclogenesis at the correct time and half too early. However, there is also a small portion colored in blue, indicating that a few members predict cyclogenesis too late. You could either mention that in the text, or choose a different lead time, e.g., -60 h.

l. 266: Could these differences be linked to different origins, i.e., the North Atlantic vs. the Mediterranean?

Fig. 10: Do the boxes shown in each small subpanel have the same spatial extent as in Fig. 2, i.e., 20° latitude by 20° longitude?

Fig. 11a: Here and in Fig. 10a, would it be possible to elaborate shortly, why p_{obj} of P decreases for the -12 h lead time? This seems to be consistent for several times along the lifecycle with exception of earliest time at cyclogenesis.

l. 425-427: You could comment on why a higher temporal and spatial resolution might matter for your analysis of Mediterranean cyclones, e.g., an improved detection of precipitation and wind extremes.

Technical comments

l. 89: You could use a comma after 'Then', i.e., 'Then, we present…', to improve readability.

l. 237-238: The sentence 'The position and orientation of the PV streamer is not unlike the one in the early phase of Medicane Zorbas in September 2018' is understandable, but it has some grammatical inconsistencies and could be expressed more clearly. Specifically, 'position and orientation" forms a compound subject, so the verb should also be plural.

References

Aragão, L., Porcù, F. Cyclonic activity in the Mediterranean region from a high-resolution perspective using ECMWF ERA5 dataset. Clim Dyn 58, 1293–1310 (2022). https://doi.org/10.1007/s00382-021-05963-x