Response to Reviewer #1

We thank the reviewers for their time and constructive comments. We have complied with most of the proposed changes. In the following, the comments made by the reviewer appear in black, while our replies are in blue and the changes in the text are in quotation marks.

General:

In this paper, the authors examine the predictability of Mediterranean cyclones by using a large reference dataset of 2853 cyclone tracks from 21 years and ensemble reforecasts for this period. This topic is very important for both research and operational scientific communities. The work is ambitious in the way that both location and intensity errors of the cyclone tracks are systematically examined, with decomposition into sub categories of cyclones according to their season of occurrence, sub-region of the Mediterranean, intensity, propagation velocity, and deepening rate. The authors find significant predictability differences among the subcategories in terms of cyclone location and intensity at different lead times. Overall, the paper is very well-structured and clearly written, with good visualizations of the data.

The manuscript will clearly provide a valuable contribution to WCD once the following concerns will be addressed. Of the major comments, the first concerns the underlying tracking methods, while the other two raise issues which are largely not addressed in the current paper, but are nonetheless key for making insightful conclusions (and are feasible given the existing datasets).

Major comments:

1. Differences in the cyclone tracking algorithm between the reference and ensemble forecast datasets:

Using non-identical tracking methods for the reference tracks and the forecasted tracks can introduce biases into the analysis. This issue should be either corrected or the differences in tracks carefully assessed for any systematic influence on the predictability metrics.

In this paper, two different tracking methods have been used to track the cyclones in two different data set. The VDG algorithm requires a reference trajectory to be given as input to begin the tracking process. Consequently, and unlike AYRAULT, the VDG algorithm cannot be used to track cyclones directly in ERA5 data without any given starting point. This characteristic makes it however very relevant to track cyclones in the reforecast fields using the reference tracks, without requiring an additional matching method that would bring more complexity. To assess the consistency of the results, the two algorithms have been tested on the same ERA5 data, given the tracks detected by AYRAULT as a reference for the VDG algorithm. Note that VDG is tuned for a temporal resolution of 6 h, while the tracks are followed hourly with the AYRAULT algorithm. For each track detected by AYRAULT and VDG, the difference in terms of location and intensity are calculated for all time-matching track points. The results are presented in Fig. R1-1a and Fig. R1-1b. For 85% of the tracks, no difference between the two techniques is observable. However, for 10% of the cases, the distance between tracks reaches almost 200 km at the time of minimum MSLP. To overcome this issue, tracks are removed from the analysis if: (1) The track is followed with AYRAULT but not with VDG in ERA5 (206 tracks). (2) If the maximal distance between two tracks reaches more than 40 km (687 tracks). With these criteria no difference is observable between AYRAULT and VDG in 99% of the dataset at any instant, for the location or for the MSLP (Fig. R1-1c and R1-1d).
All figures and results have been recalculated using this new dataset of 1960 tracks. The same conclusions are drawn, and the main results remain consistent with the previous version. All values presented in the final text are also modified according to the new robust results. A dedicated part on the verification method has been added (now Section 2.5).

2. Prediction of cyclones per se:

In addition to the systematic investigation of the predictability of location and intensity, it is even more fundamental to understand if the mere existence of the cyclone (or track) is captured by the ensemble. By construction of the verification technique against a reference track, this aspect required a separate quantification,

such as tracking the mean number of members having a cyclone at all (orange line in Fig. 5). In my view, if this aspect is developed, it could greatly enhance the take-home messages from the paper. For example, one can readily decompose this metric (fraction of ensemble members with a cyclone) to the different cyclone categories.

Thank you for your suggestion. A dedicated section has been added in the article (now Section 5.1).

3. Consideration of the cyclone lifetime relative to cyclogenesis/peak intensity:

Currently, the examination of the cyclone tracks predictability does not consider one of the most important aspects, which is the cyclone stage in the lifecycle, namely relative to cyclogenesis/peak time/lysis. Instead, the large variability in Fig. 5 envelops both the variability among ensemble members (for a given cyclone track point), and different cyclones, with all subcategories and at all track times (i.e., the time relative to genesis/peak etc.). It can be insightful to look into this variability, and see how Figs. 5 and 6 differ when decomposing the analysis to different track times. I understand that, by construction, the genesis point is more predictable when forced to start from a vicinity of the reference, but I hope the authors can think of a way to still address this aspect.

The choice to keep all track times has been made to ensure the categories to be sufficiently populated for robust results: unfortunately, the low frequency of reforecast basetimes (twice a week) does not allow decomposing the results in the different cyclone stages. We keep the suggestion for future work, depending on the possible availability of reforecasts with more frequent basetimes.

Minor comments:

Line 1 and 20: use of “beneficial” is unclear here

Replaced by ”necessary”

Line 4: “characterize cyclone predictability”: mention the timescale / range you consider here

Correction: ”medium-range” added here

Line 22: add “negative” before “impacts”

Done

Paragraph ending in line 31: the paragraph currently misses a description of heat lows over land.

Added

Line 47: mention that Baumgart et al. considered hemispheric-wide simulations

Correction: ”Earlier work by ”Zhang2007” in an idealized baroclinic wave simulation, and by ”Baumgart2019” in hemispheric-wide simulations of PV structures, identified three phases in forecast error growth”

Line 108: 1) the usage of data on the 700 hPa level comes as a surprise and was not mentioned before. Please clarify. 2) Also on line 111: is the smoothed relative vorticity field evaluated at both 850 and 700 hPa? 3) Line 120: “. . . maximum remains the track point. . . ” - on which level?

1) Correction: ”The main idea of AYRAULT is to track cyclones firstly in the relative vorticity field at 850 hPa. The horizontal wind is then used at both 700 hPa and 850 hPa to choose the best following tracking point in the direction of cyclone propagation. Finally, the track points are paired with the mean sea level pressure (MSLP) field. In the following, a time step is denoted by t, the relative vorticity field at 850 hPa by ζ, the zonal and meridional wind fields by u and v, respectively”. 2) and 3) The relative vorticity is explicitly defined as to be at 850 hPa only: ”the relative vorticity field at 850 hPa by ζ”

Lines 116-117: how is the selection made? Do you mean that closer and more intense is favoured?

Clarified: ”the strongest one”

Lines 130-131: it is unclear then why not to use a lower confidence interval to obtain a comparable number of tracks. Please also mention how many tracks are obtained here?

Correction: ”First, the latter contains only 206 tracks in the highest confidence level, which is not enough for a systematic study. At the mean confidence level (consensus of 5 over 10 algorithms), on the 2001-2021 period and with the same thresholds used here on the pressure and on the location of cyclones, 1231 tracks are detected in ”Flaounas2023”, against 2853 with the AYRAULT algorithm.”

Line 164: “combination” = mean?

Yes. Correction: ”The tracking method in VDG is then based on a combination of past movement and steering flow vector V_{av} defined as the layer average of the local wind fields at 850 hPa and 700 hPa.”

Fig. 3a: it is unclear how the subset density is normalized (% of the strongest 10% of all cyclones?), same for Fig. 4.

Yes, it is in % of the strongest 10% of all cyclones. Correction: ”(a) Relative frequency of Mediterranean cyclones, defined as the 10 % deepest cyclones having a track point within a radius of 100 km.”

Lines 305-306: please clarify if the “number of members in which a cyclone is found” means that there is a cyclone track point within a 3.5-degree radius near the reference track?

No, when the VDG algorithm found a track point at time t corresponding to the reference track, it can move independently forward at time $t + 1$. Added in the description of the VDG algorithm: ”At initialisation time, a ζ maximum is searched for in the reforecast field in the neighbourhood of the reference track calculated in ERA5. The tracking in VDG is then independent of the reference track, and is based on a combination of past movement and steering flow vector...”

Fig. 5: Bringing the orange line to the front can make it more readable

True, but the highlight is put here on the distribution, not on the precise number of members. Moreover, this curve is now presented in the next section.

Line 377: good prediction of W. Med cyclones: it can be expected because they are mostly slow-moving, so errors in location are not expected to arise compared to faster moving systems.

Yes, detailed later in the section on the motion speed categories.

Line 389: add “36 h,” before 42.

Modified by the new results: ”66 h, 90 h, 108 h and 132 h lead time,”

Fig. 10c,d and accompanying text: this view is confusing, as it does necessarily capture the cyclone at the timing of the maximum deepening (12 h until max. intensity). It will be more interesting to isolate only this section of the track and examine its predictability metrics.

See response to major comment 3

Line 428: Need to reconcile this statement with the best predictability of W. Med cyclones in terms of track location. It will be good to come back to this issue in the conclusions section.

The issue is addressed in the motion speed part. ”Indeed, these quasi-stationary lows are in a vast majority concentrated in the Gulf of Genoa in the West Mediterranean, which is the region where the cyclone location is the best predicted (see Fig. 9). This result has to be compared with the poor predictability of the West Mediterranean cyclones’ intensity, suggesting the existence of at least two different types of cyclones in this

particular region. The first is made of slow cyclones with a good predictability in terms of location and a fair predictability in terms of intensity. The second one is made of fast moving cyclones with a poor predictability in terms of intensity and a fair predictability in terms of location. ”

Line 480: cyclogenesis: this stage is no directly shown here (see also major comment 3).

Replaced by: ”Winter cyclones are indeed less accurately predicted”

Technical corrections: Done

Line 54: robuster => more robust

Line 70: systematical => systematic

Line 107: in => into

Lines 192/384/392: delete “for its/their part” without loss of information

Line 209: spell out “CDFE”

Line 213: add “.” after “applied”

Line 247-248: “do not enter over the sea” - unclear wording

Line 306: on => in

Line 315: of => to

Fig. 5 caption: in function => as a function

Line 382: “the both” - delete “the”

Line 392: firsts => first

Line 399: add “visualized” after “previously”

Line 406: every => all

Line 436: statically => statistically

Line 473: a same area => the same area

Line 491: firsts => first

Additional figures

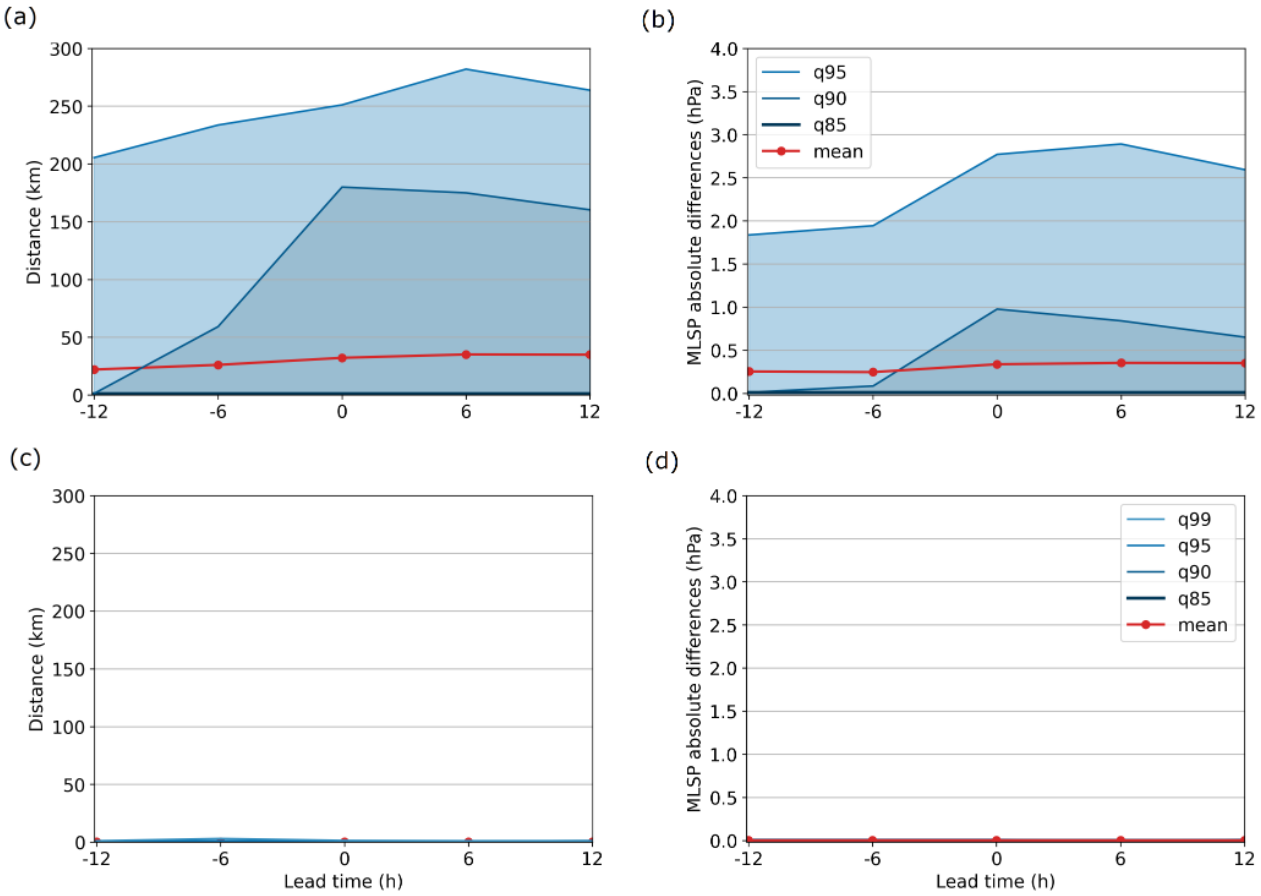

Figure R1-1: Comparison between Ayrault and VDG tracks obtained from ERA5: distance (a and c) and absolute MSLP difference (b and d). (a) and (b) show the differences in the original datasets, while (c) and (d) show the differences after removing tracks with a distance reaching more than 40 km (final data sets). 0 h lead time corresponds to the time of minimum MSLP in VDG.