

## Response to Reviewer #2, 2nd Round

We thank the reviewer for his/her time and constructive suggestions. We comply with 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 the revised version of their manuscript, the authors have made great efforts to take into account the general comments made in the first review.

They rephrased the title to more accurately reflect the specific aspects addressed in their predictability study. As for the comments on storm nature, they emphasize that their focus is on the extratropical storms and argue that medicanes would only contribute with a negligible fraction. I would like to clarify that my comment was rather meant not to simply distinguish into extratropical storms and medicanes, but to analyse a bit more how the full spectrum looks like, *i.e.*, including the hybrid stages. It would have been interesting to also condition on the storm nature, because regional differences may be related to it. Even though the authors have refrained from taking up this aspect, I think that at least the clarification made throughout the paper that the majority of the storms are ‘extratropical’ is helpful for the reader.

The main methodological concern, raised by both reviewers, namely the inconsistency between tracking algorithms, has been seriously addressed. The authors answered many questions related to it and clarified it in the manuscript as well. An extra analysis has been done, applying the VDG algorithm to the ERA5 dataset, resulting in the additional section 2.5, which presents the results of the comparison of the two tracking techniques.

Since the revision has improved the quality of the paper in a good way, and the authors have replied to the review comments in a satisfactory manner, I now recommend the paper for publication after consideration of the minor specific comments below.

*For future revisions, when you prepare a response to a reviewer, please give line number(s) of the text you cite from your revised manuscript in the response to reviewers. Otherwise, one has to do the search for the position in the revised document oneself. Thanks!*

### Specific comments:

1. 104: Add “(125-km horizontal resolution)” after “model data”.

1. 110 of the track changes document: Done

1. 143: Please add again the word “reforecasts” after “ensemble”.

1. 148 of the track changes document: Changed for “Data for the predicted tracks: the IFS ensemble reforecasts”, to avoid the repetition. 2.4. title also changed for consistency: “Tracking method for the predicted tracks: the VDG algorithm”

1. 163: Please check my comment to line 158 in the previous version of your manuscript: “l. 158: Replace ‘track predicted cyclones’ by ‘tracks predicted for cyclones’.” In the tracked changes document, I can see that you (probably accidentally) removed the sentence you indicated as corrected in your reply to my comments.

1. 168 of the track changes document: We find that the following sentence: “This characteristic is particularly useful when it comes to tracks predicted for cyclones that were previously identified in the reference data set.” is not easily understandable. Changed to: “This characteristic is particularly useful when it comes to detect cyclones in the reforecasts from the location of the reference tracks.”

1. 188: Replace “non-identical” by “different”.

1. 195 of the track changes document: Done

1. 189-190: To make the reader aware of it, please add a sentence on the fact that your comparison on the ERA5 dataset between VDG and AYRAULT has a limitation in terms of storm occurrence. VDG can only identify occurrence if it was found by AYRAULT before.

The purpose of this paragraph is not to compare two independent tracking techniques, but to see if the tracks detected in VDG and AYRAULT are indeed the same cyclones. Therefore, we do not think that repeating the limitations of VDG is relevant in this part.

1. 312-313: I'm not convinced by the statements you make in response to my comment (l. 296-297 in the original version of the manuscript). I don't see why it would be "very unlikely" that extratropical cyclones enter multiple of the regions you defined throughout their lifetime. If they are embedded in Rossby waves, they can easily get steered over long distances. And, again, I recommend not using the phrase "suggesting two different processes of cyclogenesis" as you are not distinguishing between different stages of a cyclone's lifetime in your study. Keep it on the "occurrence"-level, instead of speculating about stages!

1. 322 of the track changes document: "Cyclogenesis" changed to "cyclones". The aforementioned comment and the answer we made (italic) are the following:

1. 296-297: As you do not distinguish between different stages of a cyclone lifetime in your study, I would refrain from linking different motion speed-categories to differences in processes of cyclogenesis. It may be that a cyclone forms in one part of a region but then moves into another part of the same region, while changing motion speed. Therefore, here and throughout the paper, I recommend to not speculate on cyclogenesis, if it is not specifically analysed.

*Indeed, the motion speed of a cyclone may vary during its lifetime. However, the areas highlighted on Figure 4a are clearly distinct from each other and it is very unlikely the same cyclone would move from one to the other. Also, note that we do not explicitly refer to the cyclogenesis area but more broadly to where cyclones arrive from (the Sahara, the Atlantic).*

We identify three arguable points:

1. The possibility for a cyclone to move from one region to another ("It may be that a cyclone forms in one part of a region but then moves into another part of the same region, while changing motion speed." in the previous comment and "I'm not convinced by the statements you make in response to my comment" in the current comment)
2. The Rossby waves steering storms over long distance ("If they are embedded in Rossby waves, they can easily get steered over long distances" in the current comment)
3. The speculative claims made in the manuscript about the processes ("I recommend not using the phrase *suggesting two different processes of cyclogenesis*", in the current comment)

For which will we attempt a better answer:

1. Our claim that "it is very unlikely the same cyclone would move from one to the other" is actually based on our experience with the AYRAULT algorithm: most of the tracks are short in space and time, and we could find just a few examples of a cyclone crossing several regions in our dataset.
2. It is true, but not for every region. The Mediterranean Sea is embedded by mountain ranges, therefore the cyclones cannot be always steered easily over long distances without discontinuity.
3. Despite our current study cannot address the link between processes and predictability, we believe it will be a very interesting study to do in the future, and we would like to suggest potential directions for this research, based on our experience. This was our intention when we used the word "suggesting". However to avoid any confusion, "different types of cyclogenesis" is changed to "different types of cyclones".

l. 327: Replace “increases of” by “increases by”.

l. 336 of the track changes document: Done

l. 329-332: I know what you want to say, but I still find it hard to read and digest this sentence. Consider checking and simplifying it further to help the reader. To me, the explanations in brackets (i.e., ...) are more confusing than helping as they address a totally different question (namely track divergence, while the main text discusses cyclone numbers changing with lead times).

The track divergence is essential, a strong (weak) track divergence would imply great (small) errors. As the track divergence increases almost always with lead time, the link between proportion of cyclones followed since early/long lead time is crucial. The whole paragraph is changed to: "The distribution of TTEs is presented for each lead time up to 144 h (Fig. ??a). Both median error and interquartile range increase as lead time increases. For instance, after 72 h lead time, 50 % of the TTEs spans from 80 km to 220 km. Interestingly, the error growth is slower than linear and seems to exhibit two phases: during the first 78 h, the median TTE increases by about 40 km per day, while it increases at a smaller rate of about 20 km per day from 84 h lead time onward. This behaviour can be explained by two different reasons. Firstly, the construction of VDG constrains the tracking to start near the reference track. Given that the median lifetime of the cyclones of our dataset is 42 h, as lead time increases, the proportion of cyclones tracked from early lead times (where the forecast track may have diverged from the reference track) decreases, compared to those tracked from longer lead times (where the forecast track remains close to the reference track). As a result, the error growth tends to slow down as lead time increases. Second, the phenomenon of error saturation also plays a role. For long enough lead times, an ensemble forecast is expected to converge to the climatological distribution. Consequently, the mean and median errors are anticipated to increase at a slower rate at long lead times and saturate ultimately at constant values."

l. 368: On your response to my comment on l. 349 in the previous version of the manuscript: the word “calibration” is often degraded to a statistical post-processing method (i.e., a methodological term), but it is actually much more than that. In statistics, it means a joint property of the predictions and the events that materialize, and is thus equivalent to the word “reliable”.

The aforementioned comment and the answer we made (*italic*) are the following:

‘reliable’ to ‘calibrated’.

*Reliability is an important aspect of ensemble verification, and the spread-error relationship is one way to evaluate it. Perfect reliability can be obtained by statistical post-processing methods (calibration) but here we want to evaluate reliability of the raw ensemble forecast.*

If the two words are equivalent, we prefer to stay on the word “reliable” for consistency.

l. 452: Given that CDFEs yields absolute values, using ‘over-predicted’ seems not appropriate then. Please reword.

l. 463 of the track changes document: The word over-predicted mentioned here does not refer to the CDFE metric, but to the errors of MSLP (not shown). Changed to: "However, it should be noted that on average, the forecast intensity of deep storms in our dataset is slightly too strong from 108 h onward (not shown), while it is slightly too weak in these two previous studies."