

Review of ‘*What determines the predictability of a Mediterranean cyclone?*’ by Benjamin Doiteau, Florian Pantillon, Matthieu Plu, Laurent Descamps, and Thomas Rieutord

The aim of this paper is to investigate the predictability of Mediterranean cyclones when conditioning on different categories, namely region, season, cyclone intensity and motion speed. In a first step, reference tracks are identified in ECMWF’s ERA5 data set and results are presented in section 3, documenting the climatology of that data set and discussing commonalities with and differences to previous studies. The reference cyclones are then tracked in ECMWF’s 11-member ensemble reforecasts to compute location and intensity errors as well as different measures for means, median and spread. These are discussed together with an assessment of forecast reliability in section 4. Finally, practical predictability is evaluated condition on the categories mentioned before, using a novel approach that defines CDFE scores. The underlying idea stems from the formulation of the CRPS and integrates the area between an error CDF and the Heaviside function with the step at 0.

The results show that the predictability of cyclone location is mainly determined by cyclone motion speed, while the intensity and deepening rate are found as the two key factors that determine how good the intensity is predicted. These results are an important first step to identifying and better understanding the main drivers for predictability of Mediterranean cyclones. However, I strongly suggest to rephrase the title as there are many more factors that determine predictability than the four aspects studied here. The paper is well structured and written and the results are generally clear. However, there are two main concerns: First, the cyclones comprise a broad range of storm natures (extra-tropical to tropical-like), which may have different predictability associated with different natures, but this is not addressed at all. Second, two different tracking algorithms are applied to the reference and reforecast data, which may unnecessarily affect the calculation of errors, and hence the results of the predictability study.

I recommend the paper for publication after consideration of the general comments below. Specific comments are also provided to help improve the paper.

General comments

1. Title: I find the words ‘What determines’ too open for what is covered in the paper. The study looks at differences in predictability based on the region, season, intensity and motion speed, respectively. And these factors are definitely important. However, the word ‘what’ could also refer to differences in precursors, the influence of the environment or physical processes (as mentioned in the outlook paragraph in the summary and conclusions section). I therefore suggest specifying the title so that it better reflects what is being investigated.
2. As described in the introduction section, there is a certain range in the nature of cyclones in the Mediterranean region, spanning from deep extratropical cyclones to tropical-like Medicanes. Therefore, it can be assumed that the predictability also depends on the nature of the storm, as different processes may play a role during formation. However, the paper makes no distinction on this point, which is why it is questionable to what extent the results are comparable to other regions (the paper often compares with the Atlantic or the whole northern hemisphere), as they may cover a different range on the spectrum of storm natures with different frequency

distributions. If examining this aspect in the reforecasts is considered to be beyond the scope, I yet recommend to add some analysis and discussion in this regard for the reference tracks, to at least document the distribution of storm natures for your data set. There are different approaches to this, but one way would be to calculate CPS metrics, for which code would be available, for example here:

https://github.com/fredericferry/era5_cps_diagram

3. Apart from this additional analysis, the term 'Mediterranean cyclone' should be defined more precisely in terms of what range of cyclone natures is spanned. Also, please check the entire manuscript to see where the term 'cyclone' can be specified more precisely in order to improve clarity on this point.
4. You introduce the two 'types' of predictability (practical vs intrinsic) in the introduction, but never use these terms thereafter again. Since your study is based on forecasts from the IFS model, I suggest to clearly state once that the predictability you are assessing is of the type 'practical', and then say: 'hereafter referred to as 'predictability' for simplicity'.
5. Cyclone detection in the Ayrault algorithm: According to lines 119-120, the resulting track may be a mixture of ζ -based and MSLP-based points. Why are tracks created based on such an inconsistent use of variables? Since the study uses error distributions of cyclone location (and intensity) as the basis to investigate predictability, I consider this point to be crucial with regard to all conclusions based on it. Could you check how different reference tracks are, when solely using ζ -based track points?
6. Why is a different tracking method used for the reforecasts? For me, it would be more obvious to use the same method for both data sets. In lines 158-160, you say that you refrained from applying the Ayrault algorithm to the reforecasts. But why did you not apply the VDG algorithm to both datasets (i.e., reforecasts and ERA5), to keep it consistent? Is there any particular reason? One downside of this inconsistency is that the 'data and methods' section becomes (in my opinion unnecessarily) longer.
7. As the word "velocity" can be used and understood in different contexts, I suggest using 'motion speed' throughout instead.
8. I suggest to harmonise the titles of sections 5.1-5.4. Either use 'Differences in X' or 'X categories'.

Specific comments

l. 1: 'in a densely populated area' is very unspecific. Rephrase so that the geographical location does not have to be implicitly deduced from the term "Mediterranean cyclones".

l. 5–6: 'this region': Make sure to have clearly stated, either here or in the first sentence, which region you are referring to.

l. 5–6: Incorrect name: ECMWF is the "European Centre for Medium-Range Weather Forecasts".

l. 16-17: "opens the way to identify the key processes..." is what a reader may expect when reading the title, but it is not addressed in the present manuscript. Therefore, I strongly recommend to specify the title as mentioned in one of the general comments.

l. 19: The first sentence of the introduction is very similar to the first sentence of the abstract. However, while Mediterranean cyclones are described as a ‘component of the climate’ in the abstract, extratropical cyclones are introduced as ‘component of weather patterns’ in the former. Is there a particular reason why they are introduced in conjunction with different temporal (and spatial) scales? If so, please clarify, otherwise I would recommend to use consistent wording.

l. 24: Replace ‘life cycle’ by ‘life time’.

l.24: The previous paragraph specifically addresses extratropical cyclones. Here you make a comparison between cyclones in the Mediterranean and other basins, not mentioning the storm’s nature. For the statements you make, are you comparing extratropical / subtropical / hybrid / tropical(-like) cyclones in the Mediterranean and other basins? Please clarify.

l. 24-36: Only with the introduction of medicanes in l. 36 you begin to discuss cyclone nature for the Mediterranean. Please make sure that a reader knows which type (or range on the spectrum) you are addressing in these lines.

l. 47: Note that Baumgart et al. (2019) did not investigate error growth in the specific context of cyclogenesis and cyclone prediction, respectively. Consider this in the text.

l. 53: Replace ‘hardly’ by ‘rarely’

l. 54: The comparison seems to be inappropriate. The purpose of the ensemble is to quantify the uncertainty beyond a purely deterministic prediction, and hence provides an additional perspective. The word ‘robuster’ (should be ‘more robust’) only makes sense if one compares a single forecast with an ensemble mean.

l. 56: There are more intensity categories below the hurricane intensity threshold. Therefore, bringing together the word ‘hurricane’ and ‘genesis’ seems a bit odd. I suggest to say ‘tropical cyclone genesis’ instead.

l. 64: Add comma before ‘but’. Also, I suggest to think of (and write about) Medicanes as a subset of the ‘Mediterranean cyclones’, instead of emphasising the lack of representativeness.

l. 66: What nature (or spectrum of cyclones) of cyclones are you referring to?

l. 69: Replace ‘firsts’ by ‘first’.

l. 70: Replace ‘systematical’ by ‘systematic’, and remove ‘as objects’.

l. 71: What type of vorticity? Relative vorticity?

l. 72: Replace ‘degree’ by ‘degrees’.

l. 74: Replace ‘and this’ by ‘, which’.

l. 76: Since this paragraph reports on results from previous studies, mainly addressing model deficiencies (‘incorrect representation’, ‘systematic slow bias’, etc.), it’s probably worth

specifying the the predictability here and in l. 69 as ‘practical predictability’. But I leave it to the authors to decide.

l. 79: What is meant by ‘with a homogenous configuration’. Please clarify.

l. 79-80: Replace ‘robust statistical signals’ by ‘statistically robust signals’.

l.81: In the main text, you have not referred to specific ensemble prediction system yet, so either remove ‘the’ before ‘ensemble’ or state which one you are going to use.

l. 84: Add ‘cyclone’ before ‘tracking’.

l. 91: Add a hyphen between ‘Medium’ and ‘Range’.

l. 91-92: Remove ‘hereafter’.

l. 95: If you have used a regular lat-lon grid, replace ‘with 0.25° horizontal resolution’ by ‘on a 0.25°x0.25° horizontal grid’, as it is relevant technical information (for the purpose of reproducing the method/results).

l. 98: Replace ‘need’ by ‘step’.

l. 100: ‘Coarser’ than what? Can you give an approximate number?

l. 101: Replace ‘required a specific tuning’ by ‘had to be adapted’, ‘Indeed’ by ‘As stated before’, ‘with’ by ‘have’ and ‘life cycles’ by ‘life time’. Also, add ‘those’ before ‘in the’.

l. 102: Add a comma before ‘and’.

l. 103: Add ‘re-’ before tuned.

l. 104: Add ‘to’ before ‘operate’.

l. 107: ‘in’ → ‘into’.

l. 108: Is there any reason why you have chosen these characteristic lengths? If so, please add a sentence stating it. Maybe also consider adding a sentence explaining the reason for the smoothing (i.e., why you apply a moving average). And since you consider a broad range of cyclones (large extratropical to smaller tropical-like), did you test whether/how your choice works equally well when tracking different cyclone types?

l. 111: To avoid confusion with the two levels used for winds, add ‘at 850 hPa’ after ‘smoothed field’.

l. 112: When retaining the strongest maxima, is this evaluated for each grid point separately on the full set of maxima identified with then 300km radius or does the dropping of maxima affect the search at grid points in the neighbourhood? Depending on the procedure, revise the description accordingly to ease reproducibility.

l. 113: ‘on’ → ‘over’, and add ‘in a given reforecast’ after ‘time steps’.

- l. 114: Replace ‘three steps’ by ‘three-step’.
- l. 115: Add ‘for time $t+1$ ’ after positions, and remove the word ‘for’.
- l. 115: How large is the neighbourhood? Please state it clearly in the text.
- l. 116: The term ‘by taking into account’ is not a precise methodological description that would allow for reproducibility. Please clarify how that distance value variation (which I assume is variance?) is used in the quality criterion, to allow the reader understand how the final position is determined.
- l. 119: Replace ‘in’ by ‘within’, and add a hyphen between ‘3°’ and ‘square’.
- l. 121: Replace the second ‘of’ by ‘if’.
- l. 121-124: In line 120 you describe that it can happen that for some ζ -based track points there is no matching MSLP points identified. In such cases ‘the ζ maximum remains the track point’. How can you check validation criteria based on MSLP, when there are track points, for which no match was found? What MSLP value is associated with such points instead?
- l. 123: Replace ‘life cycle’ by ‘life time’.
- l. 124: Swap words: ‘secondary local’ \rightarrow ‘local secondary’, and replace ‘northern’ by ‘Northern’.
- l. 125: Replace ‘that enter into’ by ‘entering’.
- l. 126: By ‘areas’, are you referring to the areas that will be defined later? If so, it can’t be used here because it has not been introduced yet. Maybe remove the word.
- l. 127: Is ‘the Mediterranean-adapted version of the Ayrault (1998) algorithm’ referring to the algorithm you are using in your work or another algorithm? To me, it sounds like it describes the adaptations you did to the Ayrault (1998) algorithm. Please clarify. And replace ‘slight’ by ‘slightly’.
- l. 129: ‘distribution’ \rightarrow ‘distributions’
- l. 131: What does ‘in the highest confidence level’ mean? Please clarify.
- l. 134: Since the reference tracks are also used as part of the predictability study, I would replace ‘predictability study’ by ‘reforecast tracks’, to be consistent with the title of subsection 2.1.
- l. 135: There is no need for a period per se, you could just run a single reforecast. Therefore, I suggest to replace ‘on a historical period spanning typically several decades’ by ‘starting from historical initial conditions’.
- l. 136-137: Given that you already state in the previous sentence, that ‘reforecasts are forecasts ...’, I don’t see the need for this sentence and would hence remove it. Together with the previous comment, it should be clear that they are run from past initial conditions

(without taking into account observations afterwards). If you remove, also remove ‘therefore’ in line 137.

l. 146-147: ‘life cycle’ → ‘life time’

l. 147: Remove ‘for’.

l. 150: Replace ‘of’ by ‘for’.

l. 150-151: Having read the last two sentences of this paragraph several times, I am not sure I’m getting your point. Please revise it to make it clearer.

l. 152: Same as for the title of section 2.3: ‘predictability study’ → ‘reforecast tracks’

l. 154: Given that you applied a different tracking method to the reforecasts, it is worth mentioning why already after the first sentence of this paragraph? I suggest to move lines 158-160 after ‘... tracking of tropical cyclones.’ (in line 154).

l. 158: Replace ‘track predicted cyclones’ by ‘tracks predicted for cyclones’.

l. 159: To make clearer that the ‘additional step for matching the cyclones’ is needed for cyclones (first identified at lead times > 0) and becomes both more important and challenging as lead time increases, I would replace ‘the cyclones’ by ‘forecasted and observed cyclones, which are first identified at lead times greater 0’. I would remove ‘subjectivity’ as there are several of objective methods in the literature addressing exactly this issue.

l. 161: Add ‘time’ after ‘initialisation’. Also, here and where necessary throughout the paper, replace ‘ ζ maximum’ by ‘ ζ -maximum’.

l. 162-163: The word ‘combination’ is used twice, replace ones.

l. 163: ‘position’ → ‘positions’.

l. 167: Replace ‘and here set at 0.5, and δt the time step’ by ‘of the forecast δt , but here set to 0.5’.

l. 170: Replace ‘using’ by ‘within’ and ‘around’ by ‘on’.

l. 175: Add the statement that you did not apply the other two requirements as in the Ayrault algorithm (i.e., a minimum track length of at least 24h and a minimum intensity of 1005 hPa that has to be reached), and maybe say why. Probably, because you know that there was a cyclone observed based on ERA5 but you don’t want to restrict your ensemble reforecasts too much. Correct?

l. 176-177: A very important information that I would suggest to tell the reader before the VDG algorithm is described in detail! I’m sure it will help avoid some questions that may arise.

l. 181-182: remove ‘with each other, whether in location or in intensity’, as it is a general condition.

l. 183: 'for each cyclone track' is ambiguous as there are observed and (for a unique cyclone usually multiple consecutive) forecast tracks. I suggest to remove it and replace 'defined' by 'calculated'. Also add 'each' after 'of' and remove 's' from 'members'.

l. 184: 'the cyclone life cycle' → 'a cyclone's life time'.

l. 184-185: Is there a particular reason why you chose to calculate spread as the mean of pairwise differences rather than the standard approach based on the ensemble mean? (By ensemble mean one would of course consider only the subset of members that have a cyclone at time t.)

l. 188-189: Remove the four words 'respectively'.

l. 190: 'defined' → 'calculated'

l. 191: remove 'an'. Also add 'forecast' after 'for each'. 'life cycle' → 'life time'.

l. 195-196: Remove the two words 'respectively'.

l. 204: remove 'resp.'.

l. 211: Replace 'CDFE(τ)' by 'CDFE(F_τ)'.

l. 213: Add '.' after 'applied'.

l. 213-214: Replace 'A higher CDFE (respectively the smaller) indicates a poorer (respectively a better)' by 'A higher (smaller) CDFE indicates a poorer (better)'.

l. 215: With your significance test, You can only test for either similarity or difference at a time, as you would need to define a different null hypothesis otherwise. Remove 'respectively', and add 's' after 'CDF' (plural).

l. 215: Replace '(respectively different)' by 'or not'.

l. 221: Add 'the' before 'toponymos'.

Figure 1: The spatial distribution is only one aspect. Isn't it the (relative?) frequency of occurrence what you are showing? This also includes the temporal aspect. Also, remove 'the' in the last sentence. For (b), add a note that the informs about the nonlinear shading levels.

l. 223-249: Following the previous comment, I would suggest to replace 'spatial distribution' by 'relative frequency'. If you change it, make sure to also change it in line 220.

l. 224: Add a comma after 'period'.

l. 228-229: Add commas before designed and after areas.

l. 231: The reference (Trigo et al. 2002) for this statement made here is only mentioned a couple of lines later. Can you include it here already?

l. 247: 'at' → 'of'

l. 248-249: Since it is not found in most of the other studies, do you have an idea of potential (maybe methodological?) reasons for this? If so, it would be worth to mention. Having read lines 125-126 again, I wonder how this condition is actually defined. Does 'enter into' refer to the domain enclosed by the box (as shown in Fig. 1b) or to the entire sea basin? If the former, does it matter whether a cyclones enters the box over land vs over sea? Please clarify.

Figure 2: Replace 'on' by 'over'. What does 'in their mature phase' mean? Is Fig. 2 only based on the most intense fragments of your reference tracks? If so, it should be stated clearly how this is filtered.

l. 267: 'rate' → 'rates'

l. 267-268: Can you add a reference for the link to the Atlantic?

l. 269-279: Here you discuss the depth of cyclones, but it is never stated in the main text (but only in the caption of Fig. 3) how these categories are defined. Please clarify either here or in the 'data and methods' section.

Figure 3: As before, consider replacing 'spatial distribution' by 'Relative frequency'. I suggest to move the definitions of the categories to the main text and then add something along the lines 'See the main text for details on the category definitions.'. For (a), add a note that the informs about the nonlinear shading levels.

l. 279: As suggested in the general comments, I would replace 'velocity' by 'motion speed', and remove 'of Mediterranean cyclones' to stay consistent with the previous subsection titles.

l. 280: Why have you chosen the median and not the mean? Is it because the distribution is so skewed? Probably worth to mention. How large are the differences to the mean? Also, 'life cycle' → 'life time'.

l. 281: Replace 'generally move from the west to the east' by 'move eastward on average' as there are cases where it moves westward (e.g. Medicane Daniel). Can you also add a statement on the meridonal component?

l. 282: Remove 'the' before 'twice'.

l. 283 & l. 285: As is section 3.1, replace 'spatial distribution' by 'relative frequency'.

Figure 4: As with Fig. 1b+3a, replace 'spatial distribution' by 'relative frequency'. Add a note that the informs about the nonlinear shading levels.

l.284-285: 'change' → 'changes', and replace 'from one velocity-based category to another highlights' by 'between the motion speed-based categories highlight'.

l. 296-297: As you do not distinguish between different stages of a cyclones life time 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.

l. 302: I suggest to use a more concise title for 4.1: 'Location and intensity errors'.

l. 303: Remove second 'in', and 'Fig.' → 'Figure'.

l. 304: ', for the all' → ' for the entire'.

l. 305: It is not clear to me which hypothesis you want to test, which is why the topic of significance is not appropriate here. Presumably you want to say that the large number leads to more stable results!?

l. 306: Add 'approximately' before 'linearly', and 'on' → 'in'.

l. 308: 'until' → 'up to'.

l. 309: '3 day' → '72 h', to be consistent with the x-axis.

l. 310: 'not fully linear': Instead, I suggest to describe how it actually is, namely 'slower-than linear'.

l. 311: As you talk about the TTE growth rate x in the previous part of the sentence, 'it increases' is misleading, as the growth rate is actually decreasing. If you replace 'it' by 'TTE', it should be fine.

l. 312 & 316: Maybe replace 'processes' by 'reasons'.

l. 314-315: I find it difficult to understand what point you are trying to make here. Please clarify and rephrase the sentence. Also, do not use 'early lead times' and 'from a few hours', as it is not clear what it means. Use 'short(er)' or 'long(er) lead times' instead. And change 'i.e.' to 'i.e.,'.

l. 316: 'deals with' → 'has to do with'.

l. 316-317: In principle, a forecast can predict anything, but what we want is that it ultimately converges to climatology. Since an ensemble forecast is considered here, 'a random climatological state ...' is not adequate wording. Maybe say: 'For long enough lead times, an ensemble forecast should ultimately converge to the climatological distribution'.

l. 318: Remove 'a' and change 'value' to 'values', as mean and median may not necessarily converge to the same value if the climatological distribution is skewed.

l. 321: What maximal growth rate do you get, when calculating it over the first 48 h only, as Picornell et al. (2011) did?

l. 322: A comparison with the "whole Northern Hemisphere" means comparing to a broader spectrum of cyclone natures (e.g. including true tropical cyclones), and thus different error characteristics (i.e., distributions). If you want to keep this comparison, you should point out

this important difference. Given the much coarser resolution of that study, I wonder whether it would be better not comparing to it.

l. 327: Remove 'In our case,'.

l. 328: 'from' → 'at'. Also, replace 'are slightly late' by 'have slow propagation speed bias'.

l. 329: 'in average' → 'on average'. Also, add 'in the IFS model' after 'forecast cyclones'.

l. 330: Remove 'For their part,'.

l. 332: Add 'ly' after 'previous', and a comma after 'CTE'.

l. 333: 'shift' → 'bias'

Figure 5: Change 'compared to ERA5 in function of the lead time' to 'relative to ERA5 as a function of lead time'. Replace 'first and third' by 'first to third'. Also, replace 'black lines' by 'black whiskers'.

l. 335: 'when looking at the mean': What is 'mean' referring to here? The ensemble mean? The mean over all lead times considered? Please clarify. If the ensemble mean, the lead time reference is missing.

l. 336: Add 'error' before 'distribution', and remove 'around the reference' as you can simply say that the MSLP errors are centered around zero.

l. 337: '3 days' → '72 h'

l. 338-339: Is there a specific reason why you are reporting the IQR at 72h lead time? If you want to include in the text, maybe describe how the IQR increases with lead time.

l. 339: Same as above: The question is whether it is reasonable to compare to this deviating range (i.e., frequency distribution) of northern hemispheric cyclone natures.

l. 340: 'indicates' → 'indicate'.

l. 343: 'event' is rather unspecific. Please clarify what it means. What about using 'cyclone-base time combination' (or 'pair' instead of 'combination')? Admittedly, it sounds more complex, but describes it more precisely.

Figure 6: Suggestion: Replace the first sentence ' by 'Spread-skill relationship at 72h lead time with blue shading representing the number of cyclone-base times populating each bin.'. In the last sentence, replace 'event' as in the previous comment, and change 'in each half part' to 'above and below the diagonal, respectively'.

l. 345-346: Remove the two words 'respectively'.

l. 348: Replace the 'in's by brackets around 'Fig. 6x', and replace 'in both cases' by 'for both aspects'.

l. 349: 'trend' → 'relationship', 'reliable' → 'calibrated'.

l. 350: Since you are make these statements for both plots at once, add about before '60%' as it is '58%' for MSLP. And as above, be more specific and replace 'event'.

l. 353: Suggestion: 'is slightly over-dispersive for most of the cyclone cases' → 'tends to be over-dispersive in most forecasts', and 'remain very poorly predicted' → 'are totally off'.

l. 355: Add 'really' before 'observed', and replace 'the different' by 'other' or 'all'.

l. 357: Add a comma after 'section'.

l. 359-360: Change 'in the spatial distribution, the seasonality' to 'depending on the region, the season'.

l. 362: Replace 'Note that the CDFE having the same unit as the variable considered, the' by 'It should be noted that the CDFE has the same unit as the variable considered. The'

l. 363: Remove two times 'respectively the', and move '(smaller)' after 'greater' in line 362.

Figure 7: Move 'at 72 h lead time' after '(b)'.

l. 364: 'On' → 'in'

l. 365: Add 'the shape of' before and 'error' the word 'its'. Also, 'is closer to' → 'better resembles'.

l. 371: 'sections' → 'subsections', remove 'the' and put 'score' in the plural. Also change 'is' to 'are'.

l. 372: Add 'considered' after 'duration'

l. 374: 'spatial distribution' → 'relative frequencies'

Figure 8+9+10+11: What does 'from every other categories' mean? Do you test against every other category separately and then mark the parts where all intersect? Or do you test against all other (i.e., remaining) categories combined? If the latter, replace 'every' by 'all'. In any case, this should be clearly explained here and in the text. Also, what is plotted when a single lead time has a significant difference (i.e., with the previous and following lead times being insignificant). Is it a circle or a point, or is it possible that nothing is drawn because there is no other point with which a line could be constructed? Also, replace 'The CDFE score' by 'CDFE scores', and change 'is' to 'are'.

Figure 8: Fig. 8c+d are not needed as they do not provide additional information. Another reason to remove them is that the significant parts change as they are probably now derived testing against only two of other regions, which is a bit confusing (as it is not mentioned in the text). Also, add the black circles in panels (a) and (b).

l. 377: Add 'in terms of cyclone track' after 'predicted'.

l. 378 As for the captions of Fig. 8-11, is it really ‘against any other’ or ‘against all other’?

l. 378-382: Truly an interesting characteristic if it can be clearly attributed to the imprint of the diurnal cycle. However, do you have more supporting evidence for this than ‘just’ the line plots? For intensity, one can expect such a relationship, but why should it be the same for location errors? It could be pure coincidence. By construction, the CDFE metric combines the effects of (a) the sharpness of the ensemble (i.e., how distant the errors of the members are spread around the median) and (b) how much the location of the entire error CDF deviates from 0. Can you check the actual CDFs to see if there are differences between 0 and 12 UTC for (a) and (b), respectively.

l. 382: ‘on’ → ‘in’

l. 382-385: If Fig. 8c+d are removed, revise sentences and merge with the text describing Fig. 8a+b.

l. 384: ‘general’ → ‘mean’

l.387-388: Significance is either found or not but can’t be ‘pronounced’ from a statistical perspective. You probably mean that it is found for fewer lead times compared to the location results. However, there are also many significant parts for the intensity results, so maybe drop the statement.

l. 383-389: ‘The Middle East ...’: Maybe you can link to Fig. 3a, which shows that there are no very intense cyclones in this region, which is certainly one reason why the intensity is easier to predict. (Maybe it is worth adding relative frequencies for other percentile ranges as panel(s) to Fig. 3, similar to Fig. 3a.)

l. 392: ‘firsts’ → ‘first’, and remove ‘once again’.

l. 395: ‘or’ → ‘and’

l. 399: change to plural: ‘cyclones’.

l. 403: ‘the difference remains’ → ‘differences remain’

l. 406: ‘seasons’ → ‘season’, ‘The CDFE score’ → ‘CDFE scores’, ‘follows’ → ‘follow’.

l. 406: If significance lines are plotted when a category is different from all other categories, how can it be that SON and MAM are almost on top of each other and yet be significantly different between 48h-102h? Do you have an explanation for this?

Figure 9: Mention here as well why results for SON and MAM are not shown in panel (a)?

l. 408: Different from each other in what sense/variable? MSLPE is shown in Fig. 9b. Are you referring to intensity itself?

l. 411 & 429: Keep using ‘categories’ as before and avoid ‘classes’ to be consistent.

l. 412: Change 'Fig. 10a' to 'Fig. 10' as it applies to all panels. Add 'Fig. 10a' after 'TTE' in line 413.

Figure 10: Remove the word 'pressure'.

l. 414: You make a general statement (about independence) but only show results for a single 12h-period (before peak intensity). Have you checked other periods of the cyclone's life time to be so sure?

l. 419: Here you are comparing to papers that 'showed a poor prediction of the intensity', but Fig. 10d presents deepening rates. I suggest to move this sentence to where Fig. 10b is discussed.

l. 421: 'day 4.5' → '108 h'. Also, where is it shown in the paper that 'in our data set is slightly over-predicted from day 4.5 onward'? Are you referring to Fig. 5b? However, the plots exhibits an over-prediction on average even before 108h.

l. 421-424: Maybe is it not just a matter of sampling cases but also of the different regions that are considered.

l. 425: 'from' → 'at'

l. 427: 'the previous part' → 'Section 5.2'.

Figure 11: Add 'moving cyclones, respectively' after 'fastest'.

l. 430: 'based on' → 'in'

l. 431: 'spatial distributions' → 'patterns of relative frequency'

l. 432: Remove 'examined in Section 3.4' as the categorisation here is applied to a different (namely the reforecast) data set. When removed, also change 'for the' → 'for a'.

l. 434: 'from the beginning of the forecast' is only true for the fast category.

l. 445: 'constant' → 'fixed'

l. 447: Inconsistent years when compared to line 95.

l. 450: 'Comparatively': Do you mean 'In comparison' or 'in contrast'? If the former, the rest of the sentence needs a comparative form (e.g. 'higher').

l. 456: 'good reliabbility of the' → 'well calibrated'

l. 457: Add 'on average' after 'observed'.

l. 458: Depending on whether you have followed the suggestion above, change 'for most of the cases' to 'in most forecasts'. Also, add 'median and mean' before 'errors'.

l. 460: In my view, the newly developed CDFE score is not a generalization of the CRPS but a different application of the same concept, namely applied to forecast errors instead of actual forecast values. Also, remove 'probablistic score' as it is part of 'CRPS' already, and replace 'case' by 'cyclone forecasts'.

l. 462-464: Here and in lines 310-311: Why did you split the error growth rate at these lead times? Is there a particular reason?

l. 462: Add 'nearly' or 'almost' before 'constant'.

l. 464: Add 'with lead time' after 'errors', and 'progressive' before 'saturation'.

l. 470: 'the key' → 'a key'

l. 471: Remove second comma.

l. 472: 'important by causing' → 'considerable, as they can cause'.

l. 480: As before, your study does not distinguish between different stages of cyclone life time, nor cyclogenesis nor cyclolysis. Therefore, replace 'cyclogenesis' by 'cyclones'.

l. 484: Add a comma after 'e.g.'.

l. 488: Very minor, but why do you reverse the order of the categorisations applied in your study?

l. 491: 'firsts' → 'first'

l. 491-492: Move 'at high lead times (several days)' to after 'cyclones', and add 'part of' after 'responsible to'.

l. 500: Add '.'

l. 501: 'Is' → 'It'