Review of "Medicane Zorbas: origin and impact of an uncertain potential vorticity streamer" by Portmann et al.

Background

The authors investigate the predictions of a September 2018 medicane in the ECMWF ensemble system. They identify ensemble members that have differing day-3 representations of a PV streamer involved in the storm's development. They track these differences back to small differences in the initial conditions and show the progression of PV spread in association with an anticyclonic Rossby wave break. The remainder of the analysis focuses on the interaction between the PV streamer and the developing cyclone: distinct precipitation and storm structures are identified in the different sub-ensembles.

Reductions in predictive skill associated with the development of sub-synoptic systems in the Mediterranean region are an important subject for investigation. Similarly, the limits of predictability imposed by PV streamer evolution and interactions between such features and lower-tropospheric circulations are not well understood. Despite these interesting fundamental underpinnings, the current submission suffers from a large number of flaws in organization, preparation and analysis. Although each of these may not be considered fatal in isolation, a significant amount of effort will be required to address all of them thoroughly. Any revised submission of this investigation will necessarily be heavily modified and constitute a new piece of work. I hope that the authors will find the comments below useful as they pursue this research.

Reviewer: Ron McTaggart-Cowan

General Comments

1. The manuscript lacks organization and logical flow. This extends from the highest level of structure (sections and subsections), down to the paragraph and even sentence level. It makes the manuscript difficult to read and follow because concepts and details are disjointed, scattered and frequently repeated throughout the text. Ordinarily, I would consider these kinds of organizational issues relatively minor and possibly within the domain of the authors' discretion; however, in this case they seriously detract from the work and make it very difficult for readers to follow the investigation. Addressing there problems will involve the rewriting of much of the text, but will yield a more focused manuscript that will likely be shortened by at least 1-2 pages.
    a. High Level
        i. The structure of the introduction is ineffective. It begins with a very cursory review of tropical transition and medicanes (including PV streamers), then switches to a Rossby wave discussion that returns to error amplification twice, and then goes back to a very short summary of Zorbas. The latter also includes thesis questions and an outline of the remainder of the study that lacks section information or internal references.

ii. I do not think that the decision to replace the standard "Data and Methods" section with "Operational ECMWF Products" is effective. It means that additional data sources (satellite imagery, lightning detection, etc) have to be described in the text body, descriptions that seriously distract from the analysis when they occur. The same is true of methodological details (e.g. LAGRANTO, which is introduced twice) and the entire discussion of tracking and the CPS (L378-L392). Descriptions of all of these sources and techniques should be centralized in a "Data and Methods" section.

iii. There are numerous forward-references to section 6 throughout the earlier sections of the manuscript. While appropriate internal referencing is a useful tool, these repeated forward-references are likely an indication of poor manuscript organization, especially when they underpin important elements of the analysis (for example, the CPS referenced in sections 3 and 4 but never shown despite a reference to "Fig. 4a", which does not exist in the submission; L159). I think that the synoptic analysis (ideally shortened from its current length by enhanced focus) should include a discussion of the medicane itself, including the CPS. I understand that the medicane is not the intended focus of the work as is repeatedly stated in the text; however, the reader could be excused for thinking that it is because of (1) the title, (2) the multiple introduction subsections that deal with TC-like features, (3) the statement on L317 that the investigation of the "development of a medicane-like system" is an objective of section 6.2, and (4) the pervasive forward-referencing to the CPS analysis in the text.

b. Medium Level

i. Each section should begin with an introduction of the section contents so that the reader has an idea of how the section fits into the larger narrative of the work. A section should not begin with a description of data used in specific figures, as does section 3. Please revise each section introduction to ensure that the reader is logically guided through the work.

ii. Each paragraph should begin with an introduction to the paragraph contents, and should conclude with a statement that relates back to the material introduced. There are very few paragraphs in the manuscript that follow this basic structure. A particularly clear example of a paragraph that ranges far too broadly occurs on L50-64. The paragraph (and subsection) starts with a description of initial condition uncertainty, moves into ensembles more generally, then into PV error growth, and ends with a discussion of tropical cyclones. This lack of focus makes the study very difficult to follow despite the fact that the investigation itself is relatively straight-forward. In this case, the subject of error growth reappears

in the paragraph starting on L73, which further adds to the confusion of the discussion. Please do not simply rewrite the paragraphs noted in this comment: the structure of virtually all paragraphs in the manuscript needs to be reconsidered and revised, an effort that will lead to rewriting of large portions of the work. The readability problems induced by the lack of logical internal paragraph are more than aesthetic in this case, and are serious enough that they significantly reduce the potential impact of this work.

    iii. I do not think that summary paragraphs at the end of a section are useful, particularly given the large number of relatively short sections in this submission. For example, the summary on L163-L167 is redundant with analysis undertaken in the previous page of the manuscript. Please remove summary paragraphs (they also appear at the ends of sections 5 and 6.1) in favour of making the text itself direct, clear and readable (see item 1.c.ii below).

c. Low Level

    i. Reference to caption-level figure details within the text is highly distracting. For example, the fact that precipitation is shown in red solid contours in Fig. 1 is referenced three times in section 3 (once erroneously as "blue contours"; L130), while the fact that QG vertical motion is shown in red contours is referenced twice in the same section. These plotting details are described in the caption, and their appearance in the text detracts from the flow of the analysis. Figure and panel references should be enough to guide readers through the discussion. Please consider removing caption information from the text body throughout the submission.

    ii. The writing style is too informal and lacks the precision required for scientific text. For example, the outline of the manuscript is described as "a journey" on L101, the analyses in section 6.2 "hint" at airstreams (L321), and the first person plural ("we") is used heavily throughout the submission. The introduction of section 6.2 (L320-324) can be summarized as, "we don't do anything thoroughly here, but we show stuff that's different and make some guesses about what that means; then in the next section we do it properly". I don't think that that kind of introduction (or approach to the analysis) will make readers want to continue to invest their time in the rest of the section. Throughout the submission, irrelevant details clutter the text (e.g. does it matter that 1800 UTC 26 September is "in the evening of the same day" on L126?), and ill-defined concepts reappear throughout the analysis (e.g. the "C-shaped" PV cutoff with a "dent" and "dent structure" on L138, 141 and 146, respectively). Every effort should be made to make the text succinct and readable, so that the analysis does not get lost in superfluous details and unnecessary bridging statements.

2. I think that cyclogenesis in cluster 3 is really interesting, but that the discussion in the current study misses the opportunity to capitalize on its uniqueness (I do not think that section 6.3 is sufficient in this respect). It looks to me like this is an excellent example of a nonlinear response / bifurcation leading to a real limit on predictability. Clusters 1 and 2 are simply phase shifts of the same cyclogenesis event. From a guidance perspective, both are reasonably useful at least in terms of situational awareness. Cluster 3, however, looks to me like the development of a different cyclone. There's an 850 hPa circulation south of Turkey in all of the groups at 1200 UTC 26 September (Fig. 8, column 2). In fact, a cyclone has already formed in this region in many of the cluster 3 members and one of the cluster 1 members (also shown in Fig. 9a). In groups 1 and 2, the low between Crete and Cyprus disappears as the PV tail promotes development along the African coast. In group 3, the pre-existing cyclone intensifies and fractures the PV streamer as the low retrogresses towards Crete (Fig. 8, column 3). By 1200 UTC 28 September, the medicane lies in the central Mediterranean in groups 1 and 2, but it is a completely different storm that is centered on Crete in group 3 (this differs from the interpretation implied by discussions on L286-L290 and L303-304 of the submission). So the relatively small difference in the location of the PV streamer axis (a linear change from west to east of the observed location) leads to a highly nonlinear response in the form of development of a new cyclone (groups 1 and 2) or intensification of an existing circulation (group 3). (Note that a couple of centers form over northern Africa in group 3 at 1200 UTC 27 Sept – Fig. 8k; these are cases in which the response to the change in PV streamer position is linear.) The theory that group 3 is fundamentally different from the others is supported by the precipitation patterns and tracks (Fig. 9; noting that the large track jumps between North Africa and Crete are unlikely to be accurate) parcel trajectories (Fig. 10) and parcel properties (Fig. 11). Because a nonlinear response / bifurcation is known to impose strong limits on predictability, identifying and describing such behaviour in this case would be an important outcome of this work. I hope that some of the length reductions achieved by improving the manuscript's focus and organization can be invested in a much more thorough analysis of this possibility.

3. The values of QG vertical motion seem too small to be very meaningful despite being described as "strong" in the text (L441). Vertical motions of ~0.5 mm/s and 1 mm/s (0.005 and 0.01 Pa/s) are plotted in Fig. 2, while values of up to ~5 mm/s (0.05 Pa/s) are plotted in Fig. 7. These values are all well within the typical rms of QG vertical motion at midlatitudes and mean vertical motions across the globe (Stepanyuk et al. 2017). If these calculations and plots are correct, then the vertical motion forcing from the upper levels is almost irrelevant to the real vertical circulations in most cases. Such weak vertical motions would need to be sustained in-place for many hours/days to have any appreciable impact on moistening or stability. For example, air in the peak ascending region in Fig. 2c ascends <10 hPa in a day in response to QG forcing, an ascent rate that is dwarfed by the 600 hPa ascent in the rising parcels near the centre. If the calculations are correct, then the relevance of the PV streamer to ascent and cyclogenesis needs to be seriously reconsidered in this case, an exercise that will likely lead to

conclusions that are completely different from those arrived at by the current submission.

4. The motivation for the case study approach adopted by the study is weakened by passages that highlight case-to-case variability, and is not supported by a clear statement of the useful aspect of the case study framework. The dominance of case-to-case variability is particularly emphasized on L38 and L84, with the latter appearing to be a direct criticism of the case study as a useful analytic tool. It is good to identify the limitations of the adopted investigation technique, but this criticism should be balanced with a clear description of what the case study approach can provide that other types of analysis (e.g. climatology) cannot.

5. The analysis of vertical coupling in section 5 is not quantitative enough to be included in the study. Despite significant discussion of Fig. 7e-h (L241-253), the strongest conclusion that is draw is that it is "most likely" that baroclinic instability is active. Even this conclusion appears to over-reach the analysis given that no baroclinic growth rates were computed. Given that the Icelandic low is not the focus of this investigation and that the left-hand column of Fig. 7 shows a convincing evolution of short-wave anomaly growth, I think that the right-hand column of Fig. 7 and the associated discussions should be removed. If this analysis is to be retained, then there needs to be a real quantification of baroclinic coupling and associated growth rates [note that the 12-18h time scale is very rapid for pure baroclinic growth, which typically has a doubling time scale on the order of a day (Hakim, Encyclopedia of the Atmospheric Sciences) and suggests that moist processes are likely to be very important].

6. The study of PV error growth by Baumgart et al. (2018) is referenced in the introduction, but not in section 5, where the left-hand column of Fig. 7 bears a striking resemblance to Fig. 3 of that work (albeit with a compressed time frame). The discussion of the importance of non-linear upper-level Rossby wave dynamics here follows closely that of Baumgart et al. (2018), so much of this description could be replaced by citations and comparisons. The Torn (2015) normalized difference is a useful measure, so compressing section 5 to focus on that metric in the context of the Baumgart et al. (2018) interpretation of this process would allow for a dramatic shortening of this section and serve to place this submission in the context of investigations by other groups.

7. Assessing the significance of the differences discussed in section 5 is important; however, the technique and in-text descriptions should be revised. Wilks (2016) provides a description of problems with the multiple-testing technique (as adopted in this study), which can lead to over-confident statements about significance. Please consider using the false discovery rate here. Additionally, the level at which the differences are considered significant is not identified in the text, and only appears in the Fig. 7 caption (is 0.05 used throughout?). Note that there is currently a reconsideration of the use of the term "significant", which appears to be leaning in favour of providing p-values rather than definitive statements about significance. I'm not very familiar with that discussion, but it might be of interest to consider during revision.

8. I am surprised not to see any references to Wiegand and Knippertz (2014), who study the representation and predictive skill of anticyclonic RWB and PV

streamer formation over the Mediterranean region in the ECMWF ensemble (i.e. an earlier version of the same system used here). That work seems so directly relevant to this study (including the conceptual diagram in Fig. 10 of that paper) that it should be leveraged heavily in this investigation, particularly in terms of putting the forecast uncertainty in this case in a broader context.

9. The numbering of clusters forces readers to remember the mapping: 1 is centered, 2 is west and 3 is east. Why not call the clusters C, W and E? Then the Fig. 8 rows could be reordered to W, C, E so that there's a progression in the columns rather than having the PV streamer location (and eventual cyclone location) jumping around.

10. Throughout the study, the "surface cyclone" is discussed by the 850 hPa heights are shown. Showing 850 hPa winds is useful, but I don't see anywhere in the manuscript that the 850 hPa heights are essential to the analysis. I think that all plots that currently show 850 hPa heights should be replaced with mean sea level pressure for consistency with the text.

11. Throughout the study, short-range ECMWF forecasts are used to estimate precipitation accumulations. To avoid model biases and potential "twinning", it would be preferable to use an independent product. The GPM IMERG is readily available and would be a better choice for this study than stitched-together IFS forecasts.

12. Advection of cold air over warm Mediterranean waters is identified as a factor that increases latent heat fluxes and promotes convection; however, this effect is not quantified in the current investigation. The OAFlux dataset covers the period of interest and is readily available for this kind of study. Please consider supporting the claims made in the manuscript with an analysis of OAFlux (or equivalent) surface flux estimates. An augmented surface flux analysis may particularly interesting if model-predicted fluxes are found to be very different between groups 1/2 and group 3 (see item 2 above). Such an analysis is essential if the categorical statements about surface fluxes currently found in the conclusions (L429) are to be retained.

13. The manuscript really needs to be clear about whether the medicane itself is a focus of the study. In multiple passages, it is stated explicitly that the medicane is not going to be investigated as part of this work (e.g. L97, L161, L320). However, much of section 6 is dedicated to the evolution of the medicane, including trajectory and CPS analyses. The title of the manuscript also emphasizes the storm morphology and will attract readers interested in medicanes. It feels as though the work was initially focused entirely on the PV streamer, and that "mission creep" has led to the introduction of more storm-scale-relevant material. Please reconsider the statements that disavow the relevance of the medicane structure for this work in an effort to remove what seems like a fairly important internal inconsistency in the manuscript.

14. Why are the ECMWF data coarsened to 1°, and how is it done? The result is very poor resolution in the graphics, and if it not done carefully, the operation could result in aliasing. Is a conservative remapping used? This is a particularly important question for the precipitation field, where the difference between

sampling/interpolation and remapping/aggregation can be enormous when the degradation of resolution is so large.

15. Most published works do not consider "medicane" a proper noun (and it is therefore not capitalized). This is analogous to "hurricane", which is only capitalized when a specific storm is discussed (e.g. "Some think that Hurricane Katrina was a category 3 hurricane at landfall"). Consider using lower case "medicane" throughout except in named reference to Medicane Zorbas.

16. The terms "air mass", "airstream" and "parcel" seem to be confused in relation to trajectory analyses (L139 and section 6.2). An "airstream" is a loosely defined concept, but I think that it would be represented by a high density of air parcel trajectories in a limited area. Then the phrase "trajectories of the airstreams" (Fig. 10 caption) doesn't make sense unless the airstream (a feature in storm-relative coordinates) is somehow tracked over time. Similarly, trajectories do not track "air masses" (L139), but parcels. The difference is important, because it is unlikely that all parcels in an "air mass" are ascending near the cyclone centre.

17. The trajectory analysis in section 6.2 is incomplete. The suggestion that moistening is occurring because of surface latent heat fluxes (L345-346) implies that the parcels are in contact with the surface; however, the vertical position of the parcels is never shown. It is also possible for parcels to be moistened by evaporation of falling precipitation or by turbulent mixing. It is therefore not demonstrated that enhanced surface fluxes are responsible for the moisture changes in groups 1 and 2. The same is true for the potential temperature analysis on L346-348: surface fluxes are only one possible reason for potential temperature increases, and only influence parcels if they are in contact with the surface (even at above-surface levels in the boundary layer, the moistening/heating mechanism would be turbulent flux convergence rather than surface fluxes per se). The lack of information about the trajectories makes it impossible for reviewers or future readers to confirm the validity of the conclusions drawn at the end of this section (L356-370).

18. Section 6.2 ends with a set of suppositions and conjectures based on an incomplete trajectory analysis (see previous item) climatological behaviour. As a result, terms such as "could favour" and "might support" are used instead of definitive statements. If the analysis and descriptions in this section cannot be made robust enough to be able to conclude these statements definitively, then this section should be removed.

19. The description of the CPS (L386-392) is insufficiently detailed to allow independent confirmation of the results (a requirement for publication). Because of the small scales of medicane structures, the hurricane-based radii are usually reduced for studies of Mediterranean storms. Was the same done here, or were the original hurricane-based values used?

20. I don't understand the "deep warm core" (DWC) analysis in Fig. 12. Take groups 2 and 3, for example. They have 12 and 18 members, respectively. The average number of DWC in group 2 is 7.2, and 7.0 in group 3 according to Fig. 12. That number is "per ensemble member", so multiplying by the relevant ensemble size yields 7.2*12=86.4 for group 2 and 126 for group 3. However, the total number of DWC steps for group 2 is given as 43, and that for group 3 is given at just 14 at

the bottom of the plot. In the text (L404) the reader is told to consider the group-3 DWC analysis "with caution, due to the small sample size". However, the average number of DWC steps per ensemble member is as large in group 3 as it is in group 2: why is the sample size so small? There seems to be something about the number of **sequential** DWC steps ("duration"), but that is never clearly stated in the text or caption. What is wrong with my interpretation of the DWC analysis?

21. Throughout the text, equivalent potential temperature gradients are used to identify both baroclinic zones and moisture gradients [L125, L132, L137 (where the 850 hPa theta-e is inappropriately used to identify a "weak surface cold front") L153 and elsewhere]. Strictly, neither of these is guaranteed by a theta-e gradient, which may arise as a result of either in isolation. If baroclinicity is important, then potential temperature (or temperature on an isobaric surface) should be shown. If moisture is important, then it should be shown. Theta-e is a very useful quantity for assessing convective potential and is a useful way to identify the warm sector for the trajectory analysis, but it does not replace the more basic fields for questions of baroclinicity and moisture.

22. There are a lot of very specific geographical references throughout the text, probably more than there need to be. I'm a geographer, but I still found myself having to look for specific place-names on maps. It would be very useful to have a new Fig. 1 that shows (at least) the storm track and labels for all place names referred to in the text.

23. The conclusions of the study are not supported by the evidence provided in the text:
    a. The "clustering" technique is not rooted in a mathematic definition and fails to guarantee the separation of the members into distinct "scenarios" as stated in the text (e.g. at L481). Is it true that there are three "distinct scenarios"? I agree that there are two (see General Comment 2), but I don't see why there are three. Groups 1 and 2 are distinguished only by the fractional overlap of the PV streamer, and there was no demonstration that there is any sort of heterogeneity in overlap space. This is a weakness in the analysis that results from the failure to use a true clustering analysis, and the decision to rely on a classification heuristic. There is no guarantee that group 1 and 2 events are separate from each other in any kind of meaningful way, and selection of a different overlap threshold (70%, for example) would result in the progressive reclassification of members from one group to the next. To demonstrate the presence of different scenarios, a true cluster analysis should be performed, and the optimal number of groups should be identified (e.g. using the "elbow method).
    b. There was no analysis of the near-surface flow induced by the PV streamer, so how is the conclusion about induced advection (L432-434) supported by the evidence provided in the submission? Particularly given the limited spatial extent of the streamer immediately prior to cyclogenesis, it is possible that the induced near-surface flow is very strong. For the arguments regarding air parcel modification by surface fluxes, the parcels approaching the centre in groups 1 and 2 must be in contact with the

surface, putting them as far as possible from the upper-tropospheric streamer.

c. I cannot see what part of the analysis is used to conclude that the group-1 PV streamer was better able to "maintain the cyclonic circulation" (unclear whether this refers to the upper- or lower-level flow) than the group-2 or group-3 features (L434-438). There appears to have been a rigorous analysis behind this statement (something that determines the number of members that meet a "condition"), but I don't know what section this analysis was described in.

d. The increase in the amplitude of the cyclonic PV anomaly from about -0.5 PVU to beyond -2.5 PVU (combined with a rapid areal expansion) over the 24-h period ending at 1800 UTC on 25 September (Fig. 7b and d) is "rapid" as stated on L442. However, as noted in item 4 above, this growth rate appears to exceed that expected for typical midlatitude baroclinic growth. It is highly likely that moist processes are involved, but because no estimates of growth rates are made in this study, it is impossible to know. It is therefore also inappropriate to conclude that the observed growth is "as expected from baroclinic instability" (L442) because the expected value remains unknown in the context of this work.

e. It is unclear to me what part of the analysis demonstrates that "the contributions of diabatic airstreams … were negligible for the uncertainty amplification in this case" (L444-445). The non-conservative evolution of the PV streamer was remarkable in this case (Fig. 2), and the impact of diabatic PV reduction in WCB outflow on ridge amplification during the upstream RWB (Fig. 7a-d) was not analyzed in the study, as far as I can tell. This statement about the role of diabatic process on forecast uncertainty (L444-445) is very strong, inconsistent with previous work, and needs to be clearly supported by the presented analysis.

Minor Comments

There are a relatively large number of grammatical errors in the submission, which I have not itemized here because of the major reworking of the text that will be required to address the issues identified above.

1. [L50] It is not clear why the introductory reference to paramterization uncertainty is useful here, where initial condition uncertainty is described in the subsequent passage. I would suggest starting this paragraph with "A major source …".

2. [L51] I don't think that "slight uncertainties in initial conditions **typically** grow" (my emphasis), because the majority of uncertainties in any given analysis project onto decaying modes in the atmosphere (Privé and Errico 2013). I think that it would be more precise to say something like, "Slight uncertainties in the initial conditions that project onto the growing modes of the atmosphere can increase in amplitude during the forecast and potentially …". You could also just replace "typically" with "can" in the current phrase.

3. [L87-L94] Suggest dropping this subsection in favour of the analysis in section 3.

4. [L95-L97] Having a clear set of objectives is a good idea, but these questions are framed in a way that is too complex to make them useful for the reader (e.g.

"what is a and what of b leads to c and d in e"). Suggest simplifying or removing these questions.

5. [L99-L105] Provide a standard outline with section references.
6. [L108] How are the ensemble members "perturbed": initial conditions, stochastic physics, SPPT, etc?
7. [L111 and elsewhere] The word "data" is plural, so "data are available", etc.
8. [L115] What climatology is used for the ACC calculation?
9. [L122] Reference to a URL is inappropriate. At the very least, an access date needs to be provided. Consider including lightning strike information on the plots, rather than making reference to external information that may not be permanently available.
10. [L152] How is conditional instability identified in this analysis?
11. [L159] Figure 4a does not exist and Fig. 4 is not the CPS.
12. [L207] Reference to Fig. 7 is out of order.
13. [L221] At what level are the differences significant?
14. [L231-L232] This sentence doesn't make sense: does the amplitude "propagate" at a different speed from the difference? Are you differentiating between phase speeds and group velocities here? Please rephrase to make this clearer.
15. [L263] The section title should be much clearer, and not read like a news headline.
16. [L274] Three different time references begin this sentence. Please determine whether it is the time relative to streamer extension, Gregorian date/time, or forecast time that is most relevant here and stick to this description of the first column of Fig. 8.
17. [L278] I don't see that cluster 3 trough is "clearly" shifted to the east of the analysis at 1200 UTC 25 September (Fig. 8i). Instead, I see a trough that is too narrow, notably on the upshear flank over Germany.
18. [L281] Why isn't significance plotted here as in the first column?
19. [L304-L305] This looks like more than just smoothing of the ensemble mean. Because averaging is a linear operation, the area-averaged ensemble-mean precipitation should match the observed values if the ensemble does not under-predict rainfall.
20. [L297] Are these SLP changes computed from the central pressures of the ensemble members, of from the ensemble mean? The search for a minimum central pressure is not a linear operation, so the results will likely be sensitive to the method. Particularly given the broad spatial distribution of group-2 centres, some/much/all of this apparent weakening may simply be the dilution of the ensemble mean if the ensemble averaging is done first.
21. [L312-315] The four lines of hypothesis here would be much better invested in the actual analysis rather than this forward-referenced supposition (I recommend the removal of this whole paragraph as noted in item 1.b.iii above).
22. [L317 and L320-L321] There seems to be an internal inconsistency here. On L317 the objective of the section is stated to be "to investigates … subsequent development of a medicane-like system". However, on L320-321 you state that you "do not identify low-level warm cores directly and do not investigate their formation in detail". Because the warm core is one of the primary structural

ingredients that distinguishes medicanes from typical Mediterranean cyclones (considering the CPS), these two statements seem to be in direct conflict.

23. [L321] What do you mean that you don't identify warm cores directly? The CPS-based warm-core detection is the basis for a large part of section 6.3.

24. [L344] Do you mean a larger increase in specific humidity in groups 1 and 2, than in group 3? This sentence suggests the opposite, likely because the intended target of the pronoun "this" is unclear (although the construction suggests group 3).

25. [L353-L355] What is the physical relevance of this comparison?

26. [L393-L399] Why bother with a set of conjectures right before performing the actual analysis? A far more direct approach would be to explain why the fractions of medicanes in each group differ, based on the analysis presented in earlier sections. The conjectures do nothing to build suspense for the big "reveal" of Table 1, and just serve to consume five lines of text unnecessarily.

27. [L399-L401] This text contains every number shown in Table 1, without offering any physical insight. Choose to present these numbers either within the text, or in a table, but not both.

28. [L413-415] How is it concluded that "the detailed interaction between the surface cyclone and the upper levels become limiting factors" for predictability? Why can't internal storm processes or air-sea exchanges be the limiting factors? Those processes have not been investigated or ruled out as limits on storm structure predictability in this analysis, as far as I can tell.

29. [Fig. 6] Can the map resolution be increased a bit? (Similar in Fig. 7 zooms.)

30. [Fig. 6] At what level are the contours significant?

31. [Fig. 6] The means are too similar to be usefully distinguished on the plot. Consider plotting the full ensemble mean only rather than solid and dashed contours.

32. [Fig. 7] Is that a reference vector between (d) and the colour bar? If so, it should be highlighted and described in the caption. If it isn't, then one should be added.

33. [Fig. 10] Use a fixed domain to ease comparison between panels.

34. [Fig. 10] What does the colour-coding of the trajectories represent (the last sentence of the caption is not clear about what is indicated "in colors")? Are different members assigned random colours? Why are there fewer cyclone positions in the groups than members within the groups? Are there multiple trajectories ending at the same point because of the degradation of the grid resolution? If so, there should be some way to represent the number of overlapping triangles (potentially the size of the triangle).

35. [Fig. 10] How does the maximum "percentage of ensemble members with an airstream occurring at the specific grid point" occur outside of the trajectory envelope? For example, the maximum departure frequency in Fig. 10b occurs poleward of any trajectory. Is it because these trajectories are actually averages of many trajectory calculations? If so, then there must be some unusual spatial distributions to obtain density maxima away from the means. How many trajectories are computed in each member?

36. [Fig. 11] Why are radii the best way to identify the blue and green lines? It would be clearer to label the blue line "center" and the green line "warm sector" because the radii are technical details rather than relevant features.

References

Privé, N. C. and Ronald M. Errico (2013) The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment, Tellus A: Dynamic Meteorology and Oceanography, 65:1, DOI: 10.3402/tellusa.v65i0.21740

Stepanyuk, O., Jouni Räisänen, Victoria A. Sinclair & Heikki Järvinen (2017) Factors affecting atmospheric vertical motions as analyzed with a generalized omega equation and the OpenIFS model, Tellus A: Dynamic Meteorology and Oceanography, 69:1, 1271563, DOI: 10.1080/16000870.2016.1271563.

Wilks, D.S., 2016: "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. *Bull. Amer. Meteor. Soc.,* **97**, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1

Wiegand, L. and P. Knippertz, 2014: Equatorward breaking Rossby waves over the North Atlantic and Mediterranean region in the ECMWF operational Ensemble Prediction System. Quart. J. Roy. Meteor. Soc., **140**, 58-71.