Authors’ response to Reviewers for WCD-2020-43

We thank both reviewers for taking the time to review and comment upon our manuscript. Following the comments we have made substantial revisions to the manuscript.

The main changes are the following :

- *A significant reduction in the length of the manuscript, which now makes it more in line with other, published, articles in this journal.*
- *We have also provided an in-depth response to the inversion of the QG equation (with some addition in the manuscript) that we hope alleviates the reviewers' concern. Some figures are inserted in the reply document showing the strong similarities between the inverted and model vertical velocities even though some differences in the amplitudes do exist. We do not expect inversions of the QG equation to entirely match up to values modelled in full GCMs but the strong resemblance between the two fields together with the reasonable amplitude of the retrieved field give confidence on the results.*
- *All the figures have been updated to take into account a correction in our code of inversion of the omega equation. A spatial filter initially applied to the Q-vector remained in our code that was useless and tended to underestimate the dynamical omega. The spatial filter has been suppressed in the revised version.*
- *Some figures of the initially submitted version using older versions of the runs with slightly different parameters have been updated. The updated figures do not change the main results.*

We provide detailed responses below with our comments and changes in green italics. Lines of the changes within the revised paper are indicated in our responses. The new manuscript with the tracked changes is also provided at the end of the reply document.

Kind regards,

David Flack, Gwendal Riviere, Ionela Musat, Romain Roehrig, Sandrine Bony, Julien Delanoe, Quitterie Cazenave and Jacques Pelon

Referees' comments are in black and authors' answers in green

Response to Anonymous Referee #1

I have more expertise in the dynamics of extratropical cyclones than in the cloud microphysics. Thus, my comments mostly deal with the general presentation of the study and that part of the content which is related to the dynamics of the system. General comments: In this study, the authors evaluate the performance of two climate models simulating the evolution of one North Atlantic extra-tropical cyclone (so called Stalactite Cyclone). The authors compare the models with two horizontal resolutions against ECMWF operative analyses. Quasi-geostrophic omega equation is applied to distinguish the dynamic and thermodynamic (diabatic) processes within the cyclone. Furthermore, airborne measurements of microphysical properties of the clouds are compared to the model simulations. I found the topic of this paper interesting and suitable for the scope of WCD journal. The language is clear and understandable. However, I think that the paper suffers from some flaws in its organization, which are mostly due to its length. Therefore, more effort is needed to make the presentation of the paper and its message clearer for the readers. I hope that my comments will help the authors in this work.

Recommendation: major revision

Major comments, which are partly connected to each other:

1. In general, I think that the authors are trying to put slightly too much content into one paper and thus the manuscript was partially hard to read and lacking a clear and consistent storyline. In my opinion, answering comprehensively to the objectives 1 and 2 listed in the introduction would suffice perfectly to the topic of this paper. In its current form the point 3 feels somewhat disjoint and was, to my opinion, particularly difficult to understand. I think that the main results from points 1 and 2 (e.g. the fact that there was no change in the relative importance of diabatic processes with increased resolution) are very interesting and worth publishing alone without focusing on the comparison of microphysical properties with the models and observations, which in turn would be suitable as its own manuscript if investigated properly. Moreover, there is room for improvements in the organisation of the paper. For instance, between the introduction and Section 3, the authors present the overview of the Stalactite Cyclone. I think that this kind of section would better belong right before Section 4, where the representation of the cyclone by the models is presented. Now, after reading all the lengthy details related to the models, observations and the equations, the reader has already forgotten the whole overview of the cyclone itself.

We have extensively edited the manuscript from the original submission to reduce its length, in particular in the methodology section. We have re-read and understood your point about point 3 and as such we have moved question 3 to question 4 and introduced a new question to act as an intermediary between points 1 and 2 and the final question (see lines 66-70 of the revised paper). We feel that this new question puts greater emphasis on the link between the model and the observations, and why we look at microphysical properties.

The new question introduced reads as follows:

“Are there any differences in the diabatic processes related to microphysical properties between the different models?”

We feel this question improves the link as first of all it determines if there are differences in the diabatic heating (previously discussed) and links them to the observations by considering diabatic heating related to microphysics in the models. Therefore we feel there is greater justification and link between Section 4.3 and the other parts of the paper. Further discussions on the importance of the last point are included in the response to the next comment.

Concerning the comment related to changing the order of section 2 and section 3, we decided not to change the order of these sections because section 3 dedicated to the presentation of the models, diagnostics and observation needs to be read after the synoptic description of the cyclone. This is needed to understand the choice of the initial time of the simulations in section 3.1 and to visualize the flight legs within the cyclone in section 3.2. As previously mentioned, we reduced the length of section 3 (from 5 pages to 3 pages), in particular section 3.3 dedicated to the presentation of the omega equation. As such, the description of the main features of the cyclone in the models (cyclone track and pressure evolution in the beginning of section 4) is not so far from the overall description of the cyclone (section 2).

2. Related to comment 1, the study as a whole is a very long read. The paper in its current form has > 11 000 words (from Abstract to Acknowledgements), 13 figures and altogether 38 pages (+ additional seven pages in the supplementary material). Although to my knowledge WCD does not explicitly limit their manuscript lengths, the longevity of the paper did make me lose my interest in reading through it at the first time. For reference, AGU journals (https://www.agu.org/Publish-with-AGU/Publish/AuthorResources/Text-requirements) are recommending up to 25 publication units (PU) for their manuscripts, where 1 PU is 500 words or 1 display element (figure or table). With the 13 figures and 2 tables, the current manuscript would be left with 15 PUs for the main text, which would correspond to 7500 words. The current length is now > 50 % longer. Please consider condensing the paper. I think the best way to do this would be to focus clearly only on objectives 1 and 2 (as suggested in comment 1), or expressing really the main results from points 1, 2 and 3 in a much more condensed way.

Using word count software indicates that the original manuscript’s word count is 9896 words, and so we base our reductions on this figure.

We appreciate that this is a long manuscript and as such we have endeavoured to reduce the length of the manuscript. We have decided to take the route of shortening whilst still discussing the three points. We chose this option as we feel that section 3 is important to be included within this manuscript, and is important work for the community as a whole, as it brings forth some concerning aspects of model behaviour, it indicates areas that could be improved, and areas that can have important feedbacks on the interpretation of cyclones within the models. Part of the implication of Section 3 is that we may have a similar cyclone in both climate models but we are getting it for different reasons and this is feeding back onto the dynamics with the different deepening rates. Thus, Section 3 helps to explain differences seen in Section 2. Indeed in the Summary of the manuscript we state “<u>Finally,</u>

<u>and arguably most critically, it warns that although climate models may produce similar cyclones they can be doing so for very different reasons and these reasons are likely to have an influence upon other areas of the climate system and the response of model cyclones to climate change.</u>" Further to this, Reviewer 2 has also indicated that the inclusion of these remaining figures is important to include within the manuscript. Thus we feel these are good justifications for keeping the discussion around point 3.

The main areas we have focused on is reducing the methodology section (including transferring the model descriptions into a table and reducing the diagnostics part), but also condensing the main results. After our reductions the word count is 7826 [which reduces our submitted manuscript page count from 38 to 30 pages: 25 of those as text pages (including data availability, the link to supplementary material, author contribution and acknowledgements)]. We think this significant reduction of words (2163 less) and page count makes the paper more readable as requested by the reviewer while keeping its scope the same.

3. The manuscript is full of acronyms which are mostly related to model names. I think that the abundant use of acronyms was one of the reasons which made my reading less enjoyable. In the abstract, you call the models with names CNRM-CM6-1 and IPSL-CM6A. In the main text, these models are referred mainly to with their atmospheric components (if I understood correctly) called LMDZ and ARPEGE. Again in the summary you change to CNRM-CM6-1 and IPSL-CM6A. This was very confusing for a reader who is not familiar with these models. The inclusion of LR or HR in some places makes the names even longer (e.g. IPSL-CM6A-LR(-HR)). I strongly suggest to use short and consistent names throughout the whole manuscript, e.g. CNRM and IPSL, with possible -LR and -HR suffixes.

We appreciate that the over use of acronyms can make reading the manuscript difficult, and that our change of acronyms may not have been as clear as we intended. However, we also note that when comparing two models we need to refer to them by their full acronyms first and shorter ones throughout the main text so this will naturally lead to a large amount of acronyms being used.

We have tried our best to reduce the number of acronyms used throughout the paper. To facilitate this we now only introduce the full length, official, climate model names in the abstract and methodology when the models are introduced. Throughout the rest of the paper we only consider the (shorter) atmospheric component model names. We use the full length atmospheric component names in the abstract (lines 3-5), the introduction of the models (lines 118-121) and briefly again in the summary (lines 401-403) to act as a reminder for the contracted acronyms we use throughout the main manuscript. Thus now we refer to LMDZ and ARPEGE throughout the manuscript (with the suffixes LR and HR used only where necessary). We have also improved the signposting of this within the main manuscript.

Minor comments

1. Title: I have a feeling that the part of the title "How Well do Models Represent the Development of Extra-Tropical Cyclones?" is a bit too vague, given that you have investigated only two models and one cyclone. Please consider having a more specific title.

The new title is now: “Representation by Two Climate Models of the Dynamical and Diabatic Processes Involved in the Development of an Explosively-Deepening Cyclone During NAWDEX”.

2. ‘L4 and thereafter: CMIP5 and CMIP6 should be written together, and not “CMIP 5” or “CMIP 6” (see e.g. https://pcmdi.llnl.gov/CMIP6/).

This has now been corrected throughout the manuscript.

3. L4: Can you write down both resolutions explicitly, or leave both away? Writing only one resolution with its numerical value (0.5, HR) left me missing the other one, because at least I am not aware of the CMIP6 native resolution (LR).

We had originally put it as CMIP6 and specified the value as the two different models had different CMIP6 resolutions (and so we did not want to increase the length of the abstract). However, we understand your point and have now referred to both resolutions within the manuscript (as it is important information to include in the manuscript). We have expressed the resolutions by giving approximate grid lengths in km (150-200 km for LR and c. 50 km for HR) (lines 4-5).

4. L84: Section 2 title “NAWDEX IOP 6: The Stalactite Cyclone” could be more descriptive, e.g. The development of the Stalactite Cyclone, or The life cycle of the Stalactite Cyclone.

This has been re-named to ‘’The Lifecycle of the Stalactite Cyclone (NAWDEX IOP 6)” (line 82).

5. L89: I was confused about the word Diabatic Rossby Vortex (DRV), because Boettcher and Wernli (2013) talk about Diabatic Rossby Waves (DRW). If they mean identical phenomena, please consider adding a clarification where you indicate that they mean the same.

There is a long-standing debate in the literature about whether or not to call the phenomenon being referred to a Diabatic Rossby Vortex or a Diabatic Rossby Wave (see the Appendix of Boettcher and Wernli (2013) for more on this debate). In the literature it is understood that both of these terms refer to the same phenomenon. We use the term Diabatic Rossby Vortex within the manuscript as we feel this term better suits what we see, i.e. a single vortex moving up the eastern seaboard of the USA (rather than a wave). We do not wish to have (or get into) a debate on which term should be used. As suggested we have added clarification (by means of a footnote, line 88) for readers to indicate they are identical phenomena so as not to confuse readers that are either not used to the term or the phenomenon.

The exact wording of the footnote is as follows
‘’In essence this is the same phenomenon as a diabatic Rossby wave (see Appendix of Boettcher and Wernli, 2013).”

6. L124: Why do you use ECMWF operational analyses for comparison, and not ERA5 reanalysis? Isn’t ERA5 considered more reliable because more observations have been

assimilated into it? However, if you have strong rationale to use ECMWF analyses, please include it in the paper.

We have used ECMWF operational analysis to keep consistency with the initialization of our model simulations and to determine whether any initial shock had occurred. A brief sentence along these lines has now been added to the start of section 3 (line 116).

The end of the sentence now reads "as a consistent baseline with the analysis state."

7. L243: It was a bit unclear for me why you split the diabatic heating into components here, because in the results (Sect. 4.4.1) you only talk about the omega/baroclinic conversion due to diabatics and not these single components.

This is a hangover from an older version that we had missed, so thank you for indicating this to us. This has now been removed as part of the requested reduction in the length of the manuscript .

In the older version we had originally decomposed the diabatic heating term into its components to consider which processes / parameterizations were providing different heating rates in the models.

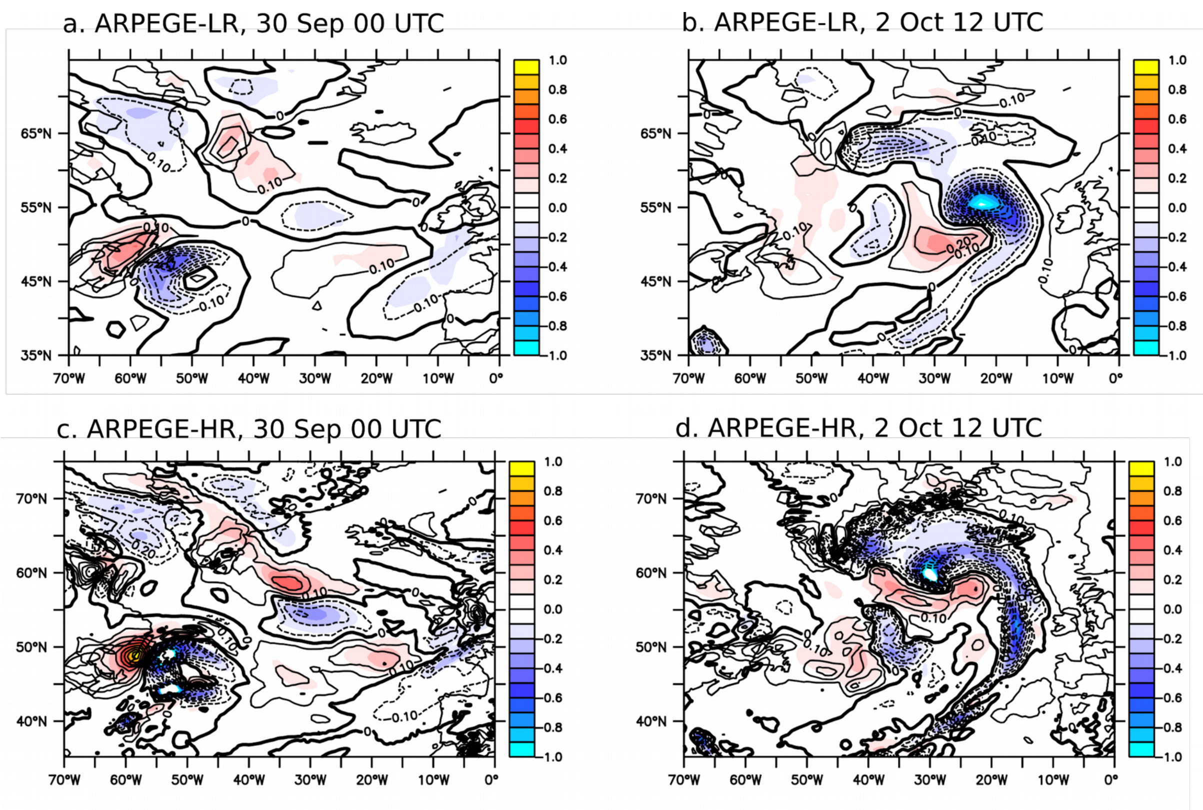
8. L245: Isn't latent heating including also the freezing and melting of ice droplets in the clouds, but you mention only condensation and evaporation. Why?

We agree that latent heating would also include the freezing of cloud and rain water droplets and melting of ice crystals/snow. The condensation and evaporation were mentioned as examples (and it was not meant to be an exhaustive list) as these are tendencies explicitly produced by the model (the latent heating was included just to give a reference meaning to the phrase large-scale condensational heating and evaporation - it was not meant as an exhaustive summary of what latent heating was). This has now been removed in relation to the comment above and the need to reduce the length of the manuscript.

9. L248: You mention that on average most of the modelled vertical motion is recovered using QG method. However, in Figure 7, there are some quite large discrepancies between Model and Inverted baroclinic conversion. How confident are you with the contributions of dynamic and diabatic processes if their sum (as indicated by Eq. 5) does not match well with the modelled baroclinic conversion? Did you verify how well the diagnosed omega (from QG equation) matches with the omega from the climate models? If yes, please consider adding a couple of sentences about the verification.

Given that we are using a full atmospheric model we do not expect the inverted QG omega to be equal to the modelled QG omega. However, we expect the QG assumption to be enough to diagnose the main sources of the cyclone development at synoptic scales.

The fact that the timing of the evolution of the inverted omega matches that of the modelled omega in Figure 7 is one reason mentioned in the original submission to provide confidence in our inversion. The other main reason relies on the strong similarities between the inverted and modelled omega (compare the contours and shadings in each panel of Fig. R1 of the *present document). The spatial resemblance is striking for both high and low resolution runs. For the low resolution runs and In some regions, the modelled omega is slightly underestimated by the inverted omega. For instance, the peak values of the inverted omega is roughly two thirds of those of the modelled omega for low-resolution run (Figs.R1a,b). For the high resolution runs, the modelled and inverted omega get roughly the same amplitude (see also the new figure 9 of the revised paper) but it varies from case to case. More information is provided in the revised version of the paper (lines 280-281, 315-316). An additional point which makes us confident in our decomposition is when we look at the effects of the two components on the total (see Fig. R2 and our reply to reviewer 2's comment). The dynamical and diabatic vertical velocities are not necessarily co-located (Fig. R2b,c) and this is only the sum of the two components that matches the model omega. It means that both components are crucial to recover the full signal.*



<u>Figure R1:</u> Inverted (QG; shaded; int: 0.1 Pa/s) and model (contoured; int: 0.1 Pa/s) vertical velocities during cyclogenesis on 30 Sep 00 UTC (a,c) and mature stages on 2 Oct 12 UTC (b,d) of the Stalactite Cyclone. a,b) are for ARPEGE-LR and c,d) for ARPEGE-HR. Similar results are obtained for LMDZ at both resolutions (not shown).

10. L252: Why do you express the omega equation twice? I think this is unnecessary and removing Eq. 2 would shorten your manuscript.

The purpose of Eq. 2 was to make the splitting of the equation into dynamic and diabatic components explicit. However, such a decomposition can be said in words and because the manuscript needed to be shortened, Eqs. (2) and (3) have been suppressed.

11. L416: How can you say that the larger ascents mainly arise from diabatic processes? For me it's very difficult to conclude that based on only the figure. Do you have some quantitative analysis behind your statement? In any case, it would be interesting to see some quantitative values related to how much the increased resolution increases the omega due to diabatic heating and the omega due to dynamics. For example, something similar as in Table 2 in Sinclair et al. (2020) (https://wcd.copernicus.org/articles/1/1/2020/)

Figure 9 shows bivariate histograms of vertical velocities as a function of vertical levels. A new version is available in the revised paper highlighting more the probabilities of extreme values. In addition, the figure slightly differs from the submitted version because a spatial filter applied to the Q-vector remained in our code that appears to be useless and tended to underestimate the dynamical omega. Considering this correction, Fig.9 led to the following conclusions. For ARPEGE, the peak values of the dynamical and diabatic omega are roughly the same while for LMDZ the diabatic omega reaches higher maxima than the dynamical omega. It shows that LMDZ has a stronger diabatic rate than ARPEGE for both resolutions (see lines 311-321 describing the new figure 9). Also, in the revised paper, all figures that use inverted omega are updated to remove the spatial filter from the inversion. Such a correction leads to slightly stronger dynamical vertical velocity but this does not change the main results.

12. Fig 1a, 2a and 7: please could you add horizontal (and vertical) grid lines so that the days are easier distinguishable from each other.

We have added tick marks on the axes of the different figures to aid in interpretation and so that the days and hours are more easily distinguishable. We have also made these additions to the corresponding supplementary material figures as well.

13. Fig 1a: in addition to the grid lines perhaps you could add shadings and small labels to the graph so that initiation and deepening phases are easier visible? And to my eye, the deepest phase of the cyclone seems to be about 6 hours earlier than what is indicated with the dotted line.

We have introduced shadings to indicate the regions of the different phases, we have also changed the deepest phase of the cyclone to a region covering both times, and added the corresponding labels to the figure.

14. Fig 1d: The labels in the thickest black contours are not seen properly. Furthermore, there seems to be some gaps in the contours in the middle of the domain.

The figure has been re-plotted to remove the gaps in the middle of the domain, and also remove the labelling of the bold contours (we feel that we can remove these labels as the other labels (and caption) are sufficient to give an idea of the spacing.

15. Fig 1e: The colorbar and its labels are too small.

The colour bar has now been increased in size.

16. Fig 9: Can you add the titles (Total, omega_diab and omega_dyn) to the plots. It would

be easier to interpret the plot when you don’t need to check from the caption what the different rows express.

Titles have now been added to this figure.

Finally, it would help the reviewers if the figures were included within the main text and not at the end of the manuscript.

In the original latex file the figures were in the correct place in the document but due to the size of figure 1 all figures were produced at the end of the manuscript. We have spent time correcting this and reducing the size of figure 1 to ensure it is now in the correct place and will allow all of the other figures to be in the correct place in the manuscript to alleviate this concern.

Response to Anonymous Referee #2

My summary: This manuscript examines the ability of GCMs to capture extratropical cyclone development and intensification using a case-study. The manuscript considers two separate models and runs each at two resolutions. A diagnostic based on the quasi-geostrophic omega equation is utilized to separate dynamic and diabatic contributions to vertical motion. The model results are compared with observations.

Recommendation: I recommend accept conditional on minor revisions. I find the paper to be well written and relatively easy to read. I appreciate the new diagnostics and the comparison with observations. There are a few details about the technical set up of analysis that need to be explained better (as line-by-line comments for details). There is an issue with the GQ omega that also needs more explanation.

Line 102: Possible verb tense disagreement for the word "interacts". I think you are talking about multiple PV anomalies, so it should be interact.

Thank you, this has now been changed (line 97).

Line 104: For the footnote, there is a typo: the word is should follow the phi symbol.

Thank you, this has now been changed (line 98).

Line 106: Placing a comma between trough and cyclonic would make the sentence easier to read.

Thank you, this has now been changed (line 100).

Line 109-110: Wouldn't Gwendal say that the cyclone participated in the NAO regime transition rather than saying it "occurred during"? I say this in jest, I'm sure you will address this later on in the manuscript.

This has now been changed by "participated in" because the ridge building ahead of the cyclone initiates the blocking and this is supported by Maddison et al (2019)'s findings (line 103).

Line 153: You write: "For both models hindcasts are initiated at 0000 UTC on 27–29 September, and 1–2 October 2016 .."

I don't understand the logic of the choices of two separate initialization times? Or to put it differently, why is there no initiation at 00 UTC on Sept 30? Also, those initiated on Oct 2 will have already had the genesis in the IC, I think. So what are they used for?

Also, you state on line 162 that hindcasts initiated on Sep 27 and 28 at low resolution did not produce the cyclone. So I'm confused as to why these dates are included in the sentence on line 153. I acknowledge that I may be misunderstanding something you've said.

This section has been substantially edited in the revised manuscript to take into account the *comment of reviewer 1 feeling the paper was too long. Thus in the new manuscript we focus on the two runs considered throughout the manuscript (29 September and 1 October). The original intention of this section was to justify the simulations considered, however we can see why this would lead to a lack of clarity.*

The first initiation time (29 Sep) is used to examine the entire life cycle of the Stalactite Cyclone to answer the following question: Are the models able to reproduce the life cycle of the cyclone once all precursors are present in the analysis ? For simulations starting at earlier times (27, 28 Sep) the LR resolutions did not produce a cyclone (due to a lack of DRV and upper-level PV interaction) and the comparison between LR and HR simulations was useless. The second initiation time (2 Oct) is used for observational comparisons to ensure similar cyclone structure and position to reality and to focus on differences in cloud microphysics. See lines 136-139.

Line 212: Typo on the word ClouSat - missing the d.

Thank you for spotting this, this is in a section of the paper that has since been removed.

Line 231: How interesting that the time-average of the static stability works in this calculation.

In QG theory the static stability appearing in the omega equation should be dependent upon the vertical only. This is the reason why a time average has been applied in addition to a horizontal average. Note that a time dependent static stability would not change much the results because the horizontal average made over the Atlantic area tends to smooth the field already.

Line 247: What are the horizontal boundary in a global model? I suppose you choose some appropriate distance?

The horizontal boundaries are the boundaries used for the inversion (so as not to invert over the whole globe) and are restricted to a region where the QG assumption makes the most sense (i.e. the mid-latitudes). The inversion box is now defined in the manuscript as 70W to 0E and 35 N to 75N (this is the area depicted in Fig. 1b, line 176). This encompasses the entire lifecycle of the cyclone and ensures that the centre of the cyclone is far enough away from the boundary throughout its lifecycle, when QG most likely applies) to show no/limited boundary effects.

Line 255: You write: “Inversion of the two previous equations allows to separate “ Sounds a bit awkward. Maybe there is a word missing between allows and to?

Yes, you are quite right there was a missing word. As part of the reduction of the manuscript based on comments from the other reviewer we have removed this section.

Line 257: What do you mean by “Vertical velocity intervenes . . ..” I don’t follow what you mean by intervenes in this context.

This has been re-worded for clarity, we meant occurs in (line 180).

Line 286: This 24-hour delay in deepening seems to run contrary to your statement on Line 217, where you wrote:

"In reality, for the hindcast that is compared against the observations in this study (initiated at 00 UTC 1 October 2016) neither timing nor positioning adjustments are required. "

I assume that this delay in the intensification that you discuss on Line 286 is in a hindcast initialized earlier than the Oct 1 date, but it would be good to include more clarity about this, either here, or in/near the sentence on Line 217.

Yes, you are correct in assuming the two statements refer to runs initialized at two different times. When we compare the observations we are using runs initiated on 1 October. We did mention this in the Methodology, however we have now made this clearer, as we appreciate it could have been easily missed. We highlight again at the start of Section 4.4 that we have changed runs to only look at the run initiated on 1 October (line 310).

Line 319-330. One thing that the table makes clear, that I don't think is mentioned: the HR models generate a stronger DRV than ECMWF. This should be stated in the text I think.

You are quite correct in saying that the DRVs in the HR models are stronger than ECMWF. We did not mention this in the text as we were focusing on whether the LR models could create a DRV - however, we do now mention this briefly in the text as suggested (line 227).

Line 325: Figure 4 is a something of a proof of concept for your dynamic/diabatic separation. Though it is might be worthwhile to mention that the dynamics and diabatic component are tightly coupled.

We now mention that the two components are tightly coupled (line 225).

Line 332: Section 4.2.2. How interesting that the LR model is more different from the HR model for the upper-level disturbance. We tend to think of diabatic forcing related to moisture parameterizations being the issue, but here that is not the case. Does this suggest that region in which the dynamics and diabatic components must interact are more sensitive to the role of resolution? Or is this an initial value problem? Or is this unique to this case-study. Alas, that is not something you can easily answer I imagine. Interesting to think – no need to provide a detailed response to this comment however.

Yes, we felt it was an interesting result too. We would say it certainly merits more investigation but would agree that there are sensitivities to the role of resolution here and potentially an initial value problem too given later runs see an improvement in the simulations. We are unsure if this is unique to this case study, but we are aware that similar (albeit not exactly the same) results occur with the subsequent cyclone that followed (which was targeted for future work). In this other case the interactions and balance between the different components of baroclinic conversion were slightly different so it is a little harder to answer.

Line 359: Section 4.2.3 Regarding the statement about these biases being the cause for the changes in track path, are you sure you can attribute it to this mechanism? If so, how?

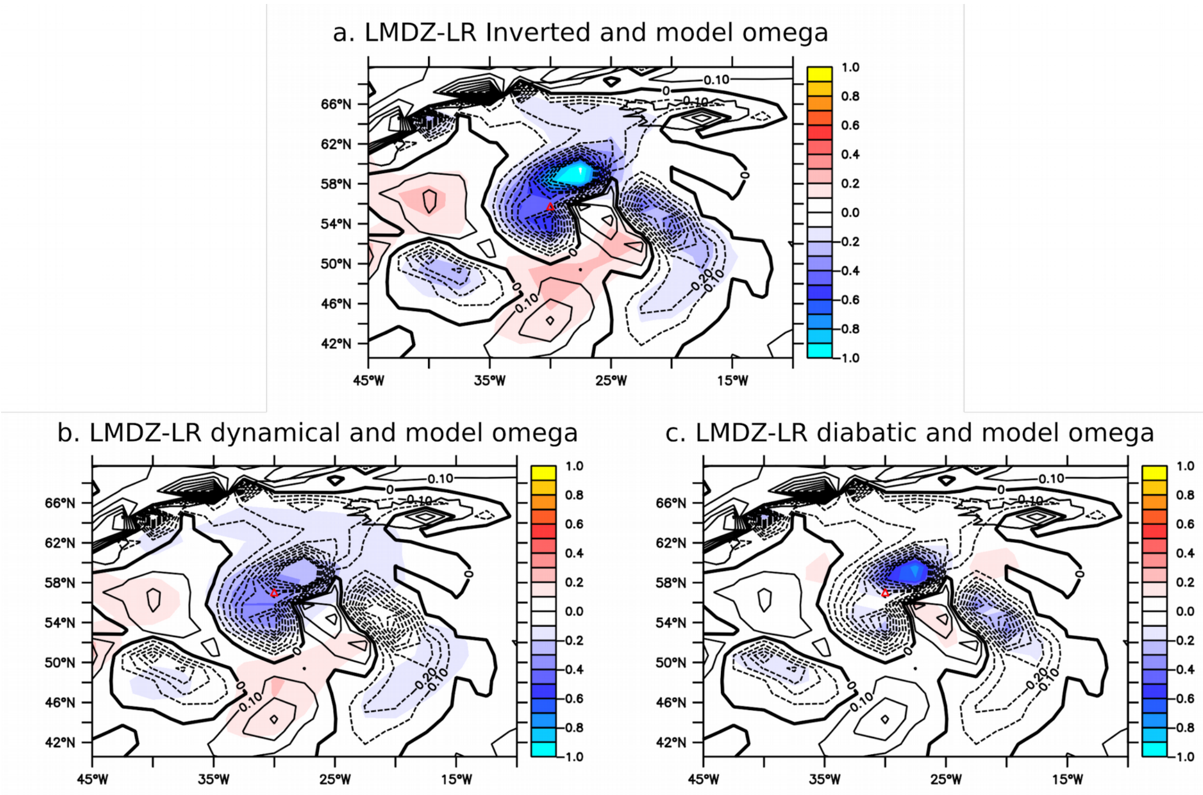
We have weakened the language used within this section, saying it is "likely". However, we do add a little more explanation to this comment as well (lines 264-265). We first note that it is straight after cyclogenesis that the track starts to deviate in the forecast when the cyclone tracks are discussed in Section 4.1. This can be seen by comparing the tracks of the HR and LR runs. Second, to understand the cyclone track we need to analyze the timing of the interaction with upper-level PV disturbances following our previous studies (Gilet et al. 2009; Oruba et al. 2012, 2013; Coronel et al. 2015). The poleward motion of midlatitude cyclones occurs during baroclinic interaction of surface cyclones with upstream upper-level troughs. The presence of the upper-level trough tends to create poleward winds near the surface cyclone center and this advects the surface cyclone perpendicularly to the mean jet axis. In the present study, the LR simulations miss the initial interaction with the small-scale upper-level disturbance and this prevents poleward advection by the latter anomaly. Baroclinic interaction with upper levels happens later in the LR runs and this likely explains the delayed poleward motion of the surface cyclone in those runs.

Line 375: You write: "The averaged quasi-geostrophic baroclinic conversion is roughly reduced by two thirds in magnitude compared to that directly calculated from the model ω but is consistent . . ."

That is quite a reduction. This is the only thing that I've read (so far) in this manuscript that makes me pause and wonder, should we be so confident in their diagnostics? There is not much discussion. Are you hoping to ignore it so that others do as well, or is there good reason to be unbothered by this difference? What is the difference between the Q-G baroclinic conversion and that using modeled omega for the cyclogenesis phase?

First our sentence might be misleading and is now written as follows : « The averaged QG baroclinic conversion roughly recovers 60--70 % of the amplitude of that directly calculated from the model omega » (line 280).
Second, following a similar comment from reviewer 1, more information is provided on the differences and similarities between the modelled omega and inverted omega. Despite the difference in amplitude of the two averaged values shown in Fig.7, the two fields are close to each other. First, we do see that the timing of the evolution of the inverted omega matches that of the modelled omega in Figure 7. Second, there are strong similarities between the inverted and modelled omega spatial patterns (compare the contours and shadings in each panel of Fig. R1 of the present document and Fig. R2.(a)). This is shown for the two models and is valid for both high and low resolution runs. Additionally, Fig. R2 shows that the dynamical and diabatic vertical velocities are not necessarily co-located (panels b and c) and this is only the sum of the two components that matches the model omega. These maps thus provide confidence in our decomposition. There is some underestimation of the amplitude of the modelled omega by the inverted omega in some regions of the low-resolution runs but this is not seen in high resolution runs (see Fig.9 of the revised paper). This information is provided in the revised version of the paper (lines 311-321).
Note that we have made a correction in our code since the initial submission. A spatial filter was initially applied to the Q-vector (at that time we did not have the Inverse of the Laplacian and we wanted to anticipate the effect of such an operator). This spatial filter remained in

our code but was useless and tended to underestimate the dynamical omega. In the revised paper, all figures that used inverted omega were updated to remove the spatial filter from the inversion. Such a correction leads to slightly stronger dynamical vertical velocity. In fact, the inverted omega is now even closer to the model one than before. The peak values are rather similar for high-resolution runs (Fig.9) even though when spatial averages are made (Fig.7), the QG baroclinic conversion roughly recovers 60-70 % of the amplitude of that directly calculated from the model omega. This difference between averaged and peak values might come from the positive regions, but this needs to be more deeply investigated.



<u>Figure R2:</u> Inverted (QG; shaded) and model (contoured) vertical velocities in LMDZ-LR run starting on 1 Oct 2016 for (a) the total inverted, i.e the sum of the dynamical and diabatic components, (b) the dynamical component and (c) the diabatic component.

Line 415: Figure 9 What do these plots look like for modeled omega?

Figure 9 includes the modeled omega histograms in the revised version. The shapes of the inverted and model omega histograms are rather similar. More discrepancies appear for LMDZ-LR where we see that the peak values of the model omega are not reproduced by the inverted omega. These values correspond to grid points near 28°W and 58°N in Fig. R2(a). Despite this underestimation of the amplitude, the location of the peak values is well reproduced by the inverted omega.

Line 487: Figure 12 is a bit disheartening isn't it? As you say, ARPEGE-LR looks the most like the observations. But the HR models more closely match the dynamics of reanalysis. So what does this mean? Does the CFAD have no relationship with the dynamics of the cyclone? Or should we raise more questions about the dynamics in reanalysis? It is a tough figure to interpret, but I am glad that you include it in the manuscript.

We would agree that Fig. 12 is a bit disheartening, but we are pleased that you feel it is important to include in the manuscript. We felt that this last section was particularly important because not only did it show strong differences between the models, but it also highlighted a problem that can be overlooked in the analysis of models. There will be some elements of dynamics involved - but these differences shown (given they are radar fields plotted) are more likely to be associated with the microphysics, which certainly combined with the previous two figures and table 2 suggest a worrying factor about condensate identification and the knock on effect on the dynamics and diabatic heating in the models (and reanalyses).

We appreciate it is a difficult figure to interpret but we believe with the inclusion of the previous discussion on the ice water content aids the interpretation (as it essentially supports those results). We also suggest that the shape is the most important factor to consider with these plots as (as stated in the manuscript) the mask is not applied to the data given that the CFAD is a direct output from the model output field rather than crafted based on the reflectivities (as in the observations). Whilst it does show some relationship to the dynamics we feel the stronger difference (and more reason for concern) is the link with the microphysics. However, based on the links between dynamics and microphysics it does raise questions about these processes and their representation in models and re-analysis, and we would strongly encourage this type of analysis to be done for both to help improve the models and re-analysis.

Representation by Two Climate Models of the Dynamical and Diabatic Processes Involved in the Development of an Explosively-Deepening Cyclone During NAWDEX

David L. A. Flack[1*], Gwendal Rivière[1], Ionela Musat[1], Romain Roehrig[2], Sandrine Bony[1], Julien Delanoë[3], Quitterie Cazenave[3], and Jacques Pelon[3]

[1]Laboratoire de Météorologie Dynamique/IPSL, Ecole Normale Supérieure, PSL Research University, Sorbonne University, École Polytechnique, IP Paris, CNRS, Paris, France
[2]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[3]LATMOS-IPSL, CNRS/INSU, University of Versailles, Guyancourt, France

Correspondence: Gwendal Rivière (griviere@lmd.ens.fr)

Abstract. The dynamical and microphysical properties of a well-observed cyclone from the North Atlantic Waveguide and Downstream impact Experiment (NAWDEX), called the Stalactite cyclone and corresponding to Intensive Observation Period 6, is examined using two atmospheric components (ARPEGE-Climat 6.3 and LMDZ6A) of the global climate models CNRM-CM6-1 and IPSL-CM6A respectively. The hindcasts are performed in "weather forecast mode", run at c. 150-200 km (LR) and c. 50 km (HR) grid spacings and initialized during the initiation stage of the cyclone. Cyclogenesis results from the merging of two relative vorticity maxima at low levels: one associated with a Diabatic Rossby Vortex (DRV) and the other initiated by baroclinic interaction with a pre-existing upper-level PV cut-off. All hindcasts produce (to some extent) a DRV. However, the second vorticity maximum is almost absent in LR hindcasts because of an underestimated upper-level PV cut-off. The evolution of the cyclone is examined via the quasi-geostrophic ω equation, which separates the diabatic heating component from the dynamical one. In contrast with some previous studies, there is no change in the relative importance of diabatic heating with increased resolution. The analysis shows that LMDZ6A produces stronger diabatic heating compared to ARPEGE-Climat 6.3. Hindcasts initialized during the mature stage of the cyclone are compared with airborne remote-sensing measurements. There is an underestimation of the ice water content in the model compared to the one retrieved from radar-lidar measurements. Consistent with the increased heating rate in LMDZ6A compared to ARPEGE-Climat 6.3, the sum of liquid and ice water contents is higher in LMDZ6A than ARPEGE-Climat 6.3 and, in that sense, LMDZ6A is closer to the observations. However, LMDZ6A strongly overestimates the fraction of super-cooled liquid compared to the observations, by a factor of approximately 50.

Copyright statement.

*Current Affiliation: Met Office, Exeter, UK

1 Introduction

Extra-tropical cyclones are one of the leading hazards in the midlatitudes but their projected behaviour under climate change remains uncertain (e.g. Harvey et al., 2012). This uncertainty lies in the location of the extra-tropical cyclones and the intensity and position of the storm track (e.g. McDonald, 2011; Zappa et al., 2013b) rather than in the total number of extra-tropical cyclones (e.g. Finnis et al., 2007; Bengtsson et al., 2009; Catto et al., 2011; Zappa et al., 2013b).

Uncertainties in climate simulations can arise from three different factors: model physics, internal variability, and forcings (e.g. Hawkins and Sutton, 2009). Therefore, to determine confidence in future projections the historical model climate is compared to observations or re-analyses (e.g. Seiler and Zwiers, 2016). Typically, the representation of cyclones in climate models is considered through statistics, e.g. number, frequency (e.g. Zappa et al., 2013a; Seiler and Zwiers, 2016). These studies generally indicate systematic limitations of coarse-resolution models rarely producing explosively-deepening cyclones, producing too many weak cyclones, and a storm track that is both too zonal and too far south.

Recently, studies have started to investigate the 3D structure in cyclones (e.g. Catto et al., 2010) and the roles of diabatic heating in climate models (e.g. Willison et al., 2013; Trzeciak et al., 2016; Sinclair et al., 2020). Willison et al. (2013) and Trzeciak et al. (2016) showed that increased resolution, compared to that of the Coupled Model Intercomparison Project (CMIP) models at the time (CMIP5), was required to improve the representation of the diabatic heating, and hence representation of the cyclone. This improved representation of diabatic heating could be important as Sinclair et al. (2020) indicated that diabatic processes could become more important in a warming climate. However, fundamental processes linked to extra-tropical cyclone formation and development need further investigation in global circulation models (GCMs). Fundamental processes linked to factors such as cyclogenesis and cyclone development are hard to examine in full length, free-running, climate simulations and could explain a lack of consideration of this to date. Therefore, to examine the representation of the physical processes in cyclone formation and development different techniques are required. These techniques include running climate model configurations in "weather forecast mode" (e.g. Phillips et al., 2004), or running short ensemble forecasts (e.g. Wan et al., 2014).

The idea of running climate model configurations in "weather forecast mode" culminated in the formation of the Transpose - Atmospheric Model Intercomparison Project (T-AMIP) experiments (Williams et al., 2013). The T-AMIP experiments are primarily used to assess whether any long-term model biases occur within the first few days of the simulations. It was hoped that, if these biases formed early in the climate simulations, model improvements to reduce those biases could be tested with less computational expense (e.g. Williams et al., 2013; Ma et al., 2013). It was further thought this application could help disentangle the origin of the model biases in a more causal way (e.g. Brient et al., 2019). The T-AMIP experiments have considered factors such as cloud cover behind fronts in extra-tropical cyclones (e.g. Williams et al., 2013); radiative feedbacks (e.g. Williams et al., 2013; Bony et al., 2013; Fermepin and Bony, 2014); 2 m temperature (e.g. Fermepin and Bony, 2014; Ma et al., 2014); precipitation (e.g. Ma et al., 2013; Fermepin and Bony, 2014; Pearson et al., 2015; Li et al., 2018); stratocumulus (e.g. Brient et al., 2019); and for use alongside random-parameter ensembles to determine structural *vs.* parameter sensitivities (e.g. Sexton et al., 2019; Karmalkar et al., 2019).

The T-AMIP type experiments can also be used as a powerful tool for considering the representation of dynamical processes in climate models. For example, Trzeciak et al. (2016) showed that climate models of resolution T127 (c. 1.1–1.5° at midlatitudes) can represent deep extra-tropical cyclones and their tracks well. This good representation was attributed to an increased importance of the diabatic heating compared to lower resolution simulations. Like Trzeciak et al. (2016), we consider the dynamical representation of extra-tropical cyclones, and the impact of resolution, in climate models. However, we focus on a single, well-observed, cyclone during the Intensive Observation Period (IOP) 6 of the North Atlantic Waveguide and Downstream impact EXperiment (NAWDEX) field campaign (Schäfler et al., 2018), which is called the "Stalactite" cyclone. This cyclone is initiated from the interaction of two features that occur on sub-grid scales of current climate models. The main deepening phase is characterized by interaction of the surface cyclone with successive synoptic-scale upper-level troughs. Here, we answer the following questions on the representation of the cyclone in climate models to provide further insights into whether climate models are producing cyclones for the correct reasons:

1. How well do climate models represent the two stages of the Stalactite Cyclone?

2. What are the relative roles of diabatic and dynamic processes in the development of the Stalactite Cyclone?

3. Are there any differences in the diabatic processes related to microphysical properties between the different models?

The NAWDEX field campaign occurred in September–October 2016 with the aim of making targeted observations of processes that numerical atmospheric models poorly represent (Schäfler et al., 2018). These observations would then be used to help determine how well the models represent these processes (e.g. Maddison et al., 2019; Oertel et al., 2019). The observations taken during the field campaign allow for an extra question to be asked in this study:

4. Can microphysical observations made during the field campaign give any useful information about the climate model's performance?

To our knowledge this study is the first time that a climate model is compared with flight data taken during a field campaign, without nudging of analyses into the simulation, and it is only feasible because of the T-AMIP protocol.

The questions asked here are of particular interest for the Stalactite cyclone as it influences the development of a blocking anticyclone over Scandinavia and marks the transition between a North Atlantic Oscillation (NAO) positive regime and a Scandinavian blocking regime over the North Atlantic European sector. Therefore, it is a particularly useful case to determine the capabilities of our current climate models.

The remainder of this paper has the following layout. The key features of the Stalactite Cyclone are discussed in Section 2. The GCMs, experimental setup, observations, and diagnostics are described in Section 3. The Stalactite cyclone's representation in the two GCMs are discussed within Section 4. A summary is made in Section 5.

2 The Lifecycle of the Stalactite Cyclone (NAWDEX IOP 6)

The Stalactite Cyclone corresponds to IOP 6 of the NAWDEX field campaign (Schäfler et al., 2018). It was an explosively-deepening cyclone that initially formed at 1800 UTC on 29 September 2016 (Fig. 1a) off the coast of Newfoundland (c. 56° W, 45° N; Fig. 1b). Cyclogenesis occurred as a result of the merging of two vorticity maxima at low levels (Fig.1c). The northern maximum over Newfoundland is formed via baroclinic interaction with an upper-level PV cut-off that extended down to the surface like a stalactite (hence the name of the cyclone). The southern maximum corresponds to a Diabatic Rossby Vortex (DRV). A DRV corresponds to an isolated positive PV anomaly rapidly travelling eastward in a moist and baroclinic region[2]. To determine if this diabatic precursor is a DRV we use the criteria set by Boettcher and Wernli (2013). All of the criteria are met in ECMWF analysis which confirms the identification of a DRV. It was formed on 27–28 September off the coast of Florida and South Carolina (not shown). The DRV was probably produced from a mesoscale convective system, as confirmed by satellite images showing cold brightness temperature (e.g. Fig. 1e).

The two low-level precursors merge into a single cyclonic vorticity maximum in a vortex roll-up by the subsequent analysis (not shown). The initial cyclogenesis phase led to a short deepening stage over 18 h as the cyclone travelled east past Newfoundland. The cyclone underwent a second, more substantial, deepening as a result of an interaction with a large-scale region of high PV at upper levels as the cyclone began to cross the North Atlantic. This region is marked by multiple regions of high PV ("B" and "C" in Fig. 1d) that are successively injected in the upper-level disturbance ("A") and interact with the Stalactite cyclone. The deepening occurred at a rate of 24.1 hPa in 24 h and so meets the criterion set in Sanders and Gyakum (1980)[3] to be classified as an explosively-developing cyclone. The explosive deepening occurred between 1800 UTC 30 September – 1800 UTC 1 October (Fig. 1a). During the interaction with the second large-scale trough, cyclonic wave breaking occurred and the cyclone re-curved towards Greenland (Fig. 1b). On reaching the coast of Greenland cyclolysis (i.e. cyclone decay) occurred; the cyclone had filled in by 0000 UTC 4 October. The cyclone posed an interesting challenge for operational numerical weather prediction models as the cyclone participated in a regime transition from an NAO positive regime to a Scandinavian blocking regime which dominated the North Atlantic European sector for the rest of the field campaign (e.g. Schäfler et al., 2018; Maddison et al., 2019). Correspondingly, there was a reduction in the forecast skill (Schäfler et al., 2018). To determine whether the climate models are correctly simulating the Stalactite Cyclone three criteria are developed from its lifecycle.

1. Initial cyclogenesis occurs as a result of the merger of a DRV and another near-surface cyclonic vortex associated with baroclinic interaction with an upper-level PV cut-off.

2. A main deepening phase associated with large-scale troughs is present.

3. A minimum pressure deepening rate of 24 hPa in 24 h during the secondary deepening phase.

If all of these criteria are met then the climate models are able to correctly represent the Stalactite cyclone. The climate models and experimental setup used are discussed in the following section.

[2] In essence this is the same phenomenon as a Diabatic Rossby Wave (see Appendix of Boettcher and Wernli, 2013).

[3] A deepening rate of 1 $\mathrm{hPa\,h^{-1}}$ for 24 h multiplied by sin(ϕ)/sin(60) to adjust to the appropriate latitude to make it equivalent to at least 1 $\mathrm{bergeron}$, where ϕ is the latitude

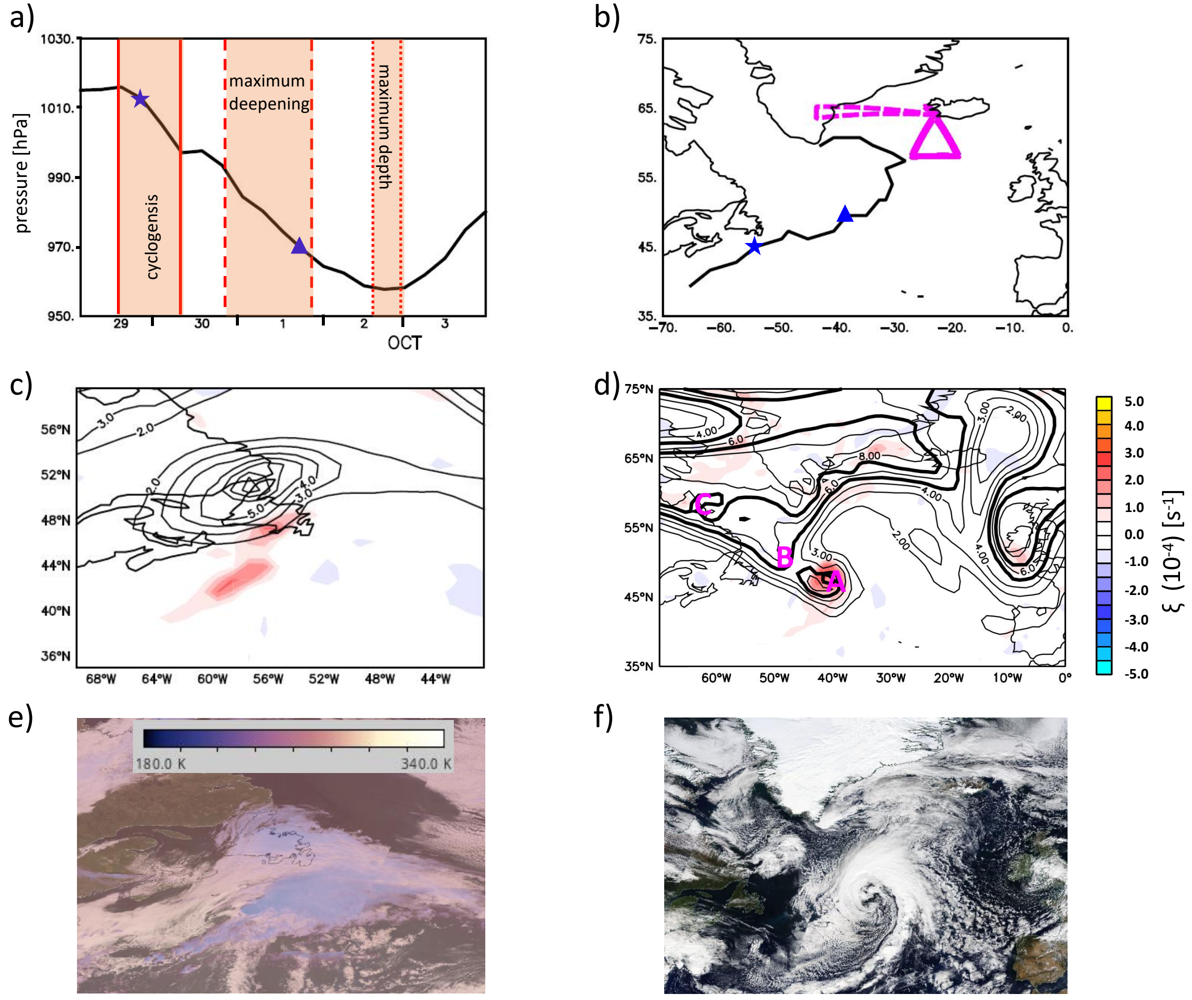


Figure 1. An overview of the Stalactite Cyclone. a) the ECMWF analysis minimum pressure evolution (black). b) the track of the Stalactite Cyclone. The magenta dashed line is the flight path of the SAFIRE Falcon-20 flight 6 and the solid is for flight 7. The star and triangle on a and b represent the timing of panels c,e and d,f respectively. c) the ECMWF analysis of the 250 hPa PV $>$ 2 PVU (contoured) and 850 hPa relative vorticity (shaded) at cyclogenesis (18 UTC 29 September) and d) as for c) but just before maximum deepening at 12 UTC 1 October. The bold lines are to indicate the PV signature of different PV regions interacting with the cyclone. “A” is upper-level signature of the Stalactite Cyclone at that time, “B” is the second PV region to interact with the cyclone, and “C” is the third. The colour scale applies to both c and d. e) A visible satellite image from MODIS on 29 September indicating the Stalactite cyclone at cyclogenesis. The brightness temperature has been overlaid and saturated at 50%. f) the visible satellite imagery from MODIS on 1 October 2016. The satellite images are produced with the courtesy of NOAA Worldview (https://worldview.earthdata.nasa.gov/).

3 Models, Observations, and Diagnostics

In this section we discuss the model setup and experimental protocol of the T-AMIP experiments (Section 3.1), the observations (Section 3.2), and diagnostics considered (Section 3.3). We also compare our simulations against the European Centre for Medium Range Weather Forecasting (ECMWF) analysis as a consistent baseline with the initiation state.

3.1 Models and experimental setup

We use two atmospheric GCMs: ARPEGE-Climat 6.3 (hereafter ARPEGE) and LMDZ6A (hereafter LMDZ) of the CNRM-Cerfacs (Centre National de Recherches Météorologiques - Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique; Voldoire et al., 2019) and IPSL (Institut Pierre Simon Laplace; Boucher et al., 2020) climate models, CNRM-CM6-1 and IPSL-CM6A, respectively. Both climate models recently contributed to CMIP6 (Eyring et al., 2016), and here we make use of the same model versions and configurations. Table 1 shows the details of the model configurations used. These GCMs are run in "weather forecast mode" to represent T-AMIP style experiments. Hereafter, ARPEGE-LR (-HR) and LMDZ-LR (-HR) refer to low-resolution (high-resolution) runs of the two models.

Table 1. Model descriptions with key parametrizations and hindcast resolutions.

	ARPEGE-Climat 6.3 (Roehrig et al., 2020)	LMDZ6A (Hourdin et al., 2020)
Low resolution (LR)	T127 (c. 150 km globally; output to 1.4°)	2.5°×1.2°
High resolution (HR)	T359 (c. 50 km; output to 1.4°)	Zoom function
Core	ARPEGE/IFS	LMDZ6A
Vertical levels	91	79
Convection	Piriou et al. (2007) and Guérémy (2011)	Rochetin et al. (2014)
Longwave Radiation	Mlawer et al. (1997)	Mlawer et al. (1997)
Shortwave Radiation	Fouquart and Bonnel (1980); Morcrette et al. (2008)	extension of Fouquart and Bonnel (1980) to six bands
Clouds and Microphysics	Lopez (2002)	Madeleine et al. (2020) and Hourdin et al. (2019) for low clouds

The LMDZ-HR configuration utilises its zoom function, in which the resolution over part of the domain is increased compared to the rest in a variable resolution configuration. Here the zoomed domain is centred at (40° E, 55° N) with a resolution equivalent to 0.33°. The resolution decreases away from the centre resulting in a resolution of approximately 0.5° over the North Atlantic and 1.1° elsewhere.

For ARPEGE, microphysics state variables and turbulent kinetic energy were initialised to zero and aerosols are prescribed from a present-day climatology. On the other hand, in LMDZ model state variables not defined in the analysis are set to zero, alongside the aerosols. All hindcasts are performed out to a leadtime of T+10 d. Furthermore, all hindcast output data is interpolated onto a pressure grid in the vertical, every 25 hPa, from 1000 hPa to 100 hPa. Output is also produced using the

CFMIP (Cloud Feedback Model Intercomparison Project) Observation Simulator Package (COSP; Bodas-Salcedo et al., 2011) for radar reflectivities from CloudSat to be compared with the observed aircraft-borne radar reflectivities from the NAWDEX field campaign.

The hindcasts are initiated at 0000 UTC 29 September, and 1 October 2016 from the ECMWF analysis (including sea surface and ice cover). The first initiation time is used to examine the entire lifecycle of the Stalactite Cyclone; the second is used for observational comparisons (Section 4.4) to ensure similar cyclone structure and position to reality. We restrict the number of hindcasts to take into account the impact of the overall synoptic situation at the time being largely unpredictable (e.g. Schäfler et al., 2018). Initial shock (e.g. Klocke and Rodwell, 2014) is checked for but is not significant. However, as a precautionary measure we do not analyse hindcasts prior to T+18 h.

3.2 Observations

During the NAWDEX field campaign the French SAFIRE Falcon aircraft operated from 1–15 October (Schäfler et al., 2018). The SAFIRE Falcon made two flights to observe the Stalactite cyclone on 2 October 2016: F6 (towards Greenland) and F7 (south of Iceland; Fig. 1b). The second flight (F7) was directly into the cyclone in the ascending branch of the associated warm conveyor belt. The first flight (F6) considered the warm conveyor belt outflow. In the main manuscript we focus on F7. The first leg of F7 (the most eastern one) was chosen because there was an overpass with CloudSat-CALIPSO track at 14:09 UTC which allows us to assess observation uncertainties by comparing airborne and satellite measurements. The payload onboard the SAFIRE Falcon included a 95-GHz Doppler cloud radar and a high-spectral resolution Doppler lidar capable of measuring at 355, 532 and 1064 nm (e.g. Delanoë et al., 2013). Measurements by these two instruments allow the retrieval of ice water content (IWC) thanks to the variational algorithm of Delanoë and Hogan (2008) updated by Cazenave et al. (2019). The combination of radar and lidar allows for the identification of the phase of the particles to be identified (e.g. super-cooled liquid, ice, liquid, etc.) using principles outlined in Delanoë and Hogan (2010). Furthermore, Doppler-derived windspeeds, and radar reflectivities are also used. Retrievals from radar products only (RASTA) and a combined radar and lidar product (RALI) are used to account for uncertainty in the measurements. Complementary information on the flight and measurements is available in Blanchard et al. (2020).

3.3 Vertical motion and baroclinic conversion budgets

Extra-tropical cyclone evolution can be considered through many methods. For example, the surface pressure tendency equation (e.g. Fink et al., 2012), through a potential vorticity framework (e.g. Davis et al., 1993), or the quasi-geostrophic (QG) vertical motion (ω) equation (e.g. Sinclair et al., 2020). Here, as in Sinclair et al. (2020), we consider the evolution through the QG ω-equation. We also consider the energetics of the cyclone through the baroclinic conversion (BC).

The QG ω-equation, that includes diabatic heating and the β term, can be written in terms of the so-called $\mathbf{Q}$ vector following Hoskins et al. (1978) and Hoskins and Pedder (1980). We use the formulation of Holton (2004) that includes the diabatic heating too:

$$\left(\sigma\nabla^2 + f_0^2\frac{\partial^2}{\partial p^2}\right)\omega_{QG} = -2\left(\nabla\cdot\mathbf{Q}\right) + f_0\beta\frac{\partial v_g}{\partial p} - \frac{R}{c_p p}\nabla^2 J, \tag{1}$$

for

$$\mathbf{Q} = -\frac{R}{p}\begin{pmatrix}\frac{\partial \mathbf{u_g}}{\partial x}\cdot\nabla T\\ \frac{\partial \mathbf{u_g}}{\partial y}\cdot\nabla T\end{pmatrix},$$

where σ is the static stability (obtained by temporally averaging the temperature across the lifetime of the Stalactite cyclone), f_0 is a reference coriolis parameter, β is the beta term in the coriolis forcing, p is the pressure, R is the specific gas constant, c_p is the specific heat, J is the rate of heating per unit mass, $\mathbf{u_g}$ is the geostrophic wind vector, T is the temperature, x and y are the positions in the meridional and zonal direction, respectively, and ω_{QG} is the vertical velocity obtained from inverting the QG ω-equation.

Equation (1) allows us to distinguish between the dynamical and diabatic contributions to the vertical motion in the cyclone. Physically the $\mathbf{Q}$ vector and the β terms represent the dynamical components of the flow and the Laplacian of the rate of heating per unit mass represents the diabatic heating.

To solve Eq. (1) the 3D Laplacian is inverted over the region 70° W – 0° E and 35° N – 75° N using Liebmann successive over-relaxation with boundary conditions such that ω is zero at 1000 hPa, 100 hPa, and all horizontal boundaries. The vertical motion is computed every 25 hPa in the vertical. Comparisons of modelled ω and ω_{QG} occur in Sec. 4.3. We also invert the dynamic and diabatic components of the ω_{QG} (ω_{dyn}, ω_{diab}) to gain further insights into the development of the cyclone.

Vertical velocity occurs in different key terms of the classical equations for the development of extratropical cyclones. We adopt the energetic framework and compute the baroclinic conversion from eddy potential energy to eddy kinetic energy within the extra-tropical cyclone (e.g. Orlanski and Katzfey, 1991; Rivière and Joly, 2006). The baroclinic conversion is proportional to the vertical heat flux and can be written as

$$BC = -h\omega'\theta',$$

where $h = (R/p)(p/p_s)^{R/C_p}$, p_s is the surface pressure and θ is the potential temperature. Primes denote the difference from the 5-day temporal average of that quantity centered over the lifecycle of the Stalactite cyclone. The results are insensitive to the definition of the temporal average provided it is made over an interval equal to or longer than the lifecycle of the cyclone to suppress the cyclone's signal. The baroclinic conversion term is mainly positive in areas following the cyclone trajectory (Rivière and Joly, 2006; Rivière et al., 2015).

We approximate BC by replacing the vertical velocity by its QG formulation Eq. (1), denoted as ω_{QG}, and keeping θ' unchanged. The approximated $-h\omega_{QG}\theta'$ is decomposed into its dynamic and diabatic components (resp. $-h\omega_{dyn}\theta'$ and $-h\omega_{diab}\theta'$) by inverting the corresponding components of vertical velocity in Eq. (1) separately.

4 Representation of the Stalactite Cyclone

Throughout this section the dynamical and diabatic representation of the Stalactite cyclone is discussed. The minimum pressure evolution and cyclone track are considered in Section 4.1. An in-depth consideration of the cyclogenesis and development occur in Sections 4.2 and 4.3, respectively. The two climate models are compared to the flight observations and discussed in relation to diabatic heating in Section 4.4.

4.1 Pressure evolution and track

The representation of the Stalactite Cyclone is first considered via an overview of the cyclone through its track and minimum sea level pressure evolution (Fig. 2). All hindcasts produce a rapidly deepening cyclone: slightly more than 24 hPa in 24 h in HR hindcasts and slightly less than 24 hPa in 24 h in LR hindcasts. However, this deepening is delayed by 24 h compared to the analysis in the LR simulations. Furthermore, the initial cyclogenesis is not as intense in LR simulations compared to the analysis. This weaker cyclogenesis results in an initially weaker cyclone compared to the analysis in both models (Fig. 2a). However, the explosive deepening in LMDZ-LR compensates for the lack of initial deepening. Conversely, ARPEGE-LR has the same secondary deepening strength as the analysis so produces a weaker cyclone. The HR hindcasts both have improved representation of the initial cyclogensis so show more realistic cyclone development in terms of pressure evolution.

The cyclone track also differs from the analysis. The difference occurs 18 h into the hindcasts. The two LR hindcasts produce a track that is too far south and has a later re-curvature so the cyclone track occurs further east compared to the analysis and HR hindcasts (Fig. 2b). The eastward shift in the track agrees with global weather forecasts prior to 29 September 2016 (e.g. Maddison et al., 2020). Given the rapid divergence of the forecast track from the analysis, differences in the cyclogenesis could be one aspect leading to the track occurring too far east as argued later in that section. The cyclogenesis being important for the cyclone track is also corroborated by the track representation having improved (i.e. no eastward shift) after the cyclone appears in the initial conditions (not shown).

The main differences to the representation of the Stalactite cyclone compared to the analysis, on initial inspection, appear to be within the cyclogenesis phase of the cyclone and the different deepening rate of LMDZ compared to ARPEGE and the analysis. These two aspects are examined further within the following subsections.

4.2 Cyclogenesis

The cyclogenesis of the Stalactite cyclone occurs on the mesoscale as the merging of two low-level vorticity precursors: a DRV coming from the subtropics and a vortex located further north baroclinically interacting with an upper-level PV cut-off (Fig. 1c). In the present section, we analyze the representation of the two precursors and their subsequent merging in the different simulations. The same vorticity fields as in Fig. 1c are shown in Fig. 3 for the different simulations. Figures 4 and 5 show the baroclinic conversion at T+18 h for both ARPEGE and LMDZ respectively and help identify the mechanisms behind the two precursors for the Stalactite Cyclone, there is a close relationship between the two components as the dynamics and diabatic processes are tightly coupled.

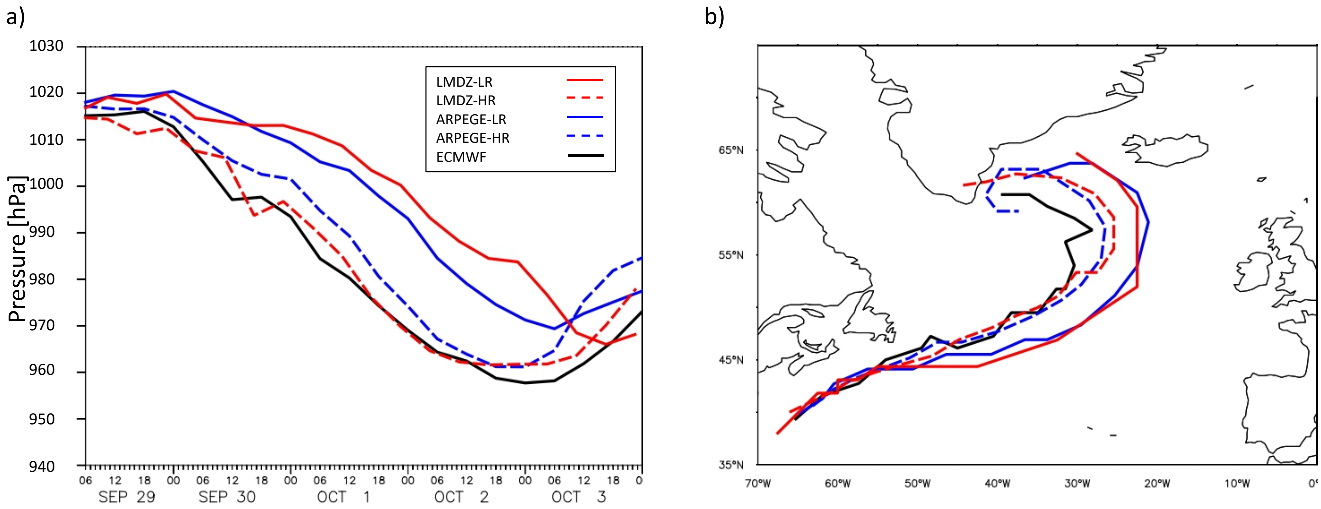


Figure 2. An overview of the Stalactite Cyclone: a) the minimum pressure evolution and b) the cyclone track. The ECMWF analyses are in black, LMDZ hindcasts in red, and ARPEGE hindcasts in blue. The LR hindcasts are the solid lines and the HR hindcasts are dashed. All hindcasts are initiated at 00 UTC 29 September 2016.

4.2.1 The Diabatic Rossby Vortex

Criteria of DRV introduced by Boettcher and Wernli (2013) have been analyzed in the different simulations. The two HR hindcasts fit all the criteria of a DRV, producing a stronger DRV than the ECWMF, which shows that 50 km grid spacing is enough to represent the DRV. The LR hindcasts meet all but two of the criteria of Boettcher and Wernli (2013): the PV intensity (for both) and propagation speed (LMDZ-LR; Table 2). However, it is encouraging to see that the LR hindcasts produce a qualitative representation of a DRV despite the coarse resolution of the models and the mesoscale nature of this self-sustaining phenomenon. The identification of the southern precursor as a DRV is confirmed by the baroclinic conversion of Figs. 4 and 5 which show that the diabatic component is almost equal to the total in the vicinity of the vortex and that the dynamical component is negligible. The DRV is more active in LMDZ-LR compared to ARPEGE-LR as the associated heating rate reaches higher values in LMDZ-LR compared to ARPEGE-LR (c.f. Figs. 4c and Figs. 5c). Vertical cross sections of the heating rates across the DRV indicate that its structure extends throughout the atmospheric column (Fig. S1 in the supplementary material) confirming the impression left by the satellite image (Fig. 1e).

4.2.2 Formation of the northern precursor via baroclinic interaction with the PV cut-off

More important differences appear between LR and HR runs in the representation of the northern precursor. In the LR hindcasts the vorticity of the northern precursor is much smaller than the vorticity of the DRV precursor (reduced by factors of 2.4 in ARPEGE-LR and 3.3 in LMDZ-LR) whereas it is only slightly smaller in HR runs (ratio of 1.6 in ARPEGE-HR and 1.3 in LMDZ-HR). Furthermore, the LR runs (Figs. 3a,c) have a more zonal PV cut-off than in the analysis (Fig. 1c) and in the two

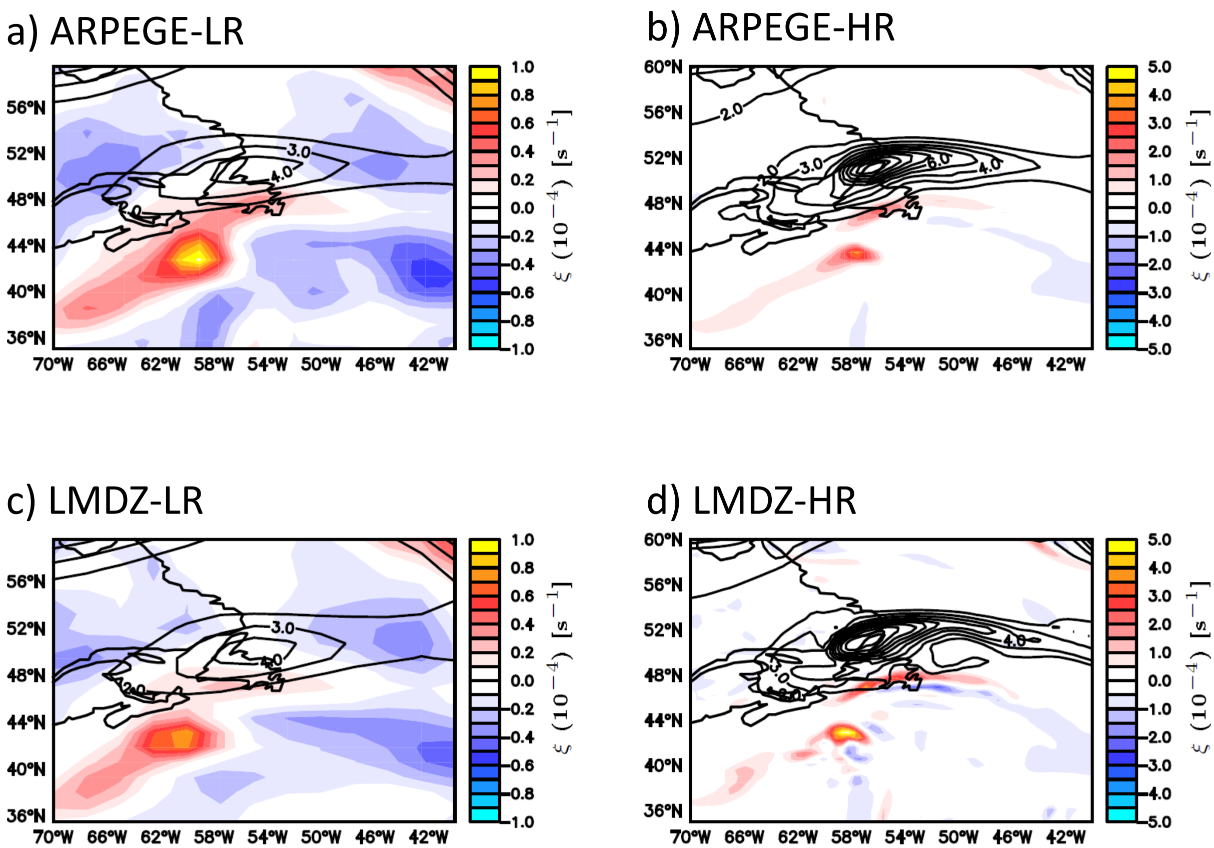


Figure 3. Hindcasts for the cyclogenesis of the Stalactite Cyclone at 18 UTC 29 September (T + 18 h) for hindcasts initiated at 00 UTC 29 September. The 250 hPa PV above 2 PVU (contoured every 1 PVU) and the 850 hPa relative vorticity (shaded) for a) ARPEGE-LR, b) ARPEGE-HR, c) LMDZ-LR, and d) LMDZ-HR. The colour scale is different between LR and HR runs.

HR runs (Figs. 3b,d). Also, the low-level northern vorticity maximum moves to the east of the cut-off in the HR runs and analysis, which is typical of strong baroclinic interaction, whereas it stays to the south of the cut-off in LR runs (Figs.3a and c).

Unlike the DRV, the northern precursor is a mixture of diabatic and dynamic processes as shown by the baroclinic conversion rates of Figs. 4 and 5. The vertical cross sections of Fig. 6 show that the dynamical component is mainly centred at upper levels, but with an equivalent-barotropic structure. This suggests that the northern precursor is forced by the vertical velocity associated with the PV cut-off which is characteristic of type-B cyclogenesis (Petterssen and Smebye, 1971). In LR hindcasts the dynamical forcing has a smaller vertical extent and is more spread out than the HR hindcasts. The dynamical forcing in LR hindcasts is located further east than the diabatic forcing (Figs. 6b and c) while the two forcings are more superimposed in HR hindcasts (Figs.6e and f, and Fig. S2 for ARPEGE). Both forcings increase with resolution by a factor of more than five

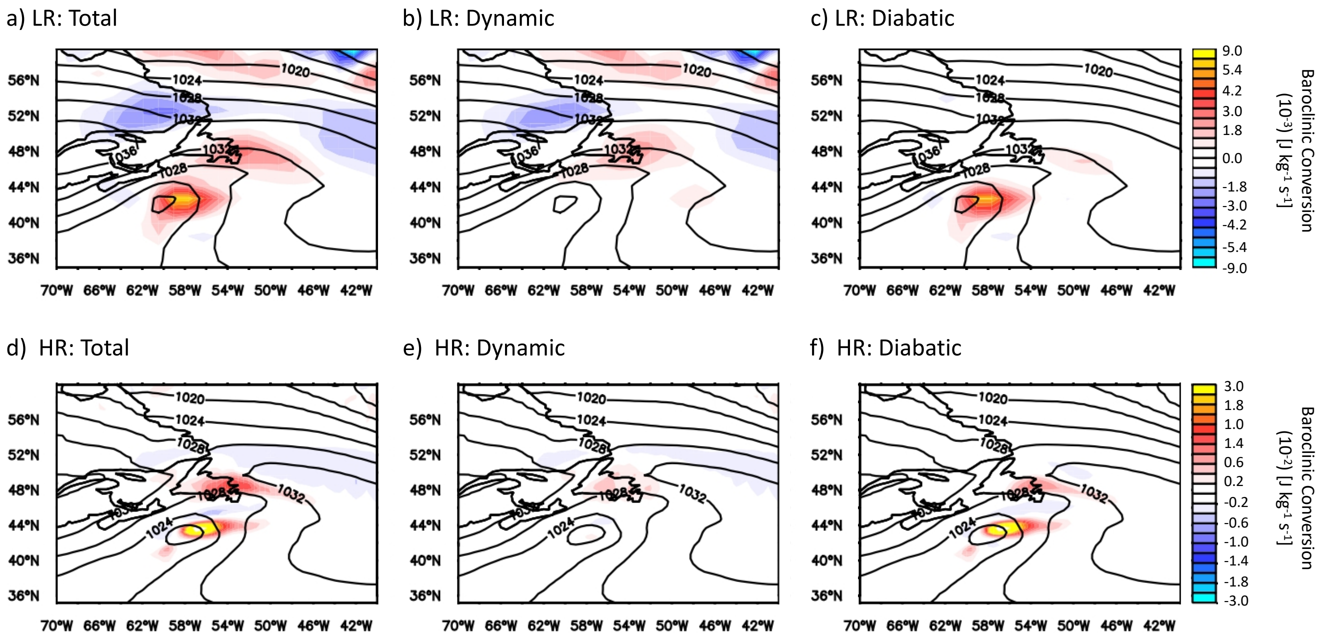


Figure 4. Vertically-averaged baroclinic conversion between 850 hPa and 300 hPa (shaded) and mean sea level pressure (contoured) at 18 UTC 29 September 2016 (T + 18 h from hindcast initiation at 00 UTC 29 September 2016) for ARPEGE hindcasts. a–c) ARPEGE-LR hindcast and d–f) ARPEGE-HR hindcast. a,d) total (dynamic + diabatic) baroclinic conversion; b,e) baroclinic conversion from dynamical processes; c,f) baroclinic conversion from diabatic processes. The colour scales refer to each row.

.

Table 2. Distance and PV criteria from Boettcher and Wernli (2013) for identifying DRVs between T+18 and T+24 from the hindcast initialised on 00 UTC 29 September. The PV criterion is based on a minimum PV value when averaged at the minimum MSLP location and the eight surrounding grid boxes, which is set to 0.8 PVU. The distance criterion is based on a minimum distance travelled by the vortex in 6 hours, which is 250 km. When the threshold is reached a ✓ is present, otherwise a ×. The DRV is described as quantitative if all the thresholds are reached and qualitative otherwise. Only criteria where at least one of the hindcasts do not meet the criteria are shown.

Model	PV (T+18) [PVU]	PV (T+24) [PVU]	850 hPa PV criterion [✓ / ×]	distance [km]	distance criterion [✓ / ×]	DRV type [Qualitative/Quantitative]
ECMWF (1.1°)	0.94	1.18	✓	370.6	✓	Quantitative
ARPEGE-LR	0.65	0.72	×	302.4	✓	Qualitative
ARPEGE-HR	1.45	1.74	✓	328.3	✓	Quantitative
LMDZ-LR	0.51	0.68	×	138.6	×	Qualitative
LMDZ-HR	2.77	2.54	✓	264.5	✓	Quantitative

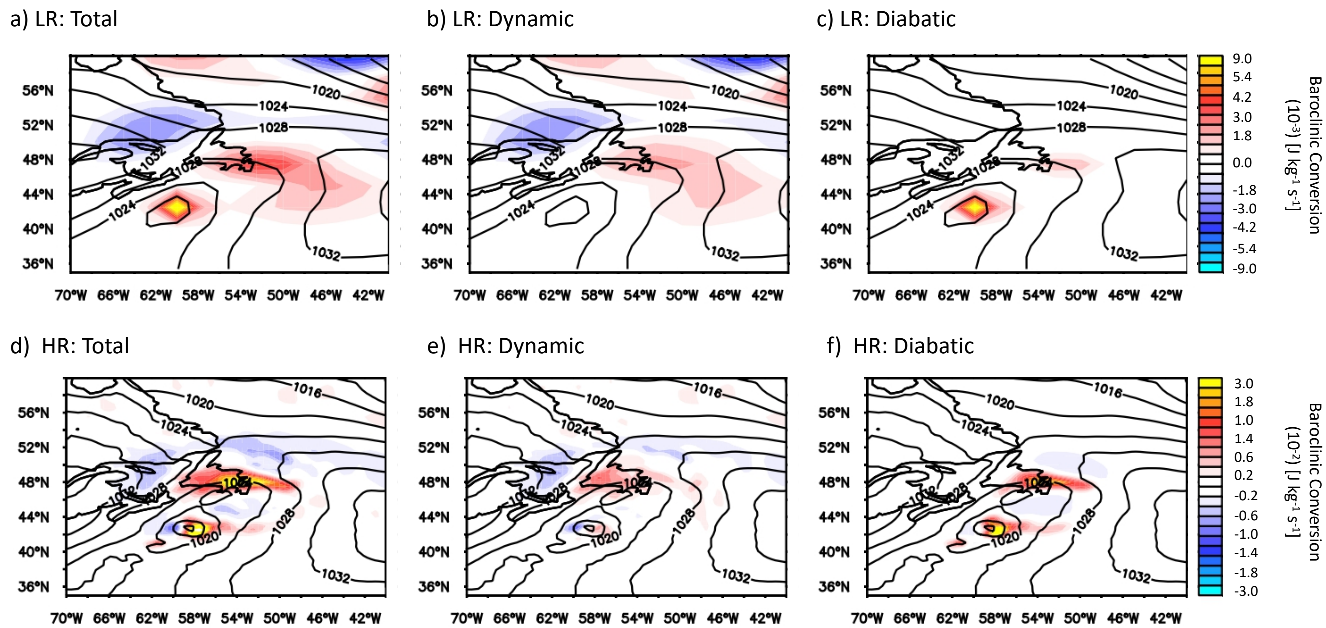


Figure 5. As in Fig. 4 but for LMDZ-LR (and -HR) hindcasts.

in the two models. However, the peak values of the diabatic baroclinic conversion exhibit a larger increase than those of the dynamical baroclinic conversion during the formation of the northern precursor (Figs. 5b,c,e,f and 6b,c,e,f).

To conclude, the northern precursor is rather poorly represented in LR compared to HR hindcasts because the less intense, and more spatially diluted, PV inside the cut-off induces a weaker dynamical forcing. An additional factor is the more active diabatic forcing in HR hindcasts in the vicinity of the northern precursor. So whilst both the dynamical and diabatic terms improve with resolution and it is difficult to determine which component matters most.

4.2.3 Merging of the two precursors

For the hindcasts shown here the merger of these two different precursors differs in timing from the analysis and between resolutions. The HR configurations (although delayed by 6 h compared to the ECMWF analysis) merge the DRV and upper-level dynamical precursor 12–18 h earlier than the LR runs (not shown). For LMDZ-LR, there is even no merging of the two precursors. The delay or absence of interaction between the two precursors likely has an impact on the track of the cyclone which was systematically located too far east in the LR runs (Fig. 2b), as the precursor merger starts the more northward movement of the cyclone in the track. This is understandable by the fact that the sooner merging is associated with a stronger upper-level forcing which is required for a cyclone to move northward perpendicularly to the jet axis (Coronel et al., 2015). There are two factors to explain the delayed or missed merging. One is the more rapidly eastward propagation of the DRV

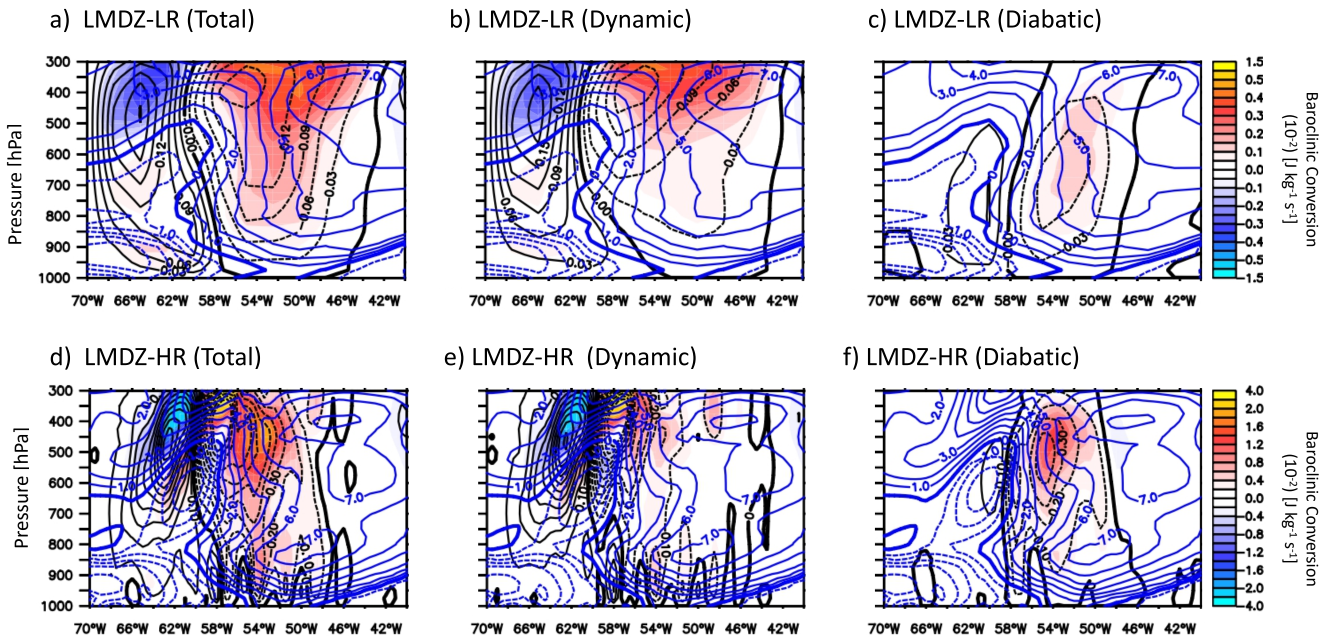


Figure 6. A vertical cross-section averaged across the northern precursor in the LMDZ hindcasts at 18 UTC 29 September 2016 (T+18 h). The baroclinic conversion (shaded), potential temperature anomaly (blue contours) and inverted ω (black contours) for a,d) total baroclinic conversion; b,e) the baroclinic conversion due to dynamic processes only; and c,f) the baroclinic conversion from diabatic processes. a–c) LMDZ-LR and d–f) LMDZ-HR. Note that the colourscales and contours are different between the LR and HR runs

.

in HR than LR runs (Fig. 3; Table 2), which is consistent with a stronger latent heating in the former runs. The second is the low-level northern precursor and the upper-level cut-off are moving less rapidly eastward in HR runs (not shown). This can be partly explained by the difference in longitude of the dynamical forcing between LR and HR hindcasts (compare Figs. 6b and e). The more rapid propagation of the DRV and less rapid motion of the northern precursor explain why the DRV is more able to catch up the northern precursor in HR runs as in the analysis.

To conclude on cyclogenesis, the LR hindcasts struggle to correctly represent the initiation of the cyclone because they miss the initial deepening of the northern small-scale low-level vortex and the roll-up of the merging two low-level vortices around the PV cut-off. However, the unexpected result is that the LR hindcasts are able to reproduce the behaviour of the DRV rather well, albeit with a smaller propagation speed.

4.3 Main deepening

The main focus of this section is the main deepening stage of the Stalactite Cyclone. Like the cyclogenesis phase the main deepening phase is considered by analysing the baroclinic conversion. The baroclinic conversion is considered either as an average over a $10° \times 10°$ area centred on the the minimum pressure of the Stalactite Cyclone (Fig. 7) or from its local maximum (Fig. S3). The averaged QG baroclinic conversion roughly recovers 60–70 % of the amplitude of that directly calculated from the model ω (Fig. 7), throughout the cyclone lifecycle. In addition, the model and QG baroclinic conversions are very similar in the timing, evolution, structure (not shown), and maximum peaks are close too (Fig. S3). This good correspondence provides confidence in our inversions and results.

In the cyclone average values (Fig. 7) the two stages of cyclone development are well separated: (i) the initial cyclogenesis stage occurring on 29–30 September (Section 4.2), and (ii) the main development stage that is dominated by the presence of a large-scale trough and an explosively-developing cyclone. The initiation stage is clearly dominated by diabatic processes. During the main deepening stage the dynamical processes begin to be more important, and more so in the HR hindcasts compared to the LR hindcasts. In the HR runs the dynamical term is even larger than the diabatic term during the whole main deepening stage. The delay in the dynamical processes compared to diabatic processes is particularly clear in LR hindcasts, suggesting a delayed forcing by the large-scale upper-level trough. Therefore, there is an increased importance of the dynamic term relative to the diabatic term with increased resolution. This ratio consistency is true for both the maximum (Fig. S3) and average values (Fig. 7) in both models and leadtimes. This ratio consistency in the main deepening stage disagrees with the previous studies of Willison et al. (2013) and Trzeciak et al. (2016). However, for the northern precursor at cyclogenesis we do agree with their studies.

Considering the dynamical processes in more detail (Figs. 8 and S4) helps to indicate the reason for the delay in the maximum deepening in the LR hindcasts compared to the analysis and the HR hindcasts. On 1 October 00 UTC, an upper-level PV signature is clearly visible above the surface cyclone in HR hindcasts while in the LR hindcast the cyclone is still mainly a DRV. The PV injection coming from the large-scale region of high PV, located to the northeast, into the upper-level disturbance interacting with the surface cyclone is delayed in the hindcasts. In the analysis, some PV injection has already occurred (Fig. 8a) but is just starting in the HR runs (Fig. 8c). The situation in ARPEGE-HR on 1 October 12 UTC (Fig. 8f) resembles more that of the analysis approximately 6 h earlier (not shown), with the cyclonic wave breaking being more advanced in the ECMWF analysis (Fig. 8d). Several studies have shown that the PV of the upper-level trough baroclinically interacting with a surface extratropical cyclone tends to advect the cyclone polewards (Rivière et al., 2012; Oruba et al., 2013; Coronel et al., 2015). Therefore the sooner nonlinear interaction of the cyclone with the large-scale upper-level PV reservoir and the sooner roll-up of the two features around each other explains the sooner deviation of the cyclone track to the north and the more westward position of the track in the analysis than in the hindcasts. For the HR hindcasts, the delay is a maximum of 6 h and the eastward shift is minimal while for LR hindcasts the delay is about 24 h and the eastward shift is more marked.

4.4 Interpretation of the difference between the models and comparison with aircraft observations

As previously said, to have cyclone features roughly at the same place in the models as in the observations, for a clean comparison, simulations initiated at 00 UTC 1 October 2016 are analysed in the present section.

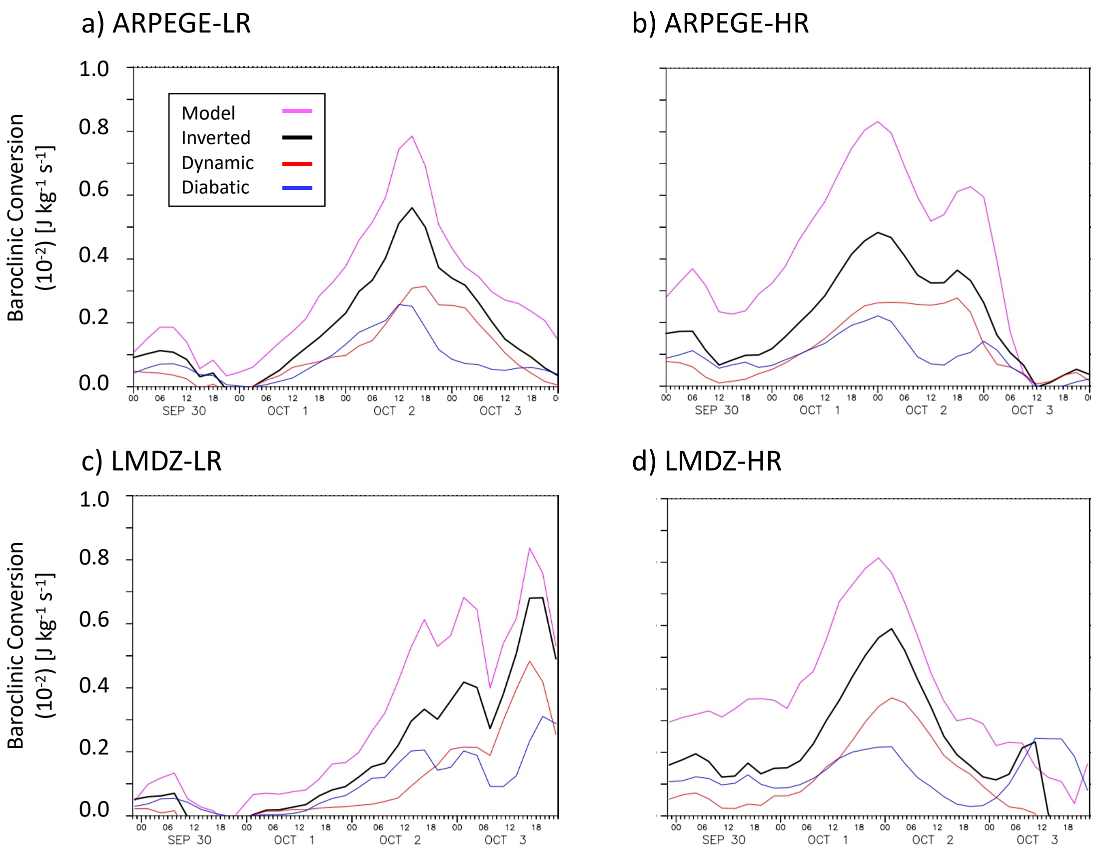


Figure 7. The evolution of the average baroclinic conversion in a $10^{\circ} \times 10^{\circ}$ box around the minimum pressure of the Stalactite cyclone. For a) ARPEGE-LR, b) ARPEGE-HR, c) LMDZ-LR, d) LMDZ-HR. The magenta line is for the baroclinic conversion calculated with the model ω; the black line is the total inverted ω; the red line the inverted ω from dynamical processes; and the blue line the inverted ω from diabatic processes. All hindcasts were initiated at 00 UTC 29 September 2016, average times are defined subtly differently in LMDZ and ARPEGE hence the 1.5 h extension in LMDZ plots. Maximum point values of baroclinic conversion are shown in Fig. S3.

4.4.1 Diabatic heating in the models

To more deeply investigate the relative contributions of dynamics and diabatic components and to assess potential differences between the models, Fig. 9 shows distributions of vertical velocities around the cyclone center for hindcasts initiated at 00 UTC 1 October 2016 but similar results occur for the hindcasts initiated at 00 UTC 29 September 2016 (not shown). Figure 9 first shows that the distribution of the model ω is rather well represented by its QG approximation ω_{QG} (Figs. 9a-d and e-h). Only some peak values of model ω near -2 Pa.s^{-1} for LMDZ-LR are missing in ω_{QG}. Second both vertical velocities increases with increased resolution (Figs. 9a,c,e,g and b,d,f,h). Distributions of ω_{QG} are rather similar in ARPEGE-LR and LMDZ-LR but the relative contributions of dynamic and diabatic parts differ between the two runs. There are more frequent strong ascents

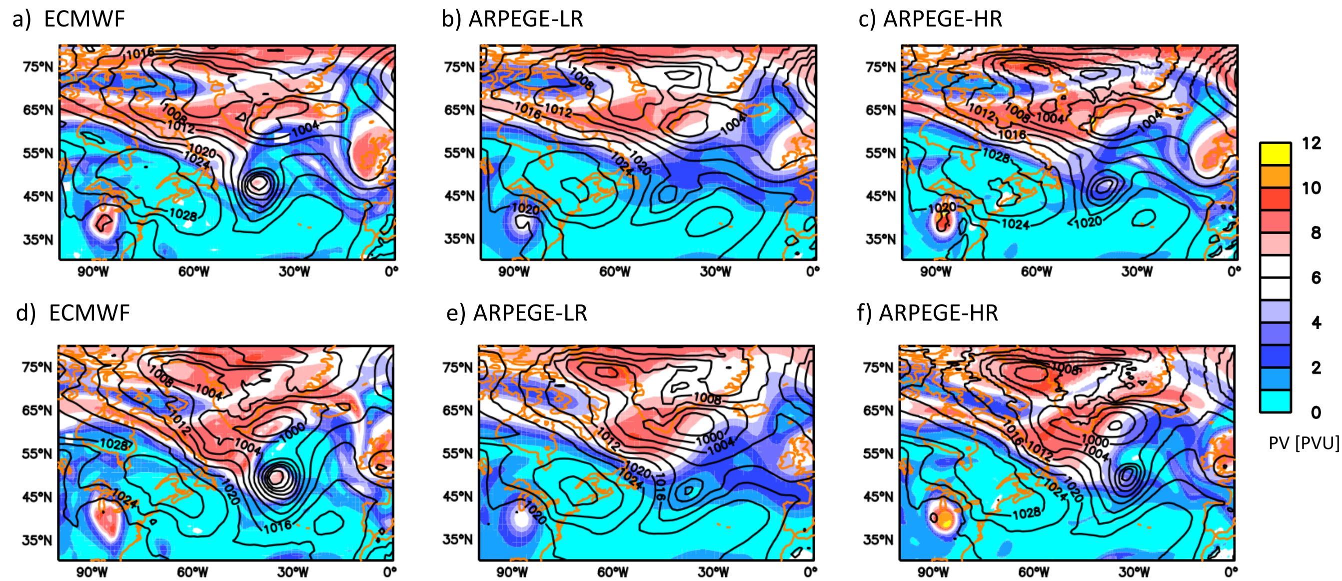


Figure 8. The 250 hPa PV (shaded) and mean sea level pressure (contoured) during the maximum deepening phase of the Stalactite cyclone. a–c) 00 UTC 1 October 2016, and d–f) 12 UTC 1 October 2016. a,d) ECMWF analysis; b,e) ARPEGE-LR hindcast;, and c,f) ARPEGE-HR hindcast. All hindcasts were initiated at 00 UTC 29 September 2016 and the colour scale applies to all plots. LMDZ-LR(-HR) plots at the same time are shown in Fig. S4.

of the diabatic component for LMDZ-LR than ARPEGE-LR (Figs. 9i,k), while the dynamical component partly offsets this difference (Figs. 9m,o). In HR hindcasts, there are largest values of ω_{QG} in LMDZ-HR compared to ARPEGE-HR (Figs. 9f,h) which is mainly due to the diabatic term.

To conclude, diabatic processes have stronger impact on vertical velocities in LMDZ than ARPEGE and the diabatic heating in the former model is stronger than in the latter. The terms that dominate the heating profiles both in ARPEGE and LMDZ are the large-scale condensational heating and convective terms (not shown). Thus, it is likely that observations of microphysical properties of the Stalactite Cyclone could be used to qualitatively determine which model has the better heating rates or structure. These comparisons are considered next.

4.4.2 Microphysical properties in the models and in observations

To determine whether observations of microphysical properties from field campaign flights can provide information on the underlying diabatic heating, the Stalactite Cyclone hindcasts are compared with flight F7 (Fig. 1b) of the SAFIRE Falcon during the NAWDEX field campaign. To ensure a fair comparison, the observation data has been linearly interpolated onto the model grid and a nearest-neighbour approach has been used to convert the model onto the flight track. Observed IWC is compared against “potential” IWC (cloud ice + snow) and “maximum” IWC [cloud ice + snow + liquid water content (LWC)] to take super-cooled liquid into account.

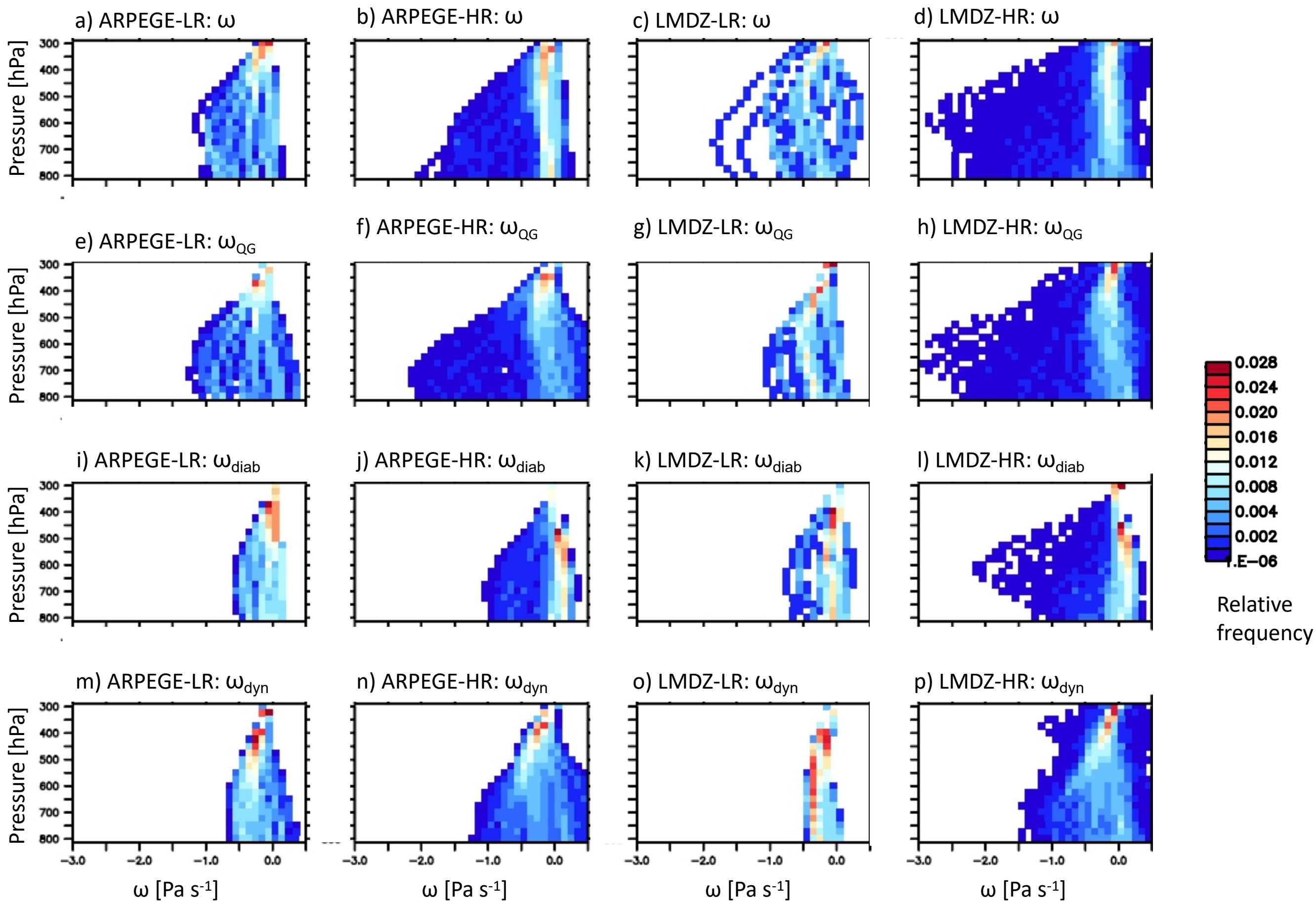


Figure 9. Bivariate histograms of vertical velocity *vs.* pressure in a 6°×6° box around the minimum pressure during the mature stage of the cyclone around maximum depth (c. 12 UTC 2 October 2016, T + 33–36 h). For a–d) modelled ω; e–h) ω_{QG}; i–l) ω_{diab}; and m–p) ω_{dyn}. For a,e,i,m) ARPEGE-LR; b,f,j,n) ARPEGE-HR; c,g,k,o) LMDZ-LR; and d,h,l,p) LMDZ-HR. The colour scale applies to all plots. The hindcasts were initated at 00 UTC 1 October 2016.

The windspeeds in the cyclone are well represented in all hindcasts with there only being a small shift in the probability density function toward smaller values by less than 5 $\mathrm{m\,s^{-1}}$ (not shown). This comparison provides confidence in the large-scale features of the cyclone. Therefore, microphysical features can be further considered. Figure 10 shows bi-variate histograms of the IWC for F7 from two observation platforms: RASTA (Fig. 10a) and RALI (Fig. 10f). There are larger values of IWC in RASTA compared to RALI because the lidar (being sensitive to smaller ice particles and smaller quantities of ice) information in RALI leads to a reduction of IWC compared to RASTA. Both platforms show the same shape with increasing values of IWC to around 600 hPa and then a uniform distribution until around 800 hPa, below which the instruments no longer detect ice clouds. The two retrieved IWC histograms provide an indication of uncertainty of the observations, which is useful to be compared with model outputs.

The model contribution to Fig. 10 consists of four rows, the first two rows showing "potential" IWC while the last two show "maximum" IWC. Comparing the first two rows (Figs. 10b–e and g–j) with the observations shows an underestimation of the model IWC. This underestimation is by a factor of 3–4, similarly to what Rysman et al. (2018) found when comparing observations and WRF simulations of Mediterranean systems. Furthermore, the peak of the model IWC distribution occurs at 700–750 hPa, 100–150 hPa lower than in the observations. There are small improvements with resolution: the HR simulations have larger IWC throughout, and particularly aloft and in the maximum values. Furthermore, there are differences between the models. The first difference is that the IWC values of LMDZ-LR are more dispersed than those of ARPEGE-LR suggesting a larger number of ice clouds at this altitude in LMDZ-LR (Figs. 10b,d, Figs. 10g,i, Figs. 11a,b). The stronger values at upper levels in LMDZ are more inline with the values given by the observations than ARPEGE (c.f. Figs. 10a–j). However, although LMDZ may be better at representing the IWC at upper levels, the overall shape of the distribution is better in ARPEGE compared to LMDZ. Indeed, the decreased IWC from 600 hPa to 300 hPa is better represented in ARPEGE. Applying the observation mask to the models (Figs. 10g–j) brings the frequencies more in line with the observations compared to without the mask by removing all the lowest values seen in the no-mask statistics. This is due to instruments not being sensitive to very small IWC, and also the models do not create discontinuities in IWC between cloudy and clear-sky regions. The comparisons between the mask (Figs. 10g–j) and no-mask (Figs. 10b–e) values implies that there are very small IWC values in the model outside of the observed region (particularly for ARPEGE-LR) indicating the horizontal structure of the cyclone is reasonable.

Is the underestimated IWC in the models due to the underestimated liquid-to-solid transition for cold temperatures or to the underestimation of condensates as a whole? To answer this question the LWC below 273 K is added to the IWC to create the last two rows ("maximum" IWC; Figs. 10k–r). Adding the LWC makes limited difference to either of the ARPEGE hindcasts (Figs. 10k,l,o,p) suggesting that either there are fewer LWC points added or the LWC points added have a small magnitude. On the other hand, adding LWC into the LMDZ definition drastically changes the shape and increases the values of total IWC at lower levels (Figs. 10m,n,q,r). The LMDZ distributions have been changed to the extent that the shape now shows more agreement with the observations than when the LWC was not taken into account. These changes in LMDZ are also apparent within Fig. 11, although the model difference is reduced at increased resolution (Figs. 11c and d). The much larger "maximum" IWC in LMDZ compared to ARPEGE over all the levels is consistent with the larger diabatic heating shown in Figs. 9e-h.

Given the change by the inclusion of LWC in the definition of the IWC it is useful to know the proportion of ice, mixed phase, and super-cooled liquid points that make up these distributions. We arbitrarily define ice points in the model to be those where the LWC component of the "maximum" IWC is less than 1$\%$ and "pure" super-cooled liquid to be points where the LWC component is greater than 99$\%$ of the "maximum" IWC, all other points are mixed phase. These results are compared with those points defined as super-cooled liquid, mixed phase and ice retrieved IWC from RALI measurements. To ensure a fair comparison between ice and super-cooled liquid water the "pure" values are combined with the mixed phase values. Table 3 shows that whilst the combined ice points exceed that of the observations (particularly for ARPEGE) the values are not unreasonable. However, when the combined super-cooled liquid water is considered the models significantly over-estimate the

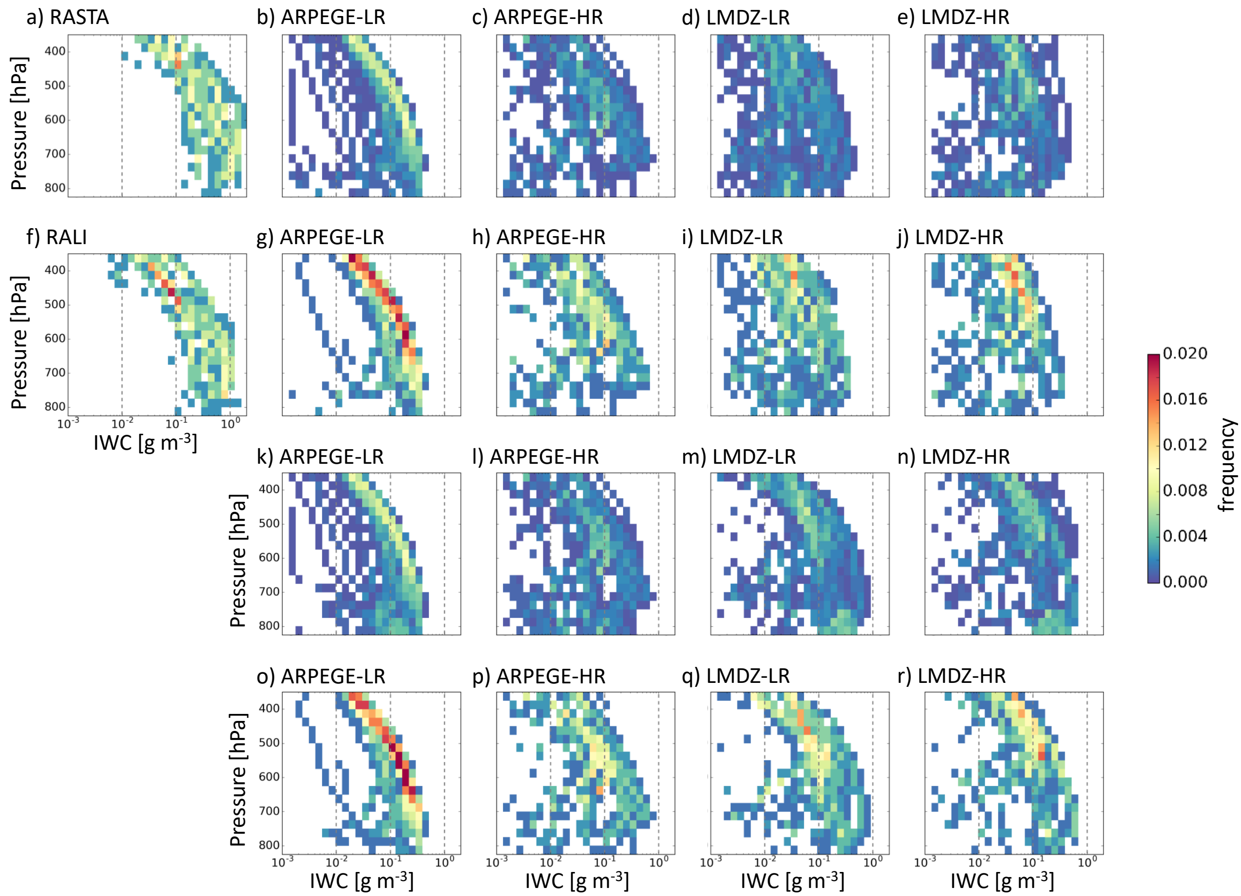


Figure 10. Bi-variate histograms of ice water content *vs.* pressure for F7 for a) RASTA observations (radar only); f) RALI (radar + lidar) observations; b–e) hindcast output using "potential" ice water content (cloud ice + snow) without applying a mask to the observations; g–j) hindcast output of "potential" ice water content and with the observation mask applied; k–n) hindcast for "maximum" ice water content (ice water content + liquid water content) without the observation mask applied; and o–r) hindcast of "maximum" ice water content with the observation mask applied. For b,g,k,o) ARPEGE-LR; c,h,l,p) ARPEGE-HR; d,i,m,q) LMDZ-LR; and e,j,n,r) LMDZ-HR. The hindcast data is initiated at 00 UTC 1 October 2016 and uses the nearest-gridpoint to the flight path from the two times surrounding the flight path (12 and 15 UTC 2 October 2016; T+36–39 h). The flight occurred from 1300–1600 UTC. The colour scale applies to all panels, and the histograms have been normalized by all points.

.

amount of super-cooled liquid points, by factors of 24–47. Considering Table 3 alongside the earlier discussion of the impact of adding LWC shows that the super-cooled liquid water being added to ARPEGE is of a smaller magnitude than that of LMDZ. It is also worth noting that although the LR hindcasts are more largely underestimating the IWC than the HR hindcasts, they are closer to the observations than the HR hindcasts in the percentage of super-cooled water.

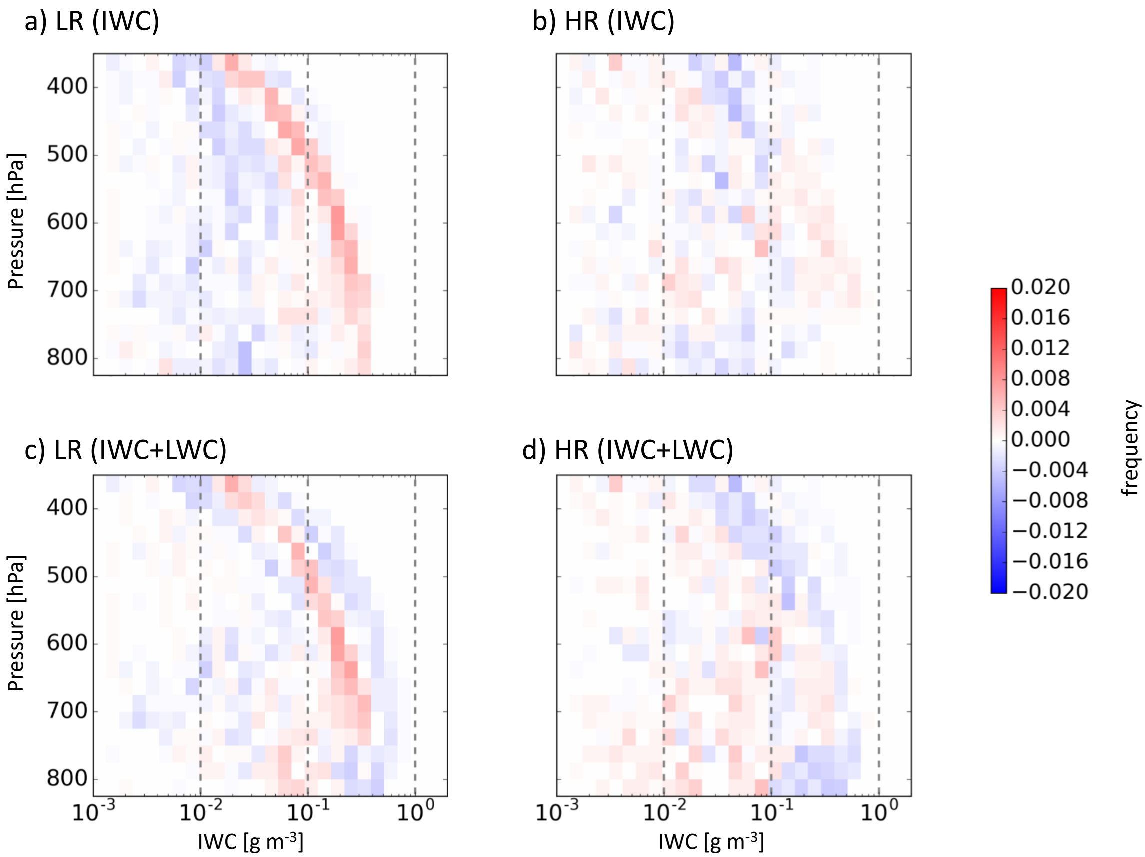


Figure 11. Difference bi-variate histograms for F7 of ice water content *vs.* pressure between ARPEGE and LMDZ for a) LR differences in "potential" ice water content (cloud ice + snow) only (Fig. 10b - Fig. 10d) ; b) HR differences in "potential" ice water content only (Fig. 10c - Fig. 10e); c) LR differences in "maximum" ice water content (ice water content + liquid water content) (Fig. 10k - Fig. 10m); and d) HR differences in "maximum" ice water content (Fig. 10l - Fig. 10n). Reds refer to ARPEGE having a larger quantity and blues for LMDZ. The colour scale applies to all panels. The hindcasts are initiated at 00 UTC 1 October 2016 and uses the nearest-gridpoint to the flight path from the two times surrounding the flight path (12 and 15 UTC 2 October 2016; T+36–39 h).

.

Radar reflectivities confirm the strong underestimation of IWC in the hindcasts (Fig. 12). The smaller values reached by LMDZ compared to ARPEGE is probably due to the larger percentage of liquid hydrometeors which induce smaller reflectivities than ice. It also confirms that the LR hindcasts outperform the HR hindcasts and ARPEGE is better than LMDZ in terms of shape of the IWC distribution. Despite a systematic underestimation of reflectivity at all levels, the ARPEGE-LR reflectivity exhibits the closest shape to the observations compared to the other three hindcasts.

Finally, to be confident in the above results, additional figures are presented in the supplementary material. Figures S5 to S7 support the above findings by doing the same analysis along flight F6. Also, a comparison between RALI and CloudSat-CALIPSO measurements has been made along the common path of flight F7 and the A-train. The CloudSat reflectivities have

Table 3. The fraction of points within F7 that have values deemed as super-cooled liquid, mixed phase and ice. “MAX”=IWC+LWC, combined super-cooled liquid = super-cooled liquid + mixed phase, and combined ice = ice + mixed phase. The hindcasts are initiated at 00 UTC 1 October 2016 and uses the nearest-gridpoint to the flight path from the two times surrounding the flight path (12 and 15 UTC 2 October 2016; T+36–39 h).

	Observations [%]	LMDZ-LR [%]	LMDZ-HR [%]	ARPEGE-LR [%]	ARPEGE-HR [%]
Super-cooled liquid (LWC > 0.99(“MAX”))	1.5	1.2	0.5	0.0	0.0
Mixed phase (0.01(“MAX”) < LWC < 0.99(“MAX”))	0.2	72.8	79.7	41.4	64.6
Ice (LWC < 0.01(“MAX”))	98.3	26.0	19.8	58.6	38.4
Combined super-cooled liquid	1.7	74.0	80.2	41.4	61.6
Combined ice	98.5	98.8	99.5	100.0	100.0

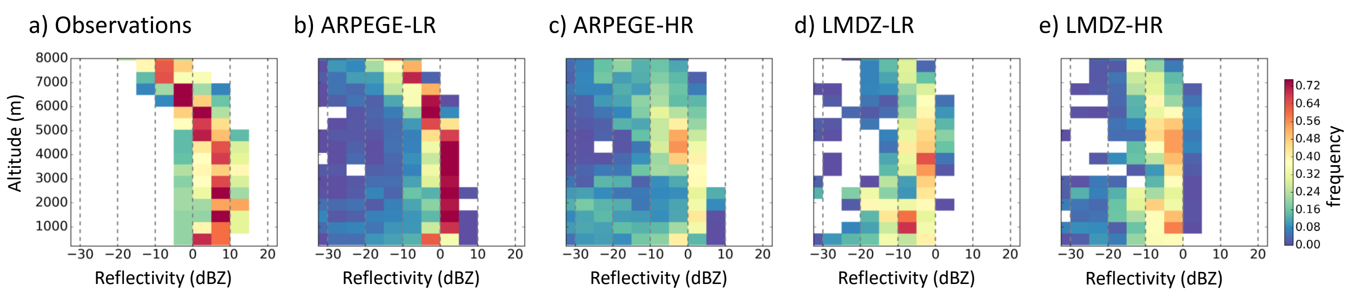


Figure 12. Contour Frequency Altitude Diagrams (CFADs) of Radar Reflectivity for F7 a) RASTA observations, b) ARPEGE-LR, c) ARPEGE-HR, d) LMDZ-LR, and e) LMDZ-HR. The hindcasts are initiated at 00 UTC 1 October 2016 and uses the nearest-gridpoint to the flight path from the two times surrounding the flight path (12 and 15 UTC 2 October 2016; T+36–39 h). The colour scale applies to all panels. No mask to the observations has been applied here.

similar structure and similar amplitude as the RALI reflectivities (Fig. S8c,d). The DARDAR and RALI target classifications tend to agree with the main discrepancies originating from the time shift and the higher noise in CALIPSO backscatter and the lower sensitivity of RASTA close to the surface. This explains why the super-cooled layer detection is consistent but the mixed phase attribution is slightly different due to the radars sensitivity (Fig.S8e,f). Despite these differences, regions of combined super-cooled liquid (super-cooled plus mixed phase) are rather similar which gives confidence in the above conclusions.

To conclude, LMDZ produces more IWC which is associated with a more intense latent heating than ARPEGE. In that sense, it is closer to the observations. However, the ratio between liquid *vs.* solid species contributing the IWC is less realistic in LMDZ than ARPEGE. Hence, it is worth noting that whilst the IWC can provide some information about the diabatic heating, caution is needed in interpreting the results as it does not provide complete information to be able to determine which of the two models produce the better heating compared to reality. However, the microphysical observations from flights during field campaigns are still useful in helping to identify the deficiencies of each model and determine what processes are linked in the models and why one of the models produces a more active cyclone compared to the other.

5 Summary

The representation of the Stalactite Cyclone in the two atmospheric GCMs, ARPEGE-Climat 6.3 (hereafter ARPEGE) and LMDZ6A (hereafter LMDZ), corresponding to the atmospheric components of the CNRM and IPSL climate models (CNRM-CM6-1 and IPSL-CM6A) has been examined in detail. The two models are run at two resolutions: one at a coarse resolution of approximately 150–200 km (LR) and the other at a higher resolution of approximately 50 km (HR). The T-AMIP protocol is used to determine how well the climate models can represent the physical processes linked to the Stalactite cyclone and how well it compares to flight observations made during the NAWDEX field campaign. The protocol gives us valuable insight into the formation of the Stalactite Cyclone.

Figure 13 shows a schematic of the many stages of the Stalactite Cyclone: from initiation as a Diabatic Rossby Vortex (DRV) initiated from a mesoscale convective system (point 0) through the merger of the DRV (point 1) and a dynamical forcing factor (point 2) at cyclogenesis (point 3), to its rapid deepening (point 4), and comparisons with the observations (point 5) round to cyclolysis. There are differences between each of the models and with the analysis at each of these points and these are summarised in the main results below. The points are numbered based on the schematic (Fig. 13).

1. All hindcasts produce a DRV to some degree of accuracy: LR hindcasts produce a qualitative DRV whereas HR hindcasts produce a quantitative DRV that meet the criteria of Boettcher and Wernli (2013).

2. All models produce an upper-level potential vorticity cut-off. However, due to its fine-scale structure, the cut-off is not as intense nor as deep in the LR hindcasts as in the HR hindcasts and analysis.

3. Due to the above the initial deepening associated with the vortex roll-up between the two precursors at cyclogenesis is weaker in LR hindcasts and the initial deepening is better represented when the resolution increases. The reduced initial deepening implies that LR versions cannot fully (dynamically) represent the Stalactite Cyclone. In particular, they do not represent the right tracks because of their too late interaction with upper-level PV reservoir.

4 (a). All hindcasts produce an explosively-deepening cyclone with near 24 hPa deepening in 24 h during the mature stage similar to the analysis. However, the strong deepening stage is delayed by 24 h in LR hindcasts.

4 (b). Diabatic heating extends throughout the troposphere during maximum deepening for both models, but is larger in LMDZ compared to ARPEGE. Increasing the resolution does not increase the relative contribution of diabatic heating to the main deepening of the cyclone unlike in previous studies (e.g. Willison et al., 2013; Trzeciak et al., 2016). Instead there

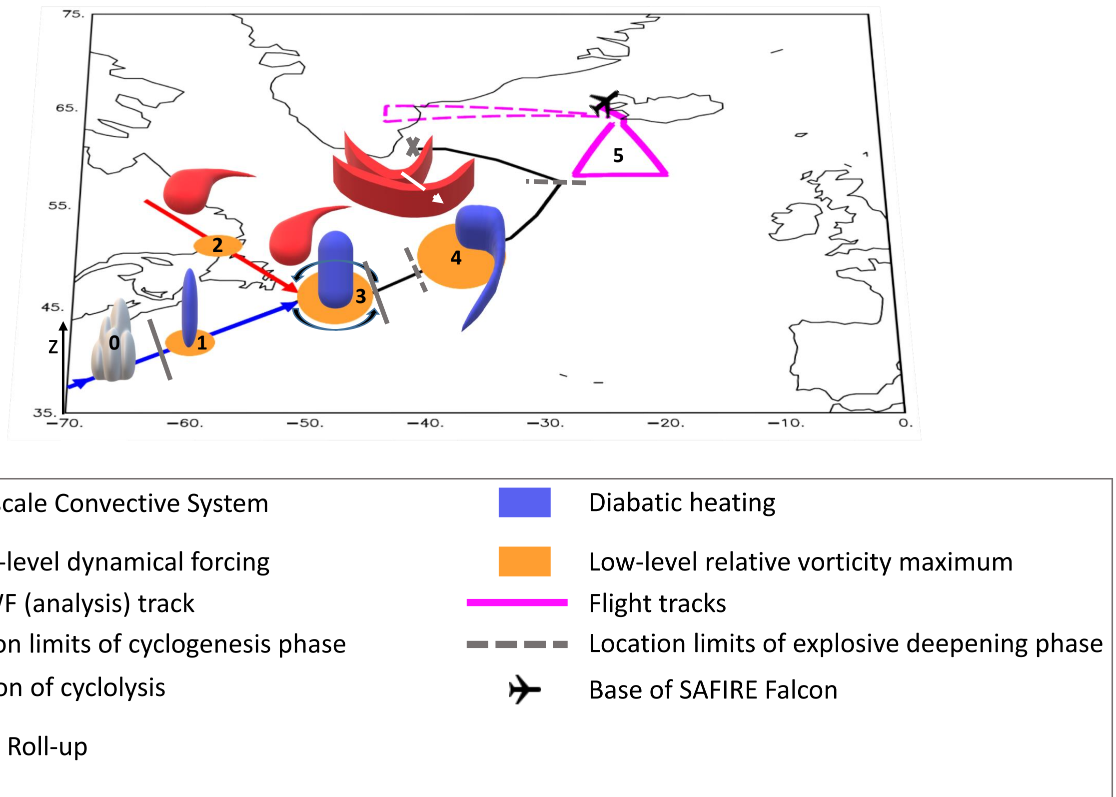


Figure 13. A schematic of the Stalactite Cyclone. 0) the Mesoscale Convective System that initiates the Diabatic Rossby Vortex (1) that travels along the blue arrow. The northern precusor (2) with upper-level PV cut-off that moves towards the diabatic Rossby vortex and initiations a roll-up between the two precursors at cyclogenesis to create the Stalactite Cyclone (3). Explosive deepening occurs as a result of strong diabatic heating throughout the column and the interaction with a series of embedded upper level high PV regions (4). Flight observations (5) indicate that ice water content is underestimated and so could have impacts on the diabatic heating and evolution of the cyclone.

.

are local increases in the diabatic heating which are particularly important for the northern precursor at cyclogenesis (Figs. 6 and S2).

5 (a). Both models and resolutions underestimate the IWC from flight observations, even when super-cooled liquid water is taken into account, by a factor of 3–4 in agreement with Rysman et al. (2018). However, the shape of the vertical distribution of IWC is in good agreement for ARPEGE. The LMDZ hindcasts only come into agreement for the shape of the distribution with the observations when super-cooled liquid water is added to the ice. When all condensates are considered, the LMDZ model presents larger values compared to ARPEGE over the whole troposphere. This larger content of

condensates is associated with larger diabatic heating, larger vertical velocities, and hence provides an explanation for the larger deepening rate in LMDZ compared to ARPEGE.

5 (b). Both models appear to substantially over-estimate the amount of super-cooled liquid water content in the cyclone. This comes as a result of an increased number of mixed phase gridpoints.

Thus, returning to the originally proposed questions and criteria for the correct representation of the Stalactite Cyclone the evidence suggests that climate models, when they are run at a coarse resolution, cannot represent the initial stage of the Stalactite cyclone but they can produce the main deepening during the mature stage. The results also indicate that improvements in dynamical processes are as (if not more) important as improvements in diabatic processes with increasing resolution. The results further show that microphysical properties can be used, with caution, qualitatively to provide indirect information on the diabatic heating in climate models. Therefore the flight observations provide (albeit not complete) an interesting insight into whether the climate models are producing the correct heating. This last topic is currently being investigated further by the authors with respect to the downstream impact of extra-tropical cyclones in climate models on subsequent ridge building.

Although the present results only apply for this particular case study[4], the results have important implications and show areas that warrant further investigation. Firstly, it shows that the T-AMIP protocol is useful for considering the physical mechanisms that occur within cyclones and their interaction with dynamics. Secondly, it shows that increasing resolution does help with the representation of cyclones such that within the next few years, when climate models will be regularly run at c. 50 km, many synoptic-scale features of the atmosphere will be dynamically well represented. Finally, and arguably most critically, it warns that although climate models may produce similar cyclones they can be doing so for very different reasons and these reasons are likely to have an influence upon other areas of the climate system and the response of model cyclones to climate change. We recommend that further research occurs into the partition of super-cooled liquid water, mixed phase and ice water in models (and the influence this has on cyclone representation) and further comparisons with observations are made in all regions as this will have a strong influence on the development of microphysical schemes in climate and weather prediction models. Therefore, whilst signs are encouraging for future versions of climate models, caution is still needed when considering current simulations of future climate scenarios and the impact of extra-tropical cyclones, particularly for regional impact-based studies.

Data availability. Data is available by contacting either D. Flack at david.flack1@metoffice.gov.uk or the corresponding author.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/wcd-0-1-2020-supplement.

[4]A second cyclone (the following cyclone; IOP 7 of NAWDEX) related to the future work shows the same results (not shown).

Author contributions. All authors contributed to the writing and editing of the manuscript as well as the scientific discussions. DLAF produced the first draft and conducted the model analysis. IM performed the IPSL-CM6A simulations and RR the CNRM-CM6-1 simulations. GR and SB designed the study. JD, QC and JP provided the observational datasets.

Competing interests. There are no competing interests present.

Acknowledgements. This work is part of the DIP-NAWDEX (DIabatic Processes in the North Atlantic Waveguide and Downstream impact EXperiment) project which is funded by the Agence Nationale de la Recherche (ANR) under grant number ANR-17-CE01-0010-01. The funding for IM for running the LMDZ T-AMIP experiments was from the European Research Council grant number 694768. The LMDZ simulations were performed using the HPC from GENCI-IDRIS (grant 0292). The airborne measurements and the SAFIRE Falcon flights received direct funding from IPSL, Météo-France, INSU-LEFE, EUFAR-NEAREX and ESA (EPATAN, contract n° 4000119015/16/NL/CT/gp). The authors wish to acknowledge use of the Ferret program for analysis and Figs. 1–9 and Figs. S1–S4 in this paper. Ferret is a product of NOAA's Pacific Marine Environmental Laboratory. (Information is available at http://ferret.pmel.noaa.gov/Ferret/). We further acknowledge the use of imagery from the NASA Worldview application (https://worldview.earthdata.nasa.gov), part of the NASA Earth Observing System Data and Information System (EOSDIS). We also wish to thank the two anonymous reviewers whose comments have helped improved the manuscript.

References

Bengtsson, L., Hodges, K. I., and Keenlyside, N.: Will Extratropical Storms Intensify in a Warmer Climate?, J. Climate, 22, 2276–2301, https://doi.org/10.1175/2008JCLI2678.1, 2009.

Blanchard, N., Pantillon, F., Chaboureau, J.-P., and Delanoë, J.: Organization of convective ascents in a warm conveyor belt, 1, 617–634, https://doi.org/10.5194/wcd-1-617-2020, 2020.

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., Zhang, Y., Marchand, R., Haynes, J. M., Pincus, R., and John, V. O.: COSP: Satellite simulation software for model assessment, Bull. Amer. Meteor. Soc., 92, 1023–1043, https://doi.org/10.1175/2011BAMS2856.1, 2011.

Boettcher, M. and Wernli, H.: A 10-yr Climatology of Diabatic Rossby Waves in the Northern Hemisphere, Mon. Wea. Rev., 141, 1139–1154, https://doi.org/10.1175/MWR-D-12-00012.1, 2013.

Bony, S., Bellon, G., Klocke, D., Sherwood, S., Fermepin, S., and Denvil, S.: Robust direct effect of carbon dioxide on tropical circulation and regional precipitation, Nature Geosci., 6, 447–451, https://doi.org/10.1038/ngeo1799, 2013.

Boucher, O., Servonnat, J., and co authors: Presentation and evaluation of the IPSL-CM6A-LR climate model, J. Atmos. Model Dev., p. submitted, 2020.

Brient, F., Roehrig, R., and Voldoire, A.: Evaluating Marine Stratocumulus Clouds in the CNRM-CM6-1 Model Using Short-Term Hindcasts, J. Atmos. Model Dev., 11, 127–148, https://doi.org/10.1029/2018MS001461, 2019.

Catto, J. L., Shaffrey, L. C., and Hodges, K. I.: Can Climate Models Capture the Structure of Extratropical Cyclones?, J. Climate, 23, 1621–1635, https://doi.org/10.1175/2009JCLI3318.1, 2010.

Catto, J. L., Shaffrey, L. C., and Hodges, K. I.: Northern Hemisphere Extratropical Cyclones in a Warming Climate in the HiGEM High-Resolution Climate Model, J. Climate, 24, 5336–5352, https://doi.org/10.1175/2011JCLI4181.1, 2011.

Cazenave, Q., Ceccaldi, M., Delanoë, J., Pelon, J., Groß, S., and Heymsfield, A.: Evolution of DARDAR-CLOUD ice cloud retrievals: new parameters and impacts on the retrieved microphysical properties, Atmospheric Measurement Techniques, 12, 2819–2835, https://doi.org/10.5194/amt-12-2819-2019, 2019.

Coronel, B., Ricard, D., Rivière, G., and Arbogast, P.: Role of moist processes in the tracks of idealized midlatitude surface cyclones, J. Atmos. Sci., 72, 2979–2996, 2015.

Davis, C. A., Stoelinga, M. T., and Kuo, Y.-H.: The integrated effect of condensation in numerical simulations of extratropical cyclogenesis, Mon. Wea. Rev., 121, 2309–2330, 1993.

Delanoë, J., Protat, A., Jourdan, O., Pelon, J., Papazzoni, M., Dupuy, R., Gayet, J.-F., and Jouan, C.: Comparison of Airborne In Situ, Airborne Radar–Lidar, and Spaceborne Radar–Lidar Retrievals of Polar Ice Cloud Properties Sampled during the POLARCAT Campaign, J. Atmos. Oceanic Technol., 30, 57–73, https://doi.org/10.1175/JTECH-D-11-00200.1, 2013.

Delanoë, J. and Hogan, R. J.: A variational scheme for retrieving ice cloud properties from combined radar, lidar, and infrared radiometer, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD009000, 2008.

Delanoë, J. and Hogan, R. J.: Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds, J. Geophys. Res.: Atmos., 115, https://doi.org/10.1029/2009JD012346, 2010.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fermepin, S. and Bony, S.: Influence of low-cloud radiative effects on tropical circulation and precipitation, J. Atmos. Model Dev., 6, 513–526, https://doi.org/10.1002/2013MS000288, 2014.

Fink, A. H., Pohle, S., Pinto, J. G., and Knippertz, P.: Diagnosing the influence of diabatic processes on the explosive deepening of extratropical cyclones, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL051025, 2012.

Finnis, J., Holland, M. M., Serreze, M. C., and Cassano, J. J.: Response of Northern Hemisphere extratropical cyclone activity and associated precipitation to climate change, as represented by the Community Climate System Model, Journal of Geophysical Research: Biogeosciences, 112, https://doi.org/10.1029/2006JG000286, 2007.

Fouquart, Y. and Bonnel, B.: Computations of solar heating of the Earth's atmosphere: A new parameterization, Beitr. Phys. Atmosph., 53, 35–61, 1980.

Guérémy, J.: A continuous buoyancy based convection scheme: one-and three-dimensional validation, Tellus A: Dynamic Meteorology and Oceanography, 63, 687–706, https://doi.org/10.1111/j.1600-0870.2011.00521.x, 2011.

Harvey, B. J., Shaffrey, L. C., Woollings, T. J., Zappa, G., and Hodges, K. I.: How large are projected 21st century storm track changes?, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL052873, 2012.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, Bull. Amer. Meteor. Soc., 90, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1, 2009.

Holton, J. R.: An Introduction to Dynamic Meteorology, 2004.

Hoskins, B. J. and Pedder, M. A.: The diagnosis of middle latitude synoptic development, Quart. J. Roy. Meteor. Soc., 106, 707–719, https://doi.org/10.1002/qj.49710645004, 1980.

Hoskins, B. J., Draghici, I., and Davies, H. C.: A new look at the ω-equation, Quart. J. Roy. Meteor. Soc., 104, 31–38, https://doi.org/10.1002/qj.49710443903, 1978.

Hourdin, F., Jam, A., Rio, C., Couvreux, F., Sandu, I., Lefebvre, M.-P., Brient, F., and Idelkadi, A.: Unified Parameterization of Convective Boundary Layer Transport and Clouds With the Thermal Plume Model, J. Atmos. Model Dev., 11, 2910–2933, https://doi.org/10.1029/2019MS001666, 2019.

Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., Jam, A., Musat, I., Idelkadi, A., Fairhead, L., Foujols, M.-A., Mellul, L., Traore, A.-K., Dufresne, J.-L., Boucher, O., Lefebvre, M.-P., Millour, E., Vignon, E., Jouhaud, J. F., Diallo, B., Lott, F., Gastineau, G., Caubel, A., Meurdesoif, Y., and Ghattas, J.: LMDZ6A: the atmospheric component of the IPSL climate model with improved and better tuned physics, J. Atmos. Model Dev., p. e2019MS001892, https://doi.org/10.1029/2019MS001892, 2020.

Karmalkar, A. V., Sexton, D. M. H., Murphy, J. M., Booth, B. B. B., Rostron, J. W., and McNeall, D. J.: Finding plausible and diverse variants of a climate model. Part II: development and validation of methodology, Climate Dyn., 53, 847–877, https://doi.org/10.1007/s00382-019-04617-3, 2019.

Klocke, D. and Rodwell, M. J.: A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models, Quart. J. Roy. Meteor. Soc., 140, 517–524, https://doi.org/10.1002/qj.2172, 2014.

Li, J., Chen, H., Rong, X., Su, J., Xin, Y., Furtado, K., Milton, S., and Li, N.: How Well Can a Climate Model Simulate an Extreme Precipitation Event: A Case Study Using the Transpose-AMIP Experiment, J. Climate, 31, 6543–6556, https://doi.org/10.1175/JCLI-D-17-0801.1, 2018.

Lopez, P.: Implementation and validation of a new prognostic large-scale cloud and precipitation scheme for climate and data-assimilation purposes, Quart. J. Roy. Meteor. Soc., 128, 229–257, https://doi.org/10.1256/00359000260498879, 2002.

Ma, H.-Y., Xie, S., Boyle, J. S., Klein, S. A., and Zhang, Y.: Metrics and Diagnostics for Precipitation-Related Processes in Climate Model Short-Range Hindcasts, J. Climate, 26, 1516–1534, https://doi.org/10.1175/JCLI-D-12-00235.1, 2013.

Ma, H.-Y., Xie, S., Klein, S. A., Williams, K. D., Boyle, J. S., Bony, S., Douville, H., Fermepin, S., Medeiros, B., Tyteca, S., Watanabe, M., and Williamson, D.: On the Correspondence between Mean Forecast Errors and Climate Errors in CMIP5 Models, J. Climate, 27, 1781–1798, https://doi.org/10.1175/JCLI-D-13-00474.1, 2014.

Maddison, J. W., Gray, S. L., Martínez-Alvarado, O., and Williams, K. D.: Upstream Cyclone Influence on the Predictability of Block Onsets over the Euro-Atlantic Region, Mon. Wea. Rev., 147, 1277–1296, https://doi.org/10.1175/MWR-D-18-0226.1, 2019.

Maddison, J. W., Gray, S. L., Martínez-Alvarado, O., and Williams, K. D.: Impact of model upgrades on diabatic processes in extratropical cyclones and downstream forecast evolution, Quart. J. Roy. Meteor. Soc., n/a, n/a, https://doi.org/10.1002/qj.3739, 2020.

Madeleine, J.-B., Hourdin, F., Grandpeix, J.-Y., Rio, C., Dufresne, J.-L., Konsta, D., Musat, I., Idelkadi, A., Fairhead, L., Millour, E., Lefebvre, M.-P., Mellul, L., Cheruy, F., Boucher, O., Vignon, E., Rochetin, N., Lemonnier, F., Touzé-Peiffer, L., and Bonazzola, M.: Improved representation of clouds in the LMDZ6A Global Climate Model, J. Atmos. Model Dev., p. submitted, 2020.

McDonald, R. E.: Understanding the impact of climate change on Northern Hemisphere extra-tropical cyclones, Climate Dyn., 37, 1399–1425, https://doi.org/10.1007/s00382-010-0916-x, 2011.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres, 102, 16 663–16 682, https://doi.org/10.1029/97JD00237, 1997.

Morcrette, J.-J., Barker, H. W., Cole, J. N. S., Iacono, M. J., and Pincus, R.: Impact of a New Radiation Package, McRad, in the ECMWF Integrated Forecasting System, Mon. Wea. Rev., 136, 4773–4798, https://doi.org/10.1175/2008MWR2363.1, 2008.

Oertel, A., Boettcher, M., Joos, H., Sprenger, M., Konow, H., Hagen, M., and Wernli, H.: Convective activity in an extratropical cyclone and its warm conveyor belt – a case-study combining observations and a convection-permitting model simulation, Quart. J. Roy. Meteor. Soc., 145, 1406–1426, https://doi.org/10.1002/qj.3500, 2019.

Orlanski, I. and Katzfey, J.: The Life Cycle of a Cyclone Wave in the Southern Hemisphere. Part I: Eddy Energy Budget, J. Atmos. Sci., 48, 1972–1998, https://doi.org/10.1175/1520-0469(1991)048<1972:TLCOAC>2.0.CO;2, 1991.

Oruba, L., Lapeyre, G., and Rivière, G.: On the Poleward Motion of Midlatitude Cyclones in a Baroclinic Meandering Jet, J. Atmos. Sci., 70, 2629–2649, 2013.

Pearson, K. J., Shaffrey, L. C., Methven, J., and Hodges, K. I.: Can a climate model reproduce extreme regional precipitation events over England and Wales?, Quart. J. Roy. Meteor. Soc., 141, 1466–1472, https://doi.org/10.1002/qj.2428, 2015.

Petterssen, S. and Smebye, S. J.: On the development of extratropical cyclones, Quart. J. Roy. Meteor. Soc., 97, 457–482, 1971.

Phillips, T. J., Potter, G. L., Williamson, D. L., Cederwall, R. T., Boyle, J. S., Fiorino, M., Hnilo, J. J., Olson, J. G., Xie, S., and Yio, J. J.: Evaluating Parameterizations in General Circulation Models: Climate Simulation Meets Weather Prediction, Bull. Amer. Meteor. Soc., 85, 1903–1916, https://doi.org/10.1175/BAMS-85-12-1903, 2004.

Piriou, J.-M., Redelsperger, J.-L., Geleyn, J.-F., Lafore, J.-P., and Guichard, F.: An Approach for Convective Parameterization with Memory: Separating Microphysics and Transport in Grid-Scale Equations, J. Atmos. Sci., 64, 4127–4139, https://doi.org/10.1175/2007JAS2144.1, 2007.

Rivière, G. and Joly, A.: Role of the Low-Frequency Deformation Field on the Explosive Growth of Extratropical Cyclones at the Jet Exit. Part II: Baroclinic Critical Region, J. Atmos. Sci., 63, 1982–1995, https://doi.org/10.1175/JAS3729.1, 2006.

Rivière, G., Arbogast, P., Lapeyre, G., and Maynard, K.: A potential vorticity perspective on the motion of a mid-latitude winter storm, Geophys. Res. Lett., 39, L12 808, 2012.

Rivière, G., Arbogast, P., and Joly, A.: Eddy kinetic energy redistribution within windstorms Klaus and Friedhelm, Quart. J. Roy. Meteor. Soc., 141, 925–938, 2015.

Rochetin, N., Grandpeix, J.-Y., Rio, C., and Couvreux, F.: Deep Convection Triggering by Boundary Layer Thermals. Part II: Stochastic Triggering Parameterization for the LMDZ GCM, J. Atmos. Sci., 71, 515–538, https://doi.org/10.1175/JAS-D-12-0337.1, 2014.

Roehrig, R., Beau, I., Saint-Martin, D., Alias, A., Decharme, B., Guérémy, J.-F., Voldoire, A., Ahmat Younous, A.-L., Bazile, E., Belamari, S., Blein, S., Bouniol, D., Bouteloup, Y., Cattiaux, J., Chauvin, F., Chevallier, M., Colin, J., Douville, H., Marquet, P., Michou, M., Nabat, P., Oudar, T., Peyrillé, P., Piriou, J.-M., Salas y Melia, D., Séférian, R., and Sénési, S.: The CNRM global atmosphere model ARPEGE-Climat 6.3: description and evaluation, J. Atmos. Model Dev., p. e2020MS002075, https://doi.org/10.1029/2020MS002075, 2020.

Rysman, J.-F., Berthou, S., Claud, C., Drobinski, P., Chaboureau, J.-P., and Delanoë, J.: Potential of microwave observations for the evaluation of rainfall and convection in a regional climate model in the frame of HyMeX and MED-CORDEX, Climate Dyn., 51, 837–855, https://doi.org/10.1007/s00382-016-3203-7, 2018.

Sanders, F. and Gyakum, J. R.: Synoptic-Dynamic Climatology of the “Bomb”, Monthly Weather Review, 108, 1589–1606, https://doi.org/10.1175/1520-0493(1980)108<1589:SDCOT>2.0.CO;2, 1980.

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J. D., McTaggart-Cowan, R., Methven, J., Rivière, G., Ament, F., Boettcher, M., Bramberger, M., Cazenave, Q., Cotton, R., Crewell, S., Delanoë, J., Dörnbrack, A., Ehrlich, A., Ewald, F., Fix, A., Grams, C. M., Gray, S. L., Grob, H., Groß, S., Hagen, M., Harvey, B., Hirsch, L., Jacob, M., Kölling, T., Konow, H., Lemmerz, C., Lux, O., Magnusson, L., Mayer, B., Mech, M., Moore, R., Pelon, J., Quinting, J., Rahm, S., Rapp, M., Rautenhaus, M., Reitebuch, O., Reynolds, C. A., Sodemann, H., Spengler, T., Vaughan, G., Wendisch, M., Wirth, M., Witschas, B., Wolf, K., and Zinner, T.: The North Atlantic Waveguide and Downstream Impact Experiment, Bull. Amer. Meteor. Soc., 99, 1607–1637, https://doi.org/10.1175/BAMS-D-17-0003.1, 2018.

Seiler, C. and Zwiers, F. W.: How well do CMIP5 climate models reproduce explosive cyclones in the extratropics of the Northern Hemisphere?, Climate Dyn., 46, 1241–1256, https://doi.org/10.1007/s00382-015-2642-x, 2016.

Sexton, D. M. H., Karmalkar, A. V., Murphy, J. M., Williams, K. D., Boutle, I. A., Morcrette, C. J., Stirling, A. J., and Vosper, S. B.: Finding plausible and diverse variants of a climate model. Part I: establishing the relationship between errors at weather and climate time scales, Climate Dyn., 53, 989–1022, https://doi.org/10.1007/s00382-019-04625-3, 2019.

Sinclair, V. A., Rantanen, M., Haapanala, P., Räisänen, J., and Järvinen, H.: The characteristics and structure of extra-tropical cyclones in a warmer climate, Wea. Climate Dyn., 1, 1–25, https://doi.org/10.5194/wcd-1-1-2020, 2020.

Trzeciak, T. M., Knippertz, P., Pirret, J. S. R., and Williams, K. D.: Can we trust climate models to realistically represent severe European windstorms?, Climate Dyn., 46, 3431–3451, https://doi.org/10.1007/s00382-015-2777-9, 2016.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, J. Atmos. Model Dev., 11, 2177–2213, https://doi.org/10.1029/2019MS001683, 2019.

Wan, H., Rasch, P. J., Zhang, K., Qian, Y., Yan, H., and Zhao, C.: Short ensembles: an efficient method for discerning climate-relevant sensitivities in atmospheric general circulation models, Geosci. Model Dev., 7, 1961–1977, https://doi.org/10.5194/gmd-7-1961-2014, 2014.

Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., Jakob, C., Klein, S. A., Senior, C. A., and Williamson, D. L.: The Transpose-AMIP II Experiment and Its Application to the Understanding of Southern Ocean Cloud Biases in Climate Models, J. Climate, 26, 3258–3274, https://doi.org/10.1175/JCLI-D-12-00429.1, 2013.

Willison, J., Robinson, W. A., and Lackmann, G. M.: The Importance of Resolving Mesoscale Latent Heating in the North Atlantic Storm Track, J. Atmos. Sci., 70, 2234–2250, https://doi.org/10.1175/JAS-D-12-0226.1, 2013.

Zappa, G., Shaffrey, L. C., and Hodges, K. I.: The Ability of CMIP5 Models to Simulate North Atlantic Extratropical Cyclones, J. Climate, 26, 5379–5396, https://doi.org/10.1175/JCLI-D-12-00501.1, 2013a.

Zappa, G., Shaffrey, L. C., Hodges, K. I., Sansom, P. G., and Stephenson, D. B.: A Multimodel Assessment of Future Projections of North Atlantic and European Extratropical Cyclones in the CMIP5 Climate Models, J. Climate, 26, 5846–5862, https://doi.org/10.1175/JCLI-D-12-00573.1, 2013b.