Manuscript WCD-2021-38

"The impact of deep convection representation in a global atmospheric model on the warm conveyor belt and jet stream during NAWDEX IOP6"

Authors' response

Replies to the Editor and reviewers are in blue. Reviewers' comments are in black.

Dear Editor,

We would like to thank the two reviewers for their very helpful and interesting comments on the manuscript. Some additional information has been inserted in the revised version on the following points:

- The discussion on the difference between the two parametrization schemes has been extended (remark from reviewer 1).
- The need to present the three simulations has been further motivated in the introduction (remark from reviewer 2). Besides, since the discussion on the difference between the two parametrization schemes has been extended following remark from reviewer 1 we do think this is important to show the difference between the three runs throughout the paper.
- Quantification of the difference between observations and model simulations has been made by computing RMS errors (remark from reviewer 1). This is now indicated in the main text.
- Welch significance tests have been computed to estimate the significance of the difference in ascending motion between trajectories of the different runs. Figure 4 has been changed accordingly.

Please find below the point by point responses to the reviewers' comments. To find the modifications, lines in the revised version are indicated in parenthesis.

Reply to James Booth (referee 1)

General comments:

I think this is an interesting study. The work is well written and thorough. It shows a pathway through which parameterized convection impacts ascending air in extratropical cyclones, and it shows that this can impact the forecast. It also affirms previous work showing that deep convection is occurring within the warm sector of the cyclones.

The one potential component that the authors might consider adding to the study is a graphical analysis of changes in the forecast that impact people, e.g., including detail about how surface conditions (winds, temperature or precipitation) differed for the forecasts from the different configurations.

Differences in surface conditions between the three runs exist at 1-day and 2-day lead times but they are rather small and no specific interesting differences have been noticed. Differences in the minimum of the mean SLP do not go beyond 2 hPa at 1-day lead time (962.9 hPa in B85, 962.0 hPa in PCMT and 961.2 hPa in NoConv) or at 2-day lead time (959.5 hPa in B85, 959.8 hPa in PCMT and 958 hPa). The total precipitation does not differ much from one run to another. It is just the partition between resolved and convective precipitation which differs much more but this not surprising according to Martinez-Alvarado et al (2014) et Booth et al. (2018). Some of this information is now inserted in the revised paper (lines 238-240). Figure 1 has been changed to show the sea level pressure at 1-day lead time and the different tracks of the minimum sea level pressure during the 2-day runs are now shown. We just observe very slight differences between them.

At lines 465-466 you discuss the differences between PCMT and B85 being largely related to schemes' closures. I think this is a good point to make, and I think you might expand a bit on the discussion. One thing that I have started to consider (over the years) when thinking about these sorts of integrations is this:

In the runs with parameterized convection, it is likely the case that resolved convection still occurs some of the time.

Since the models all have the same resolution, I imagine that there are some times when both resolved and parameterized convection occur at the same time in the model integrations that have the large-scale convection parameterization turned on - perhaps in adjacent grid cells.

The PCMT integration shows similar isolated patches of heating as in NoConv (Figure 5 of the paper) suggesting that resolved convection occurs in the former run. In B85 such patches are less visible. So we expect that the co-occurrence of resolved and parametrization convection may more likely happen in PCMT than in B85.

I wonder if PCMT more closely resembles NoConv (e.g., Figure 3) because the trigger mechanism for PCMT to activate is stricter than that of B85. (I dont know that it is, I am just speculating). This would mean that in PCMT we actually see the result of a combination of resolved and parameterized convection, whereas in B85 we see more influence from parameterized convection because the scheme is more likely to fire off before any resolved convection can take place. This might be something that you consider mentioning, or if you disagree, perhaps add some discussion stating your case.

The humidity flux convergence used by the B85 closure has two components, one coming from the resolved scale horizontal fluxes and the other from turbulent fluxes. The resolved-scale fluxes are expected to be strong in presence of synoptic-scale forcing as in the inflow regions of the warm conveyor belts. So it is not surprising to get more triggered convection in such cases with strong synoptic-scale forcing when the closure is based on moisture convergence rather than when it is based on CAPE (PCMT convection scheme). In pure convective situations when there is no significant synoptic-scale forcing as for instance during summertime convection over land CAPE is expected to get higher values and in that case convection is less sensitive to humidity convergence (Yano et al. 2013). The discussion has been extended to insert this information (lines 495-501 of the revised paper).

Two other questions/comments, just for your consideration:

(1) For the sake of the objectives of this study, it is good that the circulation of the storms is not very different despite the different convection schemes. Presumably, the circulation in the tropics is much more sensitive to the switching out of the convection schemes. I am guessing that the short length of the integrations prohibits those differences in the tropics from impacting the storm dynamics significantly? Perhaps a comment on this would be useful to the reader, or maybe I am missing some detail about the model configurations.

Influence of tropical convection on midlatitude is expected to occur via large-scale Rossby wave trains at medium range time scales (a week or beyond) (Hoskins and Karoly, 1981; Stan et al. 2017). Therefore, since the paper is focused on 1 to 2-day lead times, there is no reason to get an influence from the tropics. We have looked at geopotential maps over the whole Northern Hemisphere during the two-day integrations and no specific patterns emerge from the tropics to the midlatitudes.

(2) In multi-year integrations of a GCM, such a study is much more difficult. We tried it and found that when we changed the convection scheme we also had to re-tune the model. The changes to the tuning included changes to the microphysics which affected the storm dynamics.

Thank you for providing this information.

Minor comments:

Lines 194-199: Here you discuss the methods of calculating the diabatic heating. Method 1: taking the derivatives of theta with respect to space and time. Method 2: using the actual diabatic heating tendencies created by the model. You state that method 1 is more accurate. This result surprises me, given that method two seems to have the exact information that you seek, as generated by the model. How do you calculate the accuracy of the two methods? Perhaps add a bit more explanation here - do you attribute the issue to the regridding, or are there other factors?

Thank you for pointing out this lack of precision. The Lagrangian trajectories are computed using the wind components available over a $0.5^{\circ} \times 0.5^{\circ}$ horizontal grid, with 50-hPa vertical grid spacing and every 15 minutes. Variations in theta along trajectories were found to be very close to the integration of the heating following method 1 based on finite differences, which is not so surprising as the finite differences are computed with the same spatial grid and the same time resolution as for the computation of the trajectories. It shows that the two computations are self consistent. In method 2, the different diabatic tendencies of the temperature, even though they are generated by the model, come from another post processing. They are first available in the stretched/rotated Gaussian reduced grid of the model and an offline interpolation was then made to compute the tendencies in the $0.5^{\circ} \times 0.5^{\circ}$ horizontal grid and 50-hPa vertical grid. The interpolation done to get diabatic temperature tendencies is not the same as the one providing the other variables (wind, temperature). Even though the two methods lead to similar heating patterns (panels a and b of figures S1 and S2 of the paper) it is not fully clear to us the reasons of some discrepancies between the results of methods 1 and 2. More information on the two methods is now provided (lines 209-214 of the revised paper).

Line 231: For figure 1, the SLP contours look identical in the 3 integrations, is that the case? Perhaps my eyes deceive me and the SLP differ for the 3 panels?

In the initially submitted paper, the SLP contours were identical because it was shown at the initial time which is the same for the 3 integrations. In the revised version, the SLP contours are now shown at 1-day lead time and are not identical even though they are very close to each other. We have also added the trajectories of the minimum SLP during the 2-day integrations and slight differences in the trajectories can be observed as well.

Line 280: Just a comment: I am a bit surprised that the heating rates are positive even at the levels close to the surface, I expected rain evaporation to lead to negative values at the lowest levels. Perhaps that is only true in the stratiform precipitation regions.

We think this is not a question of convective vs stratiform precipitation because the total precipitation is dominated by the stratiform/resolved component in the two integrations including parametrized convection. Evaporative cooling exists below the heating area but its values are small compared to the positive heating rates and more localized in space. This is the reason why we do not necessarily see the evaporative cooling when the heating rates are spatially averaged. In Figure 1 of the present document, we show a modified version of Figure 5 of the paper with a cross section of the heating rates made at 45°N without latitudinal average. Areas of negative heating rates below the positive ones exist but have less amplitude and exhibit more localized patterns. These areas are more visible for PCMT and NoConv than B85.

Lines 284, 287: In these sentences, you refer to general areas as warm conveyor belt regions. So, is the terminology that you are using specifying a difference between WCB regions and WCB trajectories? I think this is a reasonable approach. But it might be useful for readers if you add some clarification in section 2.2.1, to specify that you are not exclusively using the word WCB to refer to a location where the air ascends by x amount in time t. Or maybe you do prefer this approach, in which case the phrasing at lines 284 and 287 is a bit confusing.

Thank you for pointing this inconsistency. The usual picture is that a warm conveyor belt is located ahead of the cold front, at least its inflow region. But it is true that we find numerous trajectories satisfying the WCB criterion (300 hPa ascent in 24 hours) not only ahead of the cold front but also along the bent-back warm front and at all levels. So to be more precise, we suppressed the wording "WCB region" and only use the terminology "WCB" when it concerns trajectories satisfying the criterion on ascents (see modifications lines 302-303 and 308).

Lines 297-298: regarding these poleward locations with more diabatic heating in the runs with convection on, this is perhaps consistent with these models having more sustained ascent (as you mention on line 274). I wonder if might take this a step farther and say that the parameterized convection leads to a larger amount of poleward flux of heat? You don't necessarily need to investigate this - just an idea.

We have looked at the poleward heat fluxes and the three integrations do not differ much from each other. The patterns are very similar and the peak values do not differ by more than 10% during the 2-day integrations. This is not surprising since the surface parameters of the depression do not differ from each other as said above.

Lines 358- 370: I cannot tell from this paragraph whether you are of the opinion that B85 or NoConv more closely resembles ERA5 at this particular snapshot in time. For me, the pattern of PV at 25W by 60N in NoConv looks a bit more like ERA5 because it has a sharper meridional gradient, whereas B85 has a more of a NE to SW tilt in that region.

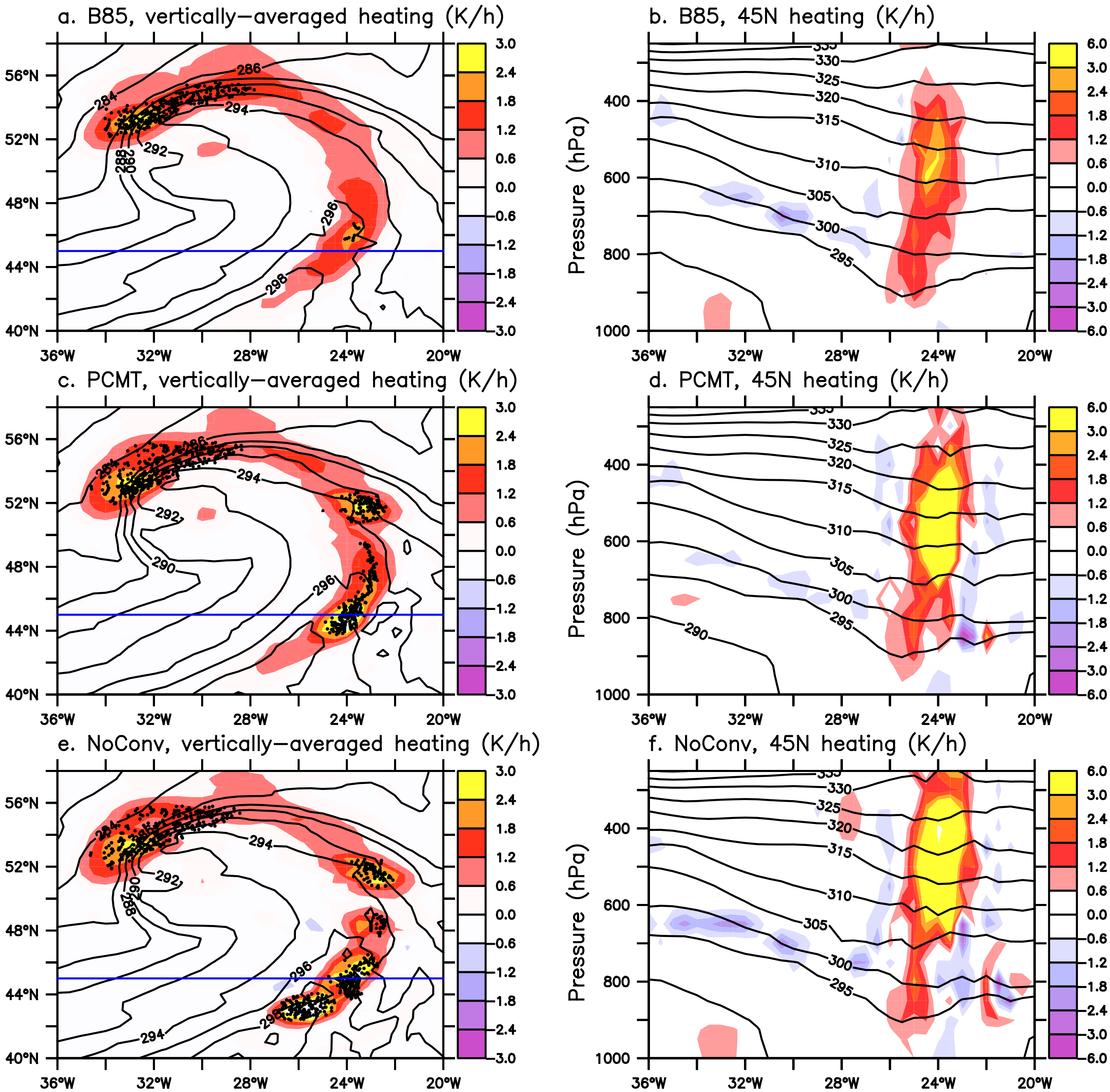

Figure 1. (Left panels) vertically-averaged heating rate between 300 and 800 hPa (shadings; units: K h^{-1}), potential temperature at 850 hPa (int: 2 K) and WCB trajectories satisfying 100 hPa ascent in 2 hours at 21 UTC 1 October for (a) B85, (c) PCMT and (e) NoConv simulations. (Right panels) heating rate at 45°N (shadings; units: K h^{-1}) and potential temperature (contours; int: 5 K) at 21 UTC 1 October for (b) B85, (d) PCMT and (f) NoConv simulations.

We do see what you mean. However, this region is not our region of interest and further west ERA5 and B85 become closer to each other and further away from NoConv (see the SE to NW oriented PV isolines in ERA5 and B85 and not in NoConv in the area 60°N-64°N and 35°W-25°W). Since more drastic differences in PV appear north of 64°N over the Greenland eastern coast, we do not comment subtle differences south of 64°N.

In lines 365-368, you discuss the 0.5-1.5 PVU ribbon in B85. As I see it, that feature is spatially continuous in B85 and therefore more distinct as compared to NoConv, and it is non-existent in ERA5. So, this also could be viewed as a region in which NoConv looks more like ERA5, perhaps.

Even though the 0.5-1.5 PVU ribbon is not as visible in ERA5 as in B85, it is there and roughly located at the same place as confirmed by the cross section of Figure 2 of the present document (see the descent of higher PV near 38°W, 300 hPa in ERA5 and B85). This blob of higher PV is displaced to the east in NoConv near 34°W. Moreover, what matters most for the location of the wind speed maximum is the PV gradient. Looking at Figure 8 of the paper, east of the 0.5-1.5 PVU ribbon the PV values decrease to negative values in both B85 and ERA5 datasets off the east coast of Greenland which is not the case in NoConv. Therefore the position of the secondary jet is over the east coast in both ERA5 and B85 while this region is marked by a minimum wind speed in NoConv (see Figure 9 of the paper). But since the 0.5-1.5 PVU ribbon in B85 has higher PV values than in ERA5 and ECMWF-IFS analysis the PV gradient is also stronger which is consistent with the fact that the secondary jet is stronger too in B85 than in ERA5 or ECMWF-IFS analysis. In other words, compared to ERA5, the secondary jet in B85 is located at the same place but is too strong. The one in NoConv has a different location. The text has been modified in the revised version to make clearer the location of the PV gradient in the different datasets (lines 394-400).

Given the spatial resolution of the model however, one could question whether ERA5 is necessarily more correct, but if we assume that it is, then I feel that the issue of which is more like ERA5 might a matter of opinion. All of which is to say, a bit more discussion here about which model integration you find to be most like reality would be useful I think. Or you can simply state what I said: it is a toss-up.

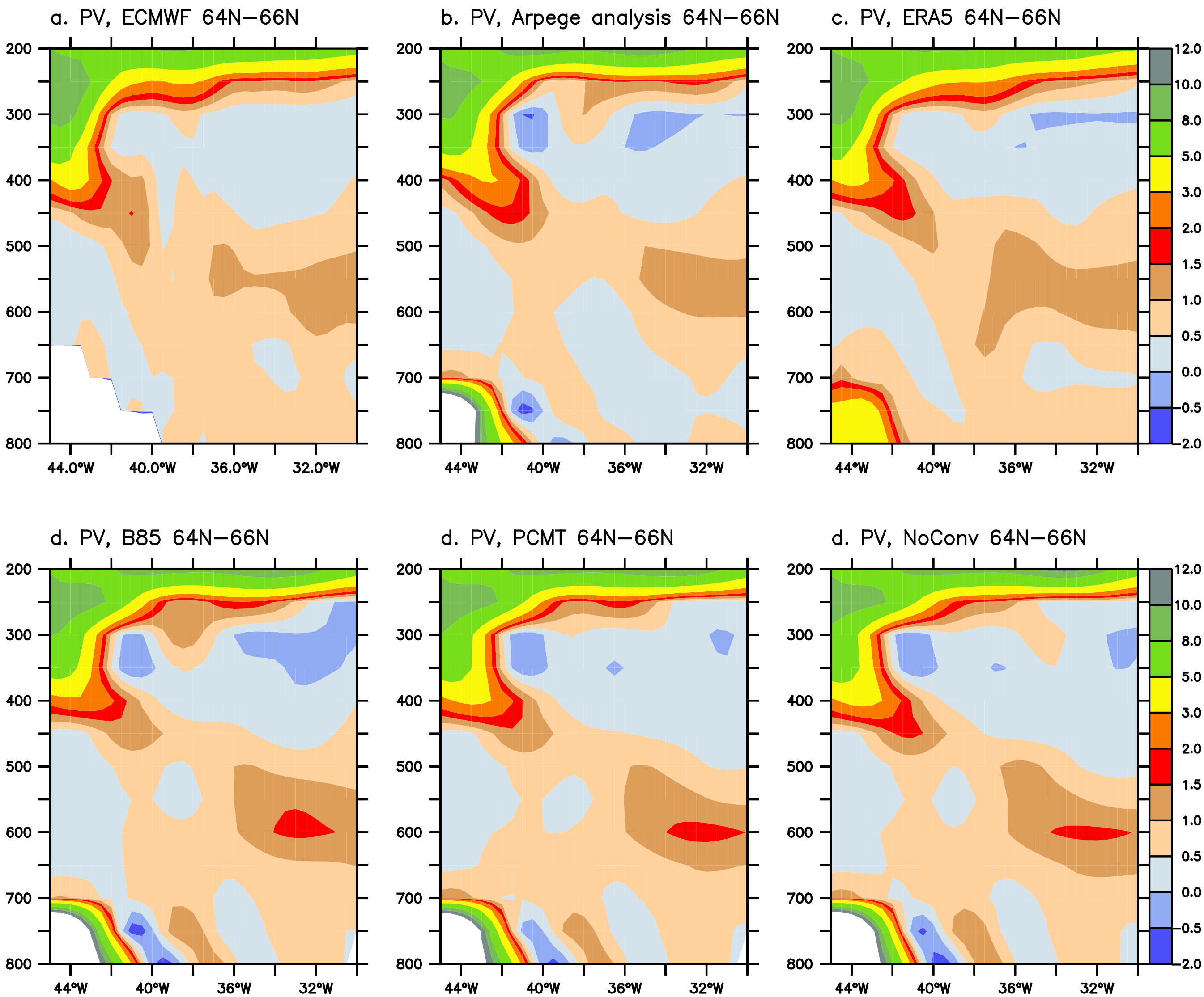

Figure 2. Latitudinally-averaged (64°N-66°N) PV for (a) ECMWF-IFS operational analysis, (b) ARPEGE operational analysis, (c) ERA5 reanalysis, (d) B85 run, (e) PCMT run and (f) NoConv run.

Our objective is to show 3 different (re)-analysis datasets because none of them can be considered as the reference. Despite important differences in the PV values shown in Fig.8, the three datasets agree about the position of the secondary jet. This is precisely that common feature which makes us confident to study and compare the three runs. Moreover, the airborne observations that are completely independent of the (re)-analysis datasets confirm the position of the secondary jet and the fact that it is better located in B85 than in PCMT or NoConv, even though it is too strong.

Lines 374-391: I appreciate that you point out times with B85 performs better (Fig 9), but other times when it performs worse (Fig 10)

Thank you for the comment.

Section 4.2: Figures 11 and 12 do well to compare with observations – always a challenge. For me, it looks like the 3 integrations are all more similar to themselves than they are to observations. But perhaps not? I wonder if you can quantifiably check this with an RMS of the difference plots or something of that nature.

Figure 3 of the present document shows the difference between B85 and PCMT integrations (panel c) and is compared to the difference between each run and the observations (panels a and b). It clearly shows that in the lower troposphere the integrations are more similar to themselves than they are to observations. In contrast, in the upper troposphere, differences between two integrations are of the same order of magnitude than differences between individual integrations and the observations. The RMS associated with the differences B85-OBS, PCMT-OBS and PCMT-B85 in regions of radar observations are 3.44, 3.75 and 1.97 m s^{-1} respectively when all vertical levels are taken into account. If we consider the upper troposphere only, it leads to 4.36, 5.00 and 3.125 m s^{-1}. Doing the same analysis but with in-situ airborne measurements, the RMS errors are 4.27 m s^{-1} for B85 and 5.43 m s^{-1} for PCMT. Therefore, in the upper troposphere, the improvement of B85 with respect to PCMT is of the order of 10% to 20% in RMSE. This information is provided in the revised version (lines 439-444).

Line 485: If you don't mention it here, perhaps the in progress Wimmer et al paper will mention the idealized work of Boutle et al. 2011. Their Figure 4 tells an interesting story about convection, whether it is seen in the real world is to be determined.

Boutle IA, Belcher SE, Plant RS. 2011. Moisture transport in midlatitude cyclones. Q. J. R. Meteorol. Soc. 137: 360–373. DOI:10.1002/qj.783

Thank you for pointing out this reference. It will be considered for the introduction of our companion paper. However their analysis concerns shallow convection while our study is focused on the sensitivity to deep convection. The shallow convection schemes are the same in the 3 integrations shown in the present paper.

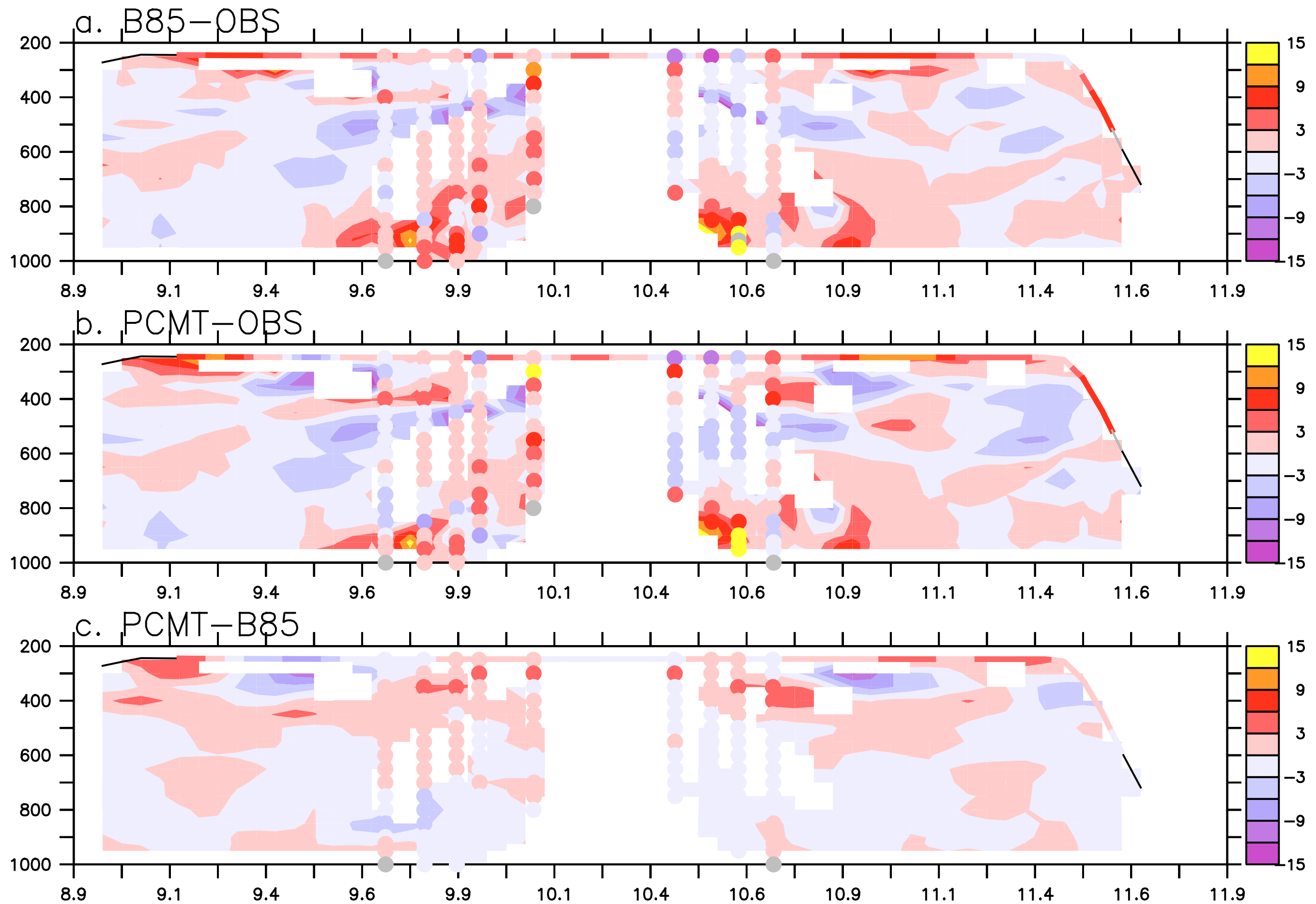

Figure 3. Difference in wind speed (units: m s^{-1}) in regions of observations: (a) B85-OBS, (b) PCMT-OBS and (c) PCMT-B85. The wind observations are composed of dropsondes, airborne in-situ measurements and Doppler radar measurements along the SAFIRE Falcon track.

Typos/technical issues:

Lines 93-94, you write: “This large-scale cyclone participated in the formation of a blocking over Scandinavia ...” The wording sounds strange to me. Should blocking be replaced by block? Or, leave the word blocking, but remove “a” that precedes it.

Done

Lines 127-128, you write: “Surface oceanic fluxes are represented by the Belamari (2005)’s scheme ...” Remove the word the, or remove the apostrophe and the “s”.

Done

Lines 286-287, you write: “However, large differences between the three runs appear in the heating rates along the WCBs and not near the bent-back warm front.” Is the word “not” supposed to be in this sentence? If so, could you instead refer to the exact location of interest instead of saying not near the bent-back warm front?

The sentence is now (lines 302-303): "However, the largest differences between the three runs appear in the heating rates ahead of the cold front (26°W-22°W; 42°N-52°N) and not near the bent-back warm front (34°W-30°W; 52°N-56°N).

Lines 338-339, you write: “Besides as they go further away from the WCB outflow region winds become more zonal and align more with the isentropic slopes as seen by comparing the orientation of the vectors with the slope of the 315 K isentropic surface near 36°W, 300 hPa in Fig.7a.” This sentence is a bit difficult to understand, and starting a sentence with Besides might be a bit too colloquial. But I leave to you to decide if you want to change it or leave it.

The sentence is now (lines 352-353): "West of the main WCB outflow region, i.e west of 30°W in the 300-400 hPa layer, winds are westward and upward. Moving to the west they become more and more horizontal and align more with the isentropic slopes (compare the orientation of the vectors with the slope of the 315 K isentropic surface in Fig. 7a)."

Lines 340-341: Verb tense issue, probably most easily fixed by changing “advection term is” to “advection terms are”

Done

Reply to referee 2

General comments

This paper presents a comparison between three simulations of a mid-latitude storm using two different convection parametrisations and no parametrisation at all. The results are useful to understand the impact of the parametrisation formulation on the evolution of the upper tropospheric flow including the jet stream. While there is a school of thought that advocates the increase of resolution in numerical simulations to avoid the need of convective parametrisations altogether, the topic remains relevant as a more efficient use of resources might still come from improving lower resolution models by improving parametrisations (including convection). Therefore I believe this paper makes an important contribution to a topic that remains largely unexplored.

The article is clearly in scope for Weather and Climate Dynamics, well-structured and well-written. However, after reading the paper I was left wondering whether the inclusion of two parametrisations was necessary at all. From the results, it is clear that the main differences arise between B85 and NoConv and that PCMT exhibits an intermediate behaviour. Thus, I wonder whether a better approach would be to concentrate on the first two to highlight these differences without the distracting element from PCMT. Otherwise, I'd encourage the authors to motivate further the need to present the three simulations throughout. A second main comment is related to the clarification of large-scale heating and other contributions. I add more details in this regard (among others) in the specific comments below.

Given these comments that in my opinion require attention, I can recommend the paper for publication in WCD, but only after these comments have been addressed.

Thank you for all the comments. Three main reasons explain why we keep showing the results of the three integrations:

- Even though PCMT has an intermediate behavior, it is interesting to quantify how far it is from the two integrations. When we started the study, we thought it will be closer to B85 than to NoConv and finally in most plots it appears to induce a mid-way response between the two other runs.

– At the end of the paper (conclusion section), we discuss the differences between PCMT and B85 using all the members of the EPS. We found the closure makes the largest difference between the two integrations. Hence, this brings new information compared to a simple comparison between B85 and NoConv.

– Providing such a comparison between B85 and PCMT is important for modellers because the two schemes are currently used in operations at Meteo-France, and for international model intercomparisons. The PCMT scheme is used in the CMIP6 version of the ARPEGE climate model (Roehrig et al. 2020), the B85 scheme is used for NWP ARPEGE operational forecasts. Such ongoing discussions on the best choice for deep convection parametrization show the need to better identify key differences between the different components of these two schemes. It is tricky to find the optimal scheme as the schemes have different skills depending on the type of weather systems and the regions. Performance of convection schemes are usually tested in classical convective situations (tropical convection, diurnal cycle over land) but in our opinion it is also important to look at them in situations where convective activity is embedded in midlatitude synoptic-scale dynamics as it exerts an influence on the predictability of the jet stream as shown here.

The introduction has been modified to better motivate the comparison between the two schemes (lines 107-112 of the revised paper).

Specific comments

L194-203: Please, give more details on the centred finite-difference scheme used. For example, were these differences computed between two output times or online as the model was run? Please also add more detail on the meaning of large-scale heating (also relevant for L14 in the abstract) and how this is different from parametrised heating. Looking at the Supplementary figures I get the impression that what is referred to as large-scale heating is the heating due to the large-scale cloud parametrisation (ormi-crophysics). Is this correct? It would be good to clarify this and the other terms in the paper.

Finite differences are computed offline with wind and temperature datasets over the $0.5° \times 0.5°$ horizontal grid with 50-hPa vertical grid spacing and a frequency of 15 minutes. This information is now explicitly given (lines 201-202).
Large-scale heating means the sensible and latent heating due to large-scale cloud microphysics. This is to be contrasted with the sensible and latent heating from the parametrized convection. In the revised version "large-scale heating" has been replaced by "resolved sensible and latent heating".

L244-245: Regarding trajectories, sometimes differences are better appreciated if the trajectories are plotted with respect to time of maximum ascent so that they are all in similar ascending stages. Plotting as a function of actual time means that trajectories at different ascent stages are averaged together. For example, the enhanced increase and decrease of PV might be due to differences in timing.

The composites of PV have been computed with respect to time of maximum ascent (see figure 4 of the present document). It shows that the decrease in PV is stronger in PCMT and NoConv than B85 during the first 12 hours but after 12 hours, the PV is still decreasing in B85 and not anymore in the other two runs. One day after maximum ascent, the PV is lower in B85 than in NoConv and PCMT. This is entirely consistent with the result of Figure 2c showing the PV composites as function of actual time. To avoid too many figures and since it does not bring new results, the PV composites as function of time of maximum ascent are not shown in the revised paper.

L280-282: I'm not sure I agree with the comment on the large difference between the three runs. If differences between the three is to be maximised, I would have chosen 3 UTC 2 October. Another period that seems interesting is 9-18 UTC 2 October during which only NoConv exhibits strong ascent. I think this might be an example where presenting only B85 and Noconv would lead to clearer presentation of results (see General comments).

The same plot as figure 5 of the paper has been made at 03 UTC 2 October to look at strong ascents (see Figure 5 of the present document). The results are not so different and we get the same overall picture: main differences in the heating ahead of the cold front, more heterogeneous pattern for NoConv and more homogeneous for B85. PCMT has a more intermediate behavior but is here slightly closer to B85 than NoConv in terms

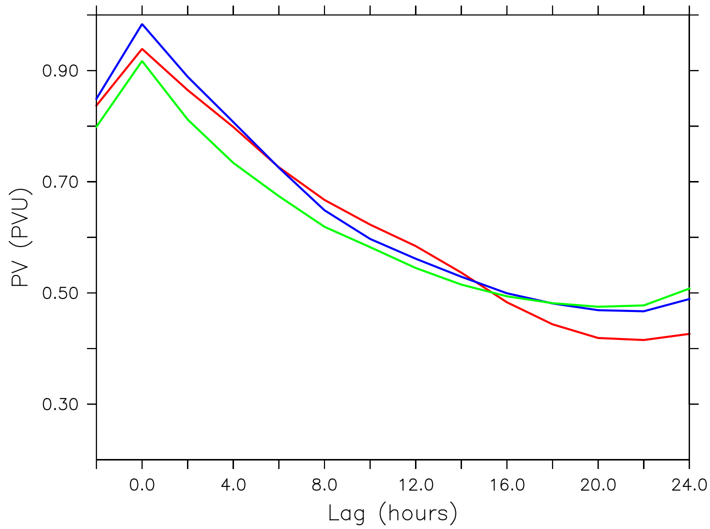

Figure 4. Time lag PV composite over WCB trajectories for B85 (red), PCMT (blue) and NoConv (green) simulations. The zero lag corresponds to the time of maximum pressure difference.

of the magnitude of the heating (panels b, d and f).

L316-318: Can the heating decomposition be explained further? Does the finite-difference heating include the parametrised heating? See also the comment to L194-203.

The heating based on finite differences reflects the Lagrangian rate of change of potential temperature and thus includes all diabatic heating processes. However, it does not provide the decomposition into distinct processes. We obtain the decomposition following another approach based on retrievals of the individual temperature tendencies from the model. As mentioned above, the two methods to get the total heating give similar patterns (see panels a and b of figures S1 and S2) but the values may largely differ in some regions. More information on the two methods and the potential reasons for their discrepancies have been provided in the revised version (lines 209-214).

Technical corrections

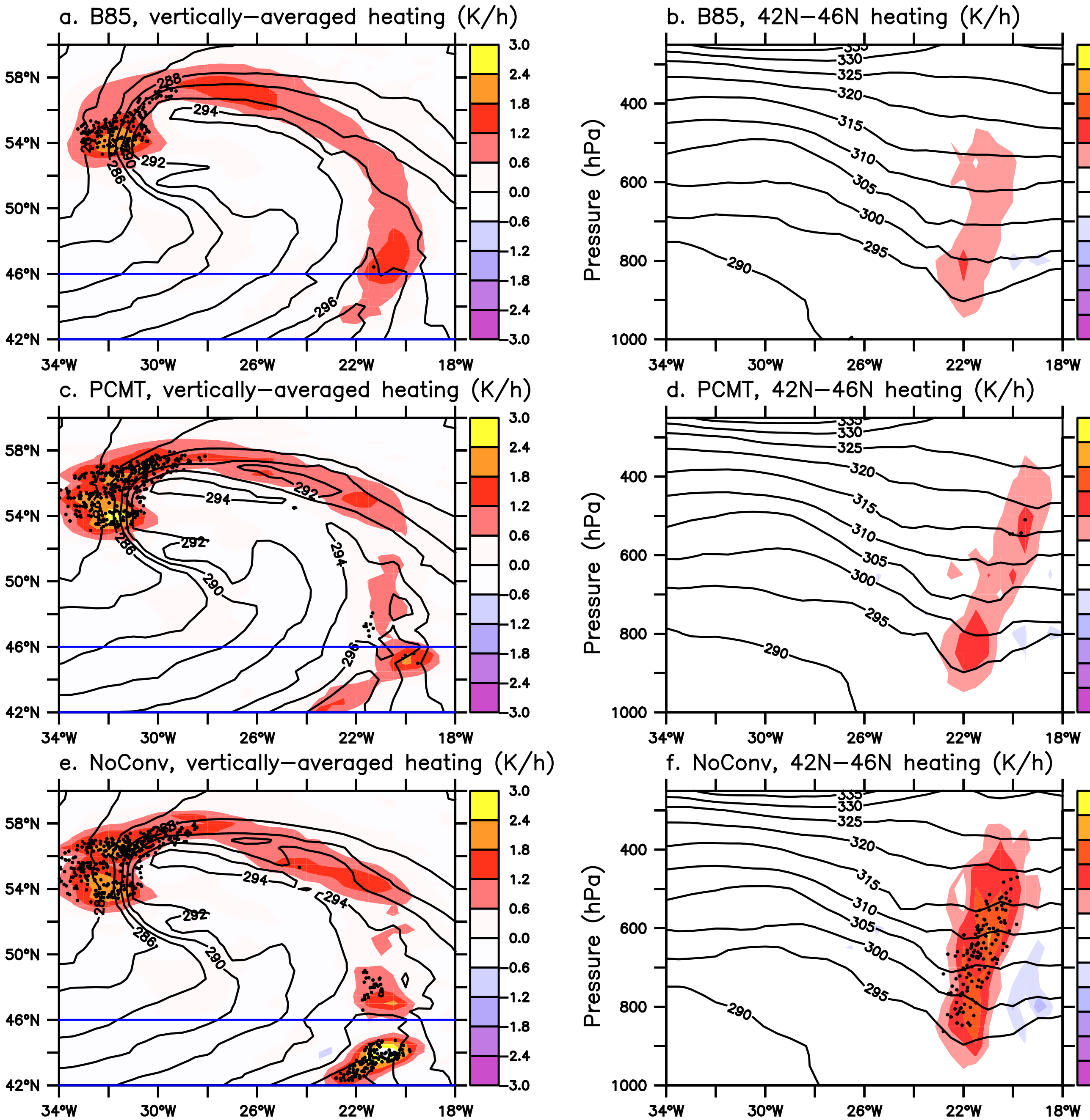

Figure 5. Same as figure 5 of the paper but at 03 UTC 2 October.

L18-19: As written, the sentence could suggest that B85 is realistic for longer than the other simulations. A simple reshuffling of words would make it clearer: "... that simulation becomes less realistic than the other ones at forecast ranges beyond 1.5 days".

Done

L72: Change ‘resolved’ for ‘resolve’.

Done

L240: ‘...slightly higher mean temperature...’ How statistically significant is this result?

A Welch test has been applied to estimate the significance of the difference between two runs. Since the datasets are correlated (the variables averaged along two adjacent WCB trajectories will have similar values), the lag-1 autocorrelation coefficient has been computed and has been used to get the effective number of trajectories for each run (see Wilks "statistical methods in the atmospheric sciences"). In terms of mean potential temperature (Figure 2b), only the difference between B85 and PCMT is statistically significant between 12 UTC 2 October and 12 UTC 3 October at 99%. In Figure 4 of the revised paper, time lags during which the vertical displacements of the three runs are significantly different are now shown by thick lines.

L261-262: I can clearly see the more important tightening between c and a,b, but I think this is less clear between a and b.

Between 00 UTC 2 Oct and 00 UTC 3 Oct the difference between B85 and PCMT is well visible by comparing the 0.2 and 0.4 K h^{-1} contours. Since differences in the vertical gradient of the heating are shown via the PV tendency term (shadings) and are clear in the upper troposphere we think it is not crucial to be precise on the vertical spacing between the different heating contours. Therefore no change has been made in that part of the text.

L263-264: Clarify that the ascent considered to classify and count the trajectories is instantaneous ascent. Is it centred at the time indicated?

Yes we consider instantaneous ascent centred at the time indicated. The sentence is now (lines 280-281): "The number of trajectories ascending by more than 100 hPa, 50 hPa and 25 hPa in 2 hours between $T - 1$h and $T + 1$h are indicated as function of time T in Fig. 2d."

L267-269: Is the dip in the number of moderately ascending trajectories in B85 between 1-3 UTC 3 October as important as the enhancement at 12 UTC 2 October? In my opinion the number of trajectories in B85 is consistently lower than in the other two simulations.

We agree with the reviewer. But now that the plot is made in terms of percentage of trajectories and not in number of trajectories the enhancement at 12 UTC 2 October for B85 becomes larger. The sentence is now (lines 283-285): "B85 has more such trajectories in proportion than the other two runs from 00 UTC to 18 UTC 2 October. Beyond 18 UTC 2 October, the proportions of moderately ascending trajectories are more close to each other between the three runs."

L300: '...significantly larger for B85...' The term 'significant' is usually restricted to denote statistical significance. Has this been tested in this case? Is there any way of adding error bars to the curves in Figure 2d? It might be worth normalising by number of trajectories as well (see comment to Figure 2d).

The word "significantly" has been suppressed here but some significance tests has been applied to other plots (see above). For that particular figure 2d the number of trajectories for each category has been normalized by the total number.

L332-335: I don't fully understand the explanation on the shift between the negative PV area and the trajectory positions, as the trajectories would include the advective effects by the non-divergent wind. Perhaps starting trajectories from the negative PV regions would be useful to clarify these results.

First note that a typo appeared in the initially submitted paper where "non divergent wind" needed to be replaced by "divergent wind". Let us come back to our explanation: a region of negative PV is not necessarily a region reached by WCB trajectories. Imagine the tropopause is in a region of westward divergent wind with stratospheric large PV to the west and tropospheric weak PV to the east as in Fig7 of the paper the advection term by the divergent wind will be mainly negative. Therefore a region not reached by WCB trajectories can have more negative PV than a region reached by WCB trajectories. As mentioned in Steinfeld and Pfahl (2019) when they comment their Fig. 9 showing a blocking onset composite, "the negative PV advection by the divergent wind on the western flank leads to a westward (and poleward) expansion of the blocking anomaly".

This is what we see in Figs. 7b,d,f of our paper for all runs. However, in B85, the divergent wind (here estimated from the ageostrophic wind) being more intense the negative PV advection is stronger than in the other two runs. The paragraph has been modified to make easier the understanding of our explanation (lines 350-356).

L360: I think ‘boundary’ would be a better term than ‘limit’ in this case.
Done

L423: To make the conclusions self-contained, give more details on the cyclone here (name, dates).
Done

L427: To make the conclusions self-contained and clearer, it would be worth including the questions here as well.
Done

L428: ‘When parameterized deep convection is inactive...’ It might be worth rewriting. As it is the phrase could mean ‘When a convection parametrisation is present, but is not being triggered...’.
The sentence is now starting as: "When parameterized deep convection scheme is turned off ..."

L440: ‘large-scale heating’ This phrase might need to be modified in response to other comments and heating definitions (see e.g. comment to L194-203).
Large-scale heating has been replaced by resolved heating (see e.g., line 334) and corresponds to the sum of sensible and latent heating coming from the large-scale cloud parametrization.

L454: Delete ‘in B85’ after ‘more realistic’.
Done

L456-458: These two sentences are slightly confusing and appear contradictory: Are the more active dynamics realistic, but too strong? Please rewrite.

The last two sentences of the paragraph have been suppressed because we were saying the same thing as in the previous sentences but this was less understandable.

L493-496: The last paragraph seems incomplete in the sense that it only gives very partial information on the contents of a future paper. In my opinion, it could be omitted.

We do think it is important to refer to the companion paper if the reader wants to know more about the difference between the runs in the WCB ascending region, which is not the focus of the present paper. The present paper is focused on WCB outflow region and upper-tropospheric effects. The paragraph has been modified to insert more precise wording.

Figure 2 and others: Use of green and red might not be ideal for colour-blind readers.

The green being bright compared to the red, the black and white print leads to clear differences between the two curves. Moreover, a test has been made with a colour-blind colleague who did not find difficulty to see the differences.

Figure 2d: While the total number of trajectories is similar, I wonder if it would be a fairer comparison to present this as a proportion of total number.

Each number has been normalized by the total number of WCB trajectories and the figure is now shown as a percentage relative to the total number.

Figure 3: Do the bold solid and dashed lines need to be coloured? Black or similar would make them clearer.

We think it makes easier the reading of the figure since we use exactly the same colours to define the three runs in the different figures: red for B85, blue for PCMT and green for NoConv.

Figure 6: One difficulty when comparing forward trajectories is that the number of trajectories at each location will depend on previous stages in their evolution. Perhaps it would be fairer to start the same trajectories at the point of interest and see their evolution towards and from that point in time.

We think forward trajectories are relevant here because our objective is to show differences in the number of WCB trajectories as a function of the different regions. It is true that the number depends on previous stages in trajectories evolution but this is what we want to highlight with this figure.

Figure 8: What do the white areas represent? Are they just regions off the colour scale? The colour scale has been changed to avoid white areas.

Figure 11: You could make the comparison between panels easier by adding lines indicating the span of panel (b) in panels (c-e).

The span of panel (b) is precisely chosen in order to have co-location in space of all the datasets. The time indicated in the x-abscissa corresponds to the local time when each aircraft was above a given region. Since the two aircraft followed each other with a 20-minutes lag (DLR Falcon being first in the zone), we do see this lag between panels (a) and (b). The model outputs are shown using the x-abscissa of panel (a). This information is now provided in the caption of the figure.