

ACP review:

This is a point to point response to the **Ref#3** comments.
Referee comments are in italic.

At first, we would like to thank the referee for all of his comments, ideas and suggestions. The manuscript has been greatly improved thanks to the time he devoted to our manuscript.

Major Concerns:

1) The results of the paper are based on a number of 30 day simulations with the MUA (Middle and Upper Atmosphere) Model. In the real atmosphere, there is substantial natural variability, and results based on a single 30 day snap shot would likely be meaningless in a statistical sense. I suspect that this model does not have much natural variability – otherwise the authors would not be able to conclude much from such short runs – but that is unclear in the current paper, which provides little insight into the background flow and no discussion of the statistics. To remedy this situation, I recommend first establishing the quality of the model, better characterizing its January climatology (for example, showing the zonal mean wind as a function of pressure and height) along with the variability (for example, the variance of the zonal mean winds). Does this model vary much at all, or is it essentially steady, seeking only to capture the climatological mean circulation. It would also be good to show the overall impact of the gravity wave drag (GWD) scheme. A panel/overlay showing the zonal mean drag as a function of pressure and latitude might help, too, giving us a better sense of the background gravity wave driving.

Answer: We absolutely agree and therefore we made major changes in the manuscript:

In section 2.1, the model, its variability and the procedure of creation of the sensitivity simulations are described in more detail. Information about the reference simulation is given in Fig.1. In Figure 3 and at the start of the Results section we now establish the time scale of the response and statistics is now included everywhere in the manuscript, where mean anomalies or differences are dealt with. Standard deviations of the zonal means in the reference simulation in the 30 days are small as described in the text (P4L10-15).

Then, how sensitive are these quantities to the forcing? At the top of page 4 the authors suggest they force planetary waves 1, 2, and 3 from ERA-I reanalysis at 1000 hPa. Do they mean climatological waves (based on what period)? What would happen if you took the waves from a given year? My concern here is that the authors need to establish that their results are robust, and wouldn't change dramatically if the climatology is altered. Varying the lower boundary would allow them to sample the natural variability of the real world; in his case they would need to run a number of simulations for each case, and could assess the statistical robustness of their conclusions.

Answer: The waves are extracted from decadal mean ERAI reanalysis and are stationary only (P3L35). The words “decadal monthly mean” have been included

(P3L30) to point that out. The model reaches a steady state after the spin-up (Fig. 3A).

More general answer from the AC1:

We agree with this comment. The overall response will be different for different background conditions. But there is a question, how reasonable it is to compare the responses of different background states to the same GWD. Naturally, each different background state means different GW sourcing and propagation conditions that should result to a different GWD. With our artificial enhancement approach we are not able to reflect this. Instead, in our simulations we have chosen a different approach. For mean January conditions we inject GWD settings stemming from our estimation based on observations (and from uncertainties especially in its meridional component) and from the mean climatology in the EA/NP region. In our paper we wanted to shed the light on the acting mechanisms and patterns of local response (e.g. build up of a positive geopotential anomaly upstream from the GWD, obstacle analogy..) that we assume will remain valid also for slightly different conditions.

Actually, we have also made simulations with slightly different background climatology and the results were almost similar (absolute agreement in the pattern, only the magnitude was slightly different in some places). We would not like to show the figures in the revised manuscript, because we think that it would have little additional value, since in our manuscript we are not exactly concerned with the precise magnitude of the response.

And finally, all figures need to acknowledge the statistics. I don't mean to be the curmudgeon who rants that a result without an error bound isn't a scientific result, but you do need to either estimate the statistical certainty, or explain that everything that is shown is robust, given the lack of variability in the model.

Answer: We absolutely agree and thank the referee for pointing out this. In a revised manuscript all of the mean (from day 7 to 30) differences and anomalies are overlaid with stippled areas of significance exceeding the 5% level (p-value estimated using the t-test). Except the new Fig. 9, where the standard deviation is dotted.

2) Following up on my first concern, it is unclear to me what these experiments are seeking to represent. Many figures (e.g. 2, 4, etc.) show the 30 day mean, which initially suggested to me that the goal was to demonstrate the steady response to the wave driving (which I presumed had occurred over this time scale). But it was not until the Fig. 5 that I realized that the response had clearly not converged over this period!

Gravity waves in the real world tend to be episodic and highly intermittent, so that the short term response is highly relevant. But if the goal is to capture the short term response, then I think the paper needs to focus on this from the start, and establish the appropriate time scale early on. This could be done, for example, by showing a Hovmoller diagram of some key quantities, such as the zonal mean wind at 6 hPa (or another key level) as a function of latitude and time, along with the evolution of the evolution of the key zonal harmonics (as in Fig. 5), but again, plotted as a function of latitude and time. The goal would be to show that the key change(s) occur on a timescale of X days (where X is with hope < 30 days!),

establishing that a short 30 day run is sufficient for the study. And then subsequent figures could focus on the key time period(s). I say periods because Fig. 5 hints that there is some oscillatory nature to the response.

Answer: The paper has been significantly improved in terms of structure, clarity and description of results. We now also make it absolutely clear what our simulations seek to represent (P6L30-38) - mean response to a monthly mean GWD distribution (as is usual from satellite observations). Following the comments made by the referee, we use Hovmöller diagrams to establish the time interval, which can be considered as quasi-steady (from ca 7 days to the end of simulation, see Fig. 3). This allows us to compute mean responses, but only for simulations with -0.5 and -10 m/s/day artificial GW induced zonal acceleration. For -70 m/s/day the simulations do not reach a steady state during the simulation. The responses in those "SSW simulations" are always presented as snapshots in selected time steps (Figs. 6, 8) or as animations (Animation 1a, b, 2 and 4). In the revised version, in addition, we present time evolution of zonal mean zonal wind to prove the SSW like nature of the vortex events (Fig. 4).

I still worry, however, that the short term response may depend a lot on the initial condition as discussed above. For example, in the real world, the propagation and breaking of gravity waves will be very different if the polar vortex is very strong vs. very weak (i.e. after a Sudden Warming). So one ideally would want to sample over different background states to robustly establish the short term response. [I assume the authors are forcing the model with some climatological mean wave forcing, but would it make a difference to use waves from a given year, etc.?)

Answer from the AC1: This is true, but the problem is that our GWD injection (value and ratio) is constant, not taking into account the background. With a little exaggeration we can say that for each time step in each ensemble member we would need another GWD enhancement values and only then we can sample for a robust response. We think that the presented results stemming from injection of constant GWD (estimated in accordance with the mean background conditions) are good enough to analyze the nature of the response as well as the discrepancy between effects of localized and zonally symmetric forcing.

I do appreciate that the authors have provided information about the time evolution in supplementary videos, but I feel that the time evolution is vital to the paper, and can't be left in the supplement.

Thanks to the referee's suggestion we use Hovmöller plots where the information on time evolution was important (Fig. 3, 4, 10). The animations in the supplement are now rather for interested readers. The paper is consistent without them and the volume of text discussing them has been minimalized. But we strongly encourage everyone to take a look at the animations, since some features are much more clear from the animation.

*3) I think it would help the paper to organize around key scientific question(s) and results. The discussion/conclusion section was more a discussion of other papers, and left me a bit confused as to what *this* paper was trying to say. In it's current*

form, the paper comes across as a bit descriptive, e.g. we tried this, and this happened. I appreciate that this is how science often moves forward, but in the conclusions, I urge them to step back and summarize how these simulations do give us new understanding.

Answer: The organization of the paper has not been changed. This means division into polar vortex, PW and residual circulation response sections. But thanks to the comments from all referees, the entire Results section has been improved. Discussion section has been restructured to be more focused and the summary of results is given P14L28-P15L24.

*I really think there is a lot of potential material here, just that the authors need to better focus the paper. Here are two key areas that could be the main result – just one or would be sufficient – and I don't mean to restrict the authors to these points. (a) Based on my own interests, I was particularly excited about the zonal mean response to zonally asymmetric wave driving. Given that downward control indicates that the time mean residual circulation depends only on the zonal mean wave driving, one might think that the zonal structure of the gravity wave driving should not matter. But since zonally asymmetric GWD induces a response in resolved waves, the *total* zonal mean wave driving depends very much on the zonal structure of the GWD. To show this, downward control analysis and more discussion of the compensation and interaction between resolved and parameterized wave driving would help.*

The authors acknowledge that nudging might limit the zonal mean response, and so drive compensation by itself. Initially I thought the nudging was done to "improve the troposphere", but upon re-reading, I realized it extends to 30 km, fairly deep into the stratosphere! How strong is the nudging in the stratosphere?

Can you estimate it's effective amplitude, and compare it to that of the applied gravity wave driving? Downward control can still be applied, but you just need to account for the torque produced by the nudging.

Answer: For a detailed response to the first paragraph, please see the AC1 - In short, this is a highly interesting topic and we want to perform such an analysis in near future.

Regarding the nudging, it is dependent on the strength of the zonal mean response. It acts on zonal mean temperatures in MUAM and the magnitude ranges from lesser than 1K/day for the reference simulation to almost 2K/day for the SSWbox simulation (See Fig. 1c in the revised manuscript). An explanation of the role of nudging in MUAM (P3L37-P4L5) and a figure of the strength of nudging for the SSWbox have been added (Fig. 1c). To minimize the effect on our results we analyze (except Figs. 7 and 8, 11 and 12) our results two layers above the nudging upper boundary.

(b) Another important and novel key result could be the impact of localized GWD on the overall resolved wave structure, following up on the Holton (1984) result that asymmetric GWD generates planetary waves. In this case, I think the time evolution of the flow is much more important. The key would be to establish how fast the resolved flow responds to the gravity wave driving, the dependence on the

background state, the linearity of the response, and so forth. These results are in the paper, but I just feel they get lost in the discussion at the end.

Note that result (a) is more about the steady/climatological response, while (b) would be more about the time evolution. Once you know the targeted result, earlier figures could help lead the way.

Answer: In the revised manuscript the section 3.2 has been rewritten to focus more on the PW generation by GWD. Also, time evolution is taken into account (new Fig. 10). But as stated in AC1, a detailed analysis of the dependence on a background state is impossible with the concept of artificial injection we are using.

4)Overall, the presentation of the paper needs to be improved. Small things, such as keeping the names of the simulations uniform and avoid non-standard acronyms (e.g. "gcu, gcv, gt"), and keeping the orientation of the latitudinal axes constant, really do help the reader. I appreciate that the first author is a student, and when I look back at my first papers, I'm embarrassed by the barely perceptible contour lines and tiny font size of the figures. So please take the comments below as suggestions on how the presentation could be improved, not as an attempt to be overly critical.

And as will come out in the detailed comments below, I think the paper relies too much on supplementary material. In my opinion, it's okay to have additional figures/movies for the curious reader, but all the key results of the paper should be within the paper.

Answer: The text in all sections underwent major revisions. The names of simulations were uniform already in the previous version - according to the naming convention (except for the SSW and specific simulations that are better characterized by shorter names): "*gcu + distribution + gcv + sign of gcv*", where *gcu* = -0.5 m/s/day is not stated in the name. This is now made clear at P4L35.

The acronyms are unchanged ("*gcu, gcv, gt*") and are properly defined in the text (P4L26-27), because we are not aware of any standard acronym for those GWD components.

Other suggestions by page:line number:

1:28-34 This first sentence about Holton 1983 is a bit confusing/vague, and then there is a giant leap of 30 years to the present.

Answer: Deleted.

1:39 Perhaps you could say "ozone and greenhouse gases" instead of "climate change gases". Also, the references here are for the mesosphere and thermosphere, but not the stratosphere. For the stratosphere, observed temperature trends have been a bit more puzzling, e.g. Thompson et al. 2012, Nature.

Answer: Deleted.

More generally, how does the second paragraph on climate change relate to the results of this paper? If you really want to cover all of climate change in the middle atmosphere, and how well models appear to simulate it, you would need a lot more

references. But I this would be taking the paper off track. It might be sufficient to shorten this paragraph and direct the reader to review papers that highlight the significance of the stratosphere (e.g. Kidston et al. Gerber et al.,) and recent analyses of the CMIP5 models, (e.g. Charlton et al. 2013 and Manzini et al. 2014), which assess the "state of the art" when it comes to modeling. I think the goal should be to quickly get across the message that the stratosphere matters, and then zero in on your topic.

Answer: The paragraph has been shortened and the reference Manzini et al. 2014, added. P1L34

2:10 "The BDC is still..."

More generally, it's my understanding that this paragraph is trying to highlight the fact that the "BDC" is a slippery creature to define. It was first discovered based on the distribution of trace gases by Brewer (1949) and Dobson (1956). It is often quantified by the residual circulation (Dunkerton 1978), which can be closely linked with the isentropic circulation. But tracers with geographically varying sources/sinks (such as ozone) are also transported by Rossby waves along isentropic surfaces, a process referred to as isentropic mixing. To understand the movement of water vapor or ozone, you need to account both for the residual mean transport and the isentropic mixing. Plumb (2002) is a good paper to highlight this. However, you can still make a lot of progress with tracer distributions in a 2D context, based on the interplay between the residual circulation and mixing. The three dimensional structure is a new frontier in research, seeking to explain the detailed 3-D structure of temperature and trace gases.

Answer: The paragraph about BDC has been rewritten based on the referee comments (P1L38-P2L7).

2:22 I would say that the discussion on pre-conditioning is still an active area of research, though "agreement" is building.

Answer: The text was changed to (P2L11-12): There is a building agreement in the literature on the role of wave activity in preconditioning sudden stratospheric warming

page 3 general comment.

There have been a few other studies that have considered the impact of localized wave torques, and they would be relevant to your discussion.

Shaw, T. A., and W. R. Boos, 2012: The tropospheric response to tropical and subtropical zonally-asymmetric torques: Analytical and idealized numerical model results. *J. Atmos. Sci.*, 69, 214-234.

Naftali Y. Cohen, and William R. Boos, 2016: Modulation of subtropical stratospheric waves by equatorial rainfall, *Geophysical Research Letters*, 43, 466-471, doi: 10.1002/2015GL067028

In addition, this paper follows up on Cohen et al. 2013 to discuss the mechanism behind compensation in greater detail. I mention it because it discusses the time scale of the response to forcing. In the stratospheric surf zone, they find it's very quick, reaching near equilibrium 5-10 days.

Naftali Y. Cohen, Edwin P. Gerber and Oliver Bühler, 2014: What drives the Brewer-Dobson circulation? *Journal of the Atmospheric sciences*, 71, 3837-3855, doi: 10.1175/JAS-D-14-0021.1

Answer: Thank you for bringing those papers into our attention. Studies Shaw and Boos (2014)-P14L10- and Cohen et al. (2014) -P3L4, P6L37, P9L37, P16L7.

First paragraph on 2.1. It might help the reader to explain a bit more about MUAM here. I gather that the model includes a troposphere, as the bottom is 1000 hPa, but is the troposphere very unrealistic, given the fairly coarse resolution? Does it have an active tropospheric circulation with synoptic variability, or is the troposphere passive, and simply there to communicate the surface planetary wave forcing up to the tropopause? Explaining a bit more detail about the nudging might be appropriate here, too. How strong is it above the tropopause?

For context, there are models that just capture the middle atmosphere, e.g. Scott and Polvani (2006), where the lower boundary condition is the geopotential height near the tropopause? Here, the lower boundary to the stratosphere is completely specified, but it was clear to me how it works in MUAM. 4:6-8 To explain my confusion above, this sentence suggest that "PW and tides" are added. Perhaps the authors mean, "develop", as they are internally generated, right? They are not specified exactly, as implied by "added". [My apologies if this is just a linguistic issue.]

4:7 Does the model spin up to a steady state? Or is it chaotic (like the real atmosphere), that it spins up eddies, etc.. and the initial condition does matter.

4:13 It might be more helpful to report the source stress than the velocity amplitude.

Answer: As we only use monthly mean reanalysis data the troposphere is not meant to be a representation of a synoptical state but rather a monthly mean climatology. This has been pointed out, now (P3L30). Furthermore, we included a sentence making clear why the troposphere is necessary and why it needs to be corrected (P3L33). The word "added" is replaced by "generated"(P4L7). An additional figure with zonal mean winds, temperature and GWD has been included (Fig. 1), standard deviations of the zonal mean within these 30 days are small as described in the text (P4L10-P4L15). Nudging is described at P3L37-P4L5.

4:35 Even though the parameters do not change, I believe that the drag can change in response to changes in the resolved flow. It might be good to emphasize this, especially if these changes are not trivial.

Answer: We agree that it should be like this, but it is not the case in our simulations, as stated at P5L4-5. In our sensitivity simulations the GW parameterization scheme is switched off and we use the GW parameterization output from the reference simulation.

4:36 Cohen et al. 2013 suggest that sharp changes in gravity wave forcing are highly likely to be compensated, as the resulting circulation wouldn't be stable otherwise.

Answer: This is now referenced at P5L9-11.

5:4 By "not covered by ... the reference run" do the authors simply mean that there is no enhanced gravity wave drag in these longitudes in the reference parameters.

Answer: The sentence has been rewritten to: There are no exceptional GWD values in the reference simulation in this region. P4L39.

5:5-12 Alexander and Rosenlof 1996 make some useful estimates of the "missing" drag that is likely explained by gravity waves. At 10 hPa, the estimate values around -1 m/s. (In general, I think we do know that net effect of gravity waves, at least in the lower stratosphere, is to decelerate the flow. Palmer et al. 1986, a pioneering study on gravity wave parameterization, added gravity waves to slow down the flow. I was therefore a bit surprised that net effect of gravity waves in Fig. 1 was generally a positive acceleration in the winter hemisphere. What level is shown here?

Answer: In Fig.2 (old Fig.1), GWD is plotted at approximately 14hPa (11th model level). We must confess that, at the start of our analysis, we were not satisfied with the performance of the MUAM GW parameterization output in the stratosphere and therefore we have chosen to modify it artificially.

Section 2.1 general comment: I may have missed it, but it was hard for me to find the vertical structure of the enhanced gravity wave drag. The horizontal structure is detailed at 4:29, but where do you explain the vertical structure? I assume the net acceleration is constant in height?

Answer: The vertical structure is now illustrated in Figs. 7,8, 11 and 12. The artificially added GWD is constant with height, but the total GW acceleration is not constant with height, because the artificial GWD is added to the reference field (very small background value). We included zonal mean latitude-height plots for GWD components from the reference simulation in Fig. 1d, e and f.

5:30-38 This is what I was trying to get out in my comment on page 2 and the BDC. It is not trivial to match the residual circulation, or even the three dimensional residual circulation, to tracer distributions because mixing plays a large role in their transport.

Answer: We deleted the part relating the residual circulation to the tracer distributions. Plots of tracer distributions are now given in the Supplement only (Fig. 1S, 2S) to give a motivation for research in the EA/NP region.

Table 1 I strongly recommend a uniform naming convention for your simulations. Why switch from Box0.1 to Box0.5 to 10Box to SSWBox. The last simulations should be Box10 and Box70 for consistency.

Also, what is "pos" in Box0.1pos supposed to indicate? I guess you mean than additional positive (northerly) wave drag has been added, but it's not clear why this makes it "pos".

Answer: We think that the names of simulations were uniform already in the previous version - according the naming convention (except for the SSW simulation and other specific simulations that are better characterized by shorter names): "*gcu + distribution + gcv + sign of gcv*", where *gcu* = -0.5 m/s/day is not stated in the name. This is now made clear at P4L35.

6:13 I don't really understand why you call this an SSW. It is true that putting a massive wave drag into the stratosphere kills the vortex, but is this really an SSW? Is it sudden, or does the vortex simply decelerate in response to the massive drag?

Answer: We have added Hovmöller diagrams to illustrate the time evolution of those observed vortex events (Fig. 4). Of course, the dynamics is visible most easily from Animation 1a and 1b in the supplement.

6:20-23 I do not understand this sentence.

6:23-24 before discussion positive/negative interference, it might be good to establish that these anomalies are indeed linear.

General comment on Figure 2: I found this figure hard to interpret. It might help to break it down a bit (or at least discuss it more slowly), to first help the reader understand the basic response of the model, and then it's sensitivity to different features.

It would help a lot to include titles above each plot, as I was constantly going up and down from the caption trying to understand what I was looking at. And the contour interval / color scale is changing all over the place. It's okay to use different color scale for the total field vs. anomaly fields, but otherwise, please fix them, so one can more easily compare panels.

Answer: The description of results shown in Fig. 5 (former Fig. 2) has been significantly improved and clarified. In Fig. 5, titles are now included.

6:33-7:4 It might be good to refocus the figure on these key results that you want to show.

It seems rather intuitive to me that the local (zonally asymmetric) response to Box gravity wave should be larger than the zonal gravity wave: the local amplitude is much larger when you focus it on a narrow region. Is the zonal mean response that much different? This seems to be a more relevant (and potentially interesting) question. As the zonal mean forcing is the same in both cases, should we expect the zonal mean response to be the same?

Answer: Titles in Fig. 5 now include also the sum of the geopotential response over the whole domain showing that the zonal mean response is indeed that different.

7:2-4 This discussion on nudging was a bit disconcerting. That's why I recommend explaining it in more detail to the reader earlier (as noted above).

Does the nudging imply that the zonal uniform response to the gravity wave drag is largely constrained, such that my questions above the zonal mean response above can't really be asked with this model?

Answer: In Fig. 1c and in the text at P3L37-P4L4 we are discussing the effect of nudging and its possible role. It acts to lower the zonal mean response, but it is not to that degree to affect the results of our analysis.

7:5-18 This discussion was confusing for me. Are the authors comparing the response to their gravity wave perturbation at 6.25 hPa with the response in to greenhouse gas forcing at 850 hPa in He et al. 2015? If so, this makes little sense. I suggest removing this paragraph entirely, or explaining why this comparison is relevant.

7:17-18 Why would you expect this? And again, how can you compare the response of the mid stratosphere to the near surface?

Answer: This paragraph has been removed.

More generally, is the response linear; i.e. if you increase the forcing by a factor of 10, does the response scale up by a factor of 10? This would be a good thing to establish.

Answer: Phrases implying the need of linearity were rephrased in the text and two sentences have been added (P8L14-17) to tackle the issue of linearity of the response. Please see AC1 for a more general response with a simple linearity analysis.

7:19-25 The connection between these results and SSWs is unclear to me. As with the preceding paragraph, I don't think it belongs in the paper.

7:26-33 As noted in my major comments, the time evolution is extremely important. And the fact that this is not a climatological (converged) response makes the comparison with global warming even more tenuous.

Answer: Both paragraphs have been removed.

7:36 If I am not mistaken, the forcing is exactly 7 times stronger than in the 10box run. Is the anomaly approximately 7 time stronger?

Answer: We didn't check this as it makes a little sense to compare these two runs (SSWbox and 10box). With certainty we can say that the response to the strongest injection of -70 m/s/day in our SSW simulations is fully nonlinear, as it contains creation of new pressure structures etc.

7:34-8:5 The authors need to show the time evolution here if they want to relate this to a SSW. How sudden is the warming? Is it simply a massive gravity wave drag destroying the vortex, or does the resolve circulation play a role in the break down of the vortex. [All this said, I'm not sure how relevant this simulation is to the real world, or to the key conclusions of this paper.]

Answer: The time evolution is shown in Fig. 4 in the revised manuscript. We didn't analyze the SSWs in detail, because this would be worth a single paper (leaving it for near future), but from Fig. 8 it seems clear that anomalous PWs play a leading role.

8:5-7 What is the vertical extent of the gravity wave driving? I presume that it does not extend to 60 km.

Answer: The artificial GWD extends from approx. 20 to 30km.

8:8-11 I'm willing to accept this is as the best default run, but I was not convinced by the discussion here. If the response is linear, then it's trivial to chose an integration. If it is not linear, then it would be good to give more motivation why this is the best case.

Answer: We are now giving a detailed argumentation why the Box0.1 simulation is chosen for majority of analyses at P5L20-30.

8:25-26 Why is this unexpected. If I understand correctly, the net acceleration of the imposed gravity wave drag is constant in height. But since there is more mass lower in the atmosphere, the effective drag is much larger at the bottom (more precisely, the net force will be proportional to pressure).

In Fig 4, I believe the authors scale the E-P flux divergence as net force, so I'd expect the response to be largest at the bottom – this is simply where perturbation force is

largest. (I suspect the response is in part a compensation, as explored by Cohen et al. 2013,14.)

Answer: This has been explained by one of the referees already during the review process to ACPD and it is not present in the ACPD version.

8:27-29 Note that the E-P fluxes are really just a diagnostic. They don't establish causality. Thus it might be better to say "The anomalies are associated with a stronger poleward ...

To establish causality, you would need to explain why the waves propagate more strongly poleward.

Answer: We are using now "is associated with", where the causality is not clear. However, the section 3.2 has been significantly rewritten and now we are giving more evidence for generation of PWs by the GWD region.

9:2-5 As noted above, it is clear that the gravity wave drag caused these changes in wave propagation (you compare with and without gravity wave drag). But it's not clear to me why this is happening?

Is it a compensating response? Or could it be interpreted with the index of refraction, that the deceleration by the gravity wave drag slows the winds, causing the resolved waves to shift and break in new places?

9:6-8 Why does this happen? Why do you need to presume that it creates poleward propagating PWs?

Answer: We are now identifying two acting mechanisms that create anomalies in the PW activity in the section 3.2: index of refraction and anomalous PW generation (absent for ring simulations).

We expect the box to create a large spectrum of waves and their modes, because it is an unbalanced force causing displacements in a balanced and predominantly zonal flow. This is now discussed at P10L9-14 .

General comment: there's too much discussion of supplementary material. If it's important, please include it in the paper.

Answer: The role of supplementary material has been diminished by the revisions we made.

9:15-17 "Probably" doesn't sound very scientific. What is the basis for this speculation? [And does the gravity wave deceleration extend below 35 km, or is it confined above this level. If you think nudging is active here, then the response of the model is probably questionable.]

9:20 By total harmonic amplitude, do you mean the RMS amplitude? Or the mean square amplitude?

9:18-26 These figures suggest that the time evolution is quite complex, and that the simulations have not converged over 30 days. If the time evolution is important, it should be explored and discussed in more detail. It's not clear to me why, for example, you get a peak response at day 6.

Answer: This paragraph and the corresponding figure have been removed.

9:27-35 I believe that the inertial gravity waves generally don't have a period of 1 day. The frequency is related to the Coriolis parameter, and so a function of latitude. Gravity wave frequencies is bounded between the Brunt-Vaisala frequency

N and the Coriolis parameter f , with interial or near inertial waves coalescing at $f = 2 \omega \sin(\text{lat})$. At (for example) 50 N, it's 16 hours, and it will only will only be a day at a single latitude in the subtropics.

Answer: Thank you very much for pointing out this. For a detailed response please see the AC1. It is most likely a product of nonlinear interaction of inertia GWs and tides- P12L3-6.

More generally, why would you expect the forcing to radiative inertial gravity waves? There's definitely something odd here. If the forcing is causing instability, I'd be quite worried about the ability of the model to resolve it, given the coarse resolution.

Answer: As written above, we expect the GWD box to create displacements (both horizontal and vertical) and then it is up to the background which waves will be supported to propagate. Also inertia GWs long enough to be resolved are anomalously generated and can propagate even to the SH (revealed by power spectral differences of anomalous modal amplitudes between different latitude bands, not shown).

10:1-2 I do not understand this argument. Is the response really periodic on longer time scales? It's unclear from just a 30 day snap shot – and the solution doesn't seem to have converged to a periodic oscillation at either latitude.

A Hovmoller diagram would allow you to show the wave amplitude as a function of latitude and time, providing valuable information in the paper that is now in the supplement.

10:23-31 It is hard to follow or understand this discussion without more evidence. What is the evidence of wave reflection? Does it show up in a change in the vertical structure of the waves?

I'm curious how these anomalous waves related to the climatological waves. Is there positive or negative interference (see Fletcher and Kushner 2011)

Answer: This subsection has been removed.

11:22-27 The time evolution is very troubling, and not sufficiently document in the text.

Note also that the shallow branch of the residual circulation is generally associated with synoptic waves breaking on the top of the subtropical jet. Are there synoptic waves in your model? If you want to explore the residual circulation in more detail, I might suggest considering a downward control analysis. To what extent are the resolved waves amplifying or compensating the anomalies associated with the artificial drag? [You could include the nudging in this analysis: if it's not too strong, it might not overwhelm the response.]

11:28-12:20 I don't think it's appropriate to spend so much of the text discussing the supplement. If this is important, please hovmoller diagrams or other means to distill into a figure that can be included in the paper.

Answer: The entire Section 3 has been revised not to rely on the supplementary animations. We prefer to present the information on residual circulation using vectors (v_{res} , w_{res}) colored by the strength of the mass flux, because then certain features nicely emerge (lee wave pattern etc.).

Fig. 8. It was unclear to me how to relate this figure to the results shown in the paper. Are the authors arguing that there is enhanced subsidence is causing the enhanced ozone column? Is so, please show this.

12:34-13:5 I think it is quite a stretch to compare MUAM vertical velocity anomalies with MIPAS CH4. Why not start by comparing with the vertical velocity in a reanalysis, such as MERRA, which I believe extends pretty high in the stratosphere. Then you could compare the same quantity.

Answer: This subsection has been removed and the figures moved to the supplement to give motivation for the EA/NP research only.

13:34-35 What discrepancy? Do you mean the differences that are only shown in supplementary figures? The "SSW" run is an extremely nonlinear case, so I don't think you should expect it to be similar, and I'm not sure how relevant it is to the real world.

14:4-12 I do not really understand the discussion in this paragraph. What positive feed-back? I suspect gravity waves break in the EA/NP region because they are produced by flow over the Tibetan plateau and instability in the storm track. They break when they reach critical levels or become convective unstable. I don't see how the temperature in the region, however, would cause them to break.

14:33-38 This would be an interesting result, but I don't think the compensation argument is really developed in the paper.

As noted in my major comments, I feel that the conclusions section become a narrative of open questions and interesting results in the field, but is not very much related to the results of the paper. This section needs to be reworked in detail. I think a shorter summary and discussion would make the paper more effective. (Overall I recommend using the conclusions section to review the key results of the paper and explain their broader context.)

Answer: We absolutely agree and the Discussion and conclusions section has been restructured, rewritten and a summary of results is now given at P14L30-P15L24.

Comments on figures: We have newly generated the figures according to the comments.