Thank you for taking the time to help improve our manuscript. We have addressed all requested points in **red**.

This manuscript by Franco-Diaz investigates convective gravity wave effects over northern Europe using data from NASA's AIRS satellite instrument, a Rayleigh-Mie radar at Kuhlungsborn, and supporting data from ECMWF operational data.

The manuscript is well-written, clear and quite an easy read, but makes interesting and well-evidenced points, showing some nice case studies. I therefore echo Reviewer 1 in their recommendation that only minor corrections are needed for publication in ACP. In particular, the paper was well-structured, the figures well-chosen, and in general most questions I thought of while reading were answered within a couple of sentences at most.

I include a series of minor comments below. I also strongly echo the comments of Reviewer 1, particularly their comments:

(1) about consistency in the description of AIRS' observational capabilities - for example, lines 95, 131, and 140 disagree on the horizontal-wavelength sensitivity of the products used, while (as reviewer 1 says) line 96 disagrees with other parts about vertical sensitivity.

and (10) about filters, as I also read it the same way

– We have modified references to the resolution throughout the manuscript to be consistent with Figure 1

In addition, I would recommend/ask that the authors:

(A) be much more specific about the ECMWF data used, particularly in the abstract where the description given ("using ECMWF") is really quite undescriptive! Reviewer 1 asks for more details of the dataset; I would also like the authors to check that they really mean ECMWF *forecast* data as they say on line 119 - do they mean operational *analysis* data? Using the operational analysis, i.e. ECMWF's best-guess of the atmospheric state, would make perfect sense, while forecast data, which is generated by running the model forward from the analysis, would be a somewhat esoteric option to use for a study like this.

- Thank you for the comment. We use the ECMWF IFS observational analysis. It has been clarified in the paper.

(B) in general, the paper does not reference many other studies which have looked at the same gravity waves in multiple datasets. Discussing a few of these in the introduction could be useful to contextualise the observational filter differences you see - I would naturally recommend our 2016 paper on orographic waves (https://amt.copernicus.org/articles/9/877/2016/amt-9-877-2016-discussion.html), but studies by many others are of course available on the same topic.

- Thank you for the suggestion. We have added a paragraph to the introduction and incorporated the reference you gave.

“Simultaneous observations from different instruments of the same gravity wave event are useful for providing insight into different portions of the gravity wave spectrum since no single instrument is capable of viewing the entire gravity wave spectrum. Each measurement technique has its strengths and limitations. Lidars have very high temporal and vertical resolution but only measure at one location. Limb sounders have good vertical but poor horizontal and temporal resolutions. Nadir viewing satellite instruments have good horizontal but poor vertical and temporal resolutions. Observations of gravity wave properties from various instrument types can differ considerably because each measurement technique is sensitive to different parts of the wave spectrum (observational filter). For example, [Wu2006] found that most of the differences in gravity wave variance distributions between different types of instruments could be related to their viewing geometry and thus their different sensitivities to various portions of the gravity wave spectrum. Similarly, [Wright2016] found that gravity wave properties for the same event over the Drake Passage measured by the nadir-viewing AIRS instrument, radiosondes, radar, and limb sounders differed significantly, sometimes being entirely uncorrelated, suggesting that the discrepancies were due to the different observational filter of each instrument. Typically, there is good agreement between instruments of the same type or that measure similar parts of the gravity wave spectrum [Wright2016, Ern2018]. Good agreement has also been shown when sampling one instrument to match the resolution of another. For example, [Preusse2000] showed that CRISTA gravity wave zonal mean variance was comparable to that of MLS if CRISTA vertical resolution was reduced to MLS vertical resolution. Understanding the full spectrum of gravity waves generated by convection requires combined analysis of instruments measuring different parts of the gravity wave spectrum and, as mentioned above, high-resolution simulations. In this study, we combine gravity wave observations over the same geographical location from two very different types of instrument: lidar and nadir-viewing AIRS. We focus on case studies of strong convective gravity wave activity observed by both instruments in the summers of 2014 and 2015.”

(C) does the 8-hour limit cutoff a chunk of permitted GW periods at this latitude? Presumably, going by the text, the 8h cutoff is chosen to avoid the mesospheric terdiurnal tide impacting the results, but in the 33-43km part of the stratosphere this isn't a major concern as it's so weak in temperature amplitude this far down in the atmosphere

- Yes, we are aware of this potential shortcoming. First, we used this filter for consistency with our previous publications of the lidar data [Baumgarten et al., JGR, 2017, and Strelnikova et al., JAS, 2021]. Indeed, the temporally filtered data will probably cut part of the inertia gravity waves. That is why we also show the vertically filtered data where these waves are included. Also, the influence of the tides in lidar data should be very small in the lower stratosphere (less than 0.5 K) [Hauchecorne et al. 2019]. We used a standard lidar data product optimized for retrievals in the mesosphere where tidal contributions are more relevant. In future work we will modify the stratospheric temperature retrieval to remove the unnecessary filter. We have added the following caveat to our manuscript on line 219:

“It is important to note that some longer period GWs will also be affected by the use of the filter.”

Some additional minor comments follow, but in general I think this is a very good paper and in my view is publishable in something very close to its current form. Good work!

Additional comments:

014 dimensions unclear - should specify these are horizontal wavelengths

- We have made the correction.

018 and also unresolved

- We have made the correction.

077 and afterwards - Aqua is a name, not an acronym, so shouldn't be all-caps

- We have made the correction.

079: would "at least two" work better here?

- We have made the correction.

085: the track-edge value would be useful as it is quite a bit lower (~40km)

- We have made the addition. The text now reads:

"This scan width is composed of 90 footprints that have a diameter of 13.5 km at nadir and increase in size off-nadir (~40 km at the edge of the scan)."

089: how is the product "special"? This is quite a vague way of describing the dataset.

- We have deleted the word "special".

093: I *think* (but a happy to be corrected) that the 4poly method needs referencing to Alexander + Barnet 2003 (JAS)

- Thank you for this comment. We have added the reference and slightly expanded the description. It reads now "The brightness temperature anomalies in this product were obtained by fitting and subtracting a fourth-order polynomial to the cross-track radiances to remove the large-scale background as well as limb brightening effects [Alexander2006].

098: brackets missing from ref

- We have made the correction.

101: what do you mean by "atmosphere" here? Couldn't one say that the clouds you're measuring are part of the atmosphere, just a wetter part?

- Sorry, we do not understand this comment.

108: Can MST radar do this?

- MST radars do measure winds in the troposphere and lower stratosphere, but no temperatures. At certain conditions (i.e. the existence of PMSE) they measure winds also in the mesopause region. There exist other radar techniques for the MLT region, but they are typically also confined to wind measurements. Lidar is the only technique that covers the full range mentioned here and is capable of temperature sounding. While it might be possible to measure winds in the stratosphere with radar, it is very difficult in practice and requires long integration times and large arrays [Maekawa et al., 1993]. The radar detects backscatter signals from strong refractive index fluctuations in the atmosphere. The refractive index of the atmosphere is a function of three components: dry air (air density), water vapor (humidity) and free electrons [Balsley and Gage 1980; see also Fig. 1 of Kato 2009]. The reason why it is very hard for radars to observe the mid- to upper stratosphere is because the radio refractive index has a minimum at altitudes of around 20-60 km as air density is decreasing and electron density is still low. [Woodman1974] shows the radar gap in the profile of echo power versus height obtained by the Jicamarca Radar in their Figure 2. We are only aware of one existing publication of radar measurements of vertical winds in the stratosphere [Maekawa et al., 1993].

Yasuyuki Maekawa, Shoichiro Fukao, Mamoru Yamamoto, Manabu D. Yamanaka, Toshitaka Tsuda, Susumu Kato, R. F. W. First observation of the upper stratospheric vertical wind velocities using the Jicamarca VHF radar. Geophys. Res. Lett. 20, 2235–2238 (1993).

Balsley, B. B. & Gage, K. S. The MST radar technique: Potential for middle atmospheric studies. Pure Appl. Geophys. PAGEOPH 118, 452–493 (1980).

Kato, S. Validity on radar observation of middle- And upper-atmosphere dynamics. Earth, Planets Sp. 61, 545–549 (2009).

Woodman, R. F. and A. G. Radar Observations of Winds and Turbulence in the Stratosphere and Mesosphere. J. Atmos. Sci. 31, 493–505 (1974).

118: what does "integrated" mean in this context?

- We changed the wording to be more clear. We removed the word “integrated” and replaced it with running mean.

“The temperature profiles are calculated as a running mean over 2 h with a 15 min shift in time and binned to a vertical resolution of 1 km [Baumgarten2017].”

Figure 1 implies hard edges, but they're quite fuzzy. Not sure how to fix it this easily, but it might be worth a think. My own papers can be just as bad though on this...

- Thank you for the suggestion. We have added a sentence to the caption to make it clearer.

“Note that the response function for AIRS for vertical wavelengths below $\sim$26 km is very small (< 1\%). While this idealized figure implies hard cut-offs at certain vertical wavelengths, the nature of the filters (whether observational or imposed) is more of a gradual transition.”

144: The transition from the case study in lines 139-144 to a more statistical study from line 144 onwards is very abrupt, and needs delineating more clearly

- We started a new paragraph to describe the statistical figure.

145: "as a function of time-averaged over an area" is ungrammatical

- It has been corrected.

Figure 2: a colourbar for the inset would be very useful

- Thank you for the comment. The colorbar has been added.

183: how long a temporal average? Could affect the results.

- It is a daily average. The clarification has been made.

“A daily average is applied, which is denoted by the over-bar above the temperatures. More details regarding lidar data processing can be found in [Baumgarten2017].”

206: they are *likely* to be related...

- Thank you for the suggestion. The sentence was changed based on the other reviewers comment.

“These events are beyond the scope of this paper because they are either generated by convective activity outside of the area we chose for this study or by other gravity wave sources, e.g., baroclinic instabilities.”

208: that's a very precise height level - is it actually a pressure level? If so might be clearer to specify that, with the height approxn in brackets afterwards, eg something like "10hPa (~16km)". Same comment for line 209.

- We changed it to say around 40 km.