Review of "Observations of OH-airglow from ground, aircraft, and satellite: investigation of wave-like structures before a minor stratospheric warming" by Wüst et al. (acp-2018-2012)

The authors study pre-SSW gravity waves from airglow and/or temperature measurements in early 2016 using measurements of four different instruments: SABER-TIMED space radiometer, GRIPS9 (Kiruna) and GRIPS14 (Alomar) ground-based spectrometers, and FAIM imager (onboard FALCON aircraft). Following the work in Wüst et al. (2016), the authors derive time variation of BV frequency at the OH layer from SABER and, in combination with GRIPS temperatures, gravity wave potential energy density. They also derived short-time series of GW spectra and propagation direction, and their time variations from two FAIM flights (one of them right before a minor SSW). They detect highest GW occurrence over mountains. They also found dominance of small-scale GW contribution a couple of weeks before the SSW, which was not the case just before the SSW. Leaning on SABER-GRIPS BV frequency evolution and ECMWF data, they concluded that the small-scale waves in the first case were due to convective instability whereas they were due to dynamical instability in the second case. They also conclude that short period waves are generated in the higher stratosphere and above.

The paper is well written and organized, although some explanations could be simpler (particularly in the discussion). The English is ok.

I recommend the manuscript for publication, once the following suggestions and comments are taken into account.

General comments

The introduction does not include a description of previous results and the state-of-the-art in the field of GWs, in particular, before or during SSWs. Also, there is not a description of the scientific interest of the results presented here. Something similar happens in the discussion, which is not put into context of results from other authors or measurements. Indeed, there are previous publications (particularly regarding large scale features) that are not mentioned here.

The authors make use of measurements of several variables from 4 different instruments. In several places in the text, it is hard to know the instrument they are referring to or the calculations they are using. That makes the reading slow. For example, Sect. 4.2.1 shows calculations of GWPED that need from GRIPS temperature anomalies and periods, but these and their estimation are not shown nor even discussed anywhere. This happens more often (see comments below) and I recommend the authors reading the manuscript carefully with this criticism in mind in order to address this issue.

GRIPS14, observing over Alomar, is not used in the analysis. Only its 15-day mean temperatures and intensities are plotted but they are not further analyzed nor used for the discussion. Some information on wave propagation direction could be extracted when combining GRIPS9 (at Kiruna) with GRIPS14 (as in Wüst et al., 2018), perhaps also in the context of FAIM measurements. In any case, the results from GRIPS14 could support (or not) those from GRIPS9 and should be analyzed in parallel here. Additionally, they start early in January and can extend the time series longer.

In the discussion section, the author's conclusions are more a consistency with the behavior expected. For example, Flight 1 is just consistent with dynamical instability as the origin for ripples and Flight 5, right before the SSW, is not. This subtle difference is important because there is not an examination of other possible sources or very strong evidence from these results behind that idea (on the one hand, it is based on the assumption that changes in brightness are only due to the generating GW; on the other hand, they only have two 2 days of measurements). That should be clear in the text.

<u>Detailed comments</u>

P2Sect.1. The introduction should be revised. The research is not put into context and the scientific scope of the paper needs to be better described. Just studying gravity waves is not an argument for a scientific paper. Please, include an explanation of the scientific interest.
P1. L18-24. Provide a small introduction of FAIM.
P1L20. Small-scales, write how small.
P1L21. Smaller aperture. How smaller?
P2.Sect.2: The instruments are poorly described. It is not easy to understand what and how exactly they measure.
Sect. 2.1: Unless you know GRIPS before reading this paper, it is not easy to know how exactly the instrument measures airglow. It is not even clear here that GRIPS is not an imager. What is the spectral resolution? Perhaps describing it here with more detail would help.
P3L7. Are these noise or systematic errors? Include a description of major sources of uncertainty.
P3L7. Include reference for temperature retrievals.
P3L11 Write observation angles for the 4 FoVs for GRIPS 9
P3L19. Shortly describe how you derive temperatures. Provide errors and error sources.
P3L25. Write the OH transitions this instrument is sensitive to.
P3L31. Please, indicate range.
P4L10. It is not clear. Are they analyzed or not?
P4L20. SABER is described in many papers. Better a reference to one of those than to a webpage that may eventually stop working.
P4L28. Remsberg et al. compared SABER v1.07 temperatures but you are using v2.0. Provide biases for v2.0, wether indicating v1.07-v2.0 comparisons or comparisons of v2.0 with other space and ground based instruments, which are already available.
P4L28-32. The authors are mixing here noise and systematic errors. Comparisons with other instruments should be commented in the context of systematic errors. SABER MLT temperature main errors are due to atomic oxygen uncertainties (Remsberg et al. 2008; Garcia-Comas et al. 2008). Also the biases strongly depend on latitude.
P4L32. For coherence, shortly comment on OH VER uncertainties.
P5L7. What do you mean by 500m negligible compared to 2000m FWHM? Please, quantify. Also note that SABER vertical sampling is several times smaller than its FOV.
P6L7. Insert 'Brünt-Vaisala (BV)'
P6L10. One really needs Wüst et al. 2016 in one hand when reading this manuscript, which is not useful. Please, shortly describe why shorter and longer than 60 min.
P6L23. For what transition?
P7L6. Could you better explain why airplane shaking prevents deriving period and phase speed? What is the error in the wavelength due to this shaking?
P7L8: Delete 'used'
P7L15: Please, clarify why you use here 87km and you mention 84km in previous section.
P7L15: Please, quantify the effects of layer altitude.
P7L19. Please, show in Fig. 1 the resulting image after applying this filter.
P8L4. According to what instrument?
P8L8. starts to rise by -> rises
P8L8. varies -> oscillates
P8L9. layer altitude
P8Sect.4.1. Fig. 2 is full of interesting things. I recommend including a more detail

description of the figure here.
P8L10. What SABER intensity is compared here? Averaged over the layer? Peak intensity? Does this choice make a difference?
P8L11. Only SABER and ALOMAR show a 4-6 day pronounced periodicity. GRIPS-9 periodicity is 9 days (one should not assume measurement for 15Feb is a maximum.
P8L12. Not in GRIPS 14.
P8L13. Include SABER OH*-temperatures. If comparable, that would somehow justify the use of SABER BV frequencies.
P8L18. Please, perform the same analysis with GRIPS 12 since it has a longer time coverage and also, if combined with GRIPS9, some information on horizontal propagation could be extracted.
P8L20. Describe here the temperature anomalies (amplitudes) you are using and how you estimated them.
P8L20. GRIPS temperature amplitudes
P8L21. There is no dashed line in Fig. 5
P8L22. Include 15-day averages in plot and discuss here in terms of fluctuations around the linear fit.
P8L24. Shortly describe criteria here.
P9L8 principal -> principle
P9L12. I guess that the authors mean an image horizontal coverage instead of spatial resolution, in contrast to the FoV used in this manuscript to refer to the spatial resolution for GRIPS. Please, homogeneize definitions.
P8L14. ... and it also varies with OH layer altitude.
P9L15. What do you mean by time difference images? Explain how you treat several images overlapping.
P9L27. in sensitive -> is sensitive
P9L28. Why is the horizontal coverage cut to 26x26?
P9L30. What do you mean by 'small-scale' here?
P9L31. But the wavelengths smaller than 15km (1/k = [1/0.1,1/0.15]) appear very strongly at 17:40-17:55. Don't they?
P10L12-13. This info is not accurate, not used, not analyzed. The reader may loose attention to the central point of the FFT analysis.
P10L16. What do you mean by this? What do you think it is causing this large mean intensity?
P10L16. What do you mean by saying this? What do you think it is causing this large intensity? For previous flight, you just mentioned that mean intensity changed too much for long wavelenghts analysis....
P10L16 maximal -> maximum
P11Sect.5. The discussion gets complicated in some paragraphs. Please, re-read and simplify (this specially holds for reasoning in pages 12-13).
P11L3. A better description of the event, including dates of SSW onset and polar vortex displacement and recovery would be more useful.
P11L3. Delete 'the' before 'January'
P11L6-7. Include reference.
P11L12. Better than 'neglecting the effect of planetary waves' (which are the responsible for the polar vortex displacement mentioned above', you could write 'We expect the following effect on the zonal means.
P11L15. Mulligan et al. is missing in the reference list. Grygalashvyly (2015) and Garcia-Comas et al. (2017) should be included in this list.
P11L16. Explain why height and thickness are not anticorrelated in Fig 3.
P11L17. Insert 'According to SABER measurements,'
P11L18. also and particularly during February 2016 (see Fig. 2 and 3).
P12L3. vertical -> horizontal
P12L10-11. This may confuse the reader. Better saying "winds in the upper stratosphere

were stronger than in the upper troposphere"
P12L11. was -> were
P12L11. Easterly winds became weaker after Jan 23rd, which, for a continuous source of GWs, should have resulted in less overall filtering and more (E) GWs propagating to the mesosphere until Jan 28th. I can only glimpse the corresponding response in potential energy density for T<60min but the enhancement on the 27th is clear. Please, discuss on that. Perhaps, analysis of the next days in GRIPS9 time series (until Feb 2nd, as in Fig.4) could help. On the other hand, the change in FAIM total number of wave events before (Fig.9) and at the onset of the SSW (Fig. 12) does not clearly show any difference. Discuss on that also.
P12, L17. Insert 'according to GRIPS9 data,' after 'Therefore'
P12L18. This is too much of a conclusion based on zonal mean winds. Note the potential longitudinal variations or the length of the time series in Fig. 6.
P12L19 had the best chance -> had best chance
P12L21. Again, you should be careful when using zonal means from Fig. 14. I do not think you can resolve measurements over Kiruna using that information alone.
P12L25 Please, rewrite sentence
P12L26. I do not agree that the wind profile is rather flat before Jan 31st. There is a wind reversal around the stratopause and in the troposphere. What can be inferred from GRIPS14 measurements?
P12-13 The conclusions the authors reach are not put into context of results from other authors here, in particular, those regarding larger scale features (e.g.., Gerrard et al., 2011).
P13L2. Insert '(see Fig. 5)'
P13L3. What 'airglow brightness maps'?
P13L2. 'Since the measurements were taken in winter'
P13L6. What do you mean by 'overall' here? Note that you may eventually have inversion layers.
P13L8. Explain here what you define as the 'grey regions' of an airglow image
P13L7: Insert 'According to ECMWF data,'
P13L19. Please, make it clear that a correlation does not always hold (as in Pautet et al.) but, on average, a positive correlation between brightness and temperature should be a fair assumption, at least from mid-autumn to mid-winter. This was shown by WINDII and SATIs (Shepherd et al., 2006) but also by SABER, instrument that you use (Garcia-Comas et al., 2017).
P13L21. Why does the temperature gradient become zero? That depends on the amplitude of the wave. Better saying 'becomes maximum'.
P13L21. The use of 'steepest' here leads to misunderstanding. Better saying 'the minimum (or, since it is negative, maximum in absolute value) temperature gradient'
P13L23. 'compared to' -> 'depending on'
P13L25. Do you mean the 'zonal wind shear'
P13L27. Could you be more precise and describe the bright airglow areas you are referring to? Legs 4 and 5? How do you know these small-scale structures are only caused by a larger dynamical instability instead of any other cause, like location or just time variation?
P13L29. Better than 'then this means' use 'then this is consistent with'
P14L2. Although I agree that causes for ripples at the onset of a SSW are more likely due to changes in static instability, I do not think this conclusion can be inferred from these measurements. Again, it seems to me just a consistency (and not a conclusion) with a smaller dynamical instability. This is in part because your assumption that the large changes in brightness are only due to the generating GW is too strong, but also because of the lack of statistics (just 2 days). Additionally, these conclusions should be put in the context of previous results, which also should be referenced here.
P14L17. Insert 'combined with SABER data'

P14L19. 'below the tropospheric jet' -> 'in the troposphere'
Fig.2-caption. L5. SABER temperature?
Fig.2-caption: Write SABER channel.
Fig.2. Instead of the hard-to-follow description in the caption, just remove non-reliable data according.
Fig. 2. Please, change color code. It is not possible to differentiate most of them from others.
Fig. 4, L4. Indicate year of campaign.
Fig. 4. For coherence with panel a), include SABER temperatures in panel b) and discuss comparisons in text.
Fig. 5. The linear fit is not completely convincing. Indicate correlation and discuss in text.
Fig. 5-caption: Insert 'SABER' or 'derived from SABER'.
Fig. 6. Why do these data end on the 30th and not Feb. 2nd, as in Fig. 4?
Fig 6. L7. GRIPS 9
Figs. 9 and 12. I think that combining these two figures, that is, including the results for the two flights in the same plots would be interesting to see.
Fig.14. Lower panel is not needed for the discussion and does not provide additional useful information. Please, remove.

References

Gerrard et al., Observations of in-situ generated gravity waves during a STE event, Atmos. Chem. Phys., 11, 11913-11917, https://doi.org/10.5194/acp-11-11913-2011, 2011.
Grygalashvyly, M.: Several notes on the OH layer, Ann. Geophys., 33, 923–930, https://doi.org/10.5194/angeo-33-923-2015, 2015.
Garcia-Comas, M., et al.: Mesospheric OH layer altitude at midlatitudes: variability over the Sierra Nevada Observatory in Granada, Spain (37N, 3W), Ann. Geophys., 35, 1151–1164, https://doi.org/10.5194/angeo-35-1151-2017, 2017.