Reply to the comments on the manuscript

"Ground-based observations of periodic temperature fluctuations in the mesopause region with periods larger than 2 days"

by Christoph Kalicinsky, Robert Reisch, and Peter Knieling

We thank the reviewers for their helpful comments and recommendations. In the following, we discuss the issues addressed by the reviewers and explain our opinions and the modifications of our manuscript.
We enumerate the comments and repeat them in bold face. The modifications of the manuscript are displayed in the marked-up manuscript version as coloured text. Deleted parts are shown in red and new or modified text parts in blue.

# 1 Comments Reviewer 2

**General comments:**
**This manuscript presents an analysis of seasonal and long-term periodicities in 30 years of mesopause region temperatures derived from hydroxyl (3-1) band spectrometer observations from Wuppertal, Germany. They use a variable sized, sliding window to apply a Lomb Scargle periodogram analysis to deseasonalized time series for each year. They analyse the main seasonal periodicities in the range of 2 to 60 days in terms of planetary Rossby wave activity. The majority of wave activity is observed in winter, with some indication of a maximum occurrence of shorter period waves at the equinoxes. Long-Term behaviour of wave activity was analysed using the standard deviation of temperature residuals and the mean amplitude of significant events as proxies. This revealed a quasi-bidecadal oscillation characterised by events with larger amplitude. The work is a worthwhile study on planetary wave activity in one of the longest hydroxyl rotational temperature data sets and yields some sensible but not unexpected results in terms of predominant wave activity. Some further work and revisions are suggested below.**

**Specific comments:**

1. **Line 14 Introduction section: In the context of Lomb Scargle analyses of long term hydroxyl rotational data sets the work from the Davis, Antarctica observatory, seems particularly relevant (for example French and Klekociuk (2011) https://doi.org/10.1029/2011jd015731 , French et al (2020) https://doi.org/10.5194/acp-20-8691-2020) but is not considered in the introduction or in the discussion on periodic fluctuations in 4.1. Including a comparison with these studies would be valuable, particularly since similar periodicities from other Antarctic sites are already discussed in the periodic fluctuations section (e.g. line 279)**
   We indeed missed to mention this important work on the time series of OH* temperatures from Davis, Antarctica. We included the publication that partly deals with planetary wave activity by French and Klekociuk (2011) in the introduction and discussion part and the corresponding positions. We further included the study by French et

al. (2005) where a detailed examination of a 14-day wave is done.
The study by French et al. (2020) is considered in the discussion of the long-term evolution. See point 10.

2. **Line 85 is there much variation in the amplitude and phase of the seasonal fit between each year? What are the ranges of the fit coefficients ? What is the difference to a seasonal fit to all years ?**
There are some variations of the fit parameters, e.g. there is a declining trend of the amplitude of the annual cycle and some further fluctuations from year to year. A fit to all years at once would not capture all of these year to year fluctuations and the long-term variations. Thus, the residuals after subtraction of such a fit would still contain some variations with implications on further analyses. We added some more information.

3. **Line 100-105 why not do a sliding window LS over the entire deseasonalized time series instead of adding 2 months to either side of each year? This paragraph also needs revising eg "all possible time windows have been of the complete time series were analysed" makes no sense.**
See point 2. One fit for all years would not be able to capture all changes in the seasonal behaviour in each year. As a consequence the fit would be not precise enough in each single year and there would be some shifts of the residuals (when the fit lies slightly above or below the best possible fit) and even long-term trends that remain in the residuals in some years (when the fit changes from below to above or vice versa). This would result in wrong or biased LSP results. We added the 2 months at each sides to be able to calculate a LSP for each possible day of the whole time series of more than 35 years and to account for all of this year to year variations of the seasonal cycle. After a deseasonalization of each single year one could stack a individual residuals together and calculate the LSPs, but this would not lead to different results. We added more information in the paragraph.

4. **The paragraph explaining the variable window length (lines 117-123) is incoherent and difficult to read and interpret. This should be revised for clarity.**
We rephrased the paragraph.

5. **The "factor" introduced in line 126 .. should be defined as cycles per window length of the period being analysed. How is this used in the analysis except to step up the window length for increasingly long periods analysed?**
We already described that the factor can be seen as cycles per window length. We rephrased the paragraph for clarity.

6. **Line 163 What is the criteria for significance here and in Fig 2 and how is this computation affected by the varying window lengths?**
The significance of the results is analysed using the false alarm probability (FAP). The FAP gives the probability that a certain peak can occur just by chance somewhere in the analysed period range. It is determined using an empirical expression where the coefficients were determined by using Monte Carlo simulations. By using this expression the significance for each peak can be evaluated depending on the window length, period range and number of data points (for more information see Kalicinsky et al. 2020).
We added this additional information to section 2.3 and section 3.

7. **Section 3.1 and Fig 3. Occurrence frequencies are counted by the centre days which meet the significance criteria. This will bias occurrence statistics to waves of longer duration. The occurrence of long period waves will necessarily occur on a larger number of days (and this is stated in Line 172). I don't think this is a valid assessment of wave occurrence statistics and the analysis needs to be revised. Normalise by the period length or count waves of each period by event.**
   We added a second version where we normalize the histogram. (See also Reviewer 1)

8. **Figure 4. This is a confusing plot as the period ranges overlap. Why not separate these ranges and do periods <10, 10-20, 20-30 and >30?**
   We reduced the plot to 2 subfigures only and added some additional information in the text (see Reviewer 1).

9. **Line 192 and Fig 5. In the discussion of observed amplitudes what is the effect of the integration of the temperature variation by a wave propagating through the 8km layer profile of the hydroxyl emission ?. Does this limit the observable frequency and amplitude of waves.**
   There might be some reduction of the amplitudes due to vertical averaging, especially for waves with smaller periods and smaller vertical wavelengths. In the majority of the cases waves with periods even below about 10 days exhibit vertical wavelengths that are a multiple of the FWHM of the OH layer (e.g. Buriti et al., 2005, Ern et al., 2013, Yamazaki and Matthias, 2019, Reisin, 2021). Very small vertical wavelengths below e.g. 20 km are only reported for a very minor portion of all observations (Reisin, 2021). Thus, the averaging effect should be small or maybe even negligible.

10. **Section 3.2 and Figure 6. In the discussion of long-term wave activity the clearly dominant periodicity in the weighted sum power spectrum at 4 years is not discussed. This has higher power than any other measure including the bi-decadal oscillation. What is this attributed to and does it align with the quasi-quadrenial oscillation of French et al (2020) https://doi.org/10.5194/acp-20-8691-2020?**
    The significance of the 4-year peak is only good with respect to the single frequency not with respect to the whole frequency range. However, as the peak occurs in all LSPs and is in the case of the weighted sum the highest in all of the periodograms, we analysed it further. It looks quite similar to the one reported by French et al. (2020) concerning the period (4 years here compared to 4.2 years) and the phase, where there seems to be a slight time lag of about 1 year with the Wuppertal oscillation preceding. We included this information to the manuscript.

**Minor corrections:** We changed the manuscript according to the minor comments where still necessary.