Referee comment: “Microbarom radiation and propagation model assessment using infrasound recordings: a vespagram-based approach” (Vorobeva et al.)

General comments

This study presents a novel method for analyzing infrasound data, the vespagram approach. The authors apply this method to 6 years of data recorded at the Norwegian array IS37 to extract microbarom arrivals. The latter are compared with simulations using a recently developed microbarom model. For comparison, the similarity index is introduced, which is based on mean-squared errors. While the manuscript title partly raises expectations on conclusions about the microbarom model performance, the manuscript essentially evaluates the vespagram approach as a method to utilize infrasound for stratospheric diagnostics. Due to its capability to scan all directions simultaneously, the advantage of this novel approach is obvious. The manuscript is generally well written and organized, but the presentation and discussion of the results need to be enhanced. In particular, Section 3.1 confronts the reader with three extensive figures within a few sentences while remaining sparse with explaining and discussing details (e.g., on remarkable features, outliers) of these figures. For instance, Figure 4 is only once briefly referred to. My specific comments indicate several features that could or should be further discussed (e.g., Figs. 2 and 3).

Given the number of comments and questions, I request a major revision, rather than a minor. Once the authors have addressed these comments and questions, this study will be an important contribution and thus certainly worthy of publication in Annales Geophysicae.

Specific comments

1) Consider revising the title of your manuscript a little (see general comment).
2) l. 16: Why do microbaroms return to the ground after penetrating the middle atmosphere (hence their potential to probe the middle atmosphere dynamics)? Briefly explain the underlying physical process.
3) l. 42: how did you determine the fixed apparent velocity of 350 m/s – from observations (using other processing techniques?) or propagation modeling (average?), or is this based on previous studies (references available? - obviously yes, but these are not cited before line 87/88). For the discussion of the results (e.g., line 205): using this fixed value, what is the corresponding standard deviation of observations at IS37 (e.g., using PMCC)? Based on this, can you roughly quantify the number of other arrivals (especially in the summer) that potentially cause discrepancies?
4) l. 75 and Fig. 1a: you could add the ARCES array to the map as this is mentioned in the text as the initially planned site for IS37; however, I am wondering if the first part of the sentence (“was initially planned … in Karasjok”) is worth to be mentioned at all. This fact is not relevant to your study but raises the question of why it was less favorable. Therefore I recommend shortening the paragraph accordingly.
5) l. 125/126: it is not necessary to repeat all references, the choice of 350 m/s was justified before; I suggest removing the second part of the sentence (beginning with “which is within …”).
6) l. 136: Landès et al. (2014) studied the global patterns of microbaroms and only discuss the potential limitations due to the lack of coastal reflections while citing Hillers et al. (2012), among others. Therefore citing that study in the way it is done here is a bit misleading. My suggestion is to modify this and add another sentence, for example:

"Studies on microseisms (e.g., Hillers et al., 2012) have demonstrated the limitations of a model that does not account for coastal reflection. These limitations have been accordingly raised in the context of microbaroms (Landès et al., 2014)."

7) l. 153/154: which of the ECMWF models in particular? If not the ERA5 reanalysis, did you interpolate the temperature and wind fields in time?
8) l. 163: remove the parenthesis (private communication with …), M. De Carlo is co-author of this study. Instead, how would the results differ if you accounted for only 3000 km? (Is it essential to account for 5000 km for providing a more realistic spectrum at IS37?)
9) Section 3.1: Here you present a lot of information (3 figures within more than 24 panels!) within the first paragraph, without much explanation. You could help the reader by focusing on Figs. 2 and 3 first. Also, I suggest that you already define Eq. 3 in Section 2; then all panels can be understood at the first occurrence of a figure in Section 3.
10) According to Section 2.1 step 5, the vespa output should be power (Pa^2), whereas in Figs. 2 and 3 the colorbar unit is Pascal (amplitude) again, correct? (also, place the units/labels to the right of the colorbars – amplitude in Pa).
11) Fig. 2b-d: in the summer, infrasound amplitudes at IS37 seem to be not relevant, whereas for the comparison (Fig. 2h) and through normalization (e-j) they certainly are (e.g., lower SI). Would a logarithmic color scale be useful in b)-d)? What is the impact of the detection threshold (noise level) of the station, especially for the summer season comparison – could this explain parts of the discrepancy between model and vespa in Fig. 2a?
12) Fig. 2j: One can recognize spots of maximum normalized power from south-easterly directions in the summer (not represented by the model). What could be their origin? There are probably not many potential sources in that direction (especially not for low frequencies).
13) Fig. 3d and particularly 3j: The vespagrams exhibit some horizontal lines (e.g., E and NW). Could these be artifacts of the vespa/beamforming processing?
14) l. 180 and Fig. 4: the median differences in direction of max. power are about 0-2 degree lower (by eye inspection) when using the smoothed model; the trend favoring the smoothed model is clearer recognized in the uncertainty ranges. However, if these uncertainties also correspond to the difference at the maximum power only, these are relatively large (not only at low frequency but also at the highest frequency band). How would you explain this?
15) Eq. 3: Please check if the equation is correctly noted. According to my understanding, the right-hand side is the definition of MSE(t). In this case, the equation should be modified to SI=1-MSE=1-(1/N)… or SI=1-MSE with MSE=(1/N)…
16) Eq. 3 / Figs. 2&3 / model output: The vespa analysis is done at a time step of 30min (1h time window), but the time step of the p2l data is 3h; do you interpolate the microbarom model output to 30min while smoothing or integrate the vespa over 3h? Do the time series in Figs. 2a and 3a (and b-g) differ in temporal resolution? What is the temporal resolution of the similarity index? Please briefly clarify in the manuscript.
17) l. 193-194: consider rephrasing this sentence towards SI instead of MSE; also, once SI has been defined in Section 2 (see comment 9), use SI for the axis labels in Figs. 2, 3, and 5, rather than 1-MSE.
18) l. 200: An SI of 0.5 corresponds to an MSE of 0.5, but the absolute difference between vespa and model must be even larger (and thus quite large!), due to the squared nature. In other words: For normalized distributions (within [0,1]), the MSE heavily weights small discrepancies instead of significant outliers, as opposed to when the absolute values exceed 1. Have you already contemplated using the mean absolute error instead?

19) Fig. 5: how are the data within a 3-day interval handled (mean/median, discrete)?
20) l. 239/240: “usually appears earlier” (3-24 hours) – this applies only to 2017 (and 2016), doesn’t it?
21) l. 250: “resulting *in* model-vespagram discrepancies” – Can you quantify these discrepancies caused by ECMWF wind along the infrasound path?
22) l. 252/253: Do vespagrams perform better than other methods such as PMCC in the context of SSW events? I am aware that this is not your point here. Nevertheless, in other sections, you correctly highlight the advantage of the vespa approach (all directions simultaneously), whereas in Fig. 6 you compare the back-azimuths of the dominant signals only – which are likely similar to the output of PMCC.

Technical corrections

- De Carlo et al. (2020) reference: this is not unique, there are two entries in the list matching this citation! Add a/b letters.
- l. 8: revealed --> reveals
- l. 9 - add “events”: sudden stratospheric warming [events].
- l. 16 - remove “back” (return or turn back are both appropriate, but return back looks like a tautology)
- l. 69 - Blanc et al. (2018) was referenced in the sentence before, could be saved here
- Fig. 1b - Sx/Sy = slowness components (I suggest you add this information to the caption, it is not defined in the text)
- l. 108: of the incoming signal
- l. 111 - remove “a” (or add a noun such as “approach” after “applied”)
- l. 124: to the square root
- l. 134 - the WW3 reference is missing in the bibliography
- l. 136: […] as described by Ardhuin et al. (2011).
- l. 153: assess --> determine (“assess” is also used in the next sentence)
- l. 154: Forecasting --> Forecasts
- l. 167: resolution of array --> array resolution
- l. 174/175 - rephrasing suggestion: Figures 2a and 3a show the maximum amplitude per time step over one year, i.e. the dominant signals in the azimuthal spectra.
- l. 178: accompanied with --> accompanied by [the] (or: combined with the)
- l. 179 - “applying” is redundant
- Fig. 2/3 - j) should be g) in order to avoid confusion when reading the caption (e.g., e-j)
- Fig. 2 caption - I assume that panels 2-4, 6, and 7 are b-d, f, and j(g), correct?
- Fig. 2 caption: similarity score --> similarity index (Eq. 3)
- Fig. 2 caption: the colormap reference is also given in the acknowledgments of the manuscript; consider removing it from the caption to focus on the essentials.
- l. 188/189 - why do you use negative back-azimuths instead of 266° (265°), 239° (245°), and 26° (34°), respectively? Please also add the degree symbol (unit).
- Fig. 4 - please add a unit to the y label (°); the figure size could be smaller in the final version (width of one column)
- l. 202: in the Arctic
- l. 212: promising
- l. 213: the analysis
- Fig. 5 caption: Multi-year comparison between vespagrams and smoothed modelled microbarom soundscapes at the IS37 station.

- Fig. 5: could you include the legend of the last panel *inside* this panel? Consider using different colors for this panel.
- l. 235: […] until late March or early April, which corresponds […]
- Fig. 6 caption: days --> onset days
- l. 243: […] addressed by Diamantakis (2014) and Smets et al. (2016).
- l. 249: […] demonstrated by Evers and Siegmund (2009) and Smets and Evers (2014) that […]
- l. 270/271 - rephrase this sentence
- General technical remark: no space between number and % as well as °N, °E, …
- General grammatical remark: I think you should add articles to a number of nouns.