General comments

- Detected events and their location:

1. I think the nature of the events should be better introduced. In particular, the frequency context is not discussed, nor is clear, which frequency range is actually analyzed. The text mentions, that the STA/LTA detector is applied to 2.5-20 Hz bandpass filtered data, while the location procedure is mentioned to happen in the 5-35 Hz band. I suggest to add details on the event's frequency content and on the frequencies used.

This is a good point. Recalling that event Class II is associated with relatively large MFP inferred source ranges, we see that the dominant frequency is (on average) lower than for the SPITS proximal Class I events (see Figure I). We also see clusters at distinct frequencies for event Class II, probably related to differences in the specific mining activities that trigger these events. Event Class I (interpreted as cryoseisms) span a reasonably wide range of dominant frequencies, though the most common seems to be ~17 Hz (see Figure I). We estimate the bandwidth as the maximum signal bandwidth across all array stations. The bandwidth containing 99% of the signal energy is quite high, since it is typically around 30 Hz and the waveforms are pre-filtered with a 5-35 Hz passband. On the other hand, the half-power (3 dB) bandwidth is quite low, rarely exceeding 5 Hz. This matches our observation that even though Class I events appear to be dominated by surface wave energy (which is typically dispersive), we don't observe the distinctive highly dispersive waveforms that one would expect for a wideband dispersive signal (see Figure 4 in the manuscript).

Given that the Nyquist frequency of these data is only 40 Hz, it is difficult to conclude whether the limited bandwidth is due to a physical effect or insufficient temporal sampling. Romeyn et al. (2021) showed that successive modes in a multimodal surface wavefield produced by frost quakes become dominant over frequency bands of ~8-10 Hz before the next mode takes over and becomes dominant. The ~5 Hz half-power bandwidth we observe for the events recorded by SPITS might indicate that a single mode is preferentially excited for a given frost quake, with the mode dominating the 15-20 Hz band being most commonly excited. While it is difficult to address this topic adequately in this manuscript (due to the low Nyquist frequency), it could be an interesting subject for further research. That certain frost quakes preferentially excite different surface wave modes would be an exciting result if it could be convincingly demonstrated. It would be even more interesting if one could develop a model to explain the physical mechanism by which such preferential excitation might occur (fracture geometry, depth, etc. could play a role).

Romeyn, R., Hanssen, A., Ruud, B. O., Stemland, H. M., and Johansen, T. A.: Passive seismic recording of cryoseisms in Adventdalen, Svalbard, The Cryosphere, 15, 283–302, https://doi.org/10.5194/tc-15-283-2021, 2021.

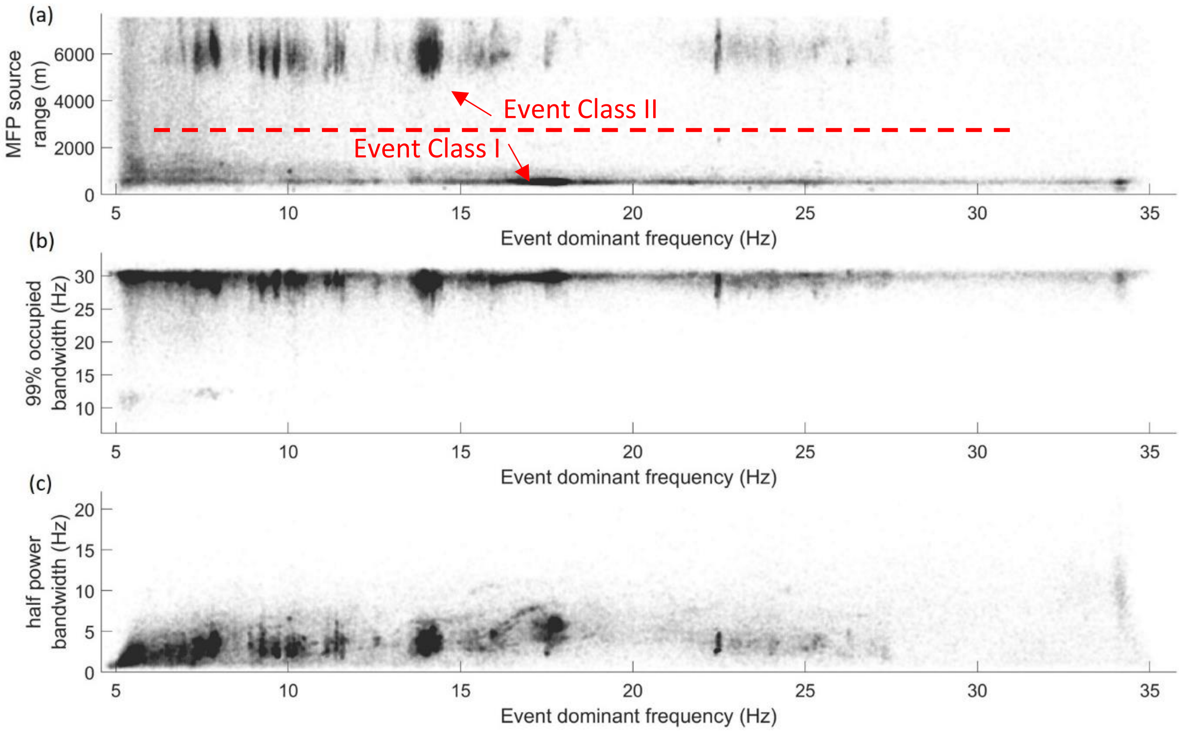

Figure I – Illustration of relationship between dominant frequency estimated by Thomson's multitaper spectral amplitude peak and (a) MFP estimated range to source, (b) bandwidth containing 99% of signal power estimated from periodogram power spectral density and (c) half-power bandwidth determined as the frequency range where power spectral density is within 3 dB of the maximum. The two event classes are most readily identified by the range to source, as annotated by the red dashed line.

In regards to the different pass-bands used in different aspects of the study. High frequency noise is important to eliminate during initial data screening since it can lead to spurious spikes in STA/LTA. When using MFP to estimate the source location of an event, spatially incoherent high-frequency random noise is not problematic and does not contribute significantly to MFP amplitude. However, If the signal contains spatially coherent high-frequency energy, this will contribute to the MFP amplitude. We found that low-frequency background noise had a higher chance of being spatially coherent. As a result, it was beneficial to eliminate the signal/noise shared bandwidth on the low frequency side, while retaining frequencies containing both signal and noise on the high frequency side for MFP. Bandpass filtering is inherently a trade-off between signal retention and noise elimination and we have attempted to optimize the pass-bands as much as possible according to the peculiarities of the STA/LTA signal detection and MFP source estimation methodologies.

This discussion is too lengthy to incorporate in the main body of the manuscript in its entirety, but we can add some of the key details from this response to sections 3.1 and 4.1 and include the figure as an appendix to the revised manuscript.

2. The authors use a cross-frequency formulation of matched-field processing to locate the seismic events. This approach favors the spatial coherence of the wavefield across frequency. However, event class I in the manuscript is interpreted to be dominated by surface wave energy, in which case dispersion should work against this spatial coherence. In my opinion, it would be interesting to compare the results of this approach with the classical formulation and a more narrow frequency band, e.g. centered around the dominant frequency of the events. In summary, I think that the robustness of the location results should be assessed.

This is also a good point. Surface wave dispersion is unlikely to play a major role for the events we consider since the amplitude spectrum is dominated by a relatively narrow band of frequencies for a given event (see response to point 1). As shown in Romeyn et al. (2021), the dispersion spectra of the multimodal surface wavefield are reasonably flat for a given mode and jump abruptly in phase velocity between modes. That the individual frost quakes recorded at SPITS are dominated by specific individual wave modes may be an explanation for the absence of highly dispersive waveforms. The MFP results are quite insensitive to the specified frequency band, particularly for the coherent-MFP formulation (see Figure II). Figure II shows that as the passband is narrowed, resolution decreases somewhat for coherent-MFP but the estimated source position remains more or less constant. The incoherent-MFP (classical formulation) performs poorer throughout and is progressively more biased towards station number 8 as the passband is narrowed. We can reject this result because Figure II-a clearly shows that the signal arrives at station 9 around the same time or slightly before station 8, so the incoherent-MFP maximum in the vicinity of station 8 (Figure II-i) must be spurious (and is likely due to interfering sidelobes).

While they appear to be robust, the accuracy of the MFP estimated source locations is worthy of further discussion, particularly in response to the reviewer's comment:

> "Line 330, Fig. 6 and especially Fig. 7: Interestingly, the three main source clusters of class I events are centered exactly around three of the array stations. This looks a bit suspicious to me, could this be an artifact in the MFP results, can you comment on this?"

In general, we would consider the MFP results as estimates that are consistent with the evidence available. The approximate vicinity of these clusters must be correct, because we have manually checked many individual events and the estimated source positions are consistent with relative arrival times and amplitudes (i.e., first arrival with strongest amplitude at the closest station, latest arrival with weakest amplitude at the furthest station etc.). However, in detail some degree of clustering/local attractor type behavior is almost guaranteed to occur due to data/model phase velocity mismatches and amplitude attenuation that doesn't exactly match the model. The error in the estimated source position will then be proportional to the amplitude spectrum weighted model/data phase velocity mismatch integrated over the signal bandwidth, convolved with the deviation from a purely geometrical spreading model of amplitude attenuation. It is very difficult to quantify this deviation *a priori* so the best way to assess MFP accuracy is by locating known test sources. As a side note, we found no location bias when locating synthetic test sources where phase velocity and amplitude variation perfectly matched those assumed by the model.

We can add a sentence to the conclusion of the revised manuscript that it would be worth carrying out additional field experiments in the future, using a set of controlled position sources, e.g. sledgehammer on steel plate, to constrain and calibrate the MFP source estimation accuracy across the study domain. While not ideal, the fact that the seismic events associated with mining activities at Gruve 7 show reasonable spatial correspondence with the location of the underground mine workings is encouraging. It's also interesting that there are only three stations that stand out as being associated with anomalously high seismicity (see Figures 6 and 7 in the manuscript). These three clusters hold up well under visual QC of the waveforms in terms of relative arrival time and amplitudes. We can also exclude that the events cluster towards all of the outermost stations in the array because there are very few events associated with the easternmost station (station 5 in Figure 4/Figure II).

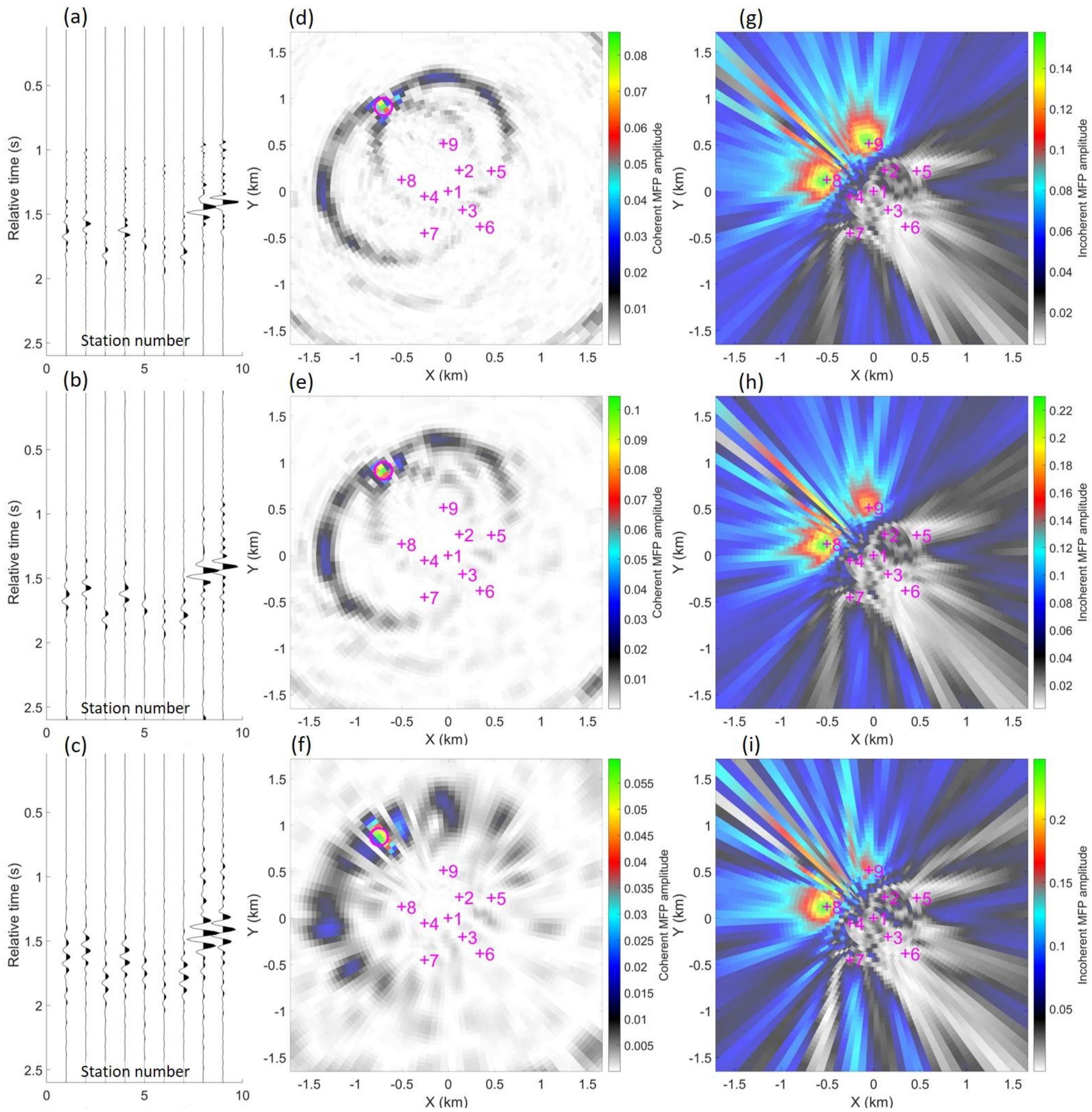

Figure II – Example of MFP source location for an event with dominant frequency of 10.2 Hz and half-power bandwidth of 6.7 Hz filtered to passbands of (a) 5-35 Hz, (b) 5-15 Hz and (c) 7.5-12.5 Hz prior to MFP. (d), (e) & (f) show corresponding coherent MFP ambiguity surfaces while (g), (h) and (i) show incoherent MFP ambiguity surfaces.

3. The aperture of the array does not seem to be ideal for the analyzed events. Given a minimum interstation distance of roughly 250 m and an aperture of 1 km , as stated in the text, the resolvable wavelength range according to Tokimatsu 1997 (see also Wathelet et al., 2008) is roughly 500-3000 m. Given the frequency range of 2.5-20 Hz (?) and the determined velocities of 1150 m/s (class I) results in considerably smaller wavelengths (while class II events seem well suited for the array aperture). I am not saying this will not work, but there should be some discussion again on the robustness of the results.

   Reference: Wathelet, M., Jongmans, D., Ohrnberger, M., & Bonnefoy-Claudet, S. (2008). Array performances for ambient vibrations on a shallow structure and consequences over Vs inversion. Journal of Seismology, 12(1), 1-19.

It is absolutely correct that the station spacing of the SPITS array is not ideally suited to sampling the wavelengths of interest in this study. We have found it very challenging to extract high-quality dispersion curves from these data using the normal phase-shift methods, in line with the Wathelet et al. (2008) study and the reviewer's calculations. This is also not a surprising result because the primary purpose of the SPITS array is to record regional or teleseismic earthquake signals that are dominated by much longer wavelengths. However, a useful property of the MFP method is that the recorded wavefield is "compared" to a model wavefield and the source position is taken as the location giving the strongest coherence between recorded and modelled wavefields. In this way MFP is less susceptible to phase wrapping issues that limit the use of small wavelengths in normal phase-shift dispersion spectrum estimation.

The array layout certainly affects the precision with which source locations can be estimated. However, as mentioned in the previous response, this is as much a function of the data-model misfit as it is a function of bandwidth, the number of array elements and their spacing. We think Figure 4 in the manuscript provides a good overview of the lobe pattern that is observed in practice for seismic sources at different ranges and azimuths (these plots are often referred to as MFP localization ambiguity surfaces). As discussed in the following response (point 4), the carefully calibrated instrument responses of the SPITS seismometers also provides a significant benefit over typical temporary geophone deployments having a similar arrangement when it comes to MFP.

4. From experience (and this comment is a bit out of curiosity), there is typically some source smearing for events outside the array such that the distance of the sources cannot be well constrained. I would expect this to happen also for the class II events, but it does not seem to be the case. Also from Fig. 4F, the distance seems to be quite well constrained. Can you comment on that?

We observed much improved source range constraint using coherent-MFP than incoherent-MFP (see revised version of Figure 4 given later in this response). In addition, some MFP implementations only use phase information, but we found that the variation of amplitudes across the array gives important additional constraint when estimating the source position. For example, two sources positioned along a common azimuth located well outside the array will have very similar relative arrival times across the array (which is why distance is relatively poorly constrained by only considering wave phase). However, the furthest source will have less variation in amplitude across the array because the array aperture is a smaller fraction of the total propagation distance. The closer source will produce a larger contrast in amplitude between the nearest and furthest array element because the array aperture is a larger fraction of the total propagation distance. The contrast in amplitudes across the array is therefore a key constraint on the source range, in particular.

In this study, we benefit substantially from the fact that the SPITS seismometers are carefully calibrated. Inconsistent ground coupling and/or poorly constrained instrument responses both work to reduce the reliability of amplitude information for temporary seismic deployments (such as Romeyn et al. 2021) where it then becomes necessary to normalize amplitudes or discard them entirely and only consider the signal phase in MFP, such as in Walter et al. (2015), Chmiel et al. (2016), etc. The relatively large interstation distances at SPITS (aperture is ~1 km) are also beneficial in this specific case, because amplitudes attenuate quite appreciably across the array for the local (<~10 km) events that we studied (so that contrast in relative amplitudes is quite well resolved).

The accuracy of the range to sources far from the array depends largely on how well the assumed geometrical spreading model represents the true amplitude variation across the array. As noted in point 2, the fact that the seismic events associated with mining activities at Gruve 7 show reasonable correspondence with the location of the underground mine workings is encouraging and indicates that the geometrical spreading model is a reasonable approximation of reality.

Romeyn, R., Hanssen, A., Ruud, B. O., Stemland, H. M., and Johansen, T. A.: Passive seismic recording of cryoseisms in Adventdalen, Svalbard, The Cryosphere, 15, 283–302, https://doi.org/10.5194/tc-15-283-2021, 2021.

Chmiel, M., Roux, P., and Bardainne, T.: Extraction of phase and group velocities from ambient surface noise in a patch-array configuration, Geophysics, 81, KS231-KS240, 2016.

Walter, F., Roux, P., Roeoesli, C., Lecointre, A., Kilb, D., and Roux, P.-F.: Using glacier seismicity for phase velocity measurements and Green's function retrieval, Geophysical Journal International, 201, 1722-1737, 2015.

Terminology: The wording could be more consistent. The text jumps around between e.g. ice wedge and segregation ice or frost quake and cryoseism. If it is not the same that is meant, please further specify each of the concepts.

This is a good point and was also brought up, in particular, in RC4. As we discussed in the author response to RC4, we suggest adding a conceptual model and revising the associated terminology to make it clearer what processes have been interpreted.

- Thermal-stress model: It is mentioned, that ignoring some of the temperature dependences (lines 244-248) results in different model formulations, that other studies used previously. Why do you chose this specific model and how would your results be affected by e.g. using the model proposed by Mellon (1997), or Podolskiy et al. (2019)?

These models can be considered a closely related family and we wanted to show clearly how they are related to one another. The simplifying assumptions made by Mellon (1997) and Podolskiy et al. (2019) were completely appropriate to those studies. Primarily, our formulation provides a convenient basis to compare those models. If we implemented the Mellon (1997) or Podolskiy et al. (2019) assumptions the modelled thermal stress would be broadly similar. However, when attempting to correlate the observed and predicted number of frost quakes at the level of detail of e.g. Figure 10, the second order contributions to thermal stress become relevant. It is important to note that the previous studies model the ground temperature profile, while we use a detailed record of in-situ borehole ground temperature measurements. This is perhaps the most significant factor that allows us to delve a little further into the fine detail of second order terms in the thermal stress model than was appropriate for, e.g., the Mellon (1997) and Podolskiy et al. (2019) studies.

Line-specific comments

Line 27: pressure release → stress release?

Yes we agree, stress release is more accurate terminology. This will be changed in the revised manuscript.

Lines 39-40: “These structures form …”. I am having trouble to understand this process, maybe consider rewriting this sentence.

The purpose is to communicate the formation of the wedge structures that lead to the surface expression of ice-wedge or sand-wedge polygons. The development of these wedge structures is well documented in the literature but we see the need to improve the clarity of the text and suggest changing to the following:

> "Ice-wedge or sand-wedge polygons are a widely observed geomorphic feature in the periglacial environment (e.g. Black, 1976; Matsuoka et al., 2004). These wedges form when water that infiltrates and freezes to ice, or wind transported sand grains, hold open an initial thermal contraction crack that subsequently becomes an enduring plane of structural weakness (Lachenbruch, 1962; Mackay, 1984; Matsuoka et al., 2004; Sørbel and Tolgensbakk, 2002)."

Lines 49-57: What's the difference between ice wedges and segregation ice. As far as I understand one can broadly distinguish them as vertical and horizontal ice structures in the subsurface, respectively? Consider to add some definition here, if applicable.

Good point. We have addressed this more extensively in the author response to RC4, where we suggest adding a conceptual model and description of the corresponding physical processes. We agree that this wasn't covered in enough detail in the initial manuscript.

Line 61: "… InSAR has used …" → has been used

Thanks for the catch, this will be changed in the revised manuscript.

line 73: "This study was motivated" → is motivated

Perhaps this is simply a stylistic preference, but past tense seems appropriate in this context given that we explain the initial study hypothesis development which precipitated the study.

line 73: sporadic? From the paper it seems there are quite many of these events?

Sporadic in the sense that the events occur at irregular intervals, in total they add up to a large population, but they are not recorded all the time or at regular intervals. We suggest that an appropriate substitute that might improve readability in this context would be "intermittent".

Line 96: Maybe add a reference after matched field processing, that describes the "broadband, coherent" approach? Because that's the special part in this study, right?

Good point, a reference to Michalopoulou (1998) would not be out of place here. On reflection, it would be better to be more general here e.g. "well suited to source localization using matched field processing" since this section is just introducing the data. The specifics of the selected matched field processing approach can then be introduced and developed in the methods section. Coherent-MFP is not really the special part of the study *per se*, since it was developed quite some time ago for ocean acoustics. However, we did find it to be a very suitable tool for this study and it is nice to be able to highlight that a technique developed within ocean acoustics can also find useful application in other fields.

Line 127: So the weather station is measuring the air temperature plus the temperature of the ground in 0.1m depth? Please clarify.

Yes, exactly. We suggest the following minor adjustment to make this clearer:

"… the Janssonhaugen Vest weather station (see Figure 1), which was installed in September 2019, includes hourly sampled records of air temperature and ground temperature at 0.1 m depth. It therefore provides a basis to compare depth and temporal sampling effects against the longer duration, more coarsely sampled P11 record."

Line 131: What do you mean by "first-pass"?

We mean that the detection is a coarse initial step in a signal identification/classification procedure that contains multiple steps (e.g. the coal mine events are subsequently separated out based on inferred source location). We fully agree that the readability can be improved here and suggest the following revised text:

> "The purpose of the event detector is to make an initial, coarse, automatic identification of short duration seismic signals, which should be distinguished from both background noise and longer duration local and regional seismic events…"

Lines 137 and following: A bit difficult to follow here – for the STA you take the envelopes and smooth them with a 1s sliding windows and for the LTA you smooth this curve once more with a 20s sliding window? Please clarify and maybe rewrite the text.

Yes, this is correct. We will clarify this in the revised manuscript.

Line 180: I think it should be the absolute value of the term after the sum. As is, it would be a complex MFP amplitude. Same for equation (7). Please check.

No, there shall not be an absolute value sign in Eqs. (4) and (7). These matrix operations are quadratic forms that guarantee a non-negative real valued output. Note that the original papers from ocean acoustics that introduced the MFP (and their variations) got it right (e.g. Michalopoulou et al. 1998), but that the (unnecessary) absolute value sign has somehow later found its way into the geophysics literature. We suggest to add the sentence:

> "The matrix operations in Eq. (4) are quadratic forms that formally guarantee a non-negative real-valued output."

Line 254-256: So you basically do a forward modeling using the measured temperature time series at a certain depth (and the parameters from Table 1) to calculate the resulting stress at this depth? If so, maybe strengthen this point here.

Yes, this is correct. We agree that this point can presented more clearly and suggest we replace the sentence

> "We assume the stress at a given depth is decoupled from the stress at adjacent depths (Mellon, 1997) and solve Eq. (12) for the temperature timeseries at the selected depth of investigation."

with

> "We solve the forward model Eq. (12) numerically to obtain a time series of the resulting horizontal stress $\sigma(z,t)$ at depth z. The crucial input to the forward model is the measured temperature time series $T(z,t)$ at depth z."

Correspondingly, we will add a short statement at Line 224 that we assume the vertical stress is zero, in line with Mellon (1997), corresponding to the assumption that the overburden is negligible in the shallow subsurface.

Line 280: So only less than 100 events were recorded by less than five stations and thus discarded?

Yes, that is correct. It makes sense to apply this check as part of a general procedure, because we were aware that instrument downtime due to maintenance or malfunction is a potential issue. It was therefore encouraging and a testament to the ongoing planned maintenance conducted by NORSAR that nearly all of the detected events were recorded by at least five seismometers. For the purpose of transparency, it is important that this is communicated, even though the number of discarded events was small.

Line 293: “are” is missing before “associated”

The “are” comes a few words earlier because of the chosen sentence formulation:

“Using coherent MFP, we find not only are these events associated with more distant sources (Figure 4f), they also have a consistent azimuth.”

But we can happily rephrase this sentence to improve readability:

“Using coherent MFP, we find that these events are associated with more distal inferred source positions (Figure 4f) that also have a consistent azimuth.”

Line 294: I see that compared to your previous study, nine seismic stations can be considered an array that coarsely samples the spatial domain, but I think this cannot be considered a general statement. Maybe relate this to your previous study.

It is clear that for these data, the SPITS array coarsely samples the spatial domain, as was commented on in major comment number three. The general statement that coherent MFP is most beneficial for arrays that coarsely sample the spatial domain (when compared with incoherent MFP) was made by Michalopoulou (1998). We see that our data support Michalopoulou’s findings and that is what we wish to communicate with this sentence. We suggest the following modification to make this clearer:

“coherent MFP decreases source localisation ambiguity for arrays that sample the spatial domain coarsely for a given range of observed wavelengths when compared to the incoherent scheme (consistent with Michalopoulou, 1998).”

Line 319: delete “due”, same in line 326.

Sure, we will remove these in the revised manuscript.

Line 341 and following: This relates to a previous comment: To calculate the stress at a certain depth, does only a single temperature time series from this particular depth enter the calculation, or does it also include the vertical temperature gradient? What does the word “combination” in line342 imply?

“Combination” highlights that the measured ground temperature field consists of two datasets, one from the weather station (0.1 m record) and one from the borehole (0.2-15 m).

The second part of the question is answered in the response to the previous comment on Line 254-256.

Line 353: "Figure 9 …" → Figure 9a. It would also be interesting to show the event rate (e.g. events/day) as a line together with the calculated stress in Fig. 9b.

In order to give the event rate, it is necessary to apply histogram binning. We prefer to represent the event rate histograms as bar rather than line plots. We will clarify in the figure and caption that the black bars in Figure 9b show the event rate binned to a 6-hour interval.

Figures

Figure 2: It took me a while to understand what's actually shown, since this is a continuous time series split into several subfigures. I would either merge the graphs of each row and/or write the year as text into the graphs, to make it easier for the reader.

This is a good point and we suggest the following revision to improve the readability of the figure. The splitting of the continuous timeseries is now mentioned explicitly in the figure caption to improve clarity.

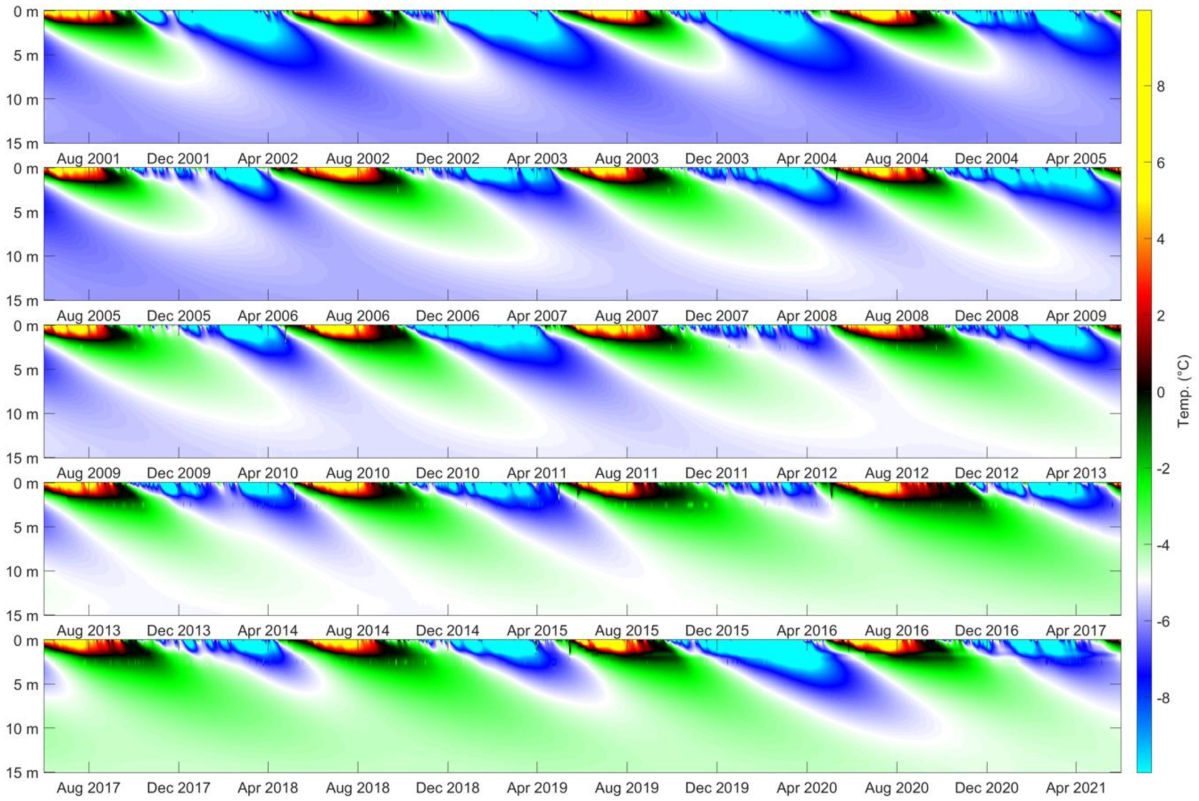

Figure 2 – Illustration of spatiotemporal borehole temperature recorded at the PACE P11 borehole on Janssonhaugen. A long-term warming trend is observed below the active layer that is subject to seasonal freeze-thaw. The continuous timeline is split across multiple figure panels.

Figure 3: a) and b) are missing, but are referenced in the text. Also, in the caption, please provide more detail on what is actually shown.

Thanks for picking up on this oversight. We suggest the following revised figure and caption to address this (which also addresses comments from RC2):

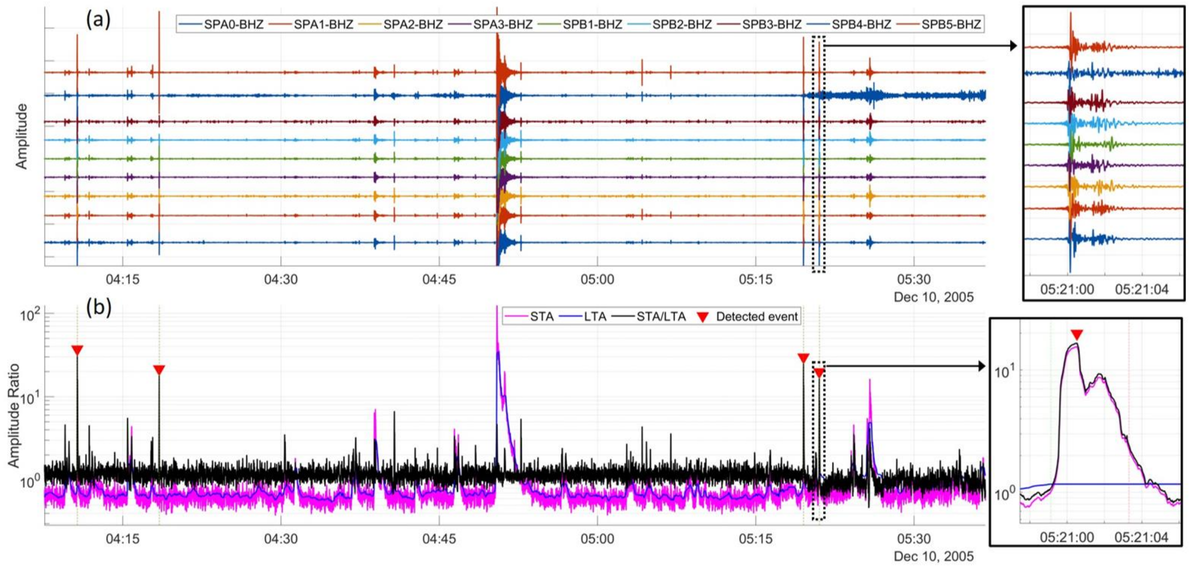

Figure 3 – Example of (a) timeseries of vertical component ground motion and (b) event detection using the STA/LTA (short term/long term average) detector. Short duration events with sufficient amplitude and array coherence are selected while longer events such as the high amplitude example at 04:50 are ignored. Inset boxes show detailed views for a specific detection.

Fig 4: The crosses of the stations are hardly visible in subplots g, h, i

Thanks for the feedback. We suggest the following modified figure to improve readability:

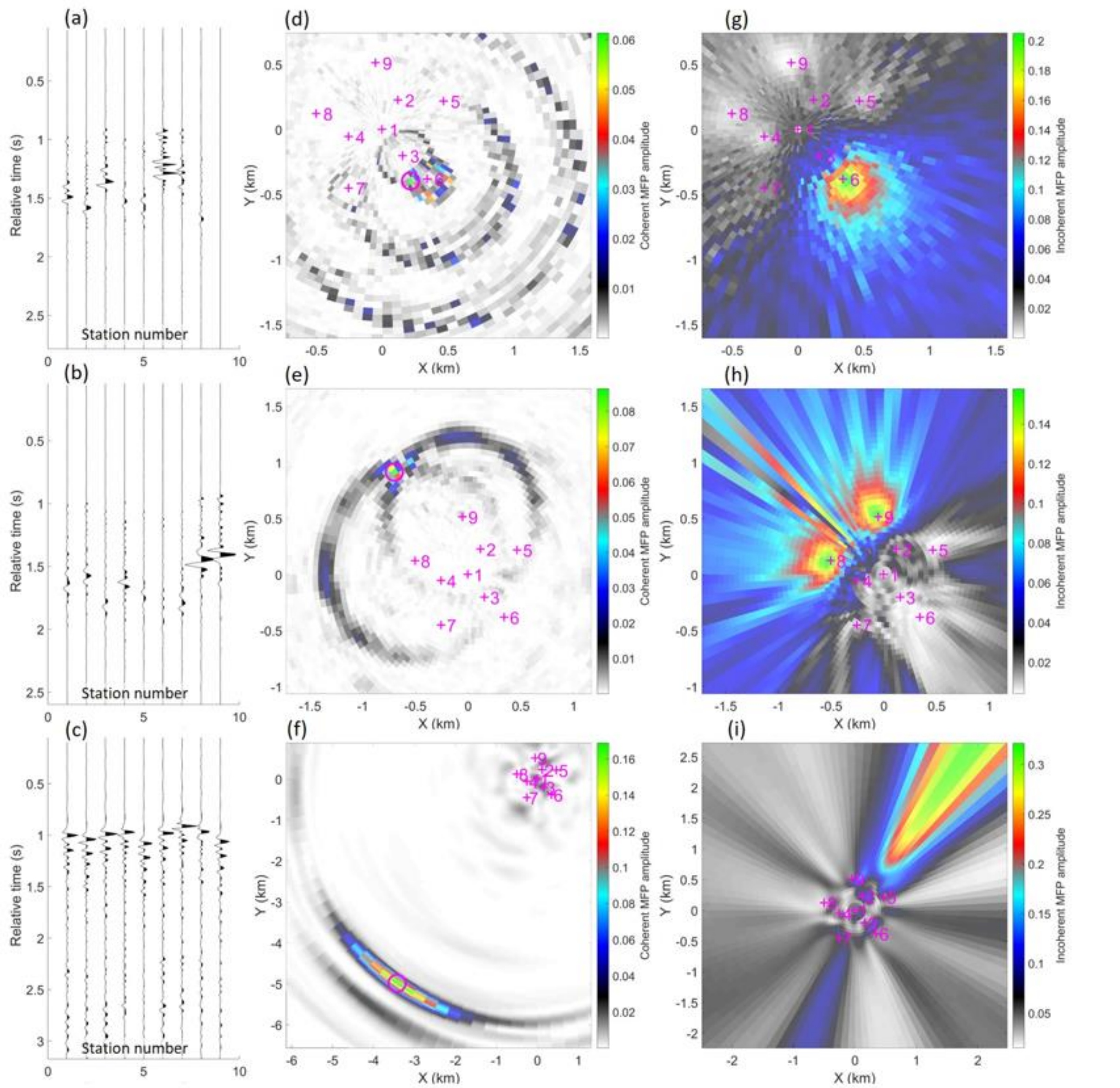

Figure 4 – (a), (b) Examples of Class I events with significant amplitude variation across the array and relatively close source position inferred by (d), (e) coherent MFP. (c) Example of a Class II event with little amplitude variation across the array and a more distal (f) coherent MFP inferred source position. (g), (h) & (i) show corresponding incoherent MFP results, Eq. (4), demonstrating the improvement gained by coherent MFP. Station numbers in (a), (b) & (c) correspond to the labels annotated on the MFP panels where the seismometer locations are marked with black crosses.

Fig 6: The seasonality is hard to see from the figure. I suggest to give the total number of events shown in each panel e.g. in their titles.

Thanks for the suggestion, we'll add this to the revised manuscript. We will also add the total number of events shown in each figure panel to Figure 5 for consistency.

Fig 7b-d: Being non-trained in this, it is difficult for me to spot the boulder producing scarps and solifluction lobes. Consider adding annotations to the images.

This is useful feedback. There is a fine line between adding annotations and obscuring the details in the images that are interpreted. Ideally one could include both fully annotated and unadorned figures, but in the interests of brevity we suggest the following revised figure to retain some of the orthophotographic details while adding a sufficient level of annotation to guide the reader's eye:

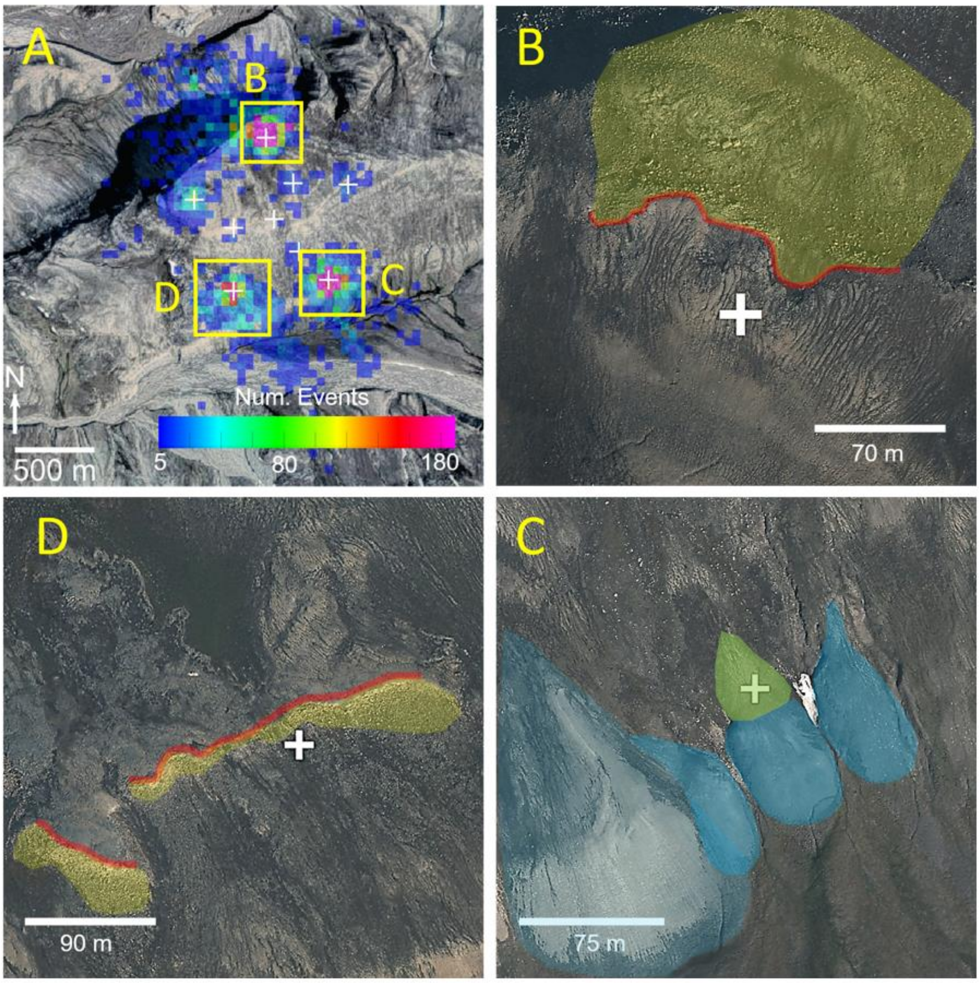

Figure 7 – (a) Dec-Feb events as plotted in Figure 6 with orthophotograph details illustrating the geomorphologic features associated with the most seismically active areas. Erosional scarps (b), (d) are annotated in red and associated boulder fields in yellow, and frozen debris/solifluction lobes (c) are annotated in blue. A faintly visible area of polygonal patterned ground is annotated in green in (c). Orthophoto © Norwegian Polar Institute (npolar.no).

Fig 10: Maybe it would be better to show the event detection rate as a line instead of the vertical bars? What is the apparent stress? Please specify.

The event detection rate is shown as a frequency histogram in Figure 10b. In Figure 10a the colored bars help to avoid clutter. It is a good point that we could improve the labelling to make it clearer that the detection rate shown by coloured bars in Figure 10a is essentially the same as the "Array local events" in Figure 10b. We suggest labelling both as "Event Class I detections" to improve this.

Apparent stress is the modelled thermal stress corrected by dissipation due to tensile cracking. It is a good point that this should be more clearly defined in section 3.3.1 (Line 264-272) and a cross reference could usefully be included in the figure caption to guide the reader. Upon reflection, "pre-fracture stress" might be more descriptive than "potential stress" and "post-fracture stress" would make a good substitute for "apparent stress".

Fig 11: I think it would be instructive to show only the class I events and again maybe as a line or as bars. You have shown that earlier, that class II events are independent of the thermal stress, so it would be better to focus on your finding that the closeby events are related to the stress.

This is good feedback. We have considered this possibility, but in order to show the class I events as lines or bars representing frequency, we would need to apply histogram binning. This is what we show in Figure 12, where we make a direct comparison to the modelled number of frost quakes accounting for tensile strength and stress release. We also think that it is quite useful that Figure 11 gives an overview of the results of the study using data in the rawest form practical.