1 General Comments

This article aims to study the hypothesis that short-duration seismic events at SPITS can be due to the frost quakes induced by thermal contraction cracking. The main interest of the study lies in the identification of seismic events that correspond to frost quake initiated by thermal stress.

Two event classes were defined by authors and the second class of seismic event was concluded due to the underground mining operations. The first event was found to be active during winter (Dec-Feb) and spring (Mar-May) and inactive in the summer time.

The article is well written and the research is interesting. A contribution from this paper is that the authors used the ground temperature measured by a series of thermistors in the numerical analysis. However, the proof of the hypothesis (especially for the frost action induced seismic events) is still not convincing. In my opinion, the paper needs to be revised to convincingly explain how frost action can initiate the dynamic response. The following comments are also needed to be addressed.

We thank the reviewer for their comments. We agree that strengthening the highlighted aspects will improve the revised manuscript and provide detailed responses to the specific comments covering these issues below.

First, the authors should include a fuller explanation of the physical mechanism of frost quakes and how the frost action could induce dynamic responses.

Thanks for this feedback. We can agree that the physical cracking mechanism should be presented in greater detail. This was also commented on by reviewer 4. As also stated in the author response to RC4, we suggest adding **Error! Reference source not found.** and the following description to more clearly illustrate the investigated cracking mechanism.

> Polygonally patterned ground indicative of cryoturbation of the active layer and/or ice wedges in the underlying permafrost are observed extensively across Janssonhaugen, except where downslope mass movements destroy or interrupt the formation of polygonal networks (Sørbel and Tolgensbakk, 2002). In this study, we observed anomalously high cryoseismicity associated with erosional scarps and a frozen debris/solifluction lobe, which are all associated with downslope mass wasting. We suggest that these downslope mass movements initiate or precondition transverse cracks or fissures to open (e.g. Darrow et al., 2016, Price, 1974). Water from rain or snowmelt infiltrates these fissures and freezes when temperatures drop below freezing (e.g. Darrow et al., 2016). Rapidly falling ground temperatures during winter then cause the surrounding ground to thermally contract. This causes the vein ice filling the frozen fissures to crack under tension, since the tensile strength of ice is lower than the surrounding ground and therefore constitutes a plane of weakness. Cracking relieves the accumulated thermal stress.
>
> This mechanism is analogous to the Lachenbruch (1962) model of thermal contraction cracking of ice wedges in permafrost. However, the cracking may occur more frequently or under milder surface cooling because 1) the frozen fissures may be pre-stressed by downslope gravitational forces making them more prone to failure and 2) the fissures extend to the ground surface where thermal stresses are largest. For the case of ice wedges, the ice wedge is located below the permafrost table (**Error! Reference source not found.**-c). There may be vein ice extending through the active layer only if the previous seasons thermal contraction crack has remained open through the summer thaw season. If the crack has closed so that vein ice is not formed during the early freezing season, initiation of thermal

contraction cracking of the ice wedge would require accumulation of thermal contraction stress at the level of the permafrost table, requiring a longer or more extreme surface cooling episode. Tensile cracks could also form in the active layer where ice veins/ice wedges are absent (as in **Error! Reference source not found.**-a), but this would require thermal contraction stresses exceeding the tensile strength of frozen ground, which is greater than that of polycrystalline ice and may therefore occur less frequently.

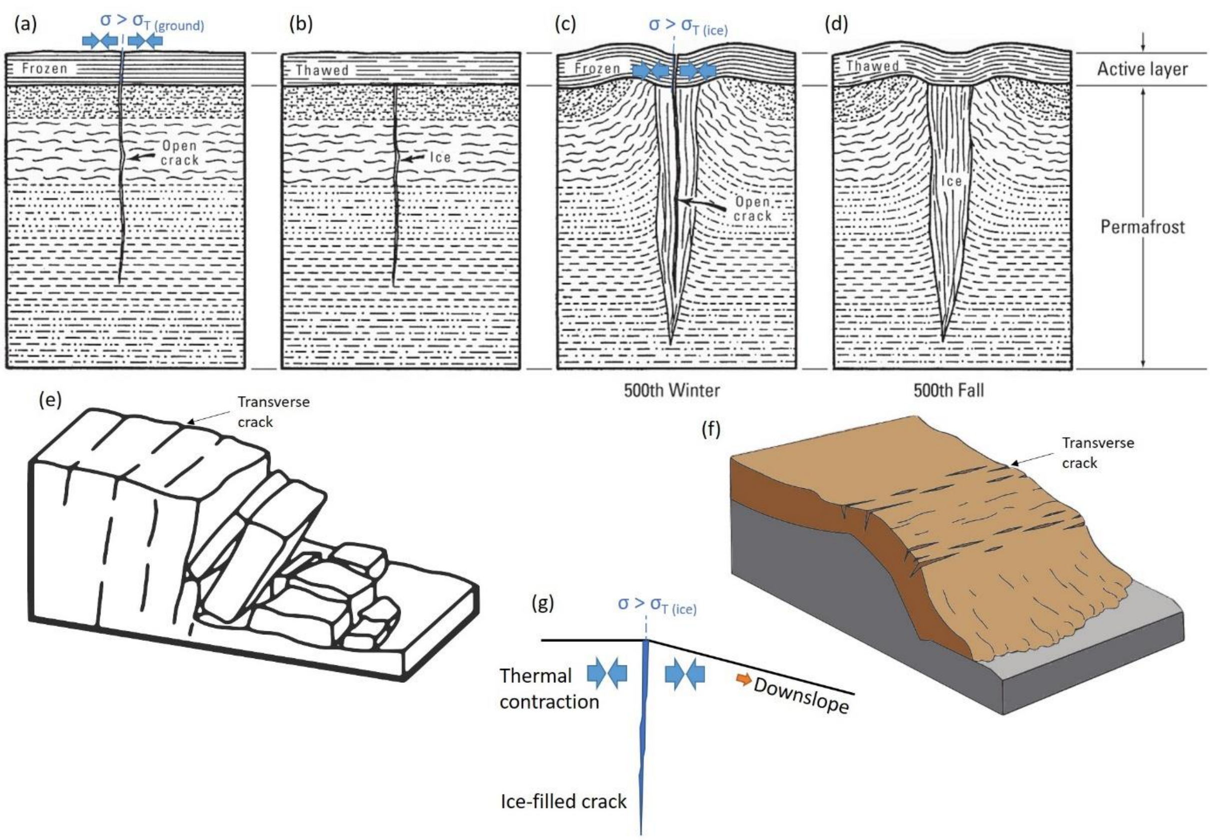

Figure I - (a-d) Lachenbruch (1962) model of ice wedge formation. (a) Thermal contraction of frozen active layer initiates a tensional crack that penetrates into permafrost, (b) meltwater infiltrates and refreezes during the thaw season, (c) ice with lower tensile strength than surrounding ground forms a plane of weakness and cracks repeatedly over many years, (d) the crack-infilling cycle causes ice wedge growth and ground surface deformation that organizes into ice wedge polygons in 2D plan view. Mass movements such as, (e) erosional scarp and (f) frozen debris/solifluction flow initiate transverse cracks/fissures which become infiltrated by water that freezes to ice. (g) Thermal contraction of the surrounding ground and downslope gravitational stress causes the ice to crack under tension.

Second, the classification of two classes of seismic events must be addressed more quantitatively, in order to classify and detect frost action related seismic events.

We have addressed why short duration signals are studied in response to several specific comments from the reviewer below. The two classes of events are ultimately distinguished based on their spatial distribution; those events located in the vicinity of the underground mine workings are assumed to originate from human activity in the mine.

Third, the mismatch between the wavelength of seismic events (could be more than 115 m) and the studying depth (top 15) needs to be better explained or addressed.

We move the specific comment to Line 299 here since it covers the same issue: Authors stated ’The mean MFP inferred propagation velocities for Class I events was 1150 m/s with a standard deviation of 1100 m/s, indicating that they are dominated by surface waves. The large standard deviation may indicate the surface waves are dispersive with different frequencies propagating at different phase velocities.' In line 132, the authors also mentioned the signal is filtered to a range of 2.5-20 Hz. This gives us a wavelength around 115 m (1150 m/s divided by 10 Hz, average frequency) and the investigation depth in this paper is only about the first 15 m. The authors should explain this mismatch.

It is difficult to understand exactly what mismatch is referred to. We present the ground temperature field measured to a depth of 15 m, that suggests a possible seismic source via thermal contraction cracking at a depth around 0.2 m. We also present seismic records of surface ground motion and show that the spatial and temporal distribution of a specific subset of seismic signals is consistent with a thermal contraction cracking source. A crack/rupture of small size can produce a seismic surface wave signal with long wavelength. Frequencies down to 3 Hz are not-uncommon in sledgehammer surveys for MASW and these are perhaps even less energetic seismic sources. There is therefore no contradiction in associating signals with relatively long wavelength with a small magnitude, shallow seismic source.

It is likely that the thermal contraction cracks also excite body waves of higher frequency. The reason we don't observe these might be that they are above the Nyquist frequency (40 Hz) and/or that attenuation is too large. However, surface waves can be intuitively explained as constructive interference of body waves trapped near the surface and the resulting horizontal wavelength can be much longer than the wavelength of the dominant frequency (corner frequency in source spectrum of ground displacement) of the body waves excited by a small magnitude seismic source.

We must emphasize that we have not conducted a MASW study where we try to estimate the shear-wave velocity depth profile of the uppermost 15 m. If this were the case, the long wavelength of the recorded seismic signals would likely be problematic.

2 Specific Comments

Line 32: I would add a few sentences to explain why ground cooling induces cracking. Is it because of the volumetric expansion during the phase change from water to ice? Or it can be due to the formation of ice lenses which is associated with the water migration during the freezing process? Can also thawing contribute to cracking?

Thanks for this feedback. It is a good point that it would give a more complete picture to introduce both the thermal contraction and segregation ice mechanisms of cracking in this section. We suggest the following formulation:

> "Frost quakes are typically observed in association with rapid air and ground cooling, in the absence of an insulating snow layer and where sufficient moisture is present for ice to form (Barosh, 2000; Battaglia et al., 2016; Matsuoka et al., 2018; Nikonov, 2010). The source of frost quakes are cracks that may be initiated by different mechanisms including thermal contraction exceeding tensile strength (e.g. Lachenbruch 1962) or by the growth of segregation ice driven by the capillary migration of water to a freezing front (e.g. Walder and Hallet 1985, Peppin and Style, 2013). In this study we focus on the thermal contraction cracking mechanism, which is consistent with the association of transient ground acceleration events with rapid cooling episodes and cold winter temperatures reported by

Matsuoka et al., (2018) and consistent with previous descriptions of frost quakes (Barosh, 2000; Battaglia et al., 2016; Nikonov, 2010; Okkonen et al. 2020)."

We are not aware of a mechanism by which thawing can directly contribute to cracking, to our knowledge desiccation cracking would be the closest. However, evidence that soil desiccation cracks produce measurable acoustic emissions or excite seismic waves seems to be limited.

Line 34: Frost heave is an upward swelling of soil due to an increasing presence of ice as it grows towards the surface (continuously delivers water to the freezing front via capillary action). I am having difficulty understanding that frost heave is ' rapid' and also ' elastic' deformation. The frost heave requires delivering unfrozen water contentiously to the freezing front via capillary force, which is likely a slow process (Darcy' s Law). It is not necessarily elastic either given its large deformation.

Good point. The term "frost heave" was poorly placed in this sentence, which could be better formulated as:

"Cryoturbation can be understood as a combination of slow creep (frost heave, e.g., Rempel, 2010) and rapid elastic cracking (frost quake) deformation of frozen ground and causes damage to roads requiring billions of dollars annually to repair in the United States alone (DiMillio, 1999)."

Line 49: What is the mechanism for the segregation ice growth in bedrock? If it is the same as the frost heave, where does the capillary force come from?

We suggest changing the order of sentences and to split up the references to previous studies, which should make this clearer. We propose the revised formulation:

> "Frost cracking driven by segregation ice growth is also an important agent of bedrock erosion in cold mountainous areas, where rockfall, active screes and high headwall erosion rates are observed in areas where frost erosion is most intense (Hales and Roering, 2009; Hales and Roering, 2007; Scherler, 2014). An important mode of crack growth in water permeable bedrock is the migration of water to form segregation ice bodies (Hallet et al., 1991; Murton et al., 2006; Walder and Hallet, 1985) that is similar to the mechanism by which ice lenses develop in freezing soil (Peppin and Style, 2013). Segregation ice growth, frost heaving and creep on slopes leads to the development of solifluction lobes and sheets (Cable et al., 2018; Matsuoka, 2001). Solifluction is broadly defined as the slow mass wasting resulting from freeze-thaw action in fine-textured soils (French, 2017; Matsuoka, 2001) and occurs due to the asymmetry between frost heaving perpendicular with the sloped ground surface and vertical subsidence upon thawing under the force of gravity."

Line 119: It would be useful to also provide the temperature distribution in the permafrost at the location of P11 (even in the supplementary information).

Since the active layer is ~2 m thick, the interval from 2 m to 15 m is permafrost. Perhaps a misunderstanding arose around the sentence (Line 123-125):

> "The ~2 m thick active layer (Christiansen et al., 2020) is sampled by thermistors at 0.2, 0.4, 0.8, 1.2, 1.6 and 2 m and there is significant inter-annual variability in the magnitude of summer warming and winter cooling."

For clarity we could add a following sentence:

> "The upper part of the permafrost at the P11 borehole location is sampled by thermistors installed at 2.5, 3, 3.5, 4, 5, 7, 10, 13 and 15 m."

The temperature distribution in the uppermost permafrost is therefore already shown in Figure 2. The permafrost deeper than 15 m is not considered relevant to this study because temperature changes occur over long inter-annual timescales (the reader is referred to Isaksen et al. 2001, who analyze the deeper interval). As shown in Figure 8 the thermal stresses are already quite small below ~5 m.

Line 131: Authors need to provide more explanation to illustrate why cryoseismic events have a shorter duration than other seismic events.

In this section of the manuscript, the events have not been interpreted as cryoseisms, we are just describing a method to isolate short duration seismic events. The reviewer's point remains valid though, and we can add a sentence to the introductory paragraph where the study hypothesis is outlined (Line 74) explaining this.

The main reason why cryoseismic events have a shorter duration than other seismic events is that the source is much closer to the array. The different propagation velocities of P, S and surface waves (dispersion) mean that the larger the source distance becomes, the more spread out the wavefield becomes and the longer the ground motion duration will be. We can add a sentence or two to the revised manuscript to address this point.

A secondary point is that duration of ground motion can be used as a metric of earthquake magnitude (e.g. Lee et al., 1972; Mousavi and Beroza, 2020; Tsumura, 1967). It is also well established that magnitude correlates with the area of rupture, length of surface rupture etc. (e.g. Wells and Coppersmith, 1994). Shorter duration seismic events can therefore also be indicative of rupture along smaller cracks, with less energy release than larger magnitude tectonic earthquakes. Such small magnitude events are only recordable locally because the amplitude of ground motion decays below the background noise level over longer distances. Romeyn et al. (2021) showed that cryoseismic events could also be recorded in Adventdalen ~10 km from Janssonhaugen. We were not able to detect these same events at SPITS (presumably because the seismic waves had attenuated below the background noise level at this range), though the same ground cooling episodes were associated with similar clusters of events with estimated source positions in closer proximity to Janssonhaugen.

Line 136: It would be useful providing any physical interpretations of the ratio of short-time averaged amplitude and long-time-averaged amplitude (explain why this ratio can be used to detect cryoseismic events). I would also add a few sentences to explain why choose Hilbert transform for the short-time-averaged amplitude and moving average for long-time-averaged amplitude.

The STA/LTA method is simply a way to automatically identify higher-amplitude transient signals embedded in background noise and has been commonly used for decades in seismology. Thermal contraction cracking rapidly releases accumulated elastic stress, exciting transient seismic waves (the energy release is not continuous) that are recorded as transient signals of ground motion. However, the transient signals detected by an STA/LTA detector could have any origin so it would not be appropriate to interpret the detections in isolation. We do filter transient signals according to their duration within this procedure, which as we have responded elsewhere is because longer duration signals correspond to longer rupture, higher magnitude, more distal earthquake sources. Importantly, the STA/LTA detections also included seismic events associated with mining activities. We were able to interpret a specific category of detections as cryoseismic events only in combination with MFP to estimate their source positions and temporal correspondence with periods of rapid cooling and increased thermal stress.

The Hilbert transform is only used to calculate the trace envelope and this is also common practice in signal processing.

To improve this section, we suggest rephrasing to:

> "Events are detected based on anomalous values of short-time-averaged (STA) amplitude divided by long-time-averaged amplitude (LTA), i.e., the classic STA/LTA approach widely used in seismology (e.g. Allen, 1982; Trnkoczy, 2009). The STA window length should be comparable to the target signal duration, while the LTA represents the background noise level. When the STA/LTA ratio exceeds a given threshold, an event is triggered. In our implementation, the STA is given by the one-second moving average smoothed trace envelope for each seismogram. The LTA is the STA further smoothed according to a 20 second period moving average."

Line 141: Authors stated that 'By visual inspection of test periods, we found that this emphasizes very local events with large amplitude variation across the array, while still ensuring that there is at least some coherency across the array'. It would be better to explicitly indicate what 'this' represents. More importantly, there should be a figure to show what authors captured by their visual inspection and prove how it emphasizes local seismic events.

We suggest re-phrasing to:

"By visual inspection of test periods, we found that the 80th percentile station-STA emphasizes short duration events with large amplitude variation across the array, while still ensuring that there is at least some coherency across the array (see Figure 3)."

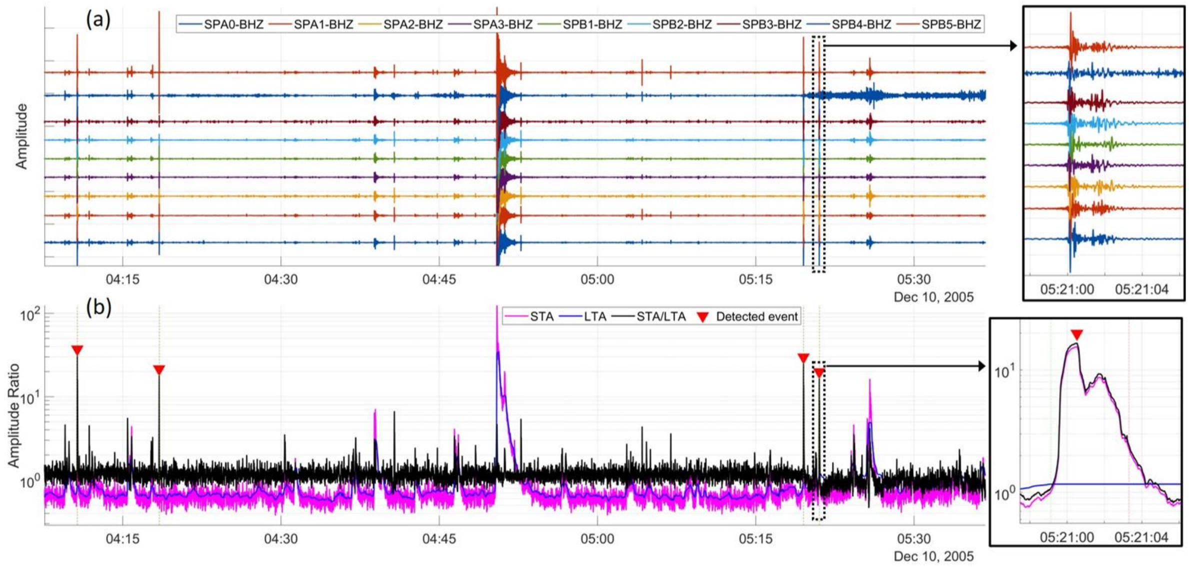

Figure 3 – Example of (a) timeseries of vertical component ground motion and (b) event detection using the STA/LTA (short term/long term average) detector. Short duration events with sufficient amplitude and array coherence are selected while longer events such as the high amplitude example at 04:50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

Figure 3 is a good illustration of the kind of events that are selected. One may also observe that events with longer duration and high amplitude are not selected and neither are short duration events with low amplitude. An important case not shown in the figure is that we also manage to ignore spurious noise spikes that are only observed on a single seismometer. We can add a note covering single channel noise spikes at Line 153.

We take the point that saying "local events" in this section of the manuscript is perhaps a leap of logic and have modified that to "short duration events". As we have responded elsewhere, duration of ground motion can be used as a metric of earthquake magnitude (e.g. Lee et al., 1972; Mousavi and Beroza, 2020; Tsumura, 1967) and source distance. It is also well established that magnitude correlates with the area of rupture, length of surface rupture etc. (e.g. Wells and Coppersmith, 1994). Shorter duration seismic events are therefore indicative of rupture along smaller cracks, and less energy release than larger magnitude tectonic earthquake. Such small magnitude events are only recordable locally because the amplitude of ground motion decays below the background noise level over longer distances. Notably, we have also used MFP to prove that the short duration signals correspond to local seismic sources.

Line 151: Authors concluded that the ratio of short-time-averaged and long-time-averaged peaks must have amplitude larger or equal to 10 and occur at least 5 seconds apart from one another. Authors need to explain how they drew this conclusion, or clearly indicate this is an assumption if that is the case.

OK, we can give a bit of extra detail here and propose the revised formulation:

"The STA/LTA threshold was set to 10. Furthermore, after a trigger, no new events are declared within 5 seconds after the STA/LTA was exceeded to avoid detecting the same events multiple times. As for the STA and LTA lengths, these parameters were found to be appropriate for detection of short-duration signals coherent across the array, while avoiding false triggers (noise bursts at single stations), by visually inspecting test periods."

Line 209: It is common to add the unit for every applicable parameter defined in the manuscript.

We follow the custom from physics literature that units are not listed when deriving the physical model. Units for every applicable parameter are listed in Table 1 which is introduced when we discuss the parametrization and numerical solution to Eq. (12).

Line 253: Authors stated that 'A significant novelty of this study is that the ground temperature profile at Janssonhaugen has been logged by a series of thermistors installed in the 15 m deep P11 borehole at 6 hr intervals since April 1999'. How did the authors use these measurements in the analysis in the entire study domain (I suspect interpolation is required)?

Good point. We can add a note to Figure 2 and Figure 8 that the measured temperatures were interpolated to regular 10 cm depth intervals using a spline interpolant. In Figures 9, 10 and 11 the measured temperature timeseries are used without interpolation.

Line 254: Authors need to elaborate: 'We assume the stress at a given depth is decoupled from the stress at adjacent depths'. How exactly did the authors decouple the stress components in space since they are dependent on each other (if assuming the soil is continuum media).

We agree that "decoupled" is not the most suitable term. We applied the standard assumption of Mellon (1997) that the vertical stress is zero (because the overburden is assumed to be negligible in the shallow subsurface). As shown by Mellon (1997), this allows the horizontal stress at a given depth to be calculated independently, even though the stresses at different depths are directly connected through the temperature profile. We will state this more clearly around Line 224 and remove the sentence highlighted by the reviewer on Line 254.

Equation 10: Can this model be used to study the thawing of frozen soils?

Yes, the model is compatible with both thermal expansion and contraction. The extent to which it could be used to study thawing of frozen soils would depend on the specific aspect one wanted to study.

Table 1: How the tensile strength is calculated? Is it always 1 MPa? Should it also be temperature dependent?

Yes, it is always 1 MPa. While the compressive strength of polycrystalline ice has been shown to vary significantly with temperature, the tensile strength seems to be a weak function of temperature and is mostly controlled by grain size (e.g. Petrovic, 2003). We can add a reference to Petrovic (2003) to Table 1 to clarify this point. We suggest the following revision to the note in Table 1:

"Currier and Schulson (1982), varies according to grain size for randomly oriented polycrystalline ice (finer grained ice stronger), insensitive to temperature (Petrovic, 2003)."

Petrovic, J. J. (2003). Review Mechanical properties of ice and snow. Journal of Materials Science, 38(1), 1-6. 10.1023/A:1021134128038

Line 288: Authors stated that 'Event class I is characterized by significant amplitude variation and arrival time differences across the array seismometers'. A quantification method (e.g., L2 distance) is needed to describe the significant amplitude variation as well as similar amplitudes (line 291).

This sentence is a description of the first-order qualitative signal characteristics that the reader may observe in Figure 4. Importantly, the amplitude variation is dealt with quantitatively in the matched field processing method (ref. Eq. (1)) where it provides an important data constraint on the estimation of source position by assuming a model of amplitude attenuation by geometrical spreading.

Also, authors predicted the source of seismic events is around 1500 m. What is the uncertainty of this prediction?

We state that the Class I events "occur in relatively close proximity, within about 1500 m of the centroid of the array", meaning that the events we interpret as belonging to this class have estimated source positions that fall approximately within a circle with a radius of 1500 m, centered over the array. This is not a prediction, but a heuristic threshold necessary to delineate the cluster in space. To clarify this, we suggest rephrasing to "occur in relatively close proximity and are clustered inside a circle with radius of ~1500 m centered over the array".

Line 303: It is difficult to determine the dominant wave type based on merely the estimated propagation velocity. Could the 1150 m/s also correspond to body wave?

Good point, a direct or refracted vertically polarized shear wave propagating through the shallow subsurface at varying depths could in theory have a similar velocity. However, we might expect body waves to have higher frequencies. Surface waves are also expected to be excited for a near-surface source and then often dominate the signal. From the waveform example included below (Figure II) it is also clear that the highest amplitude is observed for a phase arriving after the P and S wave, i.e., most likely a Rayleigh wave.

To address the comment, we suggest re-phrasing from:

"The mean MFP inferred propagation velocities for Class I events was 1150 m/s with a standard deviation of 1100 m/s, indicating that they are dominated by surface waves."

To:

"The mean MFP inferred propagation velocities for Class I events was 1150 m/s with a standard deviation of 1100 m/s, implying a relatively shallow propagation path."

We could also include Figure II as an appendix showing illustrative three-component seismograms for the two event classes. There is some indication of phase rotation between the vertical and radial components for the frost quake (Figure II-a), which would indicate a Rayleigh wave. The Gruve 7 event that is interpreted to be dominated by P wave energy (Figure II-b) has vertical and radial components that are more similar to one another. We may also observe that the largest amplitudes for the Gruve 7 event correspond to the first arrival (P-wave), whereas the largest amplitudes for the frost quake arrive later (Rayleigh wave).

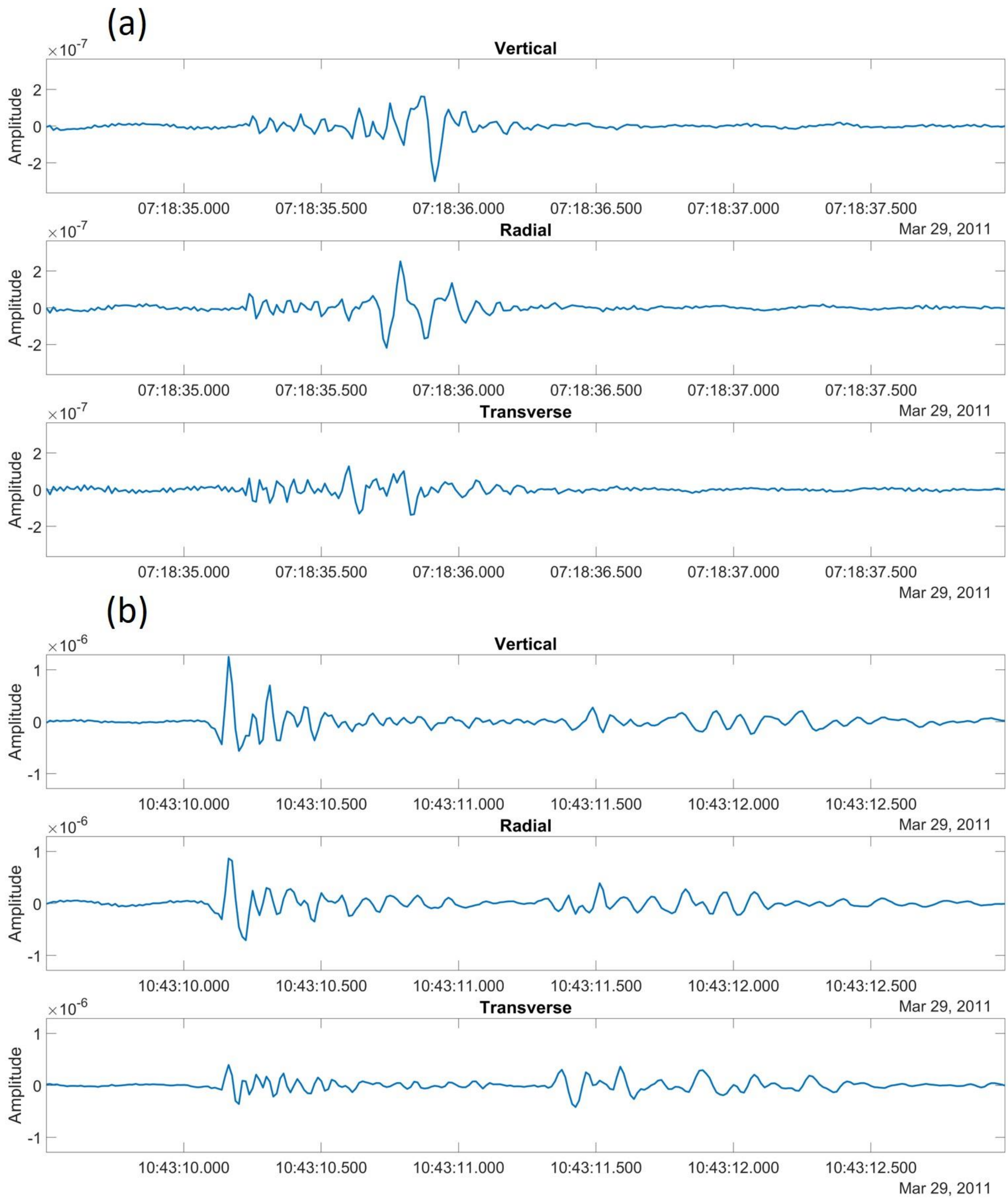

Figure II – Three-component seismograms from SPITS station SPA0 for (a) an example of event Class I (frost quake) and (b) an example of event Class II (Gruve 7 mining activity). The signals are bandpass filtered to 2.5-35 Hz and the horizontal components of ground motion (measured in NS and EW orientations) were rotated to radial and transverse to the source bearings estimated by MFP (see Figure 4e and 4f).

Figure 3: It is difficult to understand why longer seismic events have high amplitude. Authors might want to elaborate on the relation between the duration and the amplitude of displacement (or velocity and acceleration).

Similar to our response to an earlier comment, duration of ground motion can be used as a metric of earthquake magnitude because of stronger coda waves (e.g. Lee et al., 1972; Mousavi and Beroza, 2020; Tsumura, 1967). Since the magnitude is larger the amplitudes can be large, but this depends on the source-receiver distance and rate of attenuation. Therefore, high amplitude events are not necessarily high-magnitude events and do not need to have a longer duration (e.g., frost quakes). More importantly, the larger the source distance, the more spread-out are the seismic signals (longer event duration) because of different propagation velocities of seismic phases. The detector is designed to separate short duration seismic signals from background noise and longer duration signals that *may* be high amplitude (Line 131). If the events of interest were always higher amplitude than all other events a simpler detection scheme based on amplitude thresholding would have been possible. To clarify why we choose to isolate only the short duration events, we suggest adding a sentence to the introductory paragraph outlining the study hypothesis (Line 74), which also covers the earlier comment relating to line 131.

References

Allen, R.: Automatic phase pickers: Their present use and future prospects, Bulletin of the Seismological Society of America, 72, S225-S242, 1982.

Lee, W. H. K., Bennett, R., and Meagher, K.: A method of estimating magnitude of local earthquakes from signal duration, Citeseer, 1972.

Mousavi, S. M. and Beroza, G. C.: A Machine-Learning Approach for Earthquake Magnitude Estimation, Geophysical Research Letters, 47, e2019GL085976, 2020.

Petrovic, J. J.: Review Mechanical properties of ice and snow, Journal of Materials Science, 38, 1-6, 2003.

Sørbel, L. and Tolgensbakk, J.: Ice-wedge polygons and solifluction in the Adventdalen area, Spitsbergen, Svalbard, Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 56, 62-66, 2002.

Trnkoczy, A.: Understanding and parameter setting of STA/LTA trigger algorithm. In: New Manual of Seismological Observatory Practice (NMSOP), Deutsches GeoForschungsZentrum GFZ, 2009.

Tsumura, K.: Determination of earthquake magnitude from total duration of oscillation, Bull. Earthq. Res. Inst., 45, 7-18, 1967.

Wells, D. L. and Coppersmith, K. J.: New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, Bulletin of the Seismological Society of America, 84, 974-1002, 1994.