Response to review
"Local station correlation: large-N arrays and DAS".

Editor:

Thank you for the reviews and the points raised on the manuscript. We are pleased that the reviewers find the paper to be an interesting and valuable addition to the field.
In the revision we have endeavoured to meet most of the points raised by the reviewers.

The two studies for the large-N array and the DAS recording were made and analysed independently at very different geographical locations. A similar line of processing was employed but adapted to the scenario for the individual experiments. This point is now made explicitly at the beginning of section 3.

Regarding the DAS dataset in particular, we have responded to the points raised by the reviewers, however there is currently a paper in preparation focusing entirely on the Bern data, including a systematic inversion. Therefore, we have kept the focus of this paper on the demonstration of the local station correlation concepts, and details such as geological structure and inversion will be covered in the separate publication.

We have also made some expansion of the text at several points, not picked up by the reviewers, where explanations could be improved. The Discussion has been expanded to include a discussion of directional effects for co-array analysis of DAS recordings, including the influence of Love waves. We have also produced a new Supplementary Material that includes some of the figures relating to the questions raised by the reviewers with text explanations

Changes are indicated in red and inserted material in blue.

Reviewer C:
Recommendation: Revisions Required

Dear Authors,
Thank you for sharing your work. It is indeed an interesting line of research that you are discussing in your manuscript. I enjoyed reading it, but I think a few points should be addressed before the manuscript can be considered for publication.

Here below you can find my comments:

Please note that the analysis for the nodal experiment in Australia and the DAS experiment in Switzerland were carried out independently using local work flows and choices of frequency band etc. based on local conditions.

In general, you provide a great level of detail on how you obtained your results, and that is very good. However, I have a few questions on the DAS example. Why do you compute crosscorrelation, and visualize, only for 100 m channel distance? Computing every pair at say 10 m or even 4 m distance should be feasible. It will highlight more of the details that can be seen in DAS recordings compared to conventional stations. You claim the issue is inter-channel sources, could you show an example of those low-quality CC panels?

We have included a figure below (Figure R1) to demonstrate the motivation behind choosing to only use channel distances up to 100 m. As shown in the figure, beyond 100 m it becomes difficult to pick one clear, coherent signal within this record section. One can also see many complex and difficult to interpret features. While some sections of the array did show clear signals to distances of 350 m, we found 100 m to be a reasonable limit for obtaining useful data along the whole DAS line.

In response to this comment, some text in the paper has also been altered to emphasise that the difficulty with the data is not that it is low-quality, but rather that the wavefield is very complex. We believe this to be as a result of the combination of subsurface conditions (unconsolidated sediment on top of bedrock, with significant concrete infrastructure on top), as well as the nature of the anthropogenic noise in the environment.

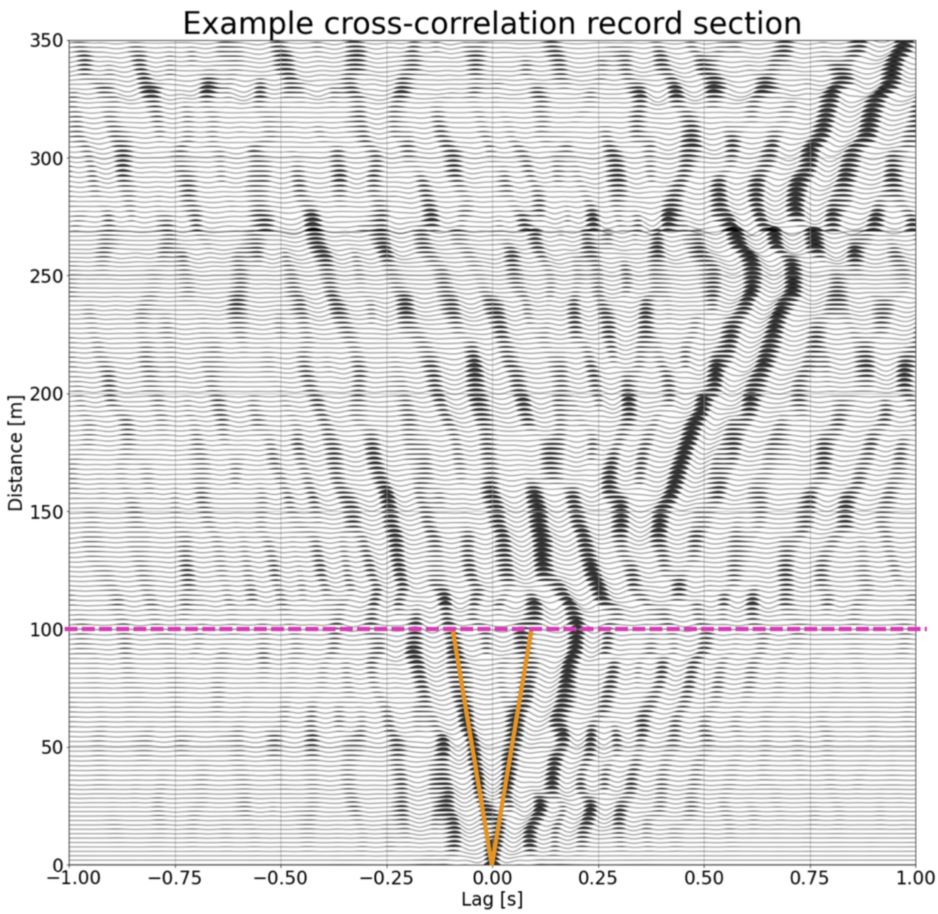

Figure R1: An example of a cross-correlation record section with channel spacing up to 350 m. The orange lines indicate the acausal and causal propagation of Rayleigh waves. The pink dashed line indicates our chosen limit of interpretation (100 m), as the signal became more complex and difficult to reliably interpret beyond this distance.

For the dispersion curve analyses, you chose different workflows between stations and DAS; could you elaborate on the reasons why? Did the approach used for the stations not work on DAS cross-correlations (CCs)? If so, please show an example of such a failure.

See note above – the results come from independent analysis of data of very different character.

Also, could you please perform the same phase-velocity picking on the DAS data? It would greatly help to corroborate the statement on the medium sampled by the DAS CCs and also a reader would be able to compare it with the conventional stations.
We have added our phase velocity picks to the DAS dispersion curves in Figure 9. Regarding comparison with conventional stations, there are no suitable stations to compare with – there were no seismic stations in the vicinity of the DAS cable during the Bern deployment, and comparison with the Australian dataset is unlikely to yield meaningful results, due to the significantly different environments and characteristics of the datasets.

Why did you use different frequency ranges between the conventional stations and the DAS data? Could you please elaborate on this?
See note above – the results come from independent analysis of data of very different character and the frequency choices reflect the local conditions.

I would move the discussion section before the field-data examples since it does not actually discuss the results in Section 3.
We have endeavoured to improve the tie of the theoretical results and the field examples and prefer to keep the Discussion at the end. This is now expanded to discuss directional issues associated with DAS.

I found your Section 2 interesting and insightful. What about Love waves? I understand within DAS data is harder to obtain such signals in CCs, but it could be something interesting to highlight in your manuscript.
The most common form of anthropogenic noise from traffic produces dominantly Rayleigh waves so attention is concentrated on Rayleigh wave for the data examples. A statement has been included in Section 2.
The Discussion in Section 4 has been expanded to examine the sensitivity of DAS arrays to both Love and Rayleigh waves – enhancement of Love waves is harder to achieve than for Rayleigh waves.

Minor comments:
- In the abstract: "For correlation work," what do you mean? Please rephrase.
Now modified to mention cross-correlation explicitly.

- Gz is not defined in equation 1.
This term represents the response of the stratified medium and is now defined explicitly.

- pag. 2 line 88: "\deltaX_12" is the difference between the distances from the sources"; which sources? From your text, it is unclear which sources are you referring to. Please, clarify this concept.
The use of "sources" was a slip – this should have been "source".
This is now corrected, and the sentence modified to improve clarity.

- In Figure 1, I would use triangles rather than purple dots. In most of the panels, they are barely visible. Also, what do the "x5" and "x10" mean in panels (d) and (f)? For total effect, do you mean when multiple sources are considered? If so, please clarify that in the caption. Panel (b), you say, distance from "the source", again which source?

The markers for the stations have been made more visible in both Figures 1 and 2.
The caption for Figure 1 has been modified and expanded to make the contributions more explicit. The representations cover a field of sources as stated in the text.
The x5, x10 labels indicate amplification factors that are now made explicit in the caption.

The caption for Figure 2 has also been modified to make it more explicit.

- Figure 3, I would use an arrow on the map rather than a green dot.

The green dot has been replaced with a red arrow by following the suggestion.

- Figure 4, The black trace is not very visible. I would move it to a panel below the 2D plot.

The averaged trace of the 2D matrix has now been moved to a separate panel b in this figure, while the original panel b has been renamed to panel c. When making the revision (comparing the averaged trace with those in Figure 5), we found a mistake in this figure. The correlation function presented was for a different station pair from an early analysis rather than the latest results for the pair of LG015-LG049. We have re-plotted the figure and double-checked other figures to make sure they are from the latest analysis and are consistent with each other as well as the text.

- Figure 5, what is the behavior as a function of the frequency of the stacked CC?

We made the same plots for two different frequency bands at 1-4 Hz and 4-9 Hz, which are shown in Figure R2 below. The similar pattern in these two frequencies bands suggest that the described behaviour is independent of the frequency.

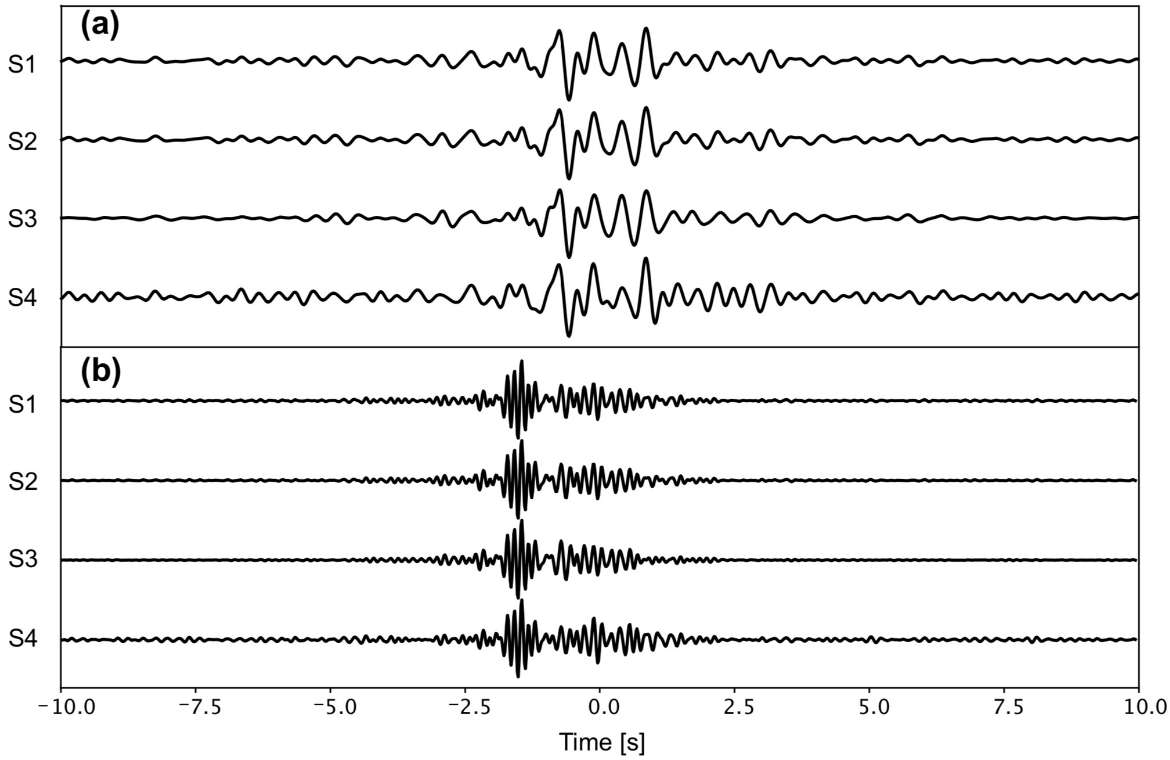

Figure R2. Same as Figure 5 in the main text, but at two different frequency bands of 1-4 Hz (a) and 4-9 Hz (b). S1 is the stack of all cross-correlation functions. S2 is the stack of all cross-correlation functions with a correlation coefficient large than 0.7 relative to S1. S3 is the stack of correlation functions over the time window of 11 am-3 pm. S4 is the stack of correlation functions over the time window of 11 pm-3 am each day.

- Figure 6b, it would be better to show 2D color plots with empty values where the measurement was not picked. It would look nicer and easier to interpret and correlate with potential subsurface structures.

We prefer the current Figure 6b because the point-based plotting avoids introducing artefacts into the images through 2D interpolation. The current Figure 6b shows nicely the period-dependent phase velocity changes across the transects, which can be visually connected with changes of subsurface structures. For example, the phase velocity at each period gradually decreases from west to east, indicating the thickening of the soft sediments/regolith layer into the basin.

Reviewer D:

Recommendation: Revisions Required

This manuscript is a valuable addition to the field, as it focuses on the effects of noise source distribution in retrieved Green's functions for DAS and N-arrays, which have been widely used in many applications. The theoretical part of this study is comprehensive and informative, providing a clear analysis of how noise distribution can affect Green's functions. Overall, it is highly relevant to many noise CC applications. I recommend publishing this study after addressing the minor revisions suggested below.

The third part of the paper deals with the observational data, which could benefit from further clarification and revisions. Although the data presented in this section are interesting and informative, they do not sufficiently explain the conclusions made in the theoretical part of the study. Therefore, I suggest that the authors revise the third section to better connect it with the theoretical.

In the revision we have endeavoured to improve the connection between the theoretical results and the practical examples.

For the large N-array data, the authors could examine various situations that would give different results according to the predictions made in the theoretical part. For example, taking a station pair in the east with a W-E station pair orientation would be favored for retrieving the signals at one side and dramatically reduced at the positive time side. Instead, taking a pair with N-S orientation would give us result with noise source off stationary zone. It would be also worth extending the comparison from a distance far away from the road to a closer distance and see their difference.

This is a good point. We find it generally true that station pairs in N-S direction show more complicated waveform and dispersion information relative to these in E-W direction. Figure R3 shows one such comparison for two station pairs of LG015-LG047 and LG015-LG049, which has similar distance (~250 m) but different geometry relative to the road. The frequency-time analysis for station pair of LG015-LG049 (E-W) reveals much coherent and simpler surface wave information compared to that for LG015-LG047 (NE-SE). However, it should be noted that significant complexity still exists in the cross-correlation functions even for LG015-LG049, and it is difficult to extract reliable dispersion information based on one single station pair. This is mostly due to the complex wavefields from the strongly localized traffic sources, which is far from approximating a diffusive condition. Therefore, we did not extend this discussion in the text but mostly focus on using array seismic methods to extract averaged dispersion information. The same conclusion also holds for station pairs closer to the roads.

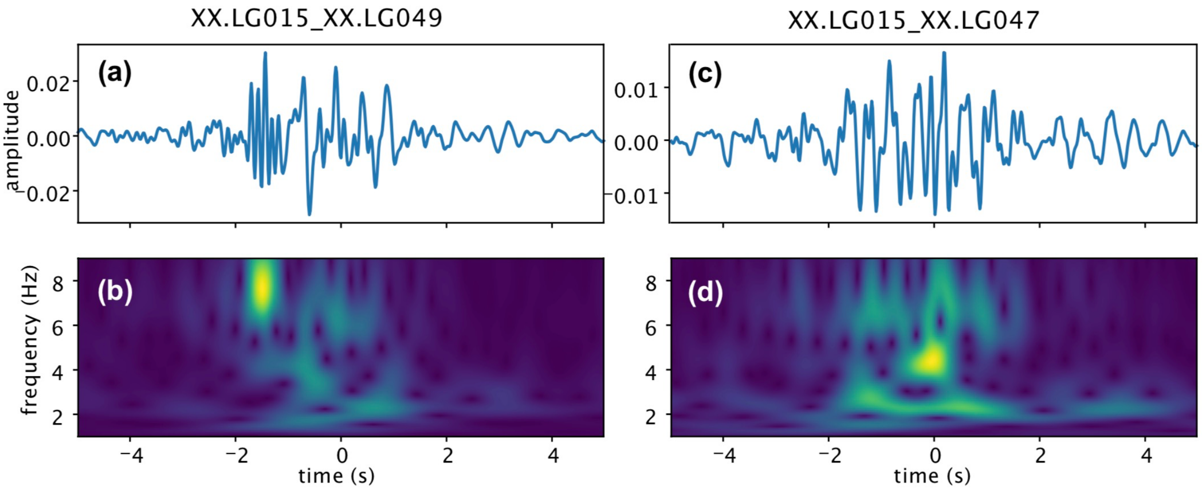

Figure R3. Comparison of cross-correlation functions and their associated frequency-time analysis for two station pairs of different geometry relative to the road. Station pair of LG015-LG049 (a-b) is almost perpendicular to the road while LG015-LG047 (c-d) has large angles relative to the road.

In addition, some stations in Fig. 3 are located at the west side of the road, what the CC result would look like, if we chose one station at the west side of the road and another paired one on the east? According to Fig. 1c,d, we would see apparent phase shift or waveform change in the CC when the traffic noise sources are located at different positions between the two stations

We would expect the phase shift or waveform change to occur for stations close to noise sources when the noise sources are relatively random in space and time. However, the busy and localized traffics on the highway (several 10s of vehicles in a minute) and the stacking over time intervals would largely eliminate any phase shift or waveform change. In other words, the correlation functions will be more dependent on the traffic conditions rather than the station location. To confirm this, Figure R4 shows the temporal evolution of the cross-correlation functions for the station pair of LG069-LG094, which two stations are located on the two sides of the road. It is clear that the correlation functions are still very coherent over time and the pattern of correlation coefficient change with time is almost perfectly in phase with that for station pair of LG015-LG049 (Figure 5 of the main text). The decoherence of the waveforms coincident with the time when the traffic conditions are reduced significantly either in the late night or holidays. However, due to the limited number of stations at the edge of the array, it is difficult to retrieve stable dispersion information from them.

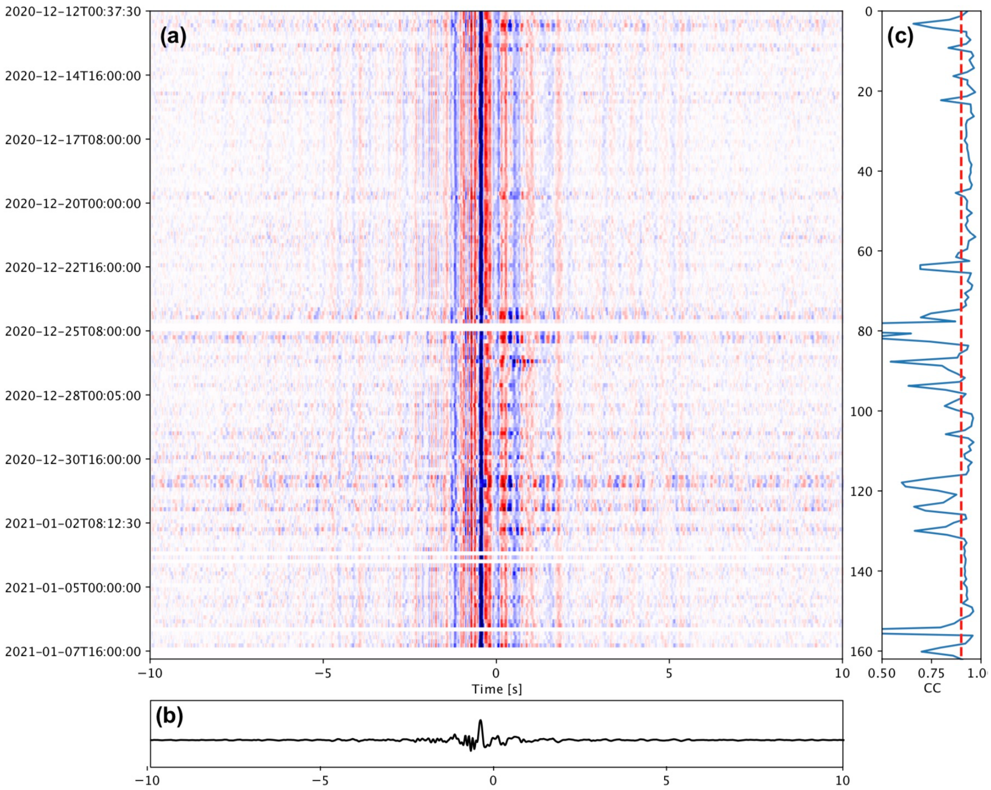

Figure R4. Same as Figure 5 in the main text but for the station pair of LG069-LG094, which are located on the two sides of the road.

Similarly, only a small part of the DAS dataset is used, e.g., only night-time one channel result shown in Fig. 8. This part can be extended using more pairs. For example, what would the result look like using daytime data? The authors claim that the DAS CC results exhibit poor quality.

We have adapted the text to emphasise that the data is not poor quality as such, but rather that the wavefield appears to be extremely complex, and features seen in the cross-correlation record sections are therefore difficult to separate and interpret.

However, according to Fig. 1f, the contribution near the station pair path is further enhanced by the DAS orientation sensitivity, meaning the secondary sources off the pair path have reduced contaminating effects.

We believe that the structure of the sensitivity kernel is likely to reduce the effect of noise sources off-line of the DAS array, and therefore one most likely does not require such a homogeneous noise source distribution as would be required when using traditional 3D sensors. However, on such local scales as the Bern array, it is important to consider the length scale of these effects. As shown if Figure 1f, secondary sources off the pair path are only significantly reduced at lateral distances of 100 – 200 m from the fibre. While conducting fieldwork in Bern, we noted a number of significant sources of noise near the DAS line. One example is the Faculty of Veterinary Medicine (VetSuisse Faculty) at the University of Bern, which consists of a number of large buildings, with a significant amount of audibly loud equipment within the complex. As shown in Figure R5, much of this faculty is situated within a distance of 20 – 200 m from the fibre, and appears to be a significant off-path source of noise.

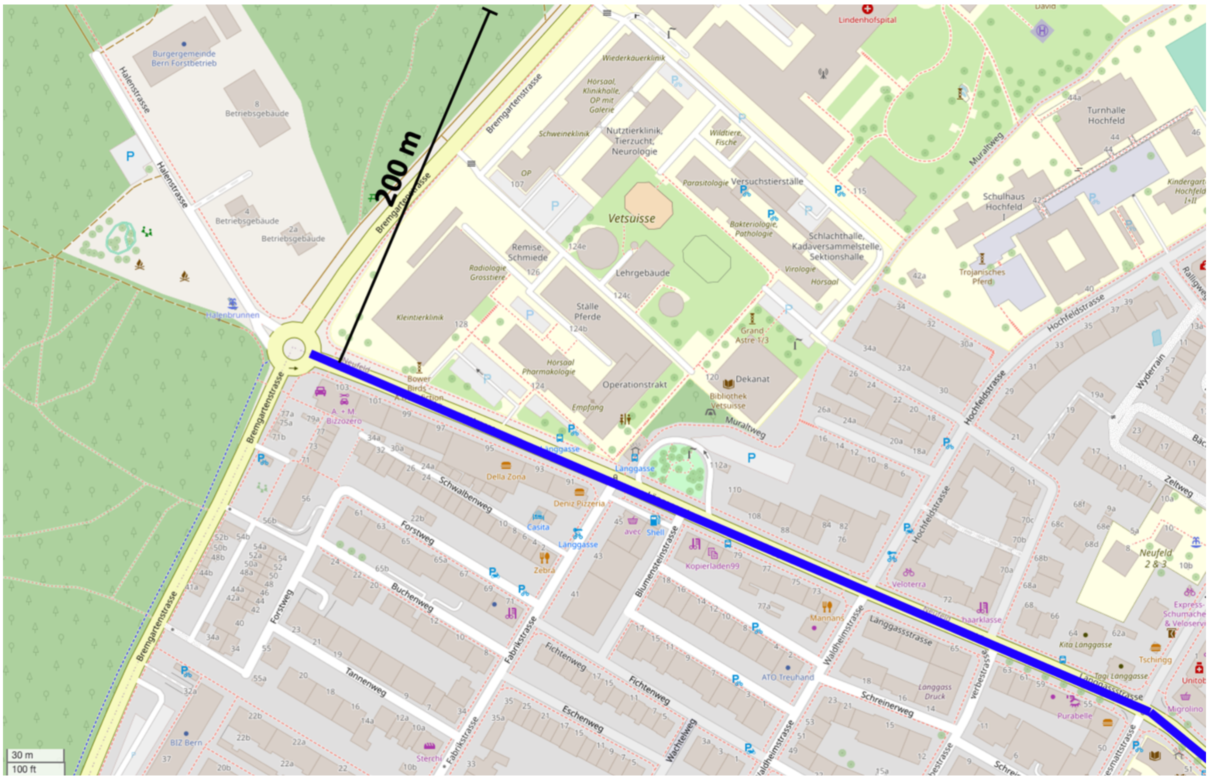

Figure R5: Map of the northern end of the DAS line, showing the proximity to the VetSuisse Faculty buildings; believed to be a significant off-line noise source.

By looking at the map, my perception is that "Bremgartenstrase" and the street along the DAS fiber are two major streets. Traffics on both of these two streets would be favored for noise CC. The "Bremgartenstrase" street have effects described at lines 308-309. For the DAS fiber street, traffics between a chosen pair would contaminate the CC results, but these cars would run along the entire street, and may still be good for noise CC. Then why we have poor CC quality? More analysis echoing with section 2 is encouraged and would further enhance the quality of this work.

While Bremgartenstrasse and Länggassstrasse are both larger than the surrounding streets, they are still relatively small 50 km/h (30 mph) roads, with single-lane traffic in each direction, and are primarily used for access to connecting streets. There are also nearby motorway routes to avoid travel along these small roads for anyone driving straight through the area. Therefore, these roads primarily bear slow-moving, stop-and-go traffic, with a number of pedestrian crossings and cars turning off of the main road. Therefore, the noise sources (vehicles) are not only moving, but they are also moving with changing velocities, and often only present along a small part of the array.

We also note that within this dataset, the seismic waves produced by cars only visibly propagate around 50 m, as shown in Figure R6. Therefore, only utilising traffic on Bremgartenstrasse is unlikely to give us any significant signal along the majority of the DAS line, particularly as only vehicles passing the end of Länggassestrasse can be expected to make a contribution.

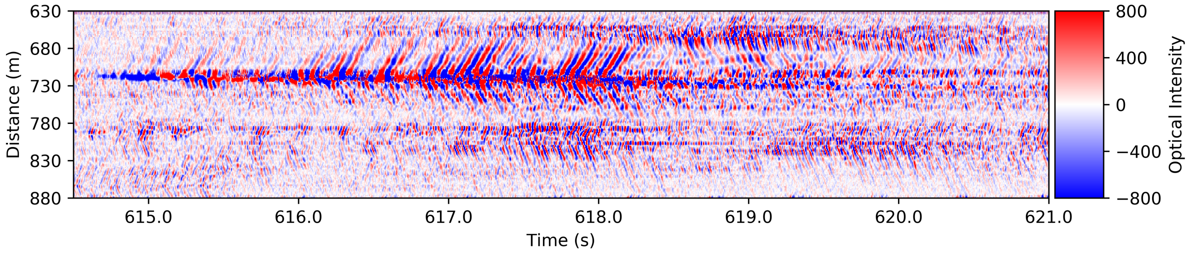

Figure R6: Example of 'fish-bone' propagation pattern produced by a car. The 'backbone' shows the velocity of car moving along the fibre; the more steeply-inclined 'bones' are the induced Rayleigh waves, propagating with much higher velocities than the car.

During the daytime period, the traffic is much higher giving rise to more moving noise sources with changing velocities and travelling different distances along the road. Such sources lying between stations are deleterious to the stacked cross-corrrelation. Also, in daytime, there are more 'undesirable' sources of noise (loud, repeated point sources), e.g. construction work including jackhammers, diggers etc.

Finally, in Fig. 9, it would be great to see more interpretation about the phase velocities. For example, are the phase velocities above 12 Hz reliable? Is the increasing-decreasing shape of phase velocity across frequency reasonable with our understanding of local geology and concrete structures? These interpretations would provide a more complete understanding of the results presented in the paper.

We have added out phase velocity picks to Figure 8, as well as expanded the text to clarify which frequencies we use and trust. We believe that the phase velocities above 12 Hz are reliable, as this is the dominant frequency range of the anthropogenic noise seen in the data, and the slow velocities are consistent with those seen in the raw data (for example in Figure R6). We do expect the velocities to be significantly higher at lower frequencies, as there is bedrock at depths of ~ 40 m.

This is a large and complex dataset, and we feel that a more in-depth discussion of this specific dataset is beyond the scope of this paper. A more detailed analysis, including systematic inversion, is in preparation.