Comments on the revised version

Some aspects of the paper have been improved considerably but the authors reply is not nearly as detailed as my comments were, and there remain open questions, which I list again below.

(1) My question whether "*instrument simulation was performed*" remains to be answered. If the 1-Hz-geophones are typical short-period velocity sensors they undergo a 180°-phase rotation at 1 Hz. According to Ringler et al. (2018, SRL 89(5), their Fig. 5) the Z-Lands have the same type of response but undergo the phase rotation at 5 Hz. If this is not accounted for, the phase responses don't cancel in the ZL-Tx-correlations of eq 3 (as opposed to the ZL-ZL, or Tx-Tx-correlations) but are projected directly into the Green's function. Since the frequency-range 2-10 Hz was essential to the MASW, I don't understand how it was possible to successfully perform MASW with ZL-Tx-correlations. Could you provide some details on the residual reduction for the different parts of the profile? Are your segments so short that you never really use ZL-Tx-correlations?

Response to (1a):
We are not sure where the reviewer got the impression from that 1 Hz geophones have been used. In fact, 4.5 Hz and 5 Hz geophones were used, which the reviewer also discussed correctly in his first revision. From the manuscript: *"Recording stations were equipped with 385 Reftek 'Texans' data loggers / 4.5 Hz 1C geophones and with 120 Fairfield ZLand 3C 5 Hz nodes at a 5 m interval."* 4.5 Hz and 5 Hz refer to the natural frequencies of the geophones, while 1C and 3C describe the number of components. These abbreviations are conventional seismological terminology. Due to the similar natural frequencies (4.5 Hz and 5 Hz), instrument simulation was not required. This is also demonstrated by the interferogram gathers which, after taking geology into account, show identical quality of arrivals for all Virtual Source (VS) and receiver (RCV) – combinations (VS-RCV: ZLAND-ZLAND, ZLAND-TEXANS, TEXAN-ZLAND, TEXAN-TEXANS; see images below). Also, the dispersion curves of the mixed gathers are of similar quality as the dispersion curves from ZLAND-only gathers and Texans-only gathers (see additional figures below)
We added this explanation in the revised manuscript.

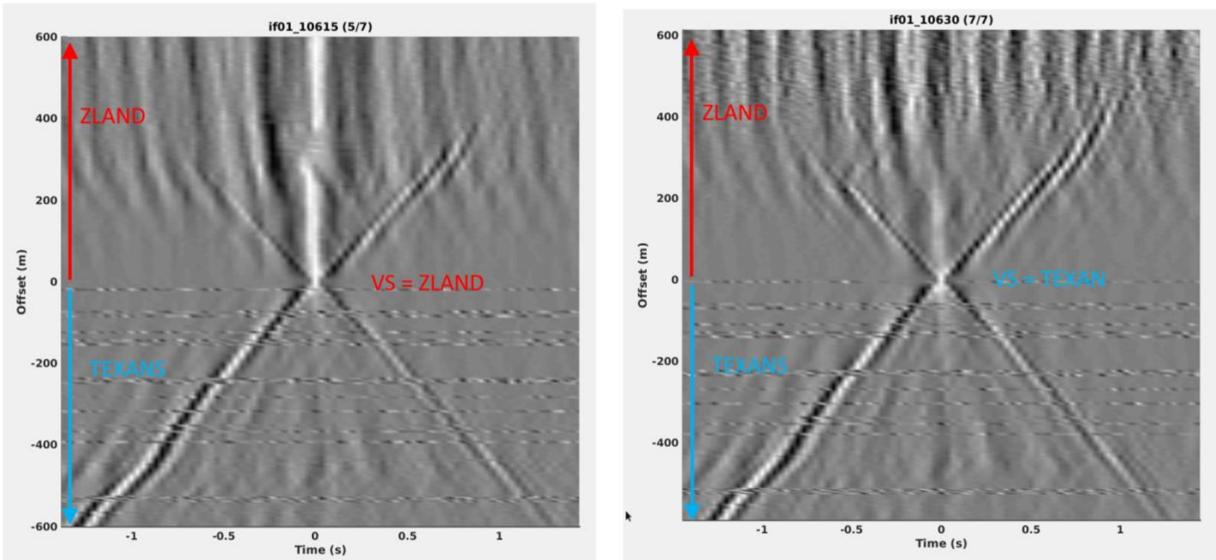
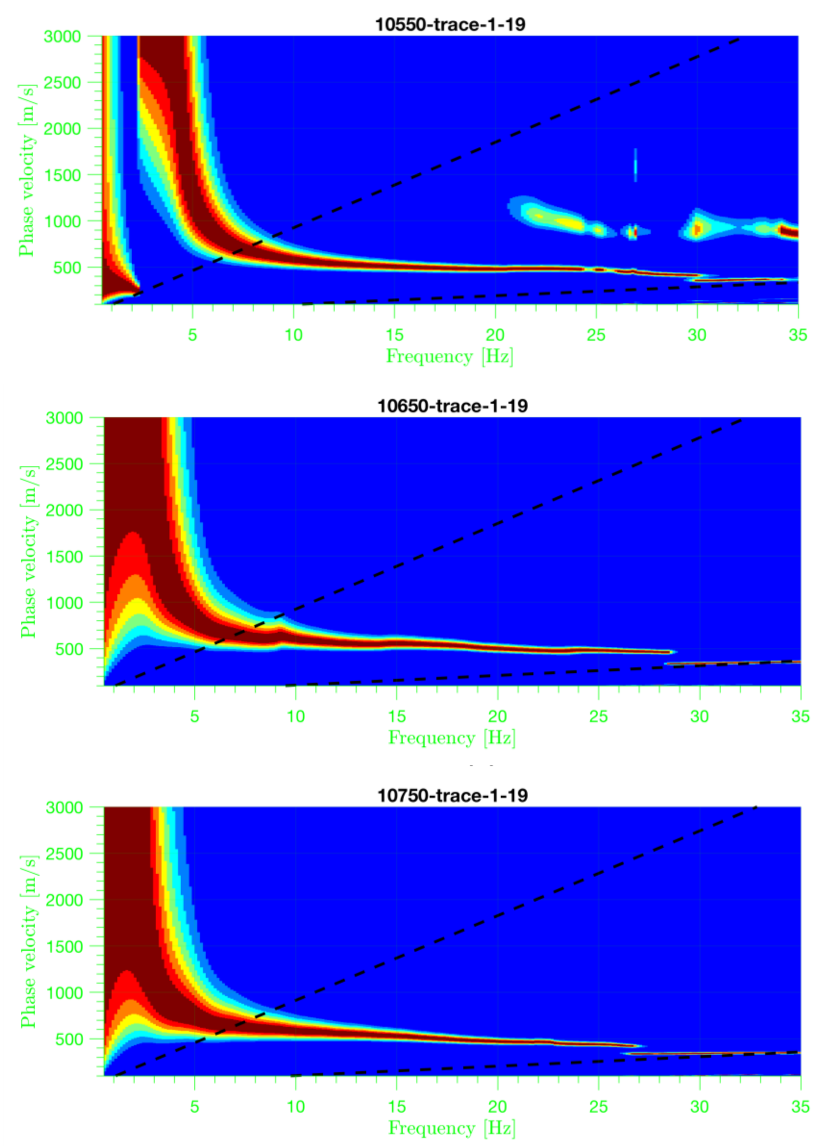


Figure above shows a comparison between virtual source gathers for different data type combinations: Left – Virtual source ZLAND, receivers to the north are LAND, recievers to the south are Texans. Right – Virtual source Texan, receivers to the north are LAND, recievers to the south are Texans.

The figures on the next page show dispersion curves from different intsrument combinations. Top: dispersion curve for a 100m-section comprising only ZLAND geophones; Middle: Top: dispersion curve for a 100m-section comprising ZLAND and Texan geophones; Bottom: dispersion curve for a 100m-section comprisong only Texan geophones



In the revised version, you write that "*the FK transform shows that the ZLand recorders have a stronger response at low frequencies (< 5 Hz*)". In contrast, their response shows that they decay ~ω^2 below 5 Hz, where the 1-Hz-geophones are probably still on the flat passband. Please clarify.

Response to (1b):

Again, given the similarity of the natural frequencies (4.5 and 5 Hz, respectively), we think that the data can be compared. However, we agree with the reviewer that the statement above is actually not accurate, as the FK transform of the entire line cannot be unambiguously separated into Texans and ZLand recorders. E.g. the apparent "stronger" response at the negative velocity branch might also be attributed to the ambient noise in the southern part of the section where Texans were deployed. As the manuscript does not aim at a detailed comparison between recording instruments, we removed this statement from the manuscript.

(2) Please reply to my previous comment "*I fully understand that there are many reasons why V_P might not be indicative for the GWT but not in this case of unconsolidated sand, where one would expect a sudden increase of V_P at the GWT from maybe 800-1200 to ~1700-1800 (as seen in the well). It would be appropriate to make a first interpretation of the GWT from a contour line in the range ~1500-1800 m/s, and then check, if such a contour line coincides with the interpretation of the GWT from V_P/V_S-ratios.*"

Response to (2):

We are not sure why the reviewer has missed the consideration of this initial remark. In revision 1, we have already included a response in the manuscript, and we have already added the contour lines in the image (see below). The remark above is part of a larger general/major comment of the first review, and we have addressed it by completely re-arranging the entire interpretation section (see reply to review below, and more importantly, the entire revised manuscript with the tracked changes).

that it is part of a 400-m-thick aquifer (13/25). Also, I find the interpretation of the GWT difficult. I fully understand that there are many reasons why V_P might not be indicative for the GWT but not in this case of unconsolidated sand, where one would expect a sudden increase of V_P at the GWT from maybe 800-1200 to ~1700-1800 (as seen in the well). It would be appropriate to make a first interpretation of the GWT from a contour line in the range ~1500-1800 m/s, and then check, if such a contour line coincides with the interpretation of the GWT from V_P/V_S-ratios.

>> We agree that the structure of the interpretation section is poor and adds more confusion than clarity. We tried to rearrange accordingly to the remarks above.
The conclusions on water saturation, GWT interpretation, and potential aquifer properties have also been revised. In particular the comment on high Vp-velocities has been addressed more clearly, also in the context of the reviewer's comment (3) below. We point out that the interpretation of "clay" below the sand describes

Specific consideration in the manuscript in revision 1:

part of the profile, where West Creek occupies the lowest topographic point. The zone dips towards the north and its top is found at ca. 120 meters depth at the presumed northern edge of the over-deepened section. A northward dipping reflector is found in a comparable depth range in the seismic image, and the P-wave velocities (1500 m/s – 1800 m/s) correspond to typical velocities of saturated near-surface sands and gravels (Knights and Endres, 2005; Everett, 2013). ~~we~~ We therefore interpret the increased V_P/V_S – ratio in the over-deepened section to represent the top of water-saturated sediments.~~–~~ Since the dip opposes the slope of the topography, this aquifer needs to be confined or it is leaking through fractured basement in the north. The latter hypothesis would be supported by the relatively low P- and S-wave velocities between profile distances 900 m to

Specific consideration in the figure in the manuscript in revision 1:

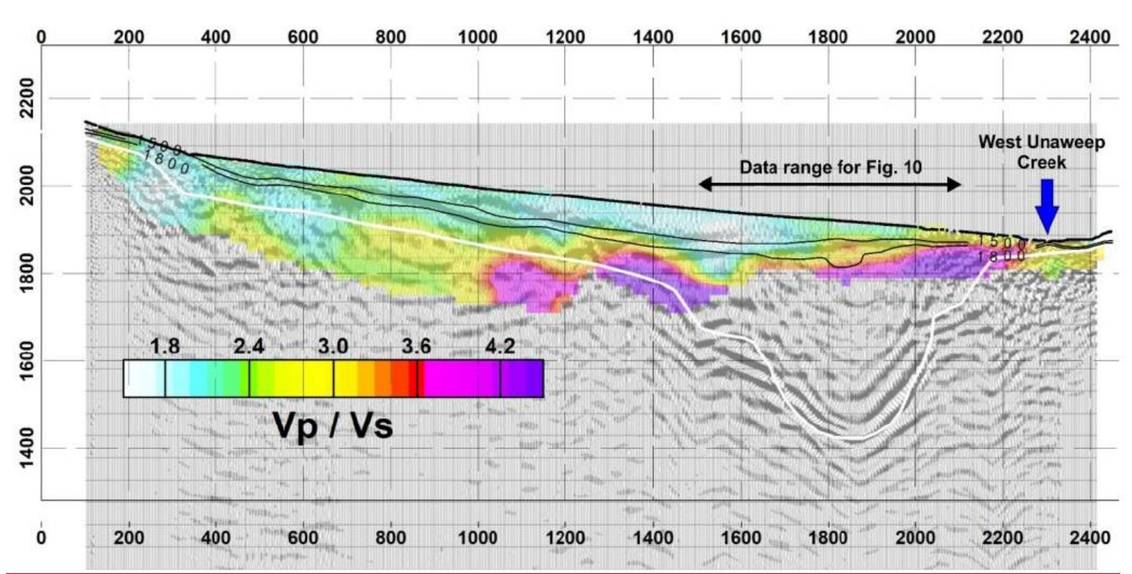


Figure 9: Vp/Vs ratio. White line: Interpreted top of the consolidated Precambrian basement based on P-wave refraction and reflection data. Thin black lines: Contour lines of the P-wave velocity model (Fig. 4) for 1500 m/s and 1800 m/s.

We however added an additional description of the Vp-velocity contours vs. Vp/Vs ratio to the manuscript.

(3) Regarding the Dix-velocities and the authors reply ("*As a result of all these uncertainties, we refrain from a detailed discussion of the lower section of the sediment fill*"): Then please remove the Vp-curve and the derived Vp/Vs-curves from the Fig. 11.(Then you'd also get rid of a problem that Vp in Figure 11 indicates the bedrock at 1600 m, where it increases sharply to 3300 m/s, while you interpret it 150 m deeper, where there is no change in Vp.)

Response to (3):

As discussed in the manuscript, the velocity jump to 3300 m/s at 1600 m elevation is speculated to represent a transition to Paleozoic sediments. There are three reasons why we think it is very unlikely that it represents bedrock: (1) Reflection processing shows the basement at the center of the U-shape to be at elevations between 1500 and 1400 m. (2) Velocity analysis (on which the value of 3300 m/s is based on) uses only reflections above the bedrock, and there are no indication from intra-basement reflections (e.g. velocity analysis cannot provide velocity information on the bedrock). (3) Refraction analysis (delay time model) shows bedrock velocities larger than 5500 m/s for the deepest part of the flat section of the basement, as outlined in section 4 of the manuscript.

We still think that the Vp-curve in the lower section is relevant for discussion, in particular in the context of possible Paleozoic sediments. The depth trend also corresponds to the Vs-curve, which taken together is supporting high velocities in the lower sediments. We agree that showing the Vp/Vs ratios of the deep section is not meaningful, and remove those Vp/Vs ratio curves from the figure.

In that context, the authors note that based on the seismic results they just got awarded an NSF grant to drill the deep sedimentary section. This will falsify or verify many of the interpretations expressed in this manuscript. We add that information to the manuscript in order to notify the scientific community of an upcoming "reality-check" of some of our speculations.

(4) Regarding Gardner: Why don't you write your comment in your paper?

Response to (4):
The comment has been added to the manuscript.

(5) Regarding the ratio profile length to wavelength: Generally, the impact from being able to use 0.3 rather than 1.5 is very significant: You can use frequencies five times smaller than according to the rule-of-thumb, meaning you can look five times deeper than others. If that Is so (?), it would be important to claim it and to analyze more carefully why that is. Naively, one would expect that empirical GF from noise, because they are tainted by non-homogeneous source distribution, require tougher standards. So, what differences in source-receiver-configurations, or dispersion measurement etc, do you refer to, and what would be their impact?

Response to (5):

As for the rule-of-thumb ratio number 1.5, it is regarded as a conservative threshold aims to avoid over-interpretation of the data. However, this ratio number (r) has been updated by many different studies, and we are definitely not the first ones to use values as low as 0.3. Park and Carnevale 2010 indicated that the maximum error in phase velocity is less than five percent (5%) for wavelengths (x) $L \le x \le 2L$, which means the optimal r could be $0.5<r<1$. Pasquet et al., 2015a,b also argued that dispersion curves can be limited down to frequencies at which the spectral amplitude of the shot gather becomes too low, which indicates r could be smaller than 0.5.

Park, C. B., & Carnevale, M. (2010). Optimum MASW Survey — Revisit after a Decade of Use. GeoFlorida 2010: Advances in Analysis, Modeling & Design, ASCE, (Gsp 199), 1303–1312.

Pasquet, S., L. Bodet, A. Dhemaied, A. Mouhri, Q. Vitale, F. Rejiba, N. Flipo, and R. Guérin, 2015a, Detecting different water table levels in a shallow aquifer with combined P-, surface and SH-wave surveys: Insights from VP/VS or Poisson's ratios: Journal of Applied Geophysics, 113,38– 50, doi: 10.1016/j.jappgeo.2014.12.005.
Pasquet, S., L. Bodet, L. Longuevergne, A. Dhemaied, C. Camerlynck, F. Rejiba, and R. Guérin, 2015b, 2D characterization of near-surface VP/VS: Surface-wave dispersion inversion versus refraction tomography: Near Surface Geophysics, 13, 315–331, doi: 10.3997/1873-0604.2015028.

There are also many other successful real data applications with $r < 1.5$ for both active and passive data sets:

r = 0.32 Figure 10a
O'Connell, D. R. H., & Turner, J. P. (2011). Interferometric Multichannel Analysis of Surface Waves (IMASW). Bulletin of the Seismological Society of America, 101(No.5), 2122–2141.

r = 0.36 Figure 9a; r = 0.15 Figure 9b
Pasquet, S., & Bodet, L. (2017). SWIP: An integrated workflow for surface-wave dispersion inversion and profiling. Geophysics, 82(6), WB47–WB61. https://doi.org/10.1190/geo2016-0625.1

r = 0.25 Figure 7
Zhang, Y., Li, Y. E., & Ku, T. (2019). Geotechnical site investigation for tunneling and underground works by advanced passive surface wave survey. Tunnelling and Underground Space Technology, 90(April), 319–329. https://doi.org/10.1016/j.tust.2019.05.003

As for the source-receiver-configurations part, I would first refer Park and Shawver 2009 work. They acquired active MASW data with different offsets and resulted in an improved depth range and imaging profiles.
Park, C. B., & Shawver, J. (2009). MASW Survey Using Multiple Source Offsets. 22nd EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems.

As for the dispersion measurement part, the poor resolution in surface wave imaging at low frequency is the key factor that affects the measured depth range and accuracy. That's why people trying to develop high resolution dispersion measurement, like Luo et al., 2007 apply high-resolution linear Radon transform for dispersion measurement; Zheng and Hu 2017 used nonlinear signal processing technique to improve resolution.

Luo, Y., Xia, J., Miller, R. D., Xu, Y., Liu, J., & Liu, Q. (2008). Rayleigh-Wave Dispersive Energy Imaging Using a High-Resolution Linear Radon Transform. Pure and Applied Geophysics, 165(5), 903–922. https://doi.org/10.1007/s00024-008-0338-4
Zheng, Y., & Hu, H. (2017). Nonlinear signal comparison and high-resolution measurement of surface-wave dispersion. Bulletin of the Seismological Society of America, 107(3), 1551–1556. https://doi.org/10.1785/0120160242

Generally, we argue that the profile length is not the only factor that will affect the measure depth range.

The remarks above have also been integrated into the manuscript.