AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE

10.1002/2017JB014419

A Shallow Seismic Velocity Model for the Groningen Area in the Netherlands

L. J. Hofman[1,2], **E. Ruigrok**[1,2], **B. Dost**[2], **and H. Paulssen**[1]

[1]Department of Earth Sciences, Utrecht University, Utrecht, Netherlands, [2]R&D Seismology and Acoustics, Royal Netherlands Meteorological Institute, De Bilt, Netherlands

Key Points:
- Seismic interferometry is applied to local event data from a dense borehole network in Groningen
- Shallow *P* and *S* wave velocity structure is determined for the Groningen area
- Borehole geophone orientations are derived from cross correlations of local event data

Supporting Information:
- Supporting Information S1

Correspondence to:
L. J. Hofman,
renshofman@gmail.com

Citation:
Hofman, L. J., Ruigrok, E., Dost, B., & Paulssen, H. (2017). A shallow seismic velocity model for the Groningen area in the Netherlands. *Journal of Geophysical Research: Solid Earth*, *122*. https://doi.org/10.1002/2017JB014419

Received 8 MAY 2017
Accepted 25 SEP 2017
Accepted article online 9 OCT 2017

Abstract

The province of Groningen in the Netherlands holds one of the world's largest natural gas fields, and it has been an important source of energy for Western Europe for many decades. The seismicity in recent years called for a better understanding of the local subsurface, and therefore a dense network of 70 boreholes was installed in early 2015. Each borehole is equipped with four geophones and a surface accelerometer. In this study, data from this network are used to determine the shallow velocity structure that is important for the quantification of the seismic hazard and accurate source localizations. Compressional and shear wave velocity profiles with uncertainties are derived for each of the 200 m deep boreholes using passive seismic interferometry applied to local event data. The resulting seismic velocity distributions are presented as contour maps for 50 m depth intervals. The maps show strong lateral variations, where areas of low V_P/V_S ratio correspond to regions of sedimentary infill. The shear wave velocities were derived using the transverse component seismograms. Because the sensor orientations of the borehole geophones were unknown, they had to be determined first. This was done using a novel method based on cross correlations between the geophones and their colocated surface accelerometer. In addition, by extensive cross-correlation analysis over the network, several installation inconsistencies were identified and resolved.

1. Introduction

The province of Groningen in the north of the Netherlands holds the largest onshore natural gas field in Europe. Production of gas from this field started in the early 1960s, shortly after its discovery. About three decades later, earthquakes started to occur that could be related to the Groningen field. Although all earthquakes to date have local magnitudes of less than 3.7, damage has been done to structures due to the shallow hypocenters (around 3 km depth) and soft soil conditions.

In 1995, a network of eight borehole stations was deployed to monitor the seismicity in the area. Since 2015, this network has been expanded with 70 new borehole stations. Each station in this new network consists of an accelerometer at the surface and an array of four 4.5 Hz geophones with a vertical spacing of 50 m (Figure 1). All instruments consist of one vertical and two horizontal components. The horizontal distance between the boreholes is in the order of 5 km. Figure 2 shows the stations of the new network on a map with their corresponding station codes. The stations are numbered from 1 to, and including, 70. Five of the stations (G12, G15, G25, G35, and G48) are excluded from this study because of technical problems. A detailed description of the seismic network in place can be found in Dost et al. (2017). Apart from being able to determine earthquake locations and magnitudes with much greater accuracy than before, borehole seismometers can be used to gain valuable information about the subsurface (e.g., Shearer and Orcutt, 1987).

With the new network in place, Groningen has become the best instrumented place to study induced seismicity by a depleting gas reservoir and the associated seismic hazard. As of 1 January 2017, 1,046 events have been detected, with 284 events larger or equal than local magnitude 1.5 since the first event in 1991. For the north of the Netherlands, detailed geologic information is publicly available from 3-D seismics and an extensive network of logged and cored boreholes (van Dalfsen et al., 2006). Furthermore, there is an extensive subsidence monitoring program in place (Bourne et al., 2014). These data are now augmented by seismic observations from more than 300 geophones and 90 accelerometers, which are all publicly available for scientific research.

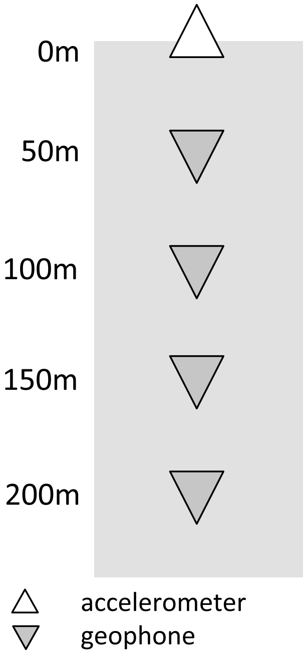


Figure 1. Typical borehole setup for the Groningen network with one surface accelerometer and four downhole geophones.

Many applications require reliable knowledge about the sensor orientations within the network. However, it is a well-known problem for borehole instruments that their orientations are unknown, as a consequence of the way that they are installed. Seismometer orientations can be estimated by evaluating data from sources at known locations. Conventional methods to determine orientations of seismometers include the use of checkshots (Anderson et al., 1987) and teleseismic events (e.g., Diez Zaldívar et al., 2016; Stachnik et al., 2012). However, the soft soil setting and the use of 4.5 Hz sensors complicate the use of teleseismic arrivals in Groningen. Therefore, instead, we devised a new method to estimate the orientations of all geophones in the network using local seismicity. This method takes advantage of consistency of the *S* wave polarizations along the boreholes and allows an estimation of the associated uncertainty for each measurement. Because the plane wave assumption is not satisfied for local events, we rely on near-vertical incidence angles and constant *S* wave polarization along the boreholes.

Using data from this new network and our derived geophone orientations we devised a method to estimate local *P* and *S* wave velocities for all of the boreholes based on passive seismic interferometry (Wapenaar et al., 2010). Seismic interferometry has been used previously for the construction of seismic velocity profiles along boreholes. Miyazawa et al. (2008) applied it to data of industrial noise. Mehta et al. (2007) extracted near-surface *P* and *S* wave velocities using data from a single event, whereas Nakata and Snieder (2012) applied seismic interferometry to a borehole network to obtain near-surface shear wave velocities along the interstation paths. They found that the technique is accurate enough to pick up, among others, seasonal near-surface velocity variations. Hillers et al. (2014) achieved a similar result using full-day cross correlations. We apply the method on local data to derive vertical velocity profiles for each borehole. We also present and implement a method to determine the associated uncertainties. V_P/V_S ratio distributions are obtained from the consistent set of V_P and V_S profiles.

A near-surface seismic velocity model is of paramount importance both for event localization and seismic hazard assessment. The latter can be described by ground motion prediction equations (GMPEs), that is, empirical models that link geophysical properties of the crust to the ground motion over the surface. The *S* wave velocity is one of the key parameters in the GMPE, since the shallow seismic velocity structure has a large influence on the site response. The latest GMPE (v4) (Bommer et al., 2017) contains an *S* wave velocity model that is derived from cone penetration tests combined with lithological maps for the uppermost 50 m. For a deeper level down to 120 m, the *S* wave velocity is estimated from inversion of surface waves (Kruiver et al., 2017). The deepest part of the model, down to a depth of 800 m, relies on a compressional wave (*P* wave) model from surface seismics and *P* wave logs combined with V_P-V_S relations established for a few wells where both *P* wave and *S* wave logging was done. However, no direct V_P and V_S measurements are available at the 70 sites of the new network.

2. Methods and Data

2.1. Geophone Orientations

The first step of the method is to evaluate the coordinate frames of the accelerometers and geophones. The polarity convention used in the accelerometers is that east, north, and upward directions are positive. For the downhole geophones, a different system is used in which the vertical component is positive in the downward direction, and the positive direction of HH1 is 90° clockwise of HH2. Both the geophone and accelerometer data are transformed to a coordinate frame where east (E), north (N), and upward are the positive directions. In the source-receiver coordinate system the radial (R) component is defined as the direction pointing from the source to the receiver, so that the positive axis points away from the source, the transverse (T) component that is positive in the direction 90° clockwise of the positive radial direction, and the vertical component is positive in the upward direction.

2.1.1. Preprocessing

Event metadata is fetched from the Royal Netherlands Meteorological Institute (KNMI) event catalog. Earthquake locations from the data set are plotted in Figure 3. The mean RMS of the travel time residuals in the origin computation is approximately 0.2 s. Details on the location procedure can be found in Spetzler and Dost (2017). A subset of the events in Figure 3 is made where $M_L \geq 1.2$. This ensures sufficient SNRs. All waveform data are corrected for the instrument response and converted to particle velocity. A band-pass filter is applied in a relatively low (3 Hz to 15 Hz) frequency band, because we focus on the *S* waves. Equation (1) shows

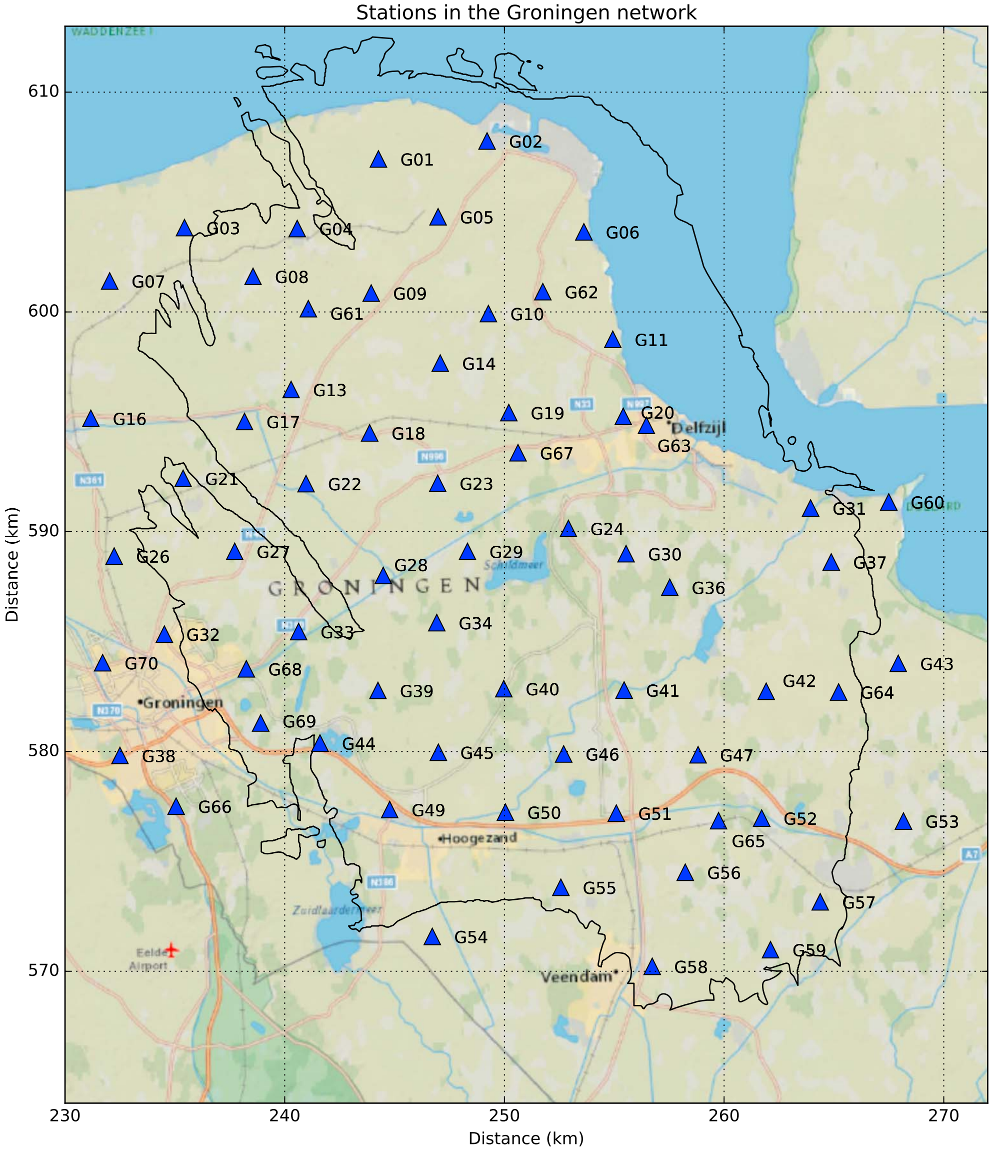


Figure 2. Map of the borehole stations of the Groningen network. Blue triangles indicate the station locations with corresponding station codes. The solid black line indicates the outline of the Groningen gas field. Coordinates are shown in kilometer within the Dutch National Triangulation Grid (Rijksdriehoekstelsel or RD).

how the accelerometer channels are rotated from (N,E) to (R,T), where α represents the event azimuth. The event azimuth α is defined as the direction from a seismic source to the seismic station measured clockwise from north. Hence, for a positive azimuth in equation (1), a clockwise rotation is performed.

$$\begin{pmatrix} \mathrm{R}(+\theta) \\ \mathrm{T}(+\theta) \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \mathrm{N}(+\theta) \\ \mathrm{E}(+\theta) \end{pmatrix} \tag{1}$$

The horizontal channels (HH1 and HH2) of the geophones are rotated by the same amount (equation (1). Because the initial orientations of the geophone channels are unknown, the resulting orientations are not (R,T), but $(\mathrm{R}+\theta,\mathrm{T}+\theta)$, where θ is the clockwise angle between north and channel 2, the angle we want to find.

2.1.2. Cross Correlations and Rotation

The orientation method is based on the waveform similarity between the horizontal channels of the geophone and the colocated accelerometer. Assuming that most of the energy on the horizontal channels comes from the SH wave, the maximum value of the cross correlation (CC) can be used as an indication of the

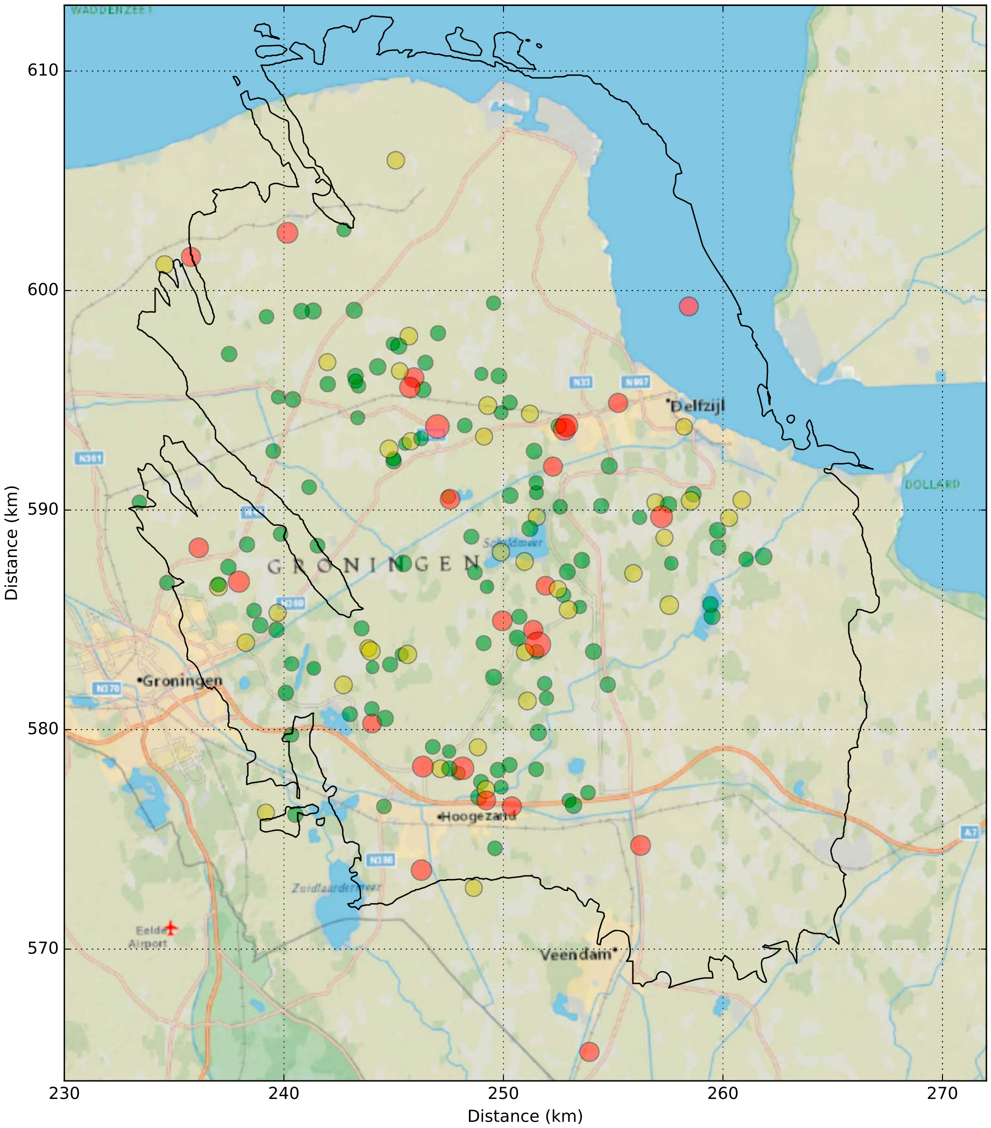


Figure 3. Map of the event locations between January 2015 and June 2016 used in this study (168). The green circles represent earthquakes with $0.0 \leq M_L < 1.0$ (106), yellow circles indicate earthquakes with $1.0 \leq M_L < 1.5$ (36) and red circles indicate earthquakes with $M_L \geq 1.5$ (26).

horizontal alignment of the channels of both instruments. Per event, cross correlations can be made between the R channels of both instruments, as well as the T channels of both instruments. However, the maximum CC then depends on both the orientation theta, as well as the lag time that is determined by the phase velocity. Because the phase velocity is still unknown at this stage, we first do a rough scan over all possible geophone orientations with 20° steps while saving the lag times corresponding to the maximum CC. The mode of these lag times is assumed to correspond to the SH wave. In the next step, the geophones channels are rotated in smaller (3°) steps, while only the CC at this specific lag time is stored. This will produce a sine function where the maximum CC is at the orientation where the geophone channels point in the same direction as the accelerometer channels. The angle corresponding to this maximum is interpreted to be θ in equation (1). The minimum CC is 180° away from the maximum, where the channels point in the exact opposite directions. An example is provided in Figure 4. We then assume that the accelerometer is properly aligned with the N and E directions, making θ the orientation of the geophone with respect to north.

The main reason for the first rough scan to find the time lag is because the maximum CC will suffer from noise when the channel orientations are perpendicular. In this case the CC related to the signal should be zero, but in practice there will always be noise-related peaks in the cross correlation, preventing smooth sine functions as in Figure 4. This also makes the calculation faster, because we only need to compute the full-length cross correlations during the rough scan, while in the smaller steps, a single lag cross correlation suffices. A measure of the uncertainty can be obtained by calculating the standard deviation of the orientations derived from all different lines in Figure 4. A way to test if the geophone orientations are correct is to closely inspect event waveforms before and after making the orientation corrections. An example is included in Figure 5.

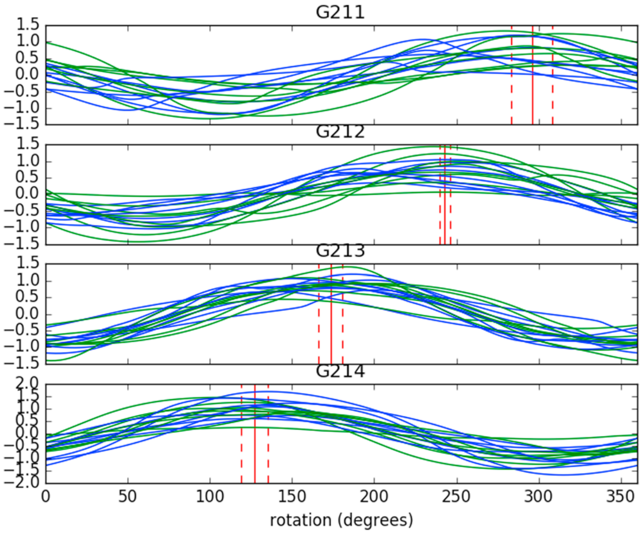


Figure 4. Orientation curves for the geophones in borehole G21 (stations G211, G212, G213, and G214). The lines indicate the maximum cross-correlation coefficient (with the surface accelerometer) as a function of counterclockwise rotation of the geophone. Blue lines are for the radial components and green lines for the transverse components. The vertical red lines indicate the average rotation ±1 standard deviation. The corresponding orientations derived from these data are presented in Table 1.

2.2. Interval Velocities

We correlate local event data recorded at different levels within each borehole to derive individual *P* wave and *S* wave velocity profiles for all stations. The cross correlation and subsequent stacking of signals recorded at two receivers from a distribution of sources can simulate the Green's function response between these two receivers. This means that the response is obtained of a "virtual source" at the location of one of the receivers that is recorded by the other. This requires an isotropic distribution of sources around the borehole, or at least enough of the seismic waves must be from within the Fresnel zones (Figure 6). This means that we assume that a part of the waves enter the borehole (near) vertically. When stacked, the cross correlations of these waves will interfere constructively while cross correlations from sources outside the Fresnel zone will cancel out (Snieder, 2004). Vossen (2016) applied the method to data of the pre-2015 network in Groningen, demonstrated the feasibility, and showed that stable results were obtained for different data sets.

2.2.1. Data Selection and Preprocessing

The events used for the velocity estimation are all events from Figure 3 with $M_L \geq 1.5$. Lower magnitude events are dismissed because they are generally recorded with lower SNRs. If no events were available within the active period of the station, the magnitude threshold is lowered to $M_L \geq 1.0$. For each of the events, 25 s of waveform data are selected, starting at the event origin time (EOT) regardless of the distance from the event to the station. The largest distance is about 35 km, resulting in a (latest) *P* wave onset of approximately 7 s after the EOT.

Due to the low velocities in the near surface in comparison with velocities near the hypocenter and below, *P* waves are primarily recorded on the vertical component. The vertical components are therefore used for the calculation of the *P* wave velocities. To determine the *S* wave velocities, we use the transverse components. This requires an orientation correction as explained in section 2.1.

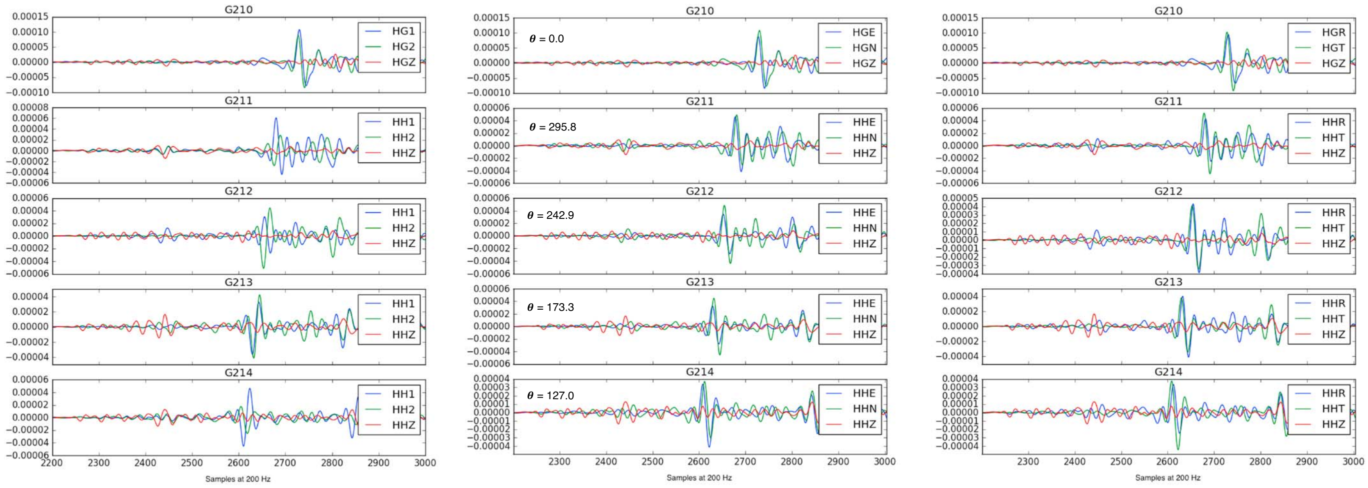
All seismograms are corrected for the instrument response and converted to particle velocity. The waveforms are band-pass filtered using a two-sided Butterworth filter (fourth order) between 3 and 25 Hz. Preliminary testing has shown that this band is appropriate for both *P* and *S* waves (Vossen, 2016). In a few cases where the cross correlations contained too much high-frequency noise by visual inspection, the upper limit of the passband was lowered to 10 Hz.

2.2.2. Cross Correlations and Stacking

The cross correlations are performed in the frequency domain, mainly because of computational efficiency. While in the frequency domain, spectral normalization (whitening) is applied to suppress the effects of an

Table 1
Geophone Orientations for Station G21

Series code	Series channel 1	Series channel 2	Series data used/Out of total	Series standard deviation
G211	25.8°	295.8°	11/16	12.7°
G212	332.9°	242.9°	12/16	3.3°
G213	263.3°	173.3°	9/16	7.0°
G214	217.0°	127.0°	10/16	8.0°



(a) **(b)** **(c)**

Figure 5. Example of three component waveforms for station G21 before the orientation correction, after the orientation correction, and after rotating to radial and transverse components. This example shows data for an M_L 1.7 event on 15 December 2015 (07:43:55). The event is almost pure south of the station, at a distance of approximately 4 km. (a) Waveforms with original (1,2,Z) channels. (b) Waveforms rotated to (E,N,Z) channels using the estimated orientations (θ). (c) Waveforms rotated to (R,T,Z) channels. As expected, the R channels correspond to the N channels and the T channels correspond to the E channels.

uneven noise distribution (Bensen et al., 2007). The amplitude of each frequency in the spectrum is normalized by the average amplitude of a 3 Hz window around that frequency.

Figure 7 shows the major steps taken from individual cross correlations to a velocity profile for station G30. In Figures 7a and 7b the cross correlations are stacked, showing the individual cross correlations in the upper frames and the stacked result in the lower frames. It illustrates that stacking enhances the signal-to-noise ratio, as the signals in the cross correlations interfere constructively, whereas the noise interferes destructively. In Figures 7c and 7d, the stacked cross correlations are shown for all geophone levels in the borehole.

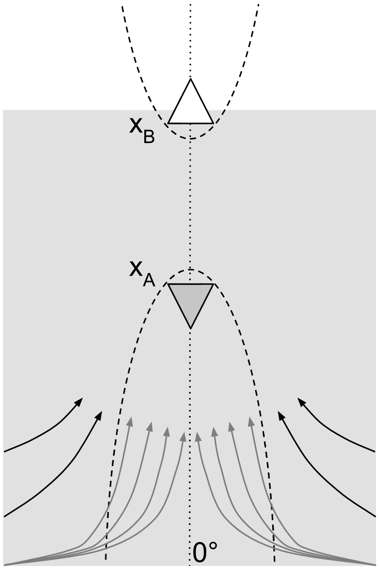


Figure 6. Schematic representation of wave raypaths (arrows) near the borehole to illustrate the notion that waves must pass through the Fresnel zone (dashed lines) to retrieve direct *P* and *S* waves between the two sensors (triangles). The grey arrows represent raypaths within the Fresnel zone, whereas the black arrows represent raypaths outside the Fresnel zone.

2.2.3. Velocity Determination and Uncertainty

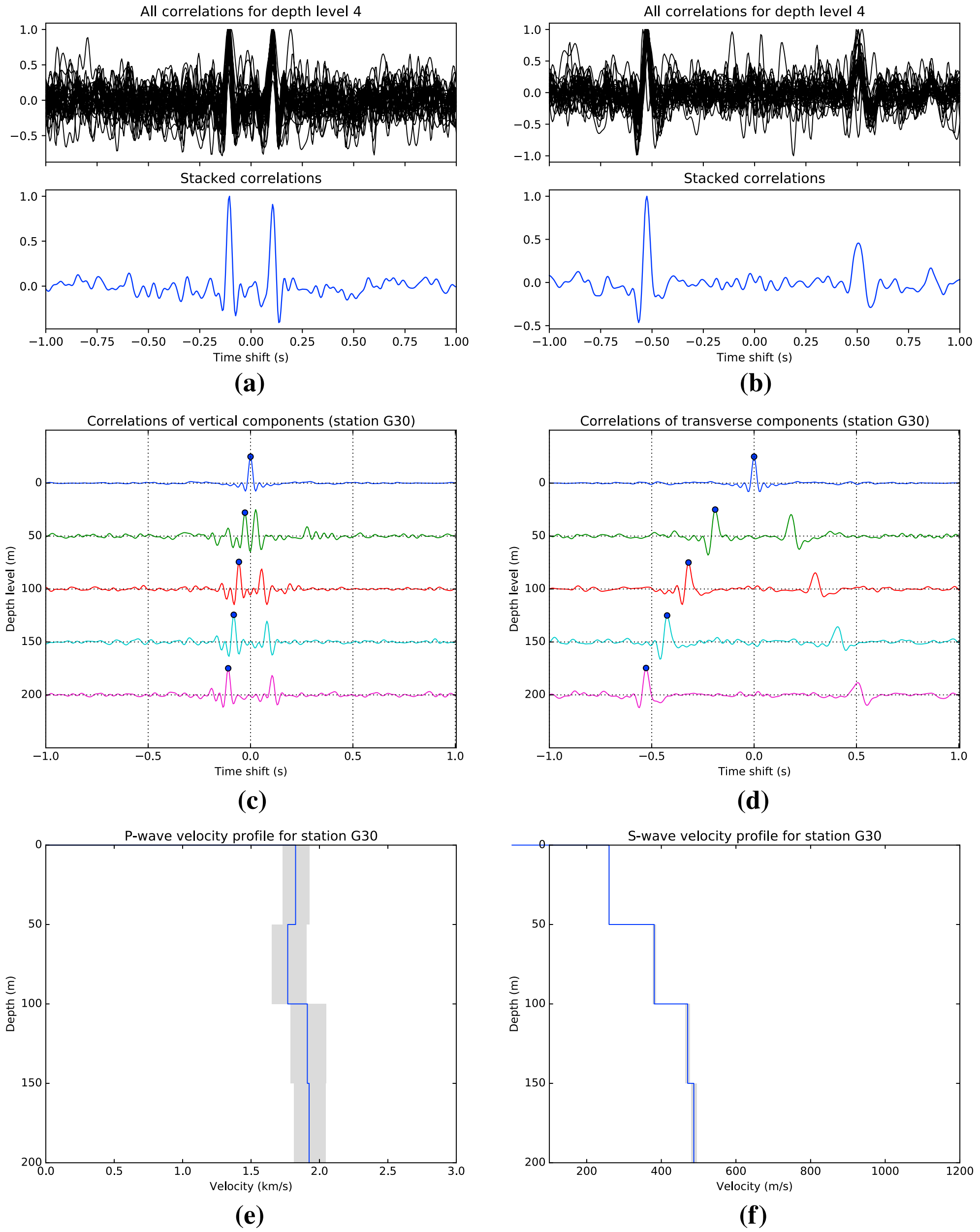
The acausal part of the cross correlations provide an interreceiver travel time measurement of the direct, upgoing wave field from the events, whereas the causal part represents the downgoing wave field. The causal part shows lower amplitudes and a slightly different frequency content compared to the acausal part. This is because the causal part is reconstructed from waves that come from below, travel through the near surface, reflect at the free surface, and propagate downward. The amplitude and bandwidth loss is caused by very low *Q* values over the top 50 m, and in some places by a highly reflective layer at a few meters below the free surface. We therefore choose to use only the acausal part of the cross correlations for the velocity determination. The stacked cross correlations are interpolated with a cubic spline function, from which the maxima are picked (Figures 7c and 7d). These picks are used to calculate the velocity profiles in Figures 7e and 7f.

The main source of uncertainty is thought to be the noise level in the data. The timing can be picked very accurately when the retrieved direct wave has a high signal-to-noise ratio (SNR). However, significant shifts may occur for poor SNR retrievals. In this section we assess how the SNR affects the timing and how the timing uncertainty maps to velocity uncertainty.

We determine the SNR for the acausal part of the retrieved cross correlations as shown in Figures 7c and 7d in decibel as

$$\mathrm{SNR} = 10\log_{10}\frac{P_S}{P_N}, \tag{2}$$

where P_S is the power of the signal and P_N the power of the noise. In the frequency band considered, the dominant frequency of the cross correlations is about 10 Hz. Henceforth, we define the signal as the time-normalized energy in a 0.1 s window around the picked lag time. We define the power of the noise as the time-normalized energy in the remaining 0.9 s of the retrieved (acausal) response.



(a) (b)

(c) (d)

(e) (f)

Figure 7. The steps from cross correlations to a velocity profile for station G30. Data from 30 local events were used for this example. (a) Stacking of cross correlations of the vertical components. (b) Stacking of cross correlations of the transverse components. (c) Stacks of vertical component cross correlations on all depth levels. The dots mark the timing of the retrieved (upgoing) direct *P* wave. (d) Stacks of transverse component cross correlations on all depth levels. The dots mark the timing of the retrieved (upgoing) direct *S* wave. (e) Resulting *P* wave velocity profile with confidence intervals. (f) Resulting *S* wave velocity profile with confidence intervals.

Thus, the noise is defined as all undulations on the retrieved seismogram other than the direct wave within the time window shown in Figures 7b and 7c. Those other undulations, whether or not part of the true subsurface response, may interfere somewhat with the direct pulse and therewith alter the timing.

For mapping SNR to timing errors we take a Monte Carlo approach, with the assumption that timing errors due to different noise realizations are normally distributed. The blue seismogram in Figures 8a–8c shows a synthetic direct wave with a similar frequency content as the Groningen data. This synthetic wave is represented by a Ricker wavelet with a dominant frequency of 10 Hz, time shifted by 0.5 s. Different levels of noise with similar frequency content as the signal are added, yielding the green seismograms in Figures 8a–8c. From each

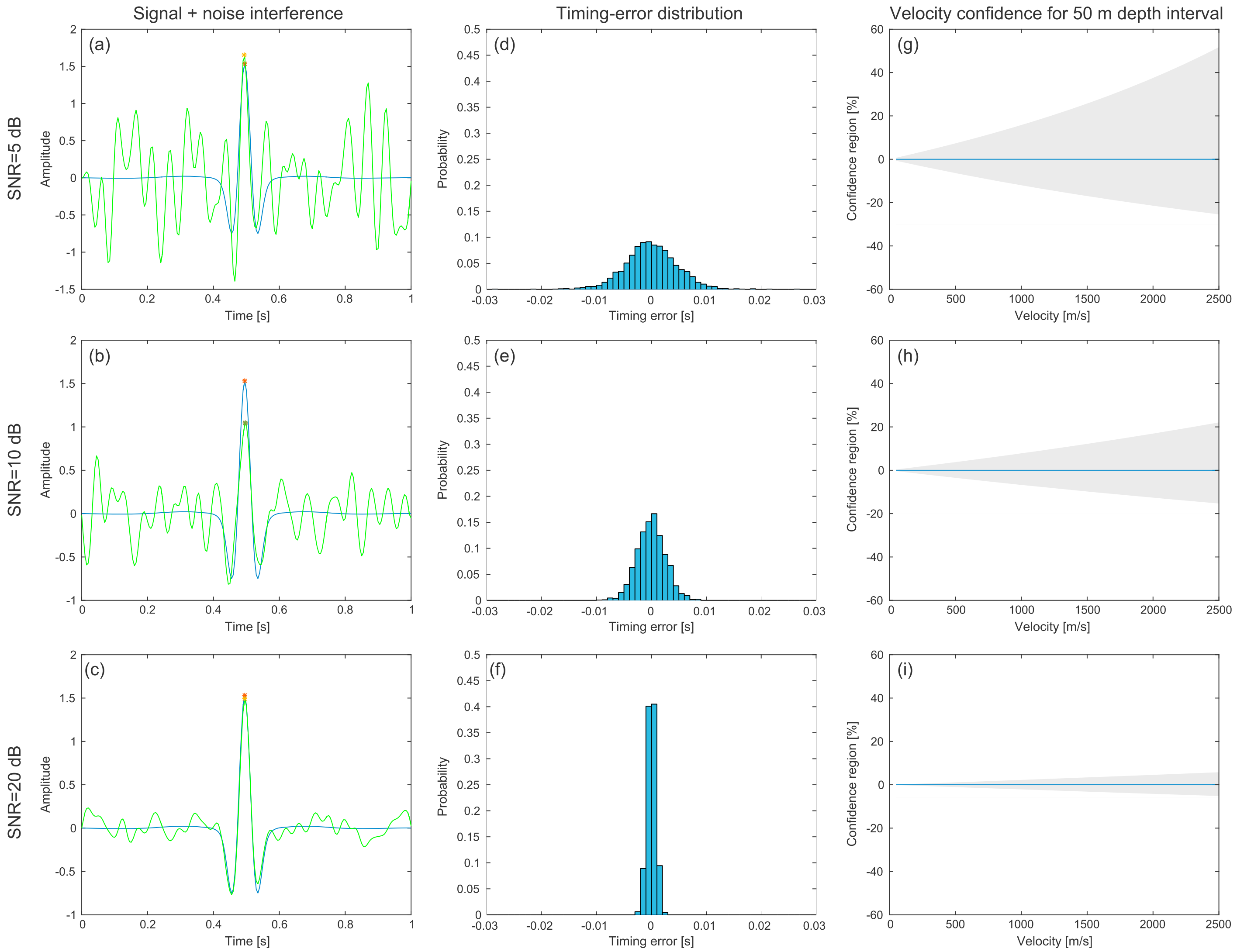


Figure 8. Error propagation for picking the timing of a (retrieved direct) wave, for varying signal-to-noise ratios (SNRs): (a, d, g) 5 dB, to (b, e, h) 10 dB, and (c, f, i) 20 dB. Figures 8a–8c depict the signal of interest (blue seismograms) and signal+noise (green seismograms), together with the picks made (stars). Figures 8d–8f show the picking-error distributions for 2000 noise realizations per SNR. From these distributions the timing-error standard deviation σ_τ is taken and propagated to a velocity confidence area. Figures 8g–8i show the upper and lower velocity error bounds enclosing the grey area, corresponding to a timing error of $-\sigma_\tau$ and $+\sigma_\tau$, respectively. It is assumed that the picks are made for two stations bounding a depth interval of 50 m and that at both stations the direct wave is retrieved with the same SNR.

noisy direct wave we pick the maximum and compute the timing error E_τ with respect to the timing of the clean signal. For each SNR we compute 2,000 noise realizations and corresponding timing errors. This yields the E_τ probability distributions as shown in Figures 8d–8f. For all the distributions, the mean is close to zero, whereas the standard deviation σ_τ decreases with increasing SNR.

The timing errors map nonlinearly to velocity errors. The velocity error is both a function of the timing τ of the pick and the error distribution quantified with σ_τ. To estimate the interval velocity between depth levels z_k and z_{k+1}, the upper and lower bounds of the velocity confidence region $v^\pm$ are given by

$$v^\pm = \frac{z_{k+1} - z_k}{\tau_k - \tau_{k+1} \mp \sqrt{\sigma_{\tau_k}^2 + \sigma_{\tau_{k+1}}^2}}. \tag{3}$$

Note that σ_{τ_0} is equal to zero because the timing of the autocorrelation peak is, by definition, zero.

In Figures 8g–8i we show the velocity confidence regions for the σ_τ found in Figures 8d–8f, for a depth interval of 50 m. To facilitate interpretation, we plot v^+ and v^- as percentage deviations from the true velocity v.

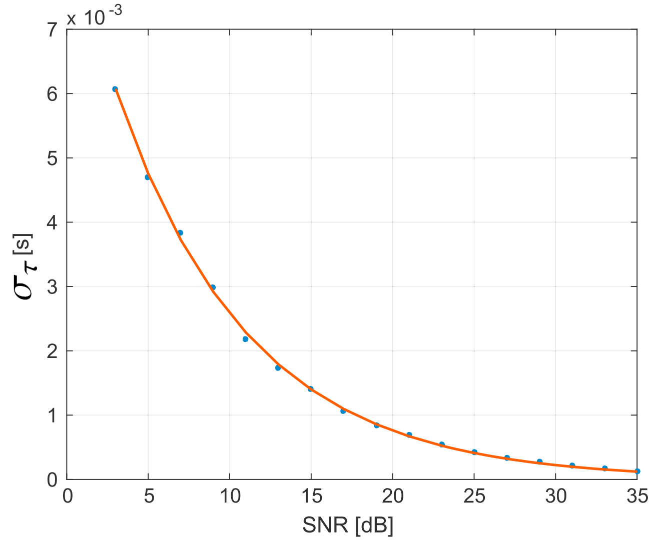


Figure 9. The standard deviation of the timing error σ_τ as function of signal-to-noise ratio (SNR), for a Ricker wavelet with a dominant frequency of 10 Hz and a frequency band spanning from 3 to 25 Hz. The σ_τ (blue dots) are obtained from the timing error distribution for 2000 noise realizations per SNR. The blue points are fitted with a decaying exponential yielding the red curve (equation (4).

As may be expected, picking errors at small lag times and with low SNR maps to large velocity errors. The velocity errors would reduce for larger depth intervals (equation (3).

We expand the above approach by mapping SNR to σ_τ for a range of realistic SNRs. As a minimum SNR we take 3 dB, which corresponds to the mean absolute amplitude of the signal being 1.41 larger than the mean absolute amplitude of the noise. This is about the lower limit SNR for which a pick can be made. As a maximum SNR we take 35 dB, since higher SNRs are unlikely reached in practice. We take a SNR sample size of 2 dB and compute σ_τ(SNR) for each sample, yielding the blue points in Figure 9. The discrete points are fitted with a decaying exponential function:

$$\sigma_\tau = 0.0088\, e^{-0.1223\ \mathrm{SNR}} \tag{4}$$

The above function can likely be extrapolated to higher SNR values. It is not valid, however, for SNR values smaller than 3 dB. For a SNR below 6 dB there is an increasing chance that a noise peak is picked, instead of an interference pattern of signal and noise. Only the latter mechanism is captured in equation (4).

For all estimated interval velocities we determine the SNR using equation (2). Subsequently, we compute the corresponding σ_τ using equation (4) and find the velocity confidence interval with equation (3). These confidence intervals are shown as the grey areas in Figures 7e and 7f and in the supporting information (Figures S1 to S63).

2.2.4. Illumination

Seismic interferometry relies on the assumption that the Fresnel zones are sufficiently illuminated. It can be reasoned that the fastest waves will travel through the deeper structures where the seismic velocity is higher than that in the shallow crust. These will then refract steeply toward the surface and arrive with near-vertical incidence.

To explicitly check whether this assumption holds, an experiment was set up where the angles of incidence are determined for different epicentral distances (Figure 10). For a single event, three stations were selected at close, intermediate, and far range. For each of the stations, the angle of incidence is calculated from the particle polarization (data variance tensor, Aster et al., 1990).

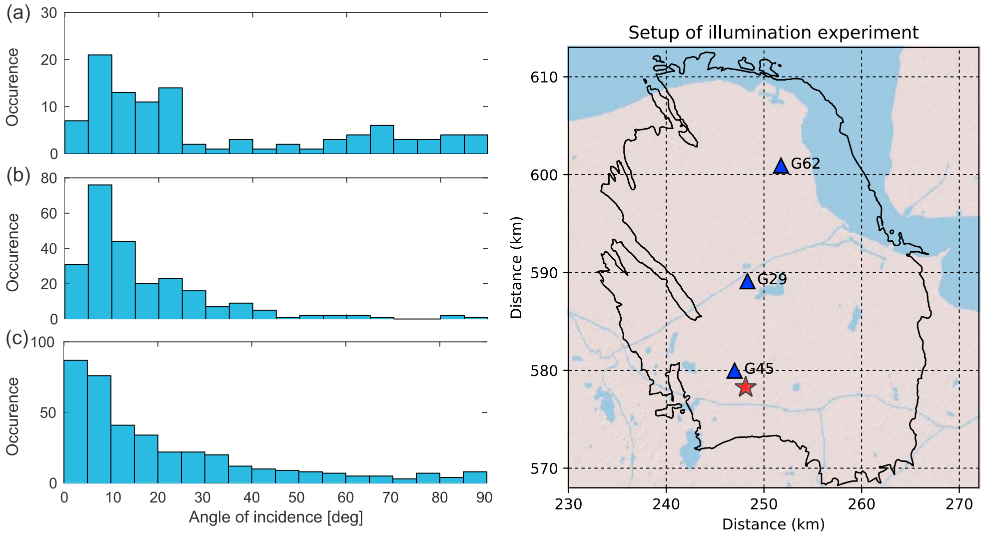


Figure 10. (left) Incidence angles of illumination distributions, at (a) close range (2.0 km), (b) medium range (10.8 km), and (c) large range (22.9 km). The angles are estimated for a moving time window that spans half the average period in the frequency band between 3 and 25 Hz. (right) Plot of event and stations used in this experiment. The data used are of an M_L 2.4 event on 25 February 2016 (22:26:30) (indicated on this map with a star), recorded by the surface instrument of stations G54, G29, and G62 (indicated on this map with triangles).

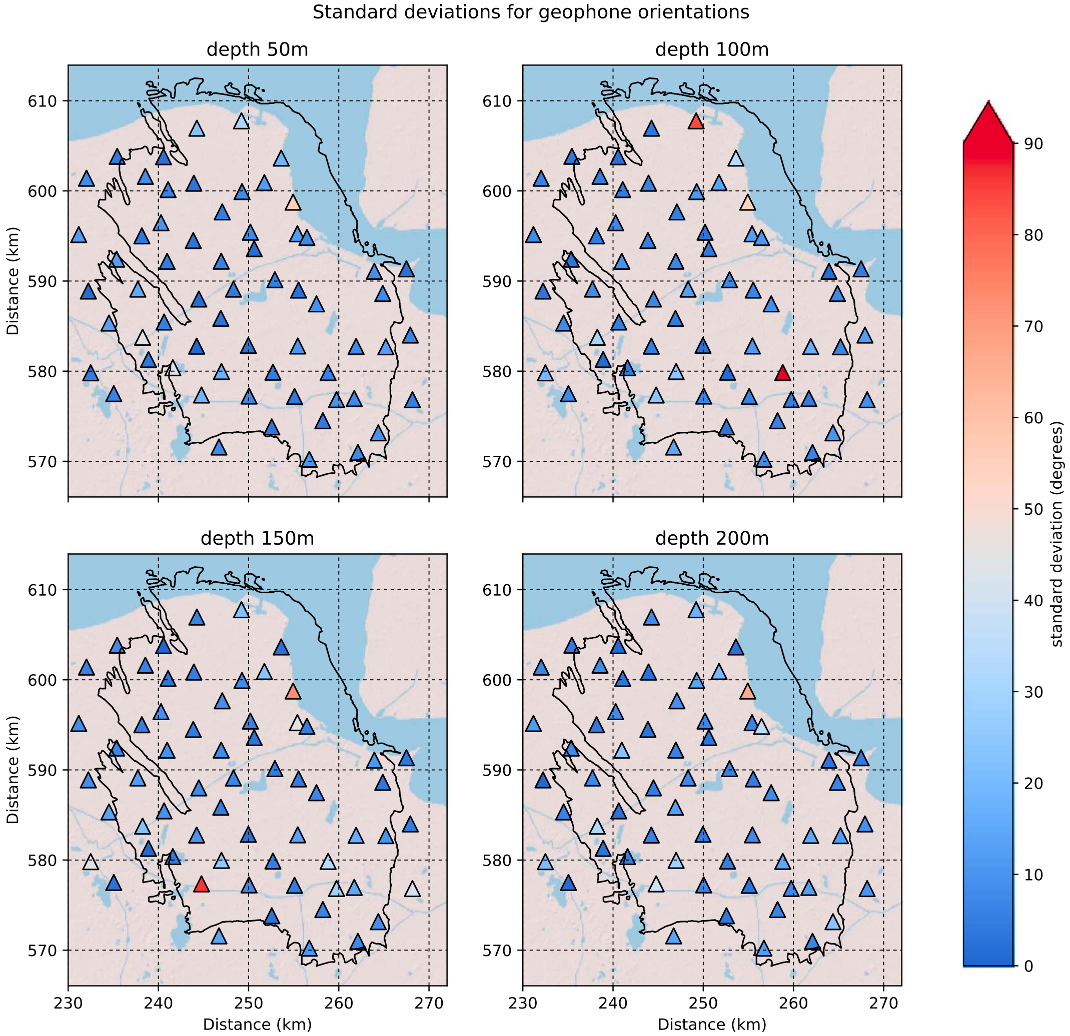


Figure 11. Standard deviations for the orientations of the geophones.

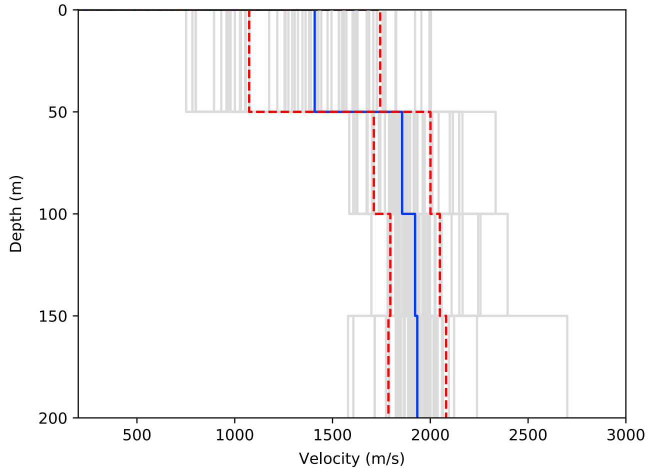


Figure 12. *P* wave velocity profiles for all 60 individual stations are drawn in light grey. The blue line indicates the mean velocity profile. The dashed red lines indicate the range covered by 1 standard deviation from the mean for each of the intervals.

A time window is selected from the first *P* to just before the direct *S* wave arrival. This time window is dominated by *P* waves. Consequently, the particle polarization can be directly translated to an angle of incidence. Nonetheless, also in this time window there is a small contribution of *P*-*S* conversions leading to a set of spurious estimations at angles larger than about 45°.

For retrieving the direct wave between receivers that are vertically aligned, the first Fresnel zone would encompass angles of incidence spanning from 0 to about 10° (depending on the frequency). From Figure 10 it can be seen that the fraction of near-vertical illumination increases with source-receiver distance. However, even the event at close range contains contributions in the 0 to 5° bin. The illumination for the actual experiment is much richer, by using many more than three events and including much of the coda waves. It is expected that the later coda contains a more diverse range of illumination, also further filling up angles near zero. Hence, lack of illumination is unlikely a source of significant errors for the Groningen data set.

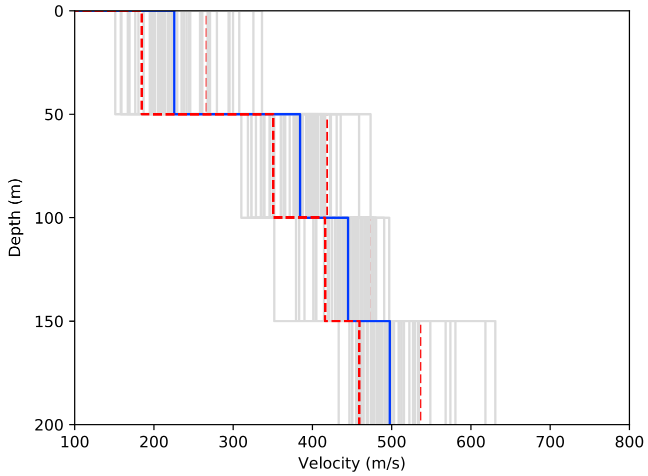


Figure 13. *S* wave velocity profiles for all 60 individual stations are drawn in light grey. The blue line indicates the mean velocity profile. The dashed red lines indicate the range covered by 1 standard deviation from the mean for each of the intervals.

3. Results

In this section, results are presented for all of the stations that were online and had data available at the time of this study (12 July 2016). Stations G12, G15, G25, G35, and G48 were not available due to technical issues. For stations G05 and G70 there were no event data available. Station G09 was left out of the results, because the velocities derived from this station were unrealistically high, and stations G38 and G44 were left out due to the large amount of noise.

3.1. Geophone Orientations

Results for the orientations of all individual geophones are provided in the supporting information (Figures S64 to S126). An example is given in Figure 4 and Table 1. The orientations are provided in degrees clockwise from north. The column “Data used” shows how many traces were used for the final calculation and how many were initially found within the station uptime The standard deviations are based on orientations calculated from the individual events used. The standard deviations for the orientations of all geophones are plotted in Figure 11.

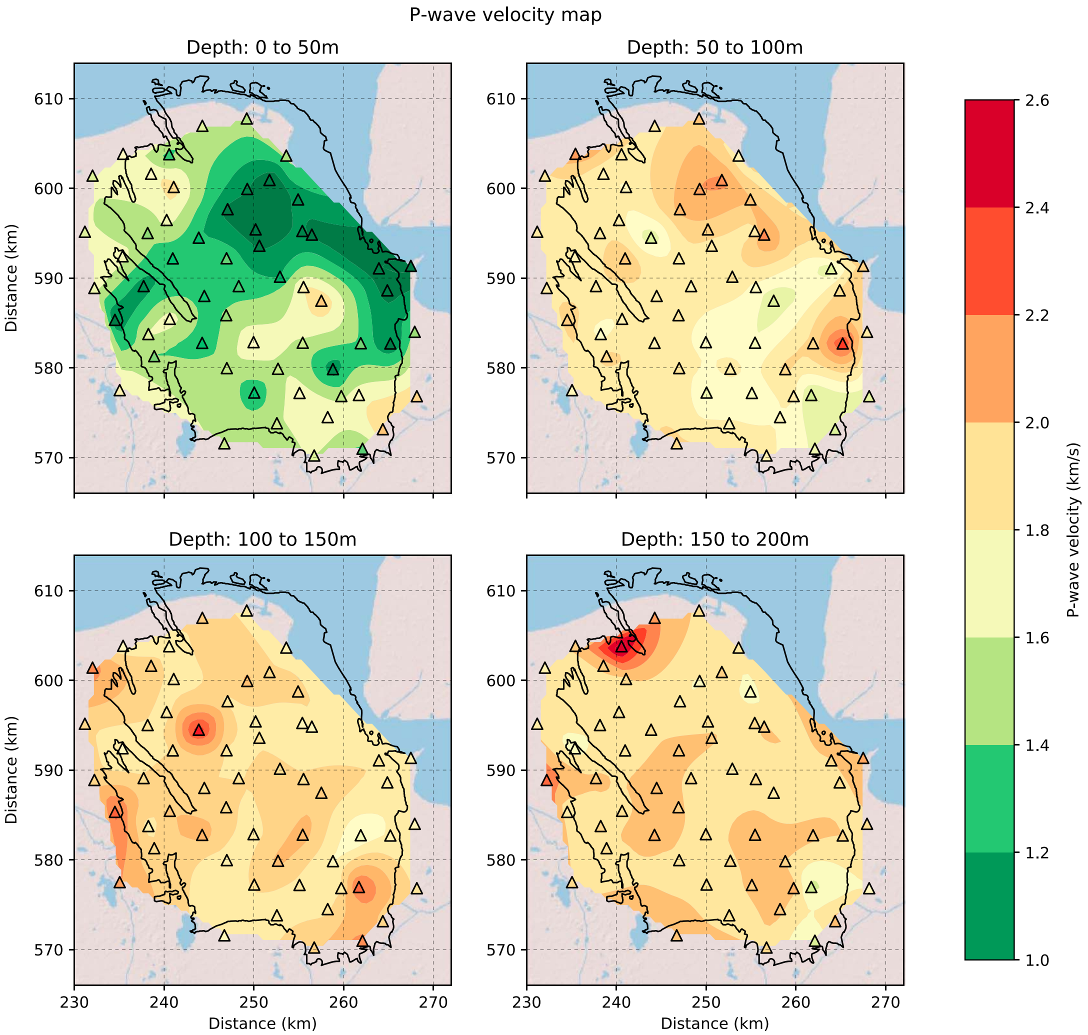


Figure 14. Map showing the *P* wave velocity for all stations in the Groningen network in four depth intervals. The colored triangles represent the values calculated from seismic interferometry. Colors in between stations are interpolated using the nearest neighbor method, exclusively for visualization purposes.

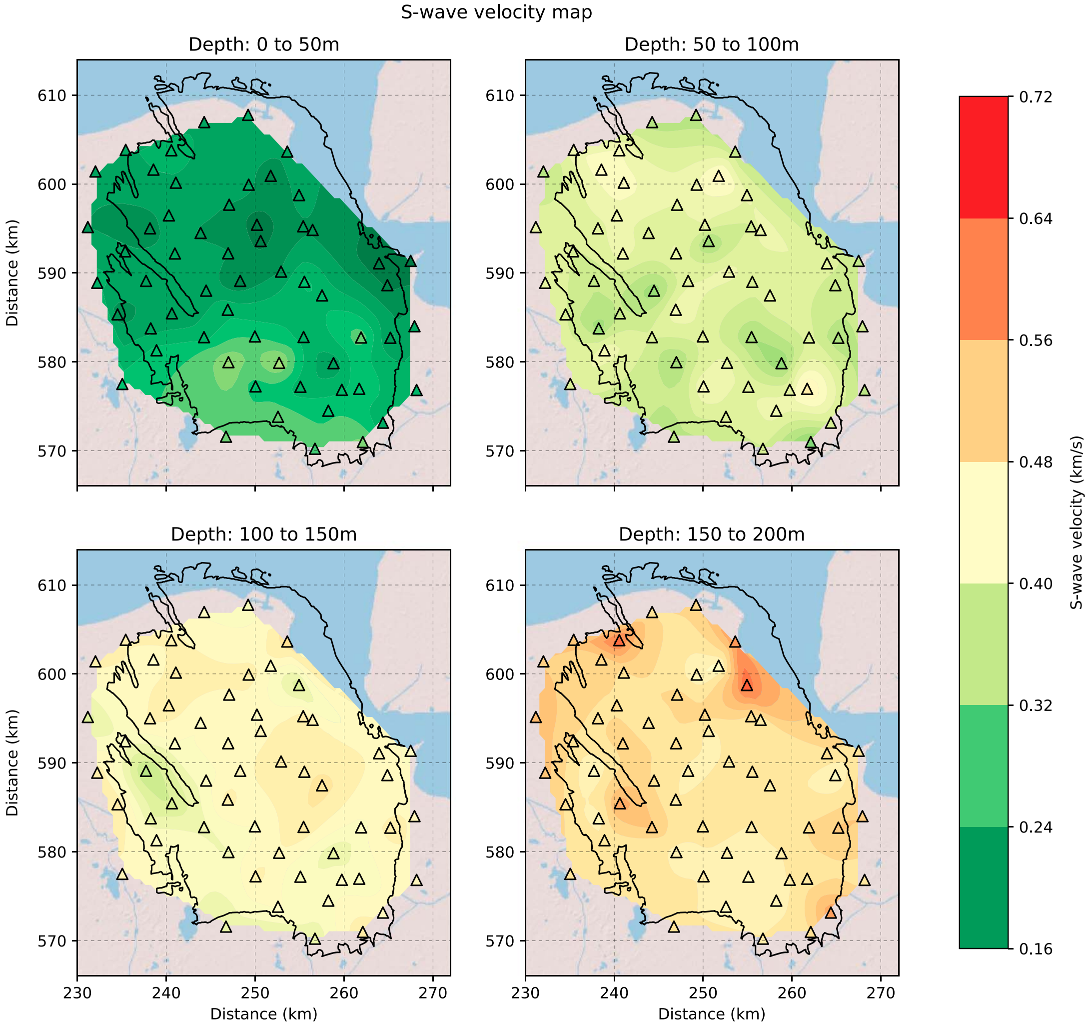


Figure 15. Map showing the *S* wave velocity for all stations in the Groningen network in four depth intervals. The colored triangles represent the values calculated from seismic interferometry. Colors in between stations are interpolated using the nearest neighbor method, exclusively for visualization purposes.

For sensors that have sufficient data available and high SNRs, we see that the orientation curves assume perfect sine functions (e.g., Figure 4). The standard deviation of the maxima of these functions is in the order of a few degrees. With decreasing SNRs the maxima of the curves are further apart. Also, the amplitude of the curves decreases and the functions become more complex. In these cases, extra tests have to be done to test the validity of the results. An example of such a test is shown in Figure 5.

3.2. Seismic Velocities

The vertical *P* wave and *S* wave velocity profiles for each of the 60 stations included in this study are shown in Figures 12 and 13, respectively. For both distributions a mean profile is also indicated together with the standard deviation from the mean.

Figures 14 and 15 show maps of the *P* and *S* wave interval velocities, where green colors represent low seismic velocities and red colors represent high seismic velocities. Colors inside the black triangles (stations) correspond with the values calculated with the method described in section 2.2. The colored areas in between stations correspond with values interpolated with a nearest neighbor method. Note that the interpolation is made exclusively for visualization purposes. These figures should not be interpreted as tomographic images, since no interstation travel time measurements are made in this study.

Figure 16 presents the average V_P/V_S ratio over the upper 200 m. This map is obtained by harmonically averaging the interval V_P/V_S ratios over the borehole for all stations. The spatial variation of the ratio is significant, varying from 4.0 to 5.2. The largest constituent within this lateral variation is the variation in *P* wave velocity

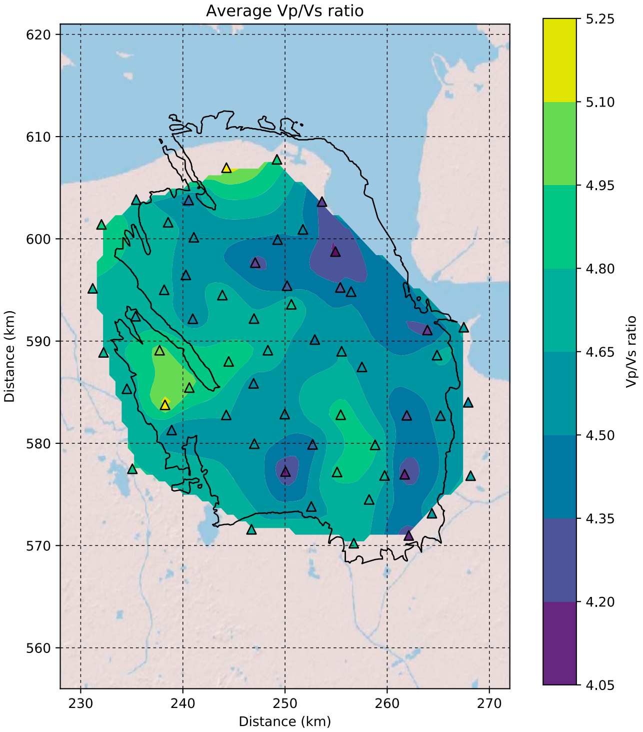


Figure 16. Ratio of the average V_P and V_S for each borehole. Like in Figures 14 and 15, the values are interpolated exclusively for visualization purposes.

over the upper 50 m (Figures 12 and 14). In the northeast, the top 50 m *P* wave velocities are very low. The areas of low V_P correspond to regions where valleys were carved out by glaciers during the Elsterian glaciation and subsequently filled with subglacial deposits of the Peelo formation (Gunnink et al., 2013).

The results for the uppermost depth interval can be compared to the V_{S30} map in Kruiver et al. (2017). Their map is based on a lithological map, where each lithology is associated with a typical seismic *S* wave velocity. The associations are based on seismic cone penetration tests for each of the lithologies and assumed to represent the average seismic *S* wave velocity for the top 30 m. This approach is completely different from the method used in this study, allowing an independent comparison of the results. In Figure 17, the V_{S30} from Kruiver et al. (2017) is plotted next to the V_{S50} results from this study.

The V_{S30} values in Kruiver et al. (2017) range from 156 m/s to 275 m/s. In Figure 13, we can see that the values for the *S* wave velocity in the upper interval plot are mostly within the range from 150 m/s to 300 m/s. This is a very good overlap, considering the different approaches taken. On average, the velocities derived for the upper 30 m (Figure 17, left) are lower than the velocities obtained for the upper 50 m (Figure 17, right). This can be expected because the shear wave velocity generally increases with depth. When comparing the maps of Figure 17, it should be noted that the values from this study are interpolated in between the stations. It therefore has lower spatial resolution and will not be able to map the geological structures that can be identified in the map by Kruiver et al. (2017). However, the large-scale lateral differences in *S* wave velocity are in agreement. An east-west line through the middle of the Groningen field at approximately 585 km N separates a zone of relatively low velocity in the north from a higher velocity area in the south. The southern part has a lower velocity zone in the northeast, around stations G31, G37, and G60. The similarity between the maps in Figure 17 means that the lateral differences in shear wave velocity in this interval can be explained by lithological differences.

Pleistocene sands with relatively high seismic velocities can be found at shallow depths in the southern part, whereas in the northern part of the area these sands are covered by lower velocity Holocene deposits (mainly clay and peat). This distinction can also be seen in the soil map from Meijles (2015). North of the city of Groningen (approximately 585 km N), the dominant soil type is clay, and south of Groningen the soil is characterized mainly by sand. A more detailed look at the sand-clay boundary in Meijles (2015) shows that the sands reach further north at two locations; just east of Groningen (around 248 km E) and around Winschoten

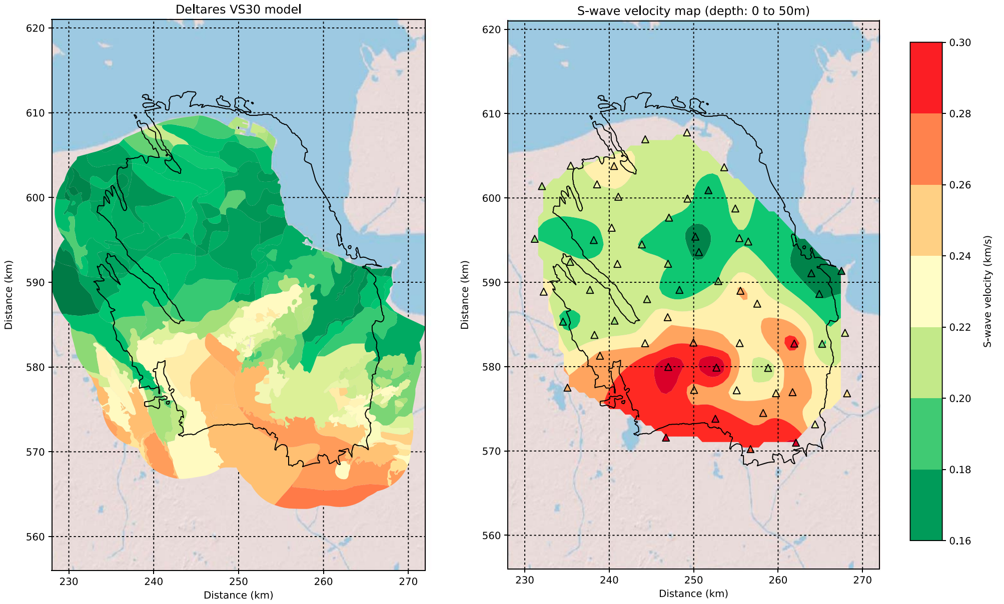


Figure 17. (left) V_{S30} map from Kruiver et al. (2017). (right) *S* wave velocities between 0 and 50 m depth from this study. The color scale is different from the one in Figure 15 to match the colors of the V_{S30} map.

(262 km E). These features correlate well with relatively high shear wave velocity zones in the *S* wave velocity map (Figure 17, right).

Kruiver et al. (2017) presented a V_{S65} model based on the (re)interpretation of surface wave data. The values in their model are average *S* wave velocities for a depth interval from 50 m to 120 m and range from 298 m/s to 453 m/s. This range matches well with the *S* wave velocities for the second depth interval (50 to 100 m) found in this study (Figure 13). The large amount of source-receiver paths used to construct the V_{S65} map in Kruiver et al. (2017) allowed a much higher spatial resolution, but the overall pattern agrees quite well.

In Figure 13, we can see that from the second to the lowest interval (50 m to 200 m depth) the average *S* wave velocity increases with a linear trend from 384 m/s to 498 m/s. In Kruiver et al. (2017) the deeper V_S structure is also assumed to increase linearly from 65 m to 350 m depth, where it reaches a stationary value of 639 m/s.

The method yields seismic velocities within the expected range, which agree well with results from previous studies that have used different approaches. The large-scale spatial variation in the seismic velocities can be explained by the local geology.

4. Discussion

4.1. Geophone Orientations

The geophone orientations presented in this study are relative to the surface accelerometer at each particular station. They can only be regarded as geographical orientations under the assumption that the surface accelerometer is accurately aligned.

A different possibility would have been to use teleseismic events instead of local events for estimating the geophone orientations. This has the advantage that the plane wave assumption is always satisfied and that there is less scattering from local structures. However, it was concluded that typical SNRs were too low for teleseismic events. This can be explained, first, by the fact that the stations are placed directly above the gas reservoir, in loose Pleistocene and Holocene sediments. Seismic noise induced by local sources remains largely trapped in the loose sediments, therewith compromising the ability of the network to measure weak waveforms induced by distant sources. Second, teleseismic events have much lower frequency than local events, making the 4.5 Hz geophones more sensitive to local rather than teleseismic seismicity.

In some cases, even local events are recorded with insufficient SNRs. Station G02, the northernmost station, (see Figure 2) suffers from very high standard deviations (Figure 11), because it is placed inside the Eemshaven harbor, surrounded by a lot of noise sources from traffic, windmills, ships, and industry.

Other cases of large standard deviations can be caused by lack of data. Some of the stations had been running only for a few (quiet) months at the time of this study, and had data only for a few events. While for some stations this does not seem to be a problem (e.g., G69, see supporting information Figure S126), for other stations it can be problematic (e.g., G11, see supporting information Figure S73). Lowering the magnitude threshold from 1.5 to 1.0 can be a solution, as more events become available, but lower magnitude events generally also have lower SNRs.

4.2. Seismic Velocity Profiles

We estimated borehole *P* and *S* wave velocity profiles by creating a virtual source at the surface sensor and picking the arrival time of the retrieved upgoing waves at all the other sensors. In an alternative implementation, the virtual source was placed at the uppermost geophone (50 m depth). This provides another way to derive the velocity in the uppermost interval, because the difference between the direct wave and the surface reflection is the travel time of the wave propagating up and down the uppermost interval. However, a highly reflecting interface in the upper 50 m, which is not uncommon, could produce a second reflection which is hard to distinguish from the surface reflection. An example of this effect can be seen in Figure S130 in the supporting information. With a virtual source at the surface, only transmitted waves are used. It was concluded that these transmitted waves are more isolated and better to pick than the reflected waves.

Initially, a few anomalies existed in both the *P* wave and *S* wave maps where the seismic velocity was extremely high in the lowest interval. The question was whether this was a natural high velocity structure like a salt diapir, or an error on the lowest geophone. The fact that they occurred on both the *P* wave and *S* wave velocity maps could not rule out either of these possibilities. Reports from the installation of the borehole stations gave the answer. The lowest geophone was placed at a depth of 185 m instead of 200 m. After studying the reports of all of the stations, there seemed to be several boreholes where such inconsistencies existed. These are listed in Table S65 in the supporting information. The report of station G09 states no such thing. However, the seismic velocities on the lowest interval of this station are also extremely high. Because no geological explanation could be found and the data quality seems to be very good, it is suspected that the lowest geophone of this station was also placed too shallow. Station G09 is not included in Figures 14 and 15.

5. Conclusions

Vertical seismic velocity profiles for both *P* and *S* waves were derived for nearly all of the borehole stations in the Groningen network using passive seismic interferometry. In previous studies, the shear velocity structure used for the GMPE depends on different techniques for different depth intervals. This study provides a single method to derive both the *P* wave and *S* wave velocities within four intervals from 0 m to 200 m depth. The resulting shear wave velocity map for the upper interval (0 to 50 m) correlates well with the seismic velocity structure derived from surface lithology, both in the range of velocities found and the spatial distribution of the velocities. For the second interval (50 to 100 m), the range of velocities found is in good agreement with results from surface wave inversion. The spatial distribution is harder to compare because of the difference in resolution, but the large-scale structure is similar. For the intervals from 100 to 150 m and 150 to 200 m, the results show an *S* wave velocity gradient that is similar to what previous studies have found. This study also presents the first detailed near-surface *P* wave velocity model for the region. Hence, no direct comparison of these results could be made with previous studies. We also found that V_P/V_S ratios show a large spatial variation over the Groningen area. This spatial variation needs to be taken into account when the arrival time difference of *P* and *S* waves is used for event location.

Together with the velocity profiles, we estimated the uncertainties for each individual station. We identified interference of signal with noise as the main cause of velocity errors, and using a Monte Carlo approach we mapped signal-to-noise ratio to a timing error standard deviation. The latter was subsequently mapped to a velocity confidence region. Especially the *P* wave velocity estimates are sensitive to timing errors, while the *S* wave velocity estimates are more robust. This is caused by significantly lower *S* than *P* wave velocities in the near surface.

Because accurate orientations were required to derive the velocities, a new method to estimate borehole instrument orientations was developed. The method is based on cross correlations with sensors at the surface. This result will be useful in the future for many other applications.

Lastly, by processing and analyzing a large amount of data from the new Groningen network, several errors and inconsistencies were detected. In some cases the connections of the geophones were confused, and in some stations the depth of the geophones deviate from the intended depths.

Acknowledgments
Waveform data were provided by the KNMI and are publicly available through the FDSN dataselect webservice (http://rdsa.knmi.nl/fdsnws/dataselect/1/). An event catalog and station metadata are available through the event webservice (http://rdsa.knmi.nl/fdsnws/event/1/), and the station webservice (http://rdsa.knmi.nl/fdsnws/station/1/). Data for the left part of Figure 17 were provided by Deltares (Kruiver et al., 2017). ObsPy, a Python framework for seismology (Beyreuther et al., 2010), was used to easily fetch and process (meta)data. The matplotlib library for Python (http://matplotlib.org/) was used to produce the figures. Background maps were downloaded from the ArcGIS online webservice (http://server.arcgisonline.com/arcgis/rest/services/). We want to thank the three anonymous reviewers for their comments and suggestions.

References

Anderson, P., Duennebier, F., & Cessaro, R. (1987). Ocean borehole horizontal seismic sensor orientation determined from explosive charges. *Journal of Geophysical Research*, *92*(B2), 3573–3579.

Aster, R. C., Shearer, P. M., & Berger, J. (1990). Quantitative measurements of shear wave polarizations at the Anza seismic network, Southern California: Implications for shear wave splitting and earthquake prediction. *Journal of Geophysical Research*, *95*(B8), 12,449–12,473. https://doi.org/10.1029/JB095iB08p12449

Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A., Lin, F., Moschetti, M., … Yang, Y. (2007). Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, *169*(3), 1239–1260.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, *81*(3), 530–533. https://doi.org/10.1785/gssrl.81.3.530

Bommer, J. J., Stafford, P. J., Edwards, B., Dost, B., van Dedem, E., Rodriguez-Marek, A., … Ntinalexis, M. (2017). Framework for a ground-motion model for induced seismic hazard and risk analysis in the Groningen gas field, the Netherlands. *Earthquake Spectra*, *33*(2), 481–498.

Bourne, S., Oates, S., Van Elk, J., & Doornhof, D. (2014). A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir. *Journal of Geophysical Research: Solid Earth*, *119*, 8991–9015. https://doi.org/10.1002/2014JB011663

Diez Zaldívar, E. R., Priolo, E., Grigoli, F., & Cesca, S. (2016). Misalignment angle correction of borehole seismic sensors: The case study of the collalto seismic network. *Seismological Research Letters*, *87*(3), 668–677. https://doi.org/10.1785/0220150183

Dost, B., Ruigrok, E., & Spetzler, J. (2017). Development of probabilistic seismic hazard assessment for the Groningen gas field. *Netherlands Journal of Geosciences*. https://doi.org/10.1017/njg.2017.20

Gunnink, J., Maljers, D., Van Gessel, S., Menkovic, A., & Hummelman, H. (2013). Digital geological model (DGM): A 3D raster model of the subsurface of the Netherlands. *Netherlands Journal of Geosciences*, *92*(1), 33–46.

Hillers, G., Campillo, M., & Ma, K.-F. (2014). Seismic velocity variations at TCDP are controlled by MJO driven precipitation pattern and high fluid discharge properties. *Earth and Planetary Science Letters*, *391*, 121–127. https://doi.org/https://doi.org/10.1016/j.epsl.2014.01.040

Kruiver, P. P., van Dedem, E., Romijn, R., de Lange, G., Korff, M., Stafleu, J., … Doornhof, D. (2017). An integrated shear-wave velocity model for the Groningen gas field, the Netherlands. *Bulletin of Earthquake Engineering*, *15*(9), 3555–3580.

Mehta, K., Snieder, R., & Graizer, V. (2007). Extraction of near-surface properties for a lossy layered medium using the propagator matrix. *Geophysical Journal International*, *169*(1), 271–280. https://doi.org/10.1111/j.1365-246X. 2006.03303.x

Meijles, E. (2015). *De Ondergrond van Groningen: Een Geologische Geschiedenis*. (pp. 24). NAM: Assen, Netherlands.

Miyazawa, M., Snieder, R., & Venkataraman, A. (2008). Application of seismic interferometry to extract P and S wave propagation and observation of shear-wave splitting from noise data at Cold Lake, Alberta, Canada. *Geophysics*, *73*(4), D35–D40.

Nakata, N., & Snieder, R. (2012). Estimating near-surface shear wave velocities in Japan by applying seismic interferometry to KiK-net data. *Journal of Geophysical Research*, *117*, B01308. https://doi.org/10.1029/2011JB008595

Shearer, P., & Orcutt, J. (1987). Surface and near-surface effects on seismic waves-theory and borehole seismometer results. *Bulletin of the Seismological Society of America*, *77*(4), 1168–1196.

Snieder, R. (2004). Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *69*, 046610.

Spetzler, J., & Dost, B. (2017). Hypocentre estimation of induced earthquakes in Groningen. *Geophysical Journal International*, *209*(1), 453–465. https://doi.org/10.1093/gji/ggx020

Stachnik, J., Sheehan, A., Zietlow, D., Yang, Z., Collins, J., & Ferris, A. (2012). Determination of New Zealand ocean bottom seismometer orientation via Rayleigh-wave polarization. *Seismological Research Letters*, *83*(4), 704–713. https://doi.org/10.1785/0220110128

van Dalfsen, W., Doornenbal, J., Dortland, S., & Gunnink, J. (2006). A comprehensive seismic velocity model for the Netherlands based on lithostratigraphic layers. *Geologie en Mijnbouw/Netherlands Journal of Geosciences*, *85*(4), 277–292.

Vossen, C. (2016). Near-surface velocity profiles of Groningen and surroundings extracted using seismic interferometry (*MSc Guided Research Report*): Universiteit Utrecht.

Wapenaar, K., Draganov, D., Snieder, R., Campman, X., & Verdel, A. (2010). Tutorial on seismic interferometry: Part 1—Basic principles and applications. *Geophysics*, *75*(5), 75A195–75A209.