Relocation of clustered earthquakes in the Groningen gas field

Lisanne Jagt[1,*], Elmer Ruigrok[1,2] & Hanneke Paulssen[1]

[1] Department of Earth Sciences, Utrecht University, Heidelberglaan 2, Utrecht, the Netherlands

[2] R&D Seismology and Acoustics, Royal Netherlands Meteorological Institute, De Bilt, the Netherlands

* Corresponding author. Email: e.a.a.m.jagt@students.uu.nl

Abstract

Previous locations of earthquakes induced by depletion of the Groningen gas field were not accurate enough to infer which faults in the reservoir are reactivated. A multiplet analysis is performed to identify clusters of earthquakes that have similar waveforms, representing repeating rupture on the same or nearby faults. The multiplet analysis is based on the cross-correlation of seismograms to assess the degree of similarity. Using data of a single station, six earthquake clusters within the limits of the Groningen field were identified for the period 2010 to mid-2014. Four of these clusters were suitable for a relocation method that is based on the difference in travel time between the P- and the S-wave. Events within a cluster can be relocated relative to a master event with improved accuracy by cross-correlating first arrivals. By choosing master events located with a new dense seismic network, the relocated events likely not only have better relative, but also improved absolute locations. For a few clusters with sufficient signal-to-noise detections, we show that the relocation method is successful in assigning clusters to specific faults at the reservoir level. Overall, about 90% of the events did not show clustering, despite choosing low correlation thresholds of 0.5 and 0.6. This suggests that different faults and/or fault segments with likely varying source mechanisms are active in reservoir sub-regions of a few square kilometres.

Keywords: Groningen, induced seismicity, multiplet analysis, relocation

Introduction

Gas extraction from the Groningen gas field in the northern part of the Netherlands has caused seismicity. It is generally accepted that depletion of a hydrocarbon reservoir causes changes in the state of stress, which, in turn, can reactivate existing faults. Prior to 2015, the Groningen gas field events were recorded by a monitoring network in the northeastern Netherlands consisting of 14 boreholes with geophones at four different depth levels and 18 accelerometers (Dost et al., 2012; Kraaijpoel & Dost, 2013). The earthquakes in Groningen are generally of a low magnitude ($M_L < 2.5$), with the strongest recorded event to date having local magnitude 3.6. With the pre-2015 network they are located with an accuracy of approximately 0.5–1 km in the horizontal plane (Kraaijpoel & Dost, 2013; Bourne et al., 2014). The location error stems from both picking errors and a simplified 1D model of the seismic velocity structure that is used for localisation. The accuracy of the locations is insufficient to associate them with active faults in the reservoir. This is where a multiplet analysis can be useful.

Seismic multiplets are sets of earthquakes that occurred on the same fault, or neighbouring parallel faults, and have a similar source mechanism (Dyer et al., 2010). Events of a multiplet therefore have seismograms with a high degree of similarity. A set of two similar events is called a doublet. The definition of a multiplet varies between studies. Arrowsmith & Eisner (2006) defined a multiplet as a set of events in which every event forms a doublet with at least one other event. According to this definition, not all events in a multiplet have to be mutually similar. An alternative definition is based on one seed event that forms doublets with all other events in a multiplet (Dyer et al., 2010). This study uses the seed event approach.

One benefit of a multiplet analysis is that P- and S-wave arrivals can be repicked more accurately, which translates to more accurate locations of earthquakes within a multiplet (Dyer et al., 2010). First, by cross-correlation, accurate relative arrival times for earthquakes in a multiplet are obtained. These are then used to determine inter-event distances with very small uncertainties. These uncertainties are largely governed by errors in the source-region velocity model, instead of errors in the
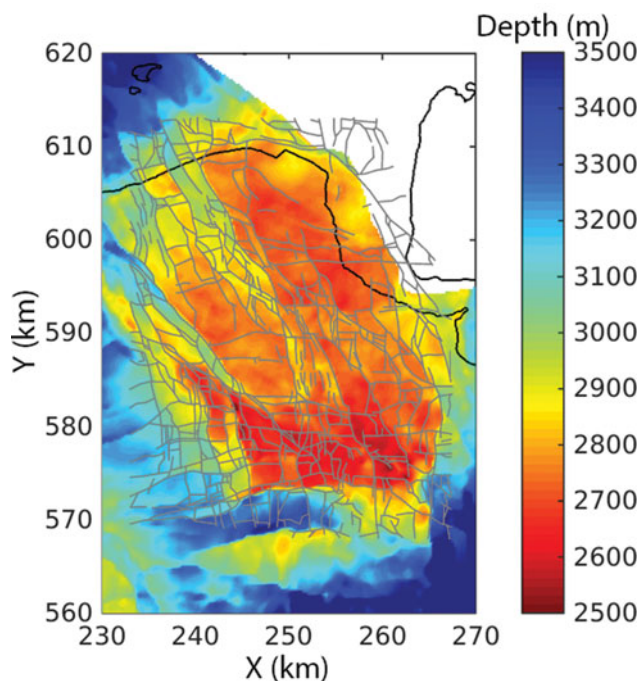
Fig. 1. *Depth of top of Rotliegend and base of Zechstein of the Groningen gas field (Van Dalfsen et al., 2006). Faults at reservoir level (grey lines) are from Bourne et al. (2014).*

Fig. 2. *Outline of the Groningen gas field (blue contour). Triangles indicate seismic stations of the pre-2015 network operated by the KNMI. Coordinates are given in kilometres within the Dutch National Triangulation Coordinates System (Rijksdriehoek). The background map is from www.openstreetmap. org.*

entire region between sources and receivers. This is because effects of source–receiver path velocity variations tend to cancel out in the relocation process. For relocation we apply a variant of the double-difference method (Waldhauser & Ellsworth, 2000). It is based on the travel time difference between the P- and S-wave arrivals of the relocated event compared to that of a master event. Uncertainties of the double-difference method are, amongst others, discussed in Moriya et al. (2003).

The first part of this paper gives a description and the results of the Groningen multiplet analysis. The second part comprises an outline of the relocation method and the results of applying this method to clusters emerging from the multiplet analysis. We show that besides improving the relative locations of events in a cluster, we can also improve their absolute locations by relocating with respect to accurately located master events that were detected with a denser seismic network that was deployed after 2014.

Study area

The reservoir rock of the Groningen gas field is the Upper Rotliegend (Permian) sandstone which varies in depth from 2600 to 3200 m (Bourne et al., 2014). Locations of faults at the reservoir level are known with relatively high accuracy through reflection seismic surveys carried out by the Nederlandse Aardolie Maatschappij (NAM). Since the main pressure perturbations are inflicted at the reservoir level, it is likely that the faults which
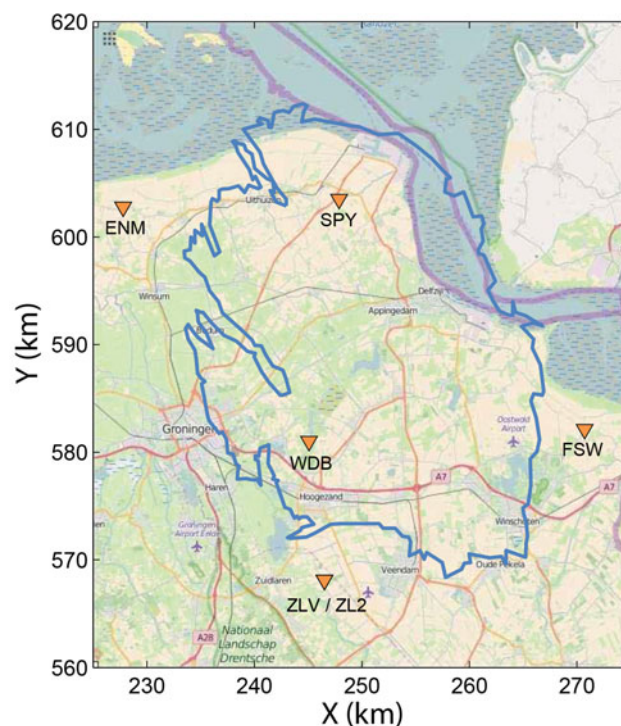
cut through the Rotliegend are the seismically active faults. Figure 1 shows the fault network and the depth of the top of Rotliegend sandstone. The outline of the gas field is approximately given by the 3000 m depth contour.

The reservoir rock is sealed by a layer of Zechstein evaporites of varying thickness. The large seismic velocity contrast between the slower reservoir rock and faster evaporites significantly affects the waveforms (Kraaijpoel & Dost, 2013). Together with the irregular shapes of the Zechstein salt domes, this complicates the location of earthquakes by picking and inverting P- and S-wave arrival times.

Multiplet analysis

Data

The multiplet analysis was carried out for data of station WDB (53.2082°N, 6.7355°E), one of five borehole stations in the direct vicinity of the Groningen gas field (Fig. 2) of the pre-2015 network deployed by the Royal Netherlands Meteorological Institute (KNMI). The station consists of four three-component borehole geophones at depths varying from 47 to 197 m below sea level. The deepest geophone of station WDB, WDB4, suffering the least from surface noise, was selected for the multiplet

analysis. We also analysed data of borehole stations ENM (53.4064°N, 6.4817°E) and SPY (53.4098°N, 6.7838°E) with a similar configuration to WDB, but, due to their closer proximity to the sea (Fig. 2), these stations had a lower signal-to-noise (S/N) ratio.

The database of induced earthquakes, provided by the KNMI (http://rdsa.knmi.nl/dataportal), includes 426 events between January 2010 and 16 June 2014, the period of this study. Of these events, 212 are recorded by WDB4 and are used for the multiplet analysis. The locations in the KNMI catalogue are obtained using the method described in Lienert et al. (1986), with a depth fixed to 3 km. From 2015 onwards, depths are estimated using P-wave arrival time differences over the densified network (Spetzler & Dost, 2016).

Method

The multiplet analysis used is described in detail in Dyer et al. (2010). A summary of the approach is presented here. In the multiplet analysis, every recorded event is cross-correlated with every other event. The outcome is a symmetric cross-correlation matrix in which every element represents the normalised cross-correlation coefficient between two events. Seed events are then identified which define the multiplets. A seed event correlates with all other events in a multiplet at a level higher than a specified threshold, henceforth referred to as the seed level. Sometimes events already assigned to a multiplet are themselves seed events for another multiplet. In that case, the events belonging to this particular seed event are included in the multiplet that has the seed event in it. The threshold for correlating events to seed events may vary with the initial conditions, such as time window, noise level and frequency band. For each set of initial conditions several seed levels are examined. The criterion for choosing a seed level is whether the multiplets translate to spatial clusters. If the seed level is too low, the events in a multiplet will not be tightly clustered and it will be hard to distinguish multiplet groups (De Meersman et al., 2009). On the other hand, if the seed level is too high, there will be no multiplets to analyse. Seed events with the largest number of events obtain the highest status. If two or more seed events have multiplets of the same size then the seed event with the highest S/N ratio obtains the highest status. Finally, the original cross-correlation matrix is sorted according to the status of the multiplets.

The multiplet analysis was applied to both the N-component (horizontal north–south) and Z-component (vertical) of the deepest geophone of station WDB. S-wave arrivals are primarily detected on horizontal components, whereas P-wave arrivals are more prominent on the verticals. The time window of the records is determined by the duration of the wave field. The majority of the direct P-waves arrive at station WDB within 0–3 s after the established origin time of the earthquake. The coda waves no longer exceed the noise level after 50–60 s. As

Schaff et al. (2004) pointed out, correlation coefficients decrease with increasing window length. However, only taking into account the part containing most energy (2–25 s) did not produce neat clusters since the coda waves need to be correlated as well when events have nearby origins. Therefore the selected time window is set to 2–45 s.

Prior to this research, the optimum frequency band for the multiplet analysis of this dataset was unknown. In near-surface boreholes, induced seismicity is detected between about 1 and 30 Hz. The greatest similarity is found for the lower frequencies (1–10 Hz), because of the extent of the study area. For small study areas, higher frequencies are used (e.g. Moriya et al., 2003). However, higher frequencies are more strongly affected by small-scale heterogeneities and scattering (Geller & Mueller, 1980), and a small disturbance in the ray path and/or source location can cause a large change in the high-frequency part of the signal. Three different frequency bands of 1–2 Hz, 1–4 Hz and 1–8 Hz were therefore investigated. The optimum frequency band turned out to be 1–8 Hz, because it takes into account the broadest frequency range of the signal, resulting in spatially tight clusters. Moreover, the largest number of clusters was obtained for this frequency band.

Results

We started the multiplet analysis with 212 events for WDB4. Events that did not have a cross-correlation coefficient >0.3 with any other event were removed to improve visualisation. After this operation, 113 and 145 events remained for the Z- and N-component, respectively. The seed level for the Z-component was set to 0.5, because it is the lowest seed level that produced the largest number of tightly spaced clusters. For the N-component the seed level was set at 0.6, because the overall cross-correlation coefficients are higher than for the Z-component. The cross-correlation matrices of the Z- and N-components of geophone WDB4, sorted after multiplet clustering, are shown in Figures 3 and 4, respectively. The red diagonal represents autocorrelation coefficients with a value of 1, and off-diagonal cross-correlation coefficients are generally small (<0.3). Multiplet analysis for the Z-component yielded one cluster of five events, two of four events and three of three events (Fig. 3). The N-component-based multiplet analysis yielded five clusters, where clusters 2–5 are identical to those obtained for the Z-component, and cluster 1 has four of the five events in common (Table 1). Cluster 6 is not identified on the N-component, possibly due to the higher seed level chosen. Figure 5 shows the traces of cluster 2 for the N-component to illustrate the waveform similarity within a multiplet. Waveform similarity in the part containing most of the energy (4–14 s) is not immediately apparent (Fig. 5b), validating our choice to cross-correlate the coda waves as well.

Thus, for the Z-component, out of all 212 analysed events recorded by geophone WDB4, 22 (10.4%) are clustered in six
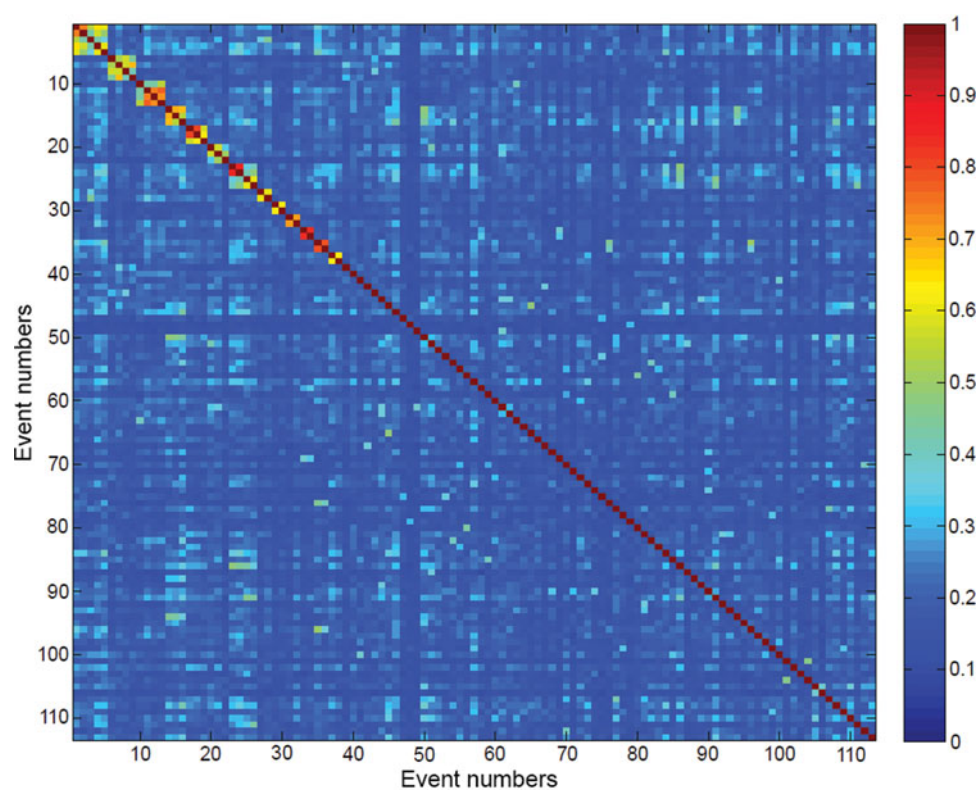
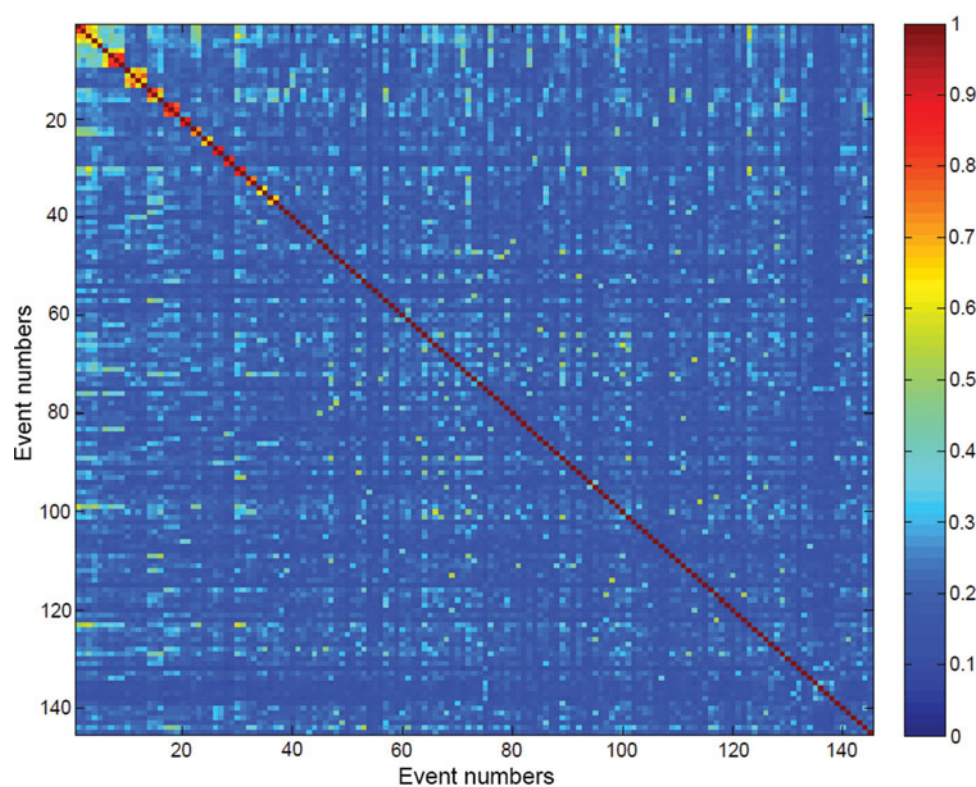*Fig. 3. Sorted cross-correlation matrix for 1–8 Hz: Z-component; geophone WDB4.*

Fig. 4. Sorted cross-correlation matrix for 1–8 Hz: N-component; geophone WDB4.
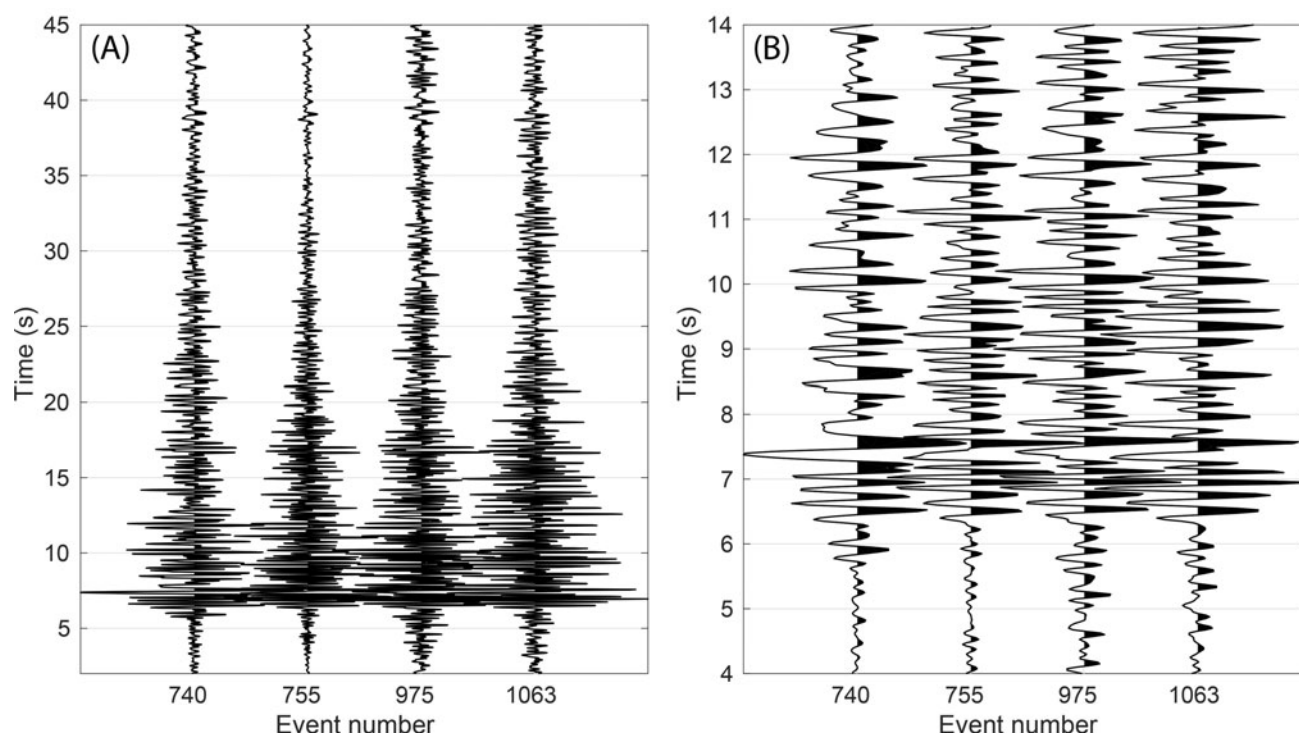
Fig. 5. N-component traces of cluster 2 for time windows of (A) 2–45 s and (B) 4–14 s.

Table 1. Event numbers in multiplets from the 1–8 Hz analysis for geophone WDB4. Event information of the event numbers is given in the Appendix. Seed events are bold.

Cluster	Z-component	N-component
1	**709**, 711, 895, 903, 1013	709, 711, **903**, 918, **1013**
2	740, **755**, 975, 1063	740, 755, **975**, 1063
3	746, 1043, 1048, **1049**	746, 1043, 1048, **1049**
4	678, 871, **1040**	**678**, 871, 1040
5	965, **987**, 1008	965, **987**, 1008
6	890, **954**, 972	

Fig. 6. Locations of Z-component-derived clusters, and non-clustered events, plotted in a map with the Groningen coastline. Sizes of the circles representing clustered events denote magnitudes. For reference, the locations of four borehole stations (orange triangles) are shown. Faults have been taken from Bourne et al. (2014).

different multiplets. Figure 6 shows the locations of the clustered and non-clustered events, together with the coastline of Groningen, the main faults at the top of the Rotliegend sandstones and locations of seismic stations. Spatial clusters can be linked to faults in this way. Note that there are many tightly located events in Figure 6 that are nevertheless not part of a cluster. This suggests that many faults or fault segments are active, with possibly varying source mechanisms, resulting in greatly varying waveforms. Despite the similarity of their seismograms and the high cross-correlation values, the events in cluster 1 appear to be relatively widely distributed (within 13 km). Furthermore, for some of the events of this cluster, it is not possible to confidently pick the P- and S-wave arrival times at other
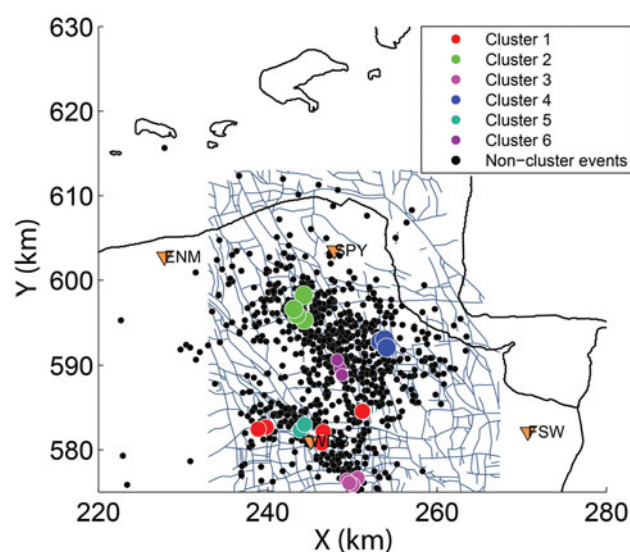
stations than that used for the multiplet analysis (WDB4). The events of this cluster are therefore not relocated.

We further quantified the similarity of the events within the different clusters by averaging the unique cross-correlation
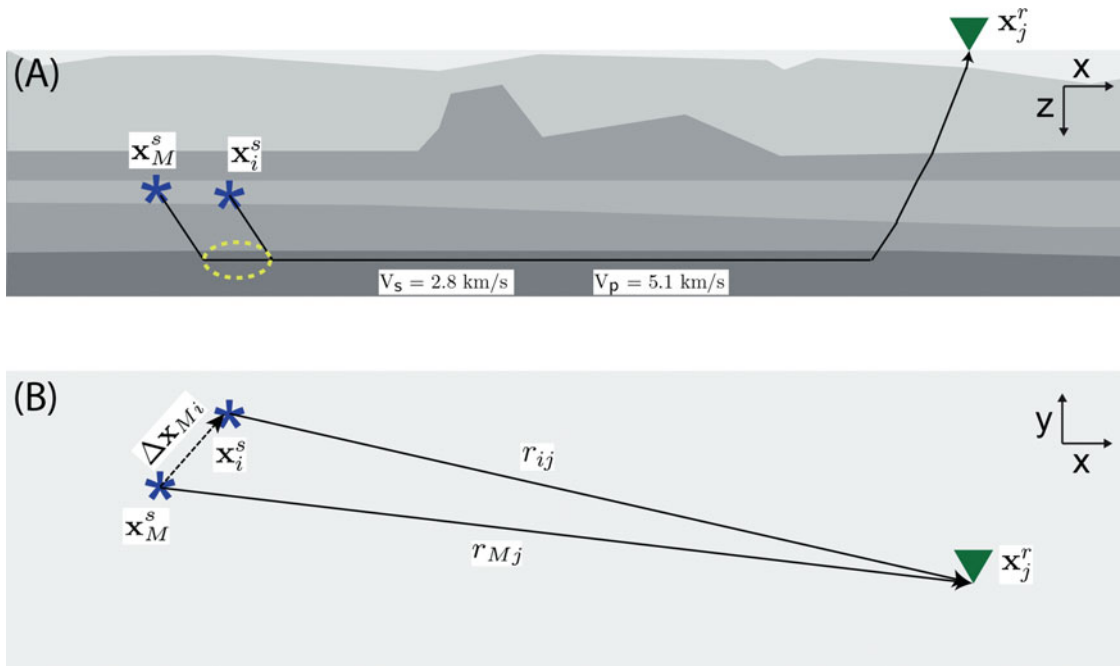
Fig. 7. *Configuration for location of a source (blue star) in a cluster x_i^s with respect to a master source x_M^s and one of the receivers (green triangle): (A) section view, (B) map view. In (A) the blue stars are at the hypocentres, in (B) they correspond to the epicentres.*

coefficients of each cluster. For a cluster with n events, there exist $n(n-1)/2$ unique cross-correlations. The highest similarity is obtained for cluster 5 (0.72), which is close to station WDB that is used for the multiplet analysis (Fig. 6). Cluster 1, the cluster with the least dense confinement of events (Fig. 6), has the smallest similarity, together with cluster 6 (both have a similarity of 0.48). Cluster 2 (0.54), cluster 3 (0.62) and cluster 4 (0.65) show values in between.

Relocation

Method

The relocation method used in this study is an adapted version of the double-difference method of Waldhauser & Ellsworth (2000). It uses the travel time difference between the P- and S-waves to locate events relative to a master event. Schematically, the ray paths of P- or S-waves for events at epicentral distances larger than roughly 5 km are shown in Figure 7. The figure presents a simplification of the velocity structure of the Groningen area based on Duin et al. (2006). The light-grey layer in Figure 7a with the two stars indicates the reservoir layer from the Rotliegend Group. The irregular layer on top of the reservoir is salt from the Zechstein Group. The Limburg Group directly below the reservoir has slightly higher velocities than the reservoir, whereas the underlying Lower Carboniferous Group (Dinantian) of compact limestones (darkest grey) is thought to have significantly higher seismic velocities. Vandenberghe et al. (1986) reported P-wave velocities of 4.9–5.6 km s^{-1} for karsti-

fied and 6.4 km s^{-1} for tight Dinantian limestones. When the distance between source and receiver is much larger than the source depth of 3 km, the first arriving P- and S-waves propagate as critically refracted head waves along this Lower Carboniferous Group or deeper strata. The P- and S-wave velocities of this refracting layer, 5.1 and 2.8 km s^{-1}, respectively, are determined by linear regression of P- and S-wave onsets of an $M = 2.4$ 2016 event that was recorded by many stations of the densified seismic network (Fig. 8). Seismic velocities are heterogeneous over the Groningen area (Duin et al., 2006), so Figure 7 is schematic.

All earthquakes are assumed to originate in the reservoir at 3 km depth. The blue stars in Figure 7 indicate the locations of an accurately determined master event, x_M^s, and the location of event i of the cluster x_i^s. Both events are detected at a receiver at location x_j^r, where j is the receiver index. Clustered events are located close to the master event, therefore the waves follow a similar path to the receiver and encounter a similar velocity structure. We want to relocate source x_i^s relative to x_M^s in the horizontal plane. The difference in horizontal distance between the source at x_i^s and master event at x_M^s as seen by the receiver at x_j^r reads

$$\Delta r_{ij} = r_{ij} - r_{Mj},$$

where r_{ij} is the horizontal distance from x_i^s to x_j^r, which is computed by taking the Euclidean length $\| x_i^s - x_j^r \|$ for the horizontal coordinates. This distance can be expressed in terms of difference in travel time for the two sources in the first-arriving P-wave multiplied by the P-wave velocity of the refracting layer:

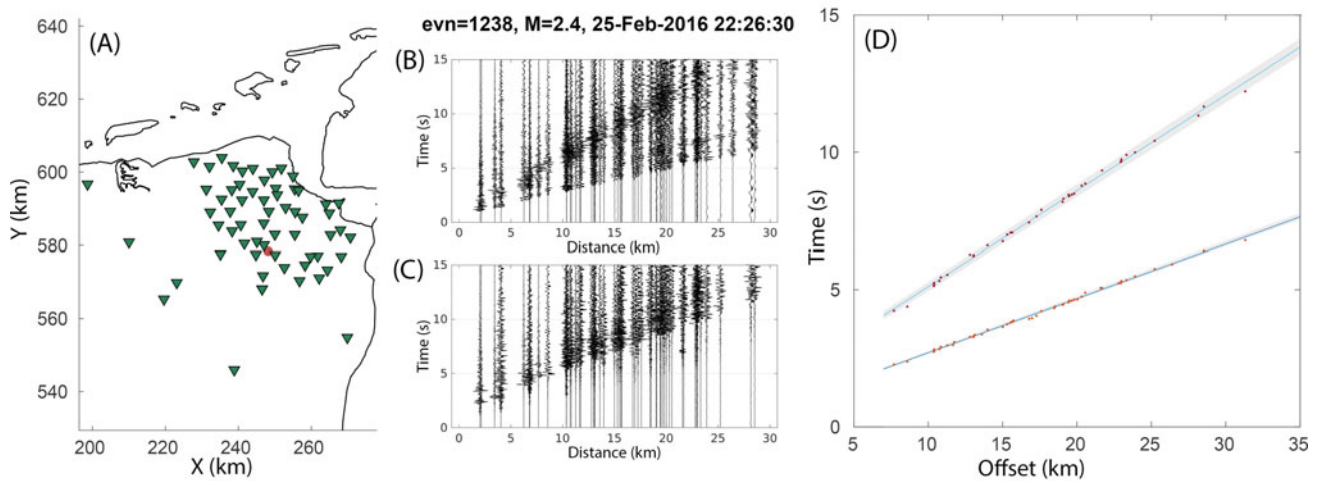$$\Delta r_{ij} = \left(t_{ij}^P - t_{Mj}^P\right) v^P, \qquad (1)$$

Fig. 8. *Estimation of average P- and S-wave head-wave velocities over the Groningen area. (a) the configuration of the 2016 M = 2.4 Froombosch event (red circle) and the stations with borehole geophones recording the event (green triangles). (b) and (c) are the vertical and transverse component event gathers, respectively, frequency bandpass filtered between 1 and 25 Hz. (d) shows picks of first P-wave and S-wave onsets between 7 and 35 km offset, which are fitted with a straight line. Behind the picks, the 95 confidence areas are shown (grey zones). For the (refracted) P-wave the mean velocity is 5.06 km s^{-1} and the 95 confidence zone is bounded by a lower and upper velocity of 5.00 and 5.13 km s^{-1}, respectively. For the (refracted) S-wave the mean velocity is 2.83 km s^{-1} and the 95 confidence zone is bounded by a lower and upper velocity of 2.79 and 2.87 km s^{-1}, respectively.*

and analogously for the first S-wave:

$$\Delta r_{ij} = \left(t_{ij}^S - t_{Mj}^S \right) v^S. \tag{2}$$

The P- and S-wave velocities need only be known for the refracting layer in the small area $\Delta \mathbf{x}_{Mi}$ indicated by the yellow ellipse in Figure 7a, and the velocities are assumed to be constant in this area. The travel time, t, is given by the difference between the arrival time, T, and the earthquake origin time, EOT, e.g. $t_{ij}^P = T_{ij}^P - EOT$. When eqn 1 is multiplied by v^S on both sides, and eqn 2 by v^P, and when the new eqn 1 is then subtracted from the new eqn 2, the dependence on the origin times vanishes:

$$\Delta r_{ij} = \frac{T_{ij}^S - T_{ij}^P - \left(T_{Mj}^S - T_{Mj}^P \right)}{\frac{1}{v^S} - \frac{1}{v^P}} = \frac{\Delta T_{ij}^{PS} - \Delta T_{Mj}^{PS}}{\frac{1}{v^S} - \frac{1}{v^P}}. \tag{3}$$

Expressing source location $\mathbf{x}_i^s$ as function of the master event location, $\mathbf{x}_i^s = \mathbf{x}_M^s + \Delta \mathbf{x}_{Mi} = (x_i^s + \Delta x_{Mi}, y_i^s + \Delta y_{Mi})$, eqn 3 can be rewritten as

$$\left(\| \mathbf{x}_j^r - \mathbf{x}_M^s - \Delta \mathbf{x}_{Mi} \| - \| \mathbf{x}_j^r - \mathbf{x}_M^s \| \right) \left(\frac{1}{v^S} - \frac{1}{v^P} \right)$$
$$= \Delta T_{ij}^{PS} - \Delta T_{Mj}^{PS}, \tag{4}$$

where the right-hand side is the difference in arrival time between the P- and S-wave of source i recorded by receiver j minus the same difference but then for the master event. To compute the relative horizontal location of source $\mathbf{x}_i^s$ with respect to $\mathbf{x}_M^s$, eqn 4 must be solved for at least two receivers to find the two unknowns Δx_{Mi} and Δy_{Mi}.

Using this approach with relative locations with respect to a nearby master event, location errors caused by a simplification of the 3D velocity structure (Van Dalfsen et al., 2006) by

a 1D model (Kraaijpoel & Dost, 2013) are largely circumvented. Furthermore, the dependence on origin time is eliminated, and erroneous arrival time picks for events with a low S/N ratio are reduced by using cross-correlations. For each cluster of events, we first select a time window around the first P-wave arrival. Then we estimate the difference in P-wave arrival time between the master source and cluster sources by cross-correlating the selected time window, yielding

$$\Delta T_{ij}^P = EOT_i + t_{ij}^P - \left(EOT_M + t_{Mj}^P \right), \tag{5}$$

where EOT_i is the initial estimate of EOT for source i. For sources for which the P-wave arrivals are not sufficiently similar, we use kurtosis with its positive derivative (Langet et al., 2014) to obtain time series that are peaked at the P-wave onsets. These peaks are highly similar for the different events and thus lead to accurate time differences after cross-correlation. We also estimate relative S-wave arrival times with cross-correlation (or the cross-correlation of the positive derivative of kurtosis, if required), yielding an equivalent expression for S-waves:

$$\Delta T_{ij}^S = EOT_i + t_{ij}^S - \left(EOT_M + t_{Mj}^S \right). \tag{6}$$

The EOTs cancel when eqn 5 is subtracted from eqn 6, and the S–P arrival time difference is obtained (right-hand side of eqn 4).

To find the relative locations of sources in the cluster with respect to the master event we do a grid search for eqn 4. We define a grid around the master event for which we forward-model the left-hand side of eqn 4. So for each potential relative location $\Delta \mathbf{x}_{Mi}$, the misfit R is computed between the estimated
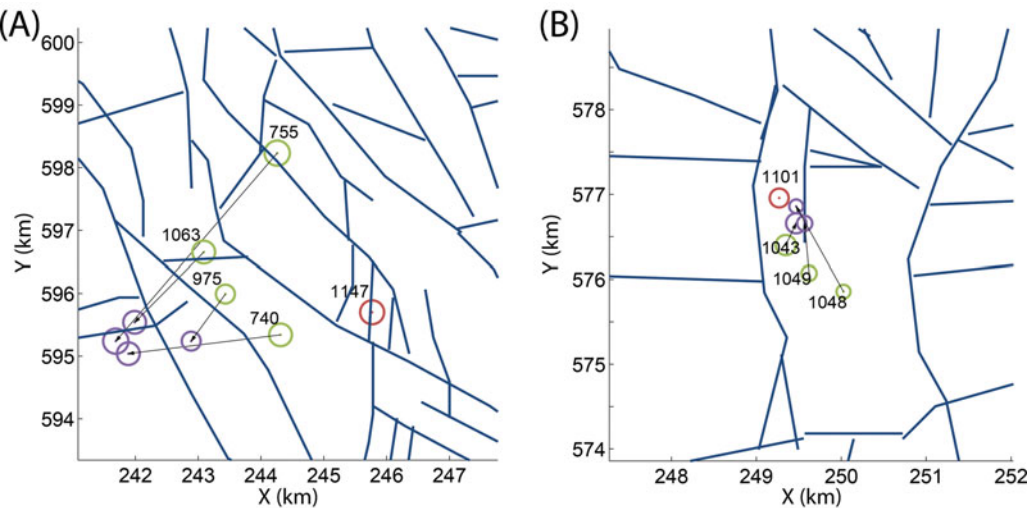
Fig. 9. Relocations for (A) cluster 2 and (B) cluster 3. Green circles represent locations estimated by the KNMI, purple circles the new locations and red circles denote locations of the master events. The size of the circles scales with the magnitude. The black arrows denote relocation vectors, and the blue lines denote identified faults taken from Bourne et al. (2014). Additional information on the numbered events can be found in the Appendix.

and recorded S–P arrival time difference:

$$R(\Delta \mathbf{x}_{Mi}) = \sqrt{\sum_{j=1}^{n} \left\{ \left(\Delta T_{ij}^{PS} - T_{Mj}^{PS} \right)^{obs} - \left(\Delta T_{ij}^{PS}(\Delta \mathbf{x}_{Mi}) - \Delta T_{Mj}^{PS} \right)^{mod} \right\}^2 / n},$$
(7)

for n receivers where both earthquake i and master event M are observed with sufficient S/N ratio.

Since a homogeneous velocity field is used, no local minima exist in this misfit function. The trial location with the smallest misfit is assigned to be the relocated source location.

▬ Results

For relocation we use master events of which the locations are known much more accurately than those of the pre-2015 monitoring network. These master events were recorded by a network that by the end of 2015 consisted of more than 60 boreholes in and around the Groningen gas field (https://www.knmi.nl/nederland-nu/seismologie/stations). The cluster events are recorded by at most five borehole stations in the region (see Fig. 2). The S–P arrival time differences are determined from the bottom geophones in the boreholes.

For cluster 2 we used event 1147 (see Appendix) to serve as master event. Boreholes SPY, ENM and WDB were used for the intra-cluster relocations. The result of the relocation is shown in Figure 9a. The KNMI estimated that one event (755) in cluster 2 was located significantly farther away from the other events. Our results show, however, that all relocated events plot on a small area close to one or two faults. The misfit (eqn 7) prior to and after relocation is given in Table 2. Event 755 shows the largest misfit reduction, of almost a factor 8.

Table 2. For each relocated event the root-mean-square error (RMSE) prior to and after relocation (eqn 7) is tabulated.

Event No	Cluster No	RMSE pre-relocation	RMSE post-relocation
740	2	0.317	0.153
755	2	0.287	0.037
975	2	0.156	0.134
1063	2	0.133	0.077
1043	3	0.034	0.017
1048	3	0.122	0.018
1049	3	0.065	0.008
678	4	0.086	0.062
871	4	0.113	0.048
1040	4	0.175	0.116
890	6	0.142	0.097
954	6	0.035	0.028

Event 1101 was selected as a master event for cluster 3. Only three of the four events of this cluster could be relocated, because the S/N ratio of one event was too low to estimate the S–P arrival time difference. Originally, the remaining events were already located relatively close together, but the relocation method with data from boreholes WDB, ZLV and FSW has brought the events even closer together and closer to the master event (Fig. 9b).

Data from boreholes WDB, ZLV and FSW were used to relocate the events of cluster 4 with master event 1142. The new locations are more tightly clustered at two nearby faults (Fig. 10a).
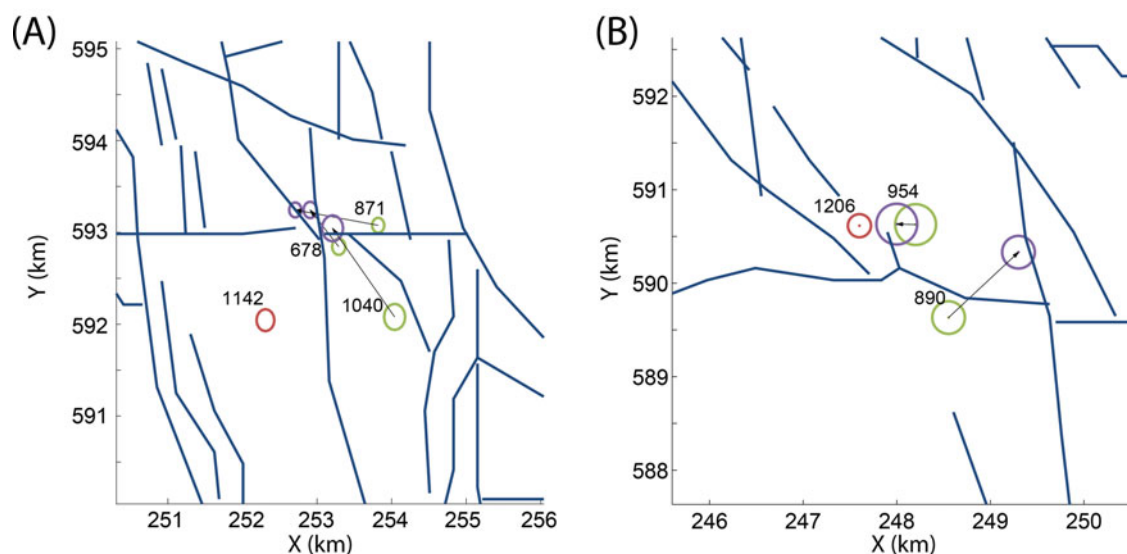
Fig. 10. Relocations for (A) cluster 4 and (B) cluster 6. Green circles represent locations estimated by the KNMI, purple circles the new locations and red circles denote locations of the master events. The black arrows denote relocation vectors and the blue lines denote identified faults taken from Bourne et al. (2014). Additional information on the numbered events can be found in the Appendix.

It was not possible to improve the relative locations of the events of cluster 5, because the S/N ratio at SPY and FSW was too low to determine the differential arrival times. The three remaining stations, ENM, WDB and ZLV, are nearly aligned in a NNW–SSE direction. Their data do not constrain the locations in E–W direction very well. The event locations could be shifted in this direction and still give a good fit to the data.

Cluster 6 comprises two events of relatively high magnitude ($M = 2.4$ and 3.0 for events 890 and 954, respectively) and one event of low magnitude ($M = 0.9$, event 972). The S/N ratio of this last event is too low to identify its P- and S-arrivals at borehole locations WDB, FSW and ENM. The remaining two events are relocated with respect to event number 1206. The events are relocated to two parallel faults as shown in Figure 10b.

The events in clusters 2, 3 and 6 are relocated with respect to nearby and high S/N-ratio events from the densified array. Only for cluster 4 could a master event be found that is also formally a seed event and hence part of the cluster. Since the multiplet analysis did not include post-2014 events, this master event was not part of cluster 4 in the first section of this paper.

Discussion and conclusion

Up to now, most multiplet analyses have been applied to identify clusters of tectonic events (e.g. Myhill et al., 2011; Lambotte et al., 2014) or induced events recorded at deep boreholes (e.g. Michelet & Toksöz, 2007; De Meersman et al., 2009). The multiplet analysis presented here shows that it is also feasible to identify clusters of induced events in the Groningen gas field, an area with a diameter of roughly 30 km. Approximately 10% of

the events recorded at station WDB between 2010 and mid-2014 could be assigned to one of six clusters of events when a relatively broad frequency range, 1–8 Hz, of the data is used. With a seed level of 0.5–0.6 the correlation between events is smaller than for induced events recorded at deep boreholes near the seismicity (e.g. Arrowsmith & Eisner, 2006). This indicates that the events are probably not always located on the same faults or that they have slightly different source mechanisms. In our relocation, only the events in cluster 3 are confined within a few hundred metres on the same fault (Fig. 9b).

Approximately 90% of the analysed events are not part of a cluster, despite the low threshold chosen (0.5) and despite small distances between events. This suggests that, even at close range, many different faults and/or fault segments are active in the Groningen field, with likely also a spread in focal mechanisms.

The cluster events recorded by the pre-2015 seismic network are relocated using master events from 2015 to March 2016 that are localised by a much denser seismic network in the Groningen area. The master events are located with only upgoing P-waves, for which velocities are well known from 3D seismics and numerous borehole logs. The relocation method is based on S–P arrival time differences of cluster events with respect to master events. Using cross-correlation, the S–P arrival time differences are determined at a minimum of two stations to allow relocation in the horizontal plane. With the ray paths of the P- and S-waves of the cluster events and the master event being very similar, the relocated events are found to be more tightly clustered than the original locations and are located close to existing faults (Figs 9, 10). The relocation yielded reductions in root-mean-square error ranging from a small fraction to a factor of 10.

The largest error in the location is thought to be caused by errors in the velocity model. The pre-2015 KNMI locations are susceptible to velocity uncertainty over large distances between source and receivers. The relocation is sensitive primarily to velocity uncertainty in a region around a source cluster. Consequently, less location error is accrued due to velocity uncertainty for the relocation than for the original location, which is sensitive to velocities on the entire source–receiver path. However, the relocation will only yield better absolute locations when there is little location uncertainty for the master event.

The similarity of the clustered events was found to be relatively low for the Groningen setting. The average similarity within the clusters was 0.58. As a consequence, we could not fully rely on cross-correlations of the original waveforms to obtain travel time differences. We therefore applied kurtosis to some of the waveforms to obtain more accurate measurements of time differences. Moreover, due to insufficient S/N ratio of the most energetic P-wave and/or S-wave package on some of the stations, we could not apply relocation to events in clusters 1 and 5 and a few events in other clusters.

We restricted ourselves to relocating the identified clustered events. The relocation approach taken, however, could be expanded to all pre-2015 events that are in the vicinity of master events located with the densified network, as our results suggest that this relocation method improves the event locations.

Acknowledgements

References

Arrowsmith, *S.J.* & *Eisner*, *L.*, 2006. A technique for identifying microseismic multiplets and application to the Valhall field, North Sea. Geophysics 71(2): V31–V40.

Bourne, *S.J.*, *Oates*, *S.J.*, *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.

De Meersman, *K.*, *Kendall*, *J.-M.* & *van der Baan*, *M.*, 2009. The 1998 Valhall microseismic data set: an integrated study of relocated sources, seismic multiplets, and S-wave splitting. Geophysics 74(5): B183–B195.

Dost, *B.*, *Goutbeek*, *F.*, *van Eck*, *T.* & *Kraaijpoel*, *D.*, 2012. Monitoring induced seismicity in the North of the Netherlands: status report 2010.

Royal Netherlands Meteorological Institute, Scientific report WR 2012-03: 1–39.

Duin, *E.J.T.*, *Doornenbal*, *J.C.*, *Rijkers*, *R.H.B.*, *Verbeek*, *J.W.* & *Wong*, *Th.E.*, 2006. Subsurface structure of the Netherlands: results of recent onshore and offshore mapping. Netherlands Journal of Geosciences / Geologie en Mijnbouw 85(4): 245–276.

Dyer, *B.C.*, *Schanz*, *U.*, *Spillmann*, *T.*, *Ladner*, *F.* & *Häring*, *M.O.*, 2010. Application of microseismic multiplet analysis to the Basel geothermal reservoir stimulation events. Geophysical Prospecting 58: 791–807.

Geller, *R.J.* & *Mueller*, *C.S.*, 1980. Four similar earthquakes in central California. Geophysical Research Letters 7: 821–824.

Kraaijpoel, *D.* & *Dost*, *B.*, 2013. Implications of salt-related propagation and mode conversion effects on the analysis of induced seismicity. Jounal of Seismology 17(1): 95–107.

Lambotte, *S.*, *Lyon-Caen*, *H.*, *Bernard*, *P.*, *Deschamps*, *A.*, *Nercessian*, *A.*, *Pacciani*, *F.*, *Bourouis*, *S.*, *Drilleau*, *M.* & *Adamova*, *P.*, 2014. Reassesment of the rifting process in the Western Corinth Rift from relocated seismicity, Geophysical Journal International 197: 1822–1844.

Langet, *M.*, *Maggi*, *A.*, *Michelini*, *A.* & *Brenguier*, *F.*, 2014. Continuous kurtosis-based migration for seismic event detection and location, with application to Piton de la Fournaise Volcano, La Réunion. Bulletin of the Seismological Society of America 104: 229–246.

Lienert, *B.*, *Berg*, *E.* & *Frazer*, *L.*, 1986. HYPOCENTER: An earthquake location method using centered, scaled and adaptively damped least squares. Bulletin of the Seismological Society of America 76(3): 771–783.

Michelet, *S.* & *Toksöz*, *M.N.*, 2007. Fracture mapping in the Soultz-sous-Forêts geothermal field using microearthquake locations. Journal of Geophysical Research: Solid Earth 112(B7).

Moriya, *H.*, *Niitsuma*, *H.* & *Baria*, *R.*, 2003. Multiplet-clustering analysis reveals structural details within the cloud at the Soultz geothermal field, France. Bulletin of the Seismological Society of America 93(4): 1606–1620.

Myhill, *R.*, *McKenzie*, *D.* & *Priestley*, *K.*, 2011. The distribution of earthquake multiplets beneath the southwest Pacific. Earth and Planetary Science Letters 301: 87–97.

Schaff, *D.P.*, *Bokelmann*, *G.H.R.*, *Ellsworth*, *W.L.*, *Zanzerkia*, *E.*, *Waldhauser*, *F.* & *Beroza*, *G.C.*, 2004.Optimizing correlation techniques for improved earthquake location. Bulletin of the Seismological Society of America, 94(2): 705–721.

Spetzler, *J.* & *Dost*, *B.*, 2016. Hypocenter estimation of induced earthquakes in Groningen. Geophysical Journal International, 209(1): 453–465.

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

Vandenberghe, *N.*, *Poggiagliolmi*, *E.* & *Watts*, *G.*, 1986. Offset-dependent seismic amplitudes from karst limestone in northern Belgium. First Break 4(5): 9–27.

Waldhauser, *F.* & *Ellsworth*, *W.L.*, 2000. A double-difference earthquake location algorithm: method and application to the northern Hayward fault, California. Bulletin of the Seismological Society of America 90(6): 1353–1368.

Appendix

Table A1. *Event information of cluster and master events (source: http://rdsa.knmi.nl/dataportal).*

#	Date (dd/mm/yyyy)	Location	Latitude (°N)	Longitude (°E)	Depth (km)	Magnitude
678	07/12/2010	Appingedam	53.313	6.862	3	1.2
709	10/05/2011	Klein Harkstede	53.224	6.657	3	1.2
711	12/05/2011	Hellum	53.239	6.828	3	1.1
740	29/07/2011	Loppersum	53.337	6.728	3	1.8
746	23/08/2011	Sappemeer	53.168	6.817	3	1.6
755	25/09/2011	Garsthuizen	53.363	6.728	3	2.0
871	22/11/2012	Appingedam	53.315	6.870	3	1.1
890	19/01/2013	Overschild	53.285	6.790	3	2.4
895	05/02/2013	Garmerswolde	53.222	6.643	3	1.5
903	11/02/2013	Woudbloem	53.218	6.758	3	1.8
918	29/03/2013	Lageland	53.233	6.717	3	0.9
954	02/07/2013	Garrelsweer	53.294	6.785	3	3.0
965	31/07/2013	Harkstede	53.221	6.718	3	1.0
972	21/08/2013	Overschild	53.278	6.793	3	0.9
975	01/09/2013	Westeremden	53.343	6.715	3	1.5
987	01/10/2013	Harkstede	53.220	6.717	3	1.1
1008	06/12/2013	Lageland	53.226	6.725	3	1.3
1013	02/01/2014	Woudbloem	53.206	6.755	3	1.4
1040	15/03/2014	Appingedam	53.306	6.873	3	1.9
1043	23/03/2014	Sappemeer	53.166	6.798	3	1.6
1048	02/04/2014	Sappemeer	53.161	6.808	3	1.1
1049	02/04/2014	Sappemeer	53.163	6.802	3	1.2
1063	16/06/2014	Westeremden	53.349	6.710	3	1.8
1101	11/01/2015	Sappemeer	53.171	6.797	3	1.5
1142	16/05/2015	Appingedam	53.306	6.847	3	1.6
1147	06/06/2015	Loppersum	53.340	6.750	3	1.9
1206	30/01/2016	Garrelsweer	53.294	6.776	3	1.7