Evaluation of a Seismic Event, 12 May 2010, in North Korea

by Won-Young Kim, Paul G. Richards, David P. Schaff, and Karl Koch

Corresponding author: Paul G. Richards
Lamont-Doherty Earth Observatory of Columbia University
61 Route 9W
Palisades, NY 10964
e-mail: richards@LDEO.columbia.edu

This paper has an electronic supplement in which are provided:

- our measurements of the log *P/S* spectral ratios for the two training sets;
- a section describing further details of our three-component linear discrimination function analysis, in particular the effectiveness of different choices of the frequency components used to apply *P/S* spectral ratios; and
- a tutorial section on the underlying ideas behind Mahalanobis methods for event classification, with particular reference to Figures 10, 13, and 15 in the main text.

Abstract We assess seismological evidence bearing on claims that North Korea conducted a small nuclear test on 12 May 2010 in the vicinity of known underground nuclear tests in 2006, 2009, 2013, and 2016. First, we use *Lg*-wave cross-correlation and more traditional methods to locate the 2010 event between about 4 and 10 km southwest of the 2009 test. Second we compare the relative sizes of regional *P*-waves and *S*-waves, using stations within 400 km of the known North Korean nuclear tests, to assess the nature of the event.

We measured *P/S* ratios at different frequencies, at first using data from the open station MDJ in northeast China, for training sets of earthquakes and of explosions. We developed a linear discriminant function (LDF), that, in application to *P/S* measured at MDJ, is most effective in separating the earthquake and explosion populations. MDJ lacks usable data for the event of interest, but we have obtained regional data from stations of the nearby Dongbei Broadband Seismographic Network (DBSN), for the event of 12 May 2010 and for nearby underground nuclear tests conducted in 2006 and 2009. When our LDF is applied to DBSN data, and to data from stations SMT and NE3C in China, the LDF values measured from *P/S* ratios from known explosions are explosion-like; but for the event of 12 May 2010 the LDF values are earthquake-like for frequencies between 6 and 12 Hz.

Our method for characterizing earthquakes and explosions on the basis of their regional signals can be widely applied. Measurements of *P/S* based upon three-component waveform data provide better discrimination power than do those based upon vertical-component data alone.

Introduction

Several papers published during 2012–2015 have claimed there is evidence for a low-yield nuclear explosion conducted in Spring 2010 in North Korea close to the site of well-known nuclear explosions in 2006, 2009, 2013, and 2016. The original claim (De Geer, 2012) was based on detection of radioxenons and radioxenon progeny nuclides at four sampling stations located 10 to 1260 km outside the borders of North Korea. Five days in April and May of 2010 were at first proposed as the candidate dates on which a nuclear test may have been carried out. Subsequent papers based on the radionuclide observations, and discussion of candidate locations from which atmospheric transport of radionuclides would fit those observations, have focused on 11 May 2010 as the likely date of a claimed low-yield nuclear test in North Korea (De Geer, 2013; Wright, 2013).

Schaff *et al.* (2012) attempted to detect seismic signals from small explosions in North Korea on the specific days in 2010 proposed by De Geer. The seismic data recorded on those days by the nearest open station of the Global Seismographic Network, namely MDJ in northeastern China, were searched, applying three-component cross-correlation methods and using high-quality signals as templates (derived from known nuclear explosions in North Korea recorded at this same station). Schaff *et al*. (2012) assessed the capability of this method of detection, and of simpler methods, all of which failed to find seismic signals that would be expected if scenarios proposed by De Geer were valid, and concluded, first, that no well-coupled underground explosion above about a ton occurred near the North Korea test site on these five days (14, 15, 16 April; 10, 11 May), and second, that any explosion would have to be very small (local magnitude less than about 2) to escape detection. An important issue emerging from this analysis, was the practical difference between two types of detection threshold, namely between

(a) estimating the present or future detection capability of a monitoring network, absent any templates similar to the signal one is attempting to detect, and (b) the detection capability for a specific time period in the past when it is known what data are available, and templates do exist. Detection capability is much better in the latter situation.

In view of these previous studies and several additional papers interpreting the radionuclide data of 2010 in terms of a low-yield nuclear test, it was of considerable interest that Zhang and Wen (2015a) reported data from a small seismic event in North Korea, occurring on 12 May 2010 at about 00:08:45 UTC, which they claimed had occurred less than 1 km from the known tests of 2009 and 2013, and which they characterized unambiguously as a small nuclear explosion. The origin time is about nine minutes later than the time window for May 2010 searched by Schaff *et al.* (2012). Zhang and Wen used data from several seismographic stations in Jilin Province, China, in the distance range approximately 80 to 200 km from the North Korea test site. At present the data these authors analyzed are not openly available. To detect the event they used a method called "match and locate", described further in Zhang and Wen (2015b), in which correlograms for different channels of seismic data are averaged using an appropriate template for each channel. The method allows templates to be drawn from more than one event, at different locations in the general source region of interest. The relative location of template events must be accurately known, and the match and locate method stacks correlograms for each point in a predetermined grid of candidate locations, using the appropriate delay for each point in the grid in a search for the grid point that maximizes the stack. In practice the method searches a 2D horizontal grid at a fixed depth (the depth of the template events, presumed to be similar to the depth of the event being sought), with templates that in the present case are provided by known underground nuclear explosions (UNEs) in North Korea. Zhang and Wen (2015a) found

a maximum mean cross-correlation (CC) value for a candidate location about 900 meters south and 200 meters west of the known UNE of 2009. They showed individual seismograms of the event as recorded on stations in Jilin Province, China, and assigned it a magnitude of 1.44 (based on *Lg*). On the basis of measured values of the spectral ratio between *P*-waves and *S*-waves they claimed it to be a very small nuclear test explosion having a yield estimated at 2.9 t (TNT equivalent), with very low uncertainty (only ± 0.8 t).

It is remarkable that the data streams and analysis methods used in modern seismology can now provide detection capability at magnitudes as low as the small event under discussion here. Note that the first UNE conducted in North Korea (on 9 October 2006) had a seismic magnitude around 4 and was widely reported to have had a low yield—roughly 1 kt. The seismic event of 12 May 2010 was about two and a half units lower in magnitude, and thus had signals about 300 times smaller than those of the small UNE of 2006. However, while Zhang and Wen are to be congratulated on finding this small seismic event, we present and analyze additional data indicating that it has characteristics different from those of a small contained single-fired explosion. We also discuss the likelihood of a somewhat different location, relative to the known North Korea UNEs. We comment on the procedure used by Zhang and Wen (2015a) for estimating yield of small underground explosions on the basis of their seismic signal strength, and argue that there is substantial uncertainty in such estimates. Our principal source of additional data has been the Dongbei Broadband Seismographic Network (DBSN), installed in mid-2004 by Dr. Kin-Yip Chun and colleagues at several sites in northeast China in Liaoning and Jilin Provinces, just north of the border with North Korea.

Signals from the seismic event of 12 May 2010 reported by Zhang and Wen (2015a) can be seen in the openly-available data acquired by a PASSCAL instrument deployment, involving

114 a temporary network of stations in northeast China known as the NECESS Array (see, e.g., Tao
115 *et al*., 2014). Ford and Walter (2015) have reported this confirmation, and they go on to make
116 the case that if the location and magnitude of the 2010 event were as characterized by Zhang and
117 Wen, then its detection at stations of the International Monitoring System (IMS) would be
118 expected, using cross-correlation methods. It should also have been detected at the open station
119 MDJ. The fact that it was not detected prior to work done by Zhang and Wen, and could not
120 subsequently be confirmed with IMS or MDJ data, is a matter we discuss and explain in terms of
121 a location for the event that is somewhat further away from prior UNEs than claimed by Zhang
122 and Wen (2015a), or if it had taken place at a significantly different source depth.
123 While it may be disconcerting that contrary interpretations have been published, with
124 reference to a claimed event of significance in the context of monitoring for nuclear explosions,
125 the larger picture is that: (a) additional resources can often be brought to bear on such claims
126 after a period of time; (b) it can then be possible to detect and characterize the nature of events
127 more than ten times smaller than has typically been reported (e.g. by the US National Academy
128 of Sciences in 2002 and 2012) as the monitoring capability for various regions; and (c) contrary
129 interpretations can potentially be resolved by on-site inspection. In the present case, new data
130 have been obtained, are now openly available, and are analyzed in this paper, indicating that the
131 12 May 2010 event exhibits the characteristics of a small natural earthquake, based on a specific
132 set of assumptions and an application of standard statistical classification methods. We re-affirm
133 a conclusion of Schaff *et al*. (2012) that no well-coupled underground explosion above about a
134 ton occurred near the North Korea test site on days we have studied.
135
136 ## Data Sources

The GSN open station MDJ has been providing continuous high-quality broadband seismic data to interested users since October 1986, and in particular acquired good seismic signals from the North Korean nuclear explosions in 2006, 2009, 2013, and 2016. It also recorded a number of chemical explosions in 1998, of yield about one or two tons, conducted in northeast China close to the border with North Korea (e.g., Song *et al.*, 2007; Zhao *et al.*, 2012), and we use these and other chemical explosion signals below, as well as MDJ records from small earthquakes, most of them within about 200 km from the North Korea nuclear test site.

As noted in the Introduction above, NECESS Array stations recorded signals from the 12 May 2010. They were weak, and only the nearest station, NE3C ($\Delta \sim 160$ km, see Figure 1), provided data of quality high enough to be used for analysis of *P/S* spectral ratios—the method we deem most useful to enable event characterization. Because the NECESS Array was deployed only from September 2009 to August 2011, it did not record any of the four well-recognized North Korea UNEs.

Most notably for the present paper we have obtained access to DBSN recordings of the 2006 and 2009 North Korea UNEs, and to the signals recorded by DBSN for the 12 May 2010 event. Details of this network are given by Chun *et al*. (2009), Chun and Henderson (2009), and Chun *et al*. (2011). It began operation in June 2004 and was removed in 2011, so it did not record the UNEs of 2013 or 2016. Furthermore, some station channels were non-operational for events of interest. Station DB09 of the network was moved 7.19 km to the east-northeast in mid-2007 and renamed DB17. But overall it is clear that these stations were carefully sited, and have provided high-quality recordings. We have used data from eight of the DBSN stations, shown in Figure 1, and their waveforms are presented and discussed in later Figures. Ground motion was

sampled at 100 sps, significantly higher than the 40 sps rate used continuously for MDJ and the NECESS Array stations.

For purposes of applying the *P/S* spectral ratio to identify a seismic event, it is desirable to have spectral data from different populations (earthquakes, explosions) in the same region, recorded at the same stations, as well as good data at these stations for the event one is trying to characterize. We do not have this situation in the present case for analyzing the 12 May 2010 event. From station MDJ we can compare *P/S* spectral ratios for UNEs and chemical explosions, and for earthquakes—though not for the 12 May 2010 event. From the NECESS Array stations we can compare *P/S* spectral ratios for earthquakes, industrial explosions, and the 12 May 2010 event. And from DBSN stations we can compare *P/S* spectral ratios for UNEs and the 12 May 2010 event. Furthermore, using the methods developed by Kim *et al*. (1997) we are able to use spectral measurements from vertical and horizontal components of recorded ground motion, which help to stabilize the *P/S* spectral ratios we report in this paper, which are based on about 500 seismograms (72 recorded at MDJ, 24 at eight stations of the DBSN, and the remainder at twenty stations of the NECESS Array). This is significantly more than the number of seismograms used for *P/S* spectral ratios reported by Zhang and Wen (2015a), for vertical components at stations SMT, that were used by them to compare the 12 May 2010 event to a pair of UNEs and three earthquakes. They did not give an objective criterion for classifying the event. We have been able to find earthquakes, in addition to those used by Zhang and Wen, that are significantly nearer to the North Korean test site, to provide a broader basis for comparison when using spectral ratios.

Estimates of location and magnitude for the 12 May 2010 event

In this section we use two methods to conclude that the 12 May 2010 event had its hypocenter within a few km of the North Korea nuclear test site, most likely southwest of the 2009 test location at a distance that we estimate to be between 4 km and 10 km. But first we note that Zhang and Wen (2015a), in application of their match and locate method, found a location only about 1 km to the southwest of the 2009 nuclear test, by maximizing the mean CC value for a stack of correlograms. The contours of their stack (their Figure 3A) appear not to have been sharply peaked at their preferred location for the event. Just how far apart the events are, of 2009 and 2010, is of interest in the context of understanding the capability of cross-correlation methods to enable detection of small events, and high-quality location estimates.

Before discussing our own estimate of the location of the 12 May 2010 event, a simple and helpful indicator of its location can be derived from the horizontal particle motion in the *P*-wave arrival used at stations with clear signals, since the motion gives an approximate azimuth from which this longitudinal wave is arriving. Figure 2 shows a particle motion plot for the horizontal components of the regional *P*-wave at NE3C (a NECESS Array station). The azimuth is appropriate for a source near the North Korea nuclear test site.

Relative location estimates based on Lg-wave cross-correlations

We used hypocentral estimates given in Zhang and Wen (2015a) for the 2006, 2009, and 2010 nuclear tests, to determine *Lg*-arrival times based on station distance and a group velocity of 3.5 km/s. Templates were used from the 2009 nuclear test since it was closer to the 2010 event than the 2006 explosion, and they were run on 10 minutes of continuous data from 2010. The correlation traces for the various stations are shown in Figure 3 for a ±10 s window centered

205 on the arrival time at each station expected for the hypocenter proposed by Zhang and Wen
206 (2015a). Five of the stations show maxima in this ±10 s window that have absolute time shifts
207 less than 1.5 s. The three closest stations show detection maxima with peak times close to zero,
208 which we take to be an indication that these are not just random detections among the stations
209 but that there is a physical basis in terms of increased signal-to-noise ratios (SNR).
210 The maximum correlation coefficients are all nevertheless quite low. The *Lg* signals are
211 not obviously visible in the raw waveforms for most stations, but detailed studies (e.g. Schaff,
212 2008) have shown that detection via cross-correlation is achievable with signals significantly
213 below the noise level. Low CC values may partly be due to the sample rate being 100 sps
214 compared to the 20 sps that we have typically used in previous studies. But this higher sample
215 rate makes the correlation values, even though low, more statistically significant because of the
216 longer duration and greater time-bandwidth product for the template window of 50 s. (Note that
217 we do correlation in the time domain so the correlation function has a sample rate of 100 sps as
218 well. And interpolation of the correlation function around the peak usually achieves precision
219 that is about $1/10^{th}$ of the sample interval.) The higher rate means there are more false alarms in
220 a given time period because they scale linearly with sample rate. To do a thorough false alarm
221 analysis and to quantify the statistical significance of these detections, continuous data records
222 longer than 10 minutes need to be analyzed. But a rough estimate of the statistical significance
223 of these detections can be made, based on the length of the lag windows searched over for a
224 given maximum, even for these short (10 minute) records. To do this we count the single
225 maximum of each of the five stations with peak lag times t_{peak} having an absolute value less than
226 1.5 s. Then we determine the nearest time t_{lag} at which the CC value exceeds the local maximum

for the lags searched over. The probability that the original maximum was a false alarm can then be estimated as p = absolute value of t_{peak} / t_{lag}. These values are shown in Table 1.

We have found from work on the seismicity of China, Parkfield (California), and Kazakhstan, that association even at just two stations is a powerful means of validating detections at single stations and reducing false alarms (Schaff, 20009; Schaff and Waldhauser, 2010; and Slinkard *et al.* 2014). The combined probability of a false alarm coincidentally occurring at two or more stations can be determined by multiplying the individual probabilities at each station, since false alarms at different stations are independent from each other. From the last column of Table 1 we find a combined probability of 1.3 E-11, indicating that the chance of a false detection occurring at all five of these stations, within a time window that includes the expected arrival of the *Lg*-wave for an event with the hypocenter determined by Zhang and Wen (2015a), is very very low—and that therefore the detections of an event on 12 May 2010 are highly likely to be true, from *Lg* signals alone.

The match and locate method of Zhang and Wen (2015b) effectively combines the detection, association, and location steps all together in one process to determine a single averaged correlation trace. They show their detection is statistically significant for two months of data. Assuming a sample rate of 100 sps and only one maximum value in this time period the probability of such a detection occurring by chance is p = 1 / 60 days / 24 hours / 3600 s / 100 sps ~ 2 x 10^{-9}. This shows the power of association in Table 1 because from just 10 minutes of data we find a lower probability of occurring by chance (once in 25 years using 100 sps) compared to one in 2 months of data from a single trace.

To estimate a relative location of the May 2010 event with respect to the known 2006 UNE, we can use the lag times of the peak values as differential travel times, as has been done in

a case study of the 1999 Xiuyan sequence in China (Schaff and Richards, 2004). We explore the sensitivity and robustness of relocating the pair of events (2009 and 2010) by adding stations of increasingly poor quality as determined by their deviation from the zero value we expect from reduced travel times of the *Lg*-arrival, based on the hypocenter in Zhang and Wen (2015). Three stations are needed as a minimum to solve for the three unknowns of the epicenter—but then since these are exact solutions, with zero residuals, no formal error bars can be computed to determine the location quality.

Figure 4 shows several location estimates for the 2010 event relative to the 2009 nuclear test, including those based on our *Lg* double difference analysis using different numbers of stations. The stars are the locations given in Zhang and Wen (2015a). The *Lg* location estimates for the four, five, and six station cases all agree with the 95% confidence error ellipse which is shown for the four-station case. But neither the three-station location nor the Zhang and Wen (2015a) location falls within this error ellipse. The four-station case has the smallest error bars and best fit for the residuals. But we find the data quality to be poorer than what we have typically used for *Lg* locations in other studies. The standard deviation of the residuals in the present case is 0.2618 s and is thus quite poor, for example by comparison with results from a study using *Lg* to obtain relative locations for 3689 events over all of China in which the average residual standard deviation was only 0.0417 s (Schaff *et al.*, 2016). The estimated 95% confidence ellipse for the *Lg* location has semi-axes 3.91 km and 0.50 km (using four stations). The azimuthal gap is 249 degrees. The standard deviation of the residuals before location based on the hypocenters in Zhang and Wen (2015a) is 0.7519 s, so there is improvement by about a factor of 3 in the residuals to explain the Dongbei data with our best *Lg* relocation.

Relative location based on picks of P- and S-arrivals

Figure 4 also shows results of a double-difference relocation of the 2010 event relative to the 2009 event using phase pick data only, for regional *P*- and *S*-waves from stations of the DBSN in the distance range 158 to 206 km. The square shows the estimated epicenter which in this case is about 7 km to the southwest of the 2009 event. It is in general agreement with the epicenter location estimated independently using *Lg*-wave correlation data, that is, lying several km to the southwest.

The RMS residuals for the locations based on *P*- and *S*-wave picks are 0.189 s for the 2009 event and 0.117 s for the 2010 event. These can be compared to the standard deviation of the residuals for the *Lg* location based on CC (0.2618 s). The standard deviation is related to the RMS but differs in the numbers of degrees of freedom. The estimated 95% confidence error bars for the *P*- and *S*-wave locations are ±2.71 km east-west and ±0.79 km north-south. The error ellipse sizes and fit of the residuals are also in general agreement for the two independent location runs, i.e. using *P*- and *S*-wave phase data and *Lg* CC data.

Additional discussion of location estimates for the 2010 event

Based on the formal error ellipses shown in Figure 4 we can say with 95% confidence that the location of the 2010 event relative to the 2009 event is no closer than 3.8 km and no further than 9.5 km to the SW. (The orientation of the *P*- and *S*-wave error ellipse should also be elongated in a NW-SE direction, like the *Lg* CC error ellipse. We were not able to obtain singular value decomposition locations using *P*- and *S*-waves so these locations are based on the LSQR method of inversion which returns only the diagonal elements of the covariance matrix in the cardinal directions.)

We can compare overall statistics for the 2010 relative event location with those of 3,689 events in and near China for which we have recently obtained relative epicenter locations based on high quality *Lg* CC measurements (Schaff *et al*., 2016). The average azimuthal gap is large in our study of all China, 205°, but the gap for the 2010 event is larger, at 249°. For all of China the average of the semi-major axes (= 0.42 km), is much smaller than for the 2010 event (= 3.91 km), and the same applies to the average of the semi-minor axes (0.09 km vs 0.50 km). The fit for the residuals for events over all of China has a standard deviation of 0.0417 s which is smaller than the sample interval of 0.05 s. The 3,689 events are located within 1 km of at least one other event. Typically, Schaff *et al*. (2016) find that *Lg* CC data lead to high-quality relative relocations only for inter-event separations less than 1 km. The fact that the 2010 event has residuals more than six times higher than residuals for the high-quality *Lg* CC data, in the study for all of China for events separated by 1 km or less, is strong additional evidence that the 2010 event is more than 1 km away from the 2009 event, besides the actual location estimate placing it about 4 km to the SW. Further but more indirect evidence that the 2010 event is more than 1 km away from the 2009 event, and/or is a different type of event such as an earthquake instead of an explosion, is that there was no CC data of sufficient quality to be used for differencing the *P*- and *S*-wave arrivals (so picks were used). Carmichael & Hartse (2016) describe how detection thresholds, based on CC methods, rise by up to a magnitude unit for sources located four km or more from one of the larger North Korean announced nuclear tests. Typically, CC data for *P*- and *S*-waves has been shown to be useful for inter-event separation distances up to 2 km in northern California (Schaff e*t al*., 2004; Schaff and Waldhauser, 2005; Waldhauser and Schaff, 2008).

We are reasonably confident in the *Lg* correlation DD location because it agrees with the location estimate of the *P*-wave picks. These two locations were arrived at completely independently with two different data types and inversions (for *Lg* we solve only for epicenter in a different program whereas for *P*-waves hypocenters are returned). The general direction from the 2009 event and the size of the ellipses agree from the two independent tests. The data is not high quality either for *Lg* or *P*, but we are fairly confident in our qualitative conclusions that the 2010 event occurs more than 1 km away from the 2009 event and likely is between 4 to 10 km to the southwest.

Another line of general evidence for greater than one km separation between the 2009 and 2010 events, is the experience in obtaining a location for the 2009 explosion relative to that of 2006. In days immediately following the 25 May 2009 nuclear test, one of us (Kim) acquired data from more than 50 regional stations in China, South Korea, and Japan, that had recorded both these first two tests by North Korea. For purposes of estimating a relative location he used 83 travel-time pairs (48 from *Pn*, 35 from *Sn*), at stations where the quality of cross-correlation of 2009 signals against those of 2006, was so low that travel time picks appear to provide the best way to measure relative arrivals. He also used 36 waveform cross-correlation pairs derived from signals where CC methods appeared to be somewhat better than differencing arrival-time picks, but CC quality was still somewhat poor, and this is the relevant point, for these two explosions were estimated to be about 2.6 km apart, as reported in June 2009 at the CTBTO's ISS conference just a few weeks after the occurrence of the 2009 explosion. Seismic events at the same depth but more than about 2 km apart have regional signals that do not cross-correlate well in bands such as 0.5 – to 5 Hz that we are using for precision studies. And there is even

greater sensitivity to depth (that is, CC values fall off faster with depth than they do with differences in epicenter).

Finally on this point, we note again a conclusion of Ford and Walter—that if the 2010 event were at the Zhang and Wen location it would likely have had signals sufficiently similar to the explosion of 2009, that the 2010 event would be detectable using cross-correlation methods on IMS data.

Magnitude estimates for the 12 May 2010 event

The magnitudes given in Table 2 are based on root-mean-squared (RMS) amplitude ratios of a 2009 UNE template trace with known magnitude (4.3 according to Zhao *et al.*, 2012) and that of the detected signal window, about 50 s in length (*P* and *S* included). The magnitude is about 1.5 based on data from DBSN stations.

Characterization of the 12 May 2010 event—earthquake or explosion

In this section we describe features in several seismograms obtained for the 12 May 2010 event. We describe and apply a linear discriminant function (LDF), derived from the *P/S* spectral ratios at station MDJ, and apply it to data from DBSN. In a later section we apply it to data from the only station (SMT) used by Zhang and Wen (2015a) to analyze spectra, and from the best station (NE3C) of the NECESS Array. Our work is done first using only the vertical component of recorded motions.; and then, where possible, using horizontal as well as vertical records. We find that the event of interest consistently has *P/S* spectral features within the preferred frequency band putting it into the earthquake population in the several different

analyses we have conducted. Our strongest conclusions in this regard, are those based on measurements made from three-component records.

We also comment on the method used by Zhang and Wen (2015a) to provide a yield estimate if the event is taken to be an explosion.

Observed data from the Dongbei Broadband Seismographic Network (DBSN)

Seismic records on 12 May 2010 from several stations of the DBSN within about 230 km from the North Korean test site show signals consistent with the origin time and location reported by Zhang & Wen (2015a). The vertical-component record section given in Figure 5 shows possible signals from the event. The waveforms shown have been passed through a third-order Butterworth filter with cut-off frequencies at 0.8 and 10 Hz. Seismograms shown in Figure 5 and later are recordings of ground velocity, since the DBSN stations were equipped with broadband seismometers flat to velocity between periods ranging from 30 or 60 s to 0.02 s (0.017 or 0.033 Hz to 50 Hz) (Chun & Henderson, 2009). At stations DB08 (Δ = 160 km, Az = 0º) and DB17 (Δ = 202 km, Az = 14º), arrival times of *P*- and *Lg*-waves are consistent with signals expected from the 12 May 2010 event. On the other hand, for signals at DB06 (Δ = 181 km, Az = 320º) and at DB05 (Δ = 193km, Az = 306º), only the vertical records show *P*-wave arrivals, and two horizontal components show very weak signals nearly comparable to the noise. At DB10 (Δ = 209 km, Az = 29º), only signals arriving with *Lg*-wave speed are visible and no *P*-waves are observed on any of the three components. Data quality varies between these stations, with indications of occasional noise bursts at DB04, DB06, and DB10 (Fig. 5). We place greater reliance on DB08, and DB09/DB17. Hence, *P*- and *Lg*-wave energy arrivals at several stations are consistent with the event on 12 May 2010 reported by Zhang and Wen (2015a).

A comparison of vertical records at DB08 from the 2006 and 2009 UNEs and the 12 May 2010 event is shown in Figure 6a. *P*- and *S*-wave arrival times indicate that the three events occurred close to each other, but the relatively stronger *S*-wave on the 12 May 2010 record suggests that they may not be the same type of source (Figure 6a). Three-component waveform data from the 2009 UNE and May 2010 events at DB17, and from the 2006 UNE recorded at DB09, suggest that the time and location of the claimed event on May 12 is broadly consistent with the report by Zhang and Wen (2015a) (see Figure 6b). However, there are questions: is the 12 May 2010 event located as close to the 2009 UNT as claimed by Zhang & Wen (2015a)? Can the seismic data tell us if the event is an underground explosion as claimed by Zhang & Wen (2015a)?

Observed data from the Northeast China Extended Seismic Array

The NECESS Array experiment deployed 127 seismographic stations in northeastern China during Sept. 2009 – Aug. 2011 and the waveform data (recorded continuously at 40 sps) are available from the IRIS Data Management Center. Only the station closest to the North Korean test site shows seismic signals from the 12 May 2010 event that appear suitable for measurement of spectra. At this station, NE3C (Δ = 164 km, Az = 350º; see Figs. 1 and 2), *P*- and *Lg*-wave arrival times on all three components appears to be consistent with those of the Dongbei network station, DB08. Figure 7 compares records at DB08 and NE3C. Arrivals from the event of interest at other NECESS Array stations are of lower quality.

Since the NECESS Array recorded waveform data in continuous mode, we had an opportunity to examine the entire two years of archive data for additional possible events from

410 the same area as the event of 12 May 2010. Using a 30 s long waveform trace containing *P*- and
411 *Lg*-waves of the 12 May 2010 event as a three-component template, we carried out waveform
412 cross-correlation detection in a search for similar signals in two years of continuous data. Using
413 a criterion of CC > 0.20 this process yielded 6,373 energetic but dissimilar detections. Visual
414 inspection of most of the detections indicated that they are likely quarry blasts, not near the
415 North Korean test site. We found no similar events for the period from September 2009 through
416 August 2011. Somewhat similar signals were detected on 6 June 2010 at 11:17:51, which were
417 also found by Ford and Walter (2015). By comparison with the 12 May event the 6 June event
418 had magnitude about 2.3, and its signals are observed at other stations. A preliminary location
419 for the event suggests that it is about 77 km northwest (Az = 292º) from the 2009 UNE.

420

421 *Characterization of the 12 May 2010 event from spectral features*

422 A previous study found eight shallow earthquakes with magnitudes 2.5 to 4.1 that
423 occurred during 1989 – 2005 within about 200 km of the North Korean test site (see Figure 1 of
424 Kim & Richards, 2007). That study also found signals at MDJ from four known chemical
425 explosions conducted in the summer of 1998 not far from the test site, as part of a PASSCAL
426 experiment led by Francis Wu (Song *et al.*, 2007). They were 1 or 2 ton (TNT equivalent)
427 single-hole shots ranging in magnitude from about 1.0 to 2.0; their signals at MDJ appear as
428 Figure 6 of Schaff *et al.* (2012).

429 In studies of earthquakes and explosions that we have conducted for northeastern China
430 and the Korean Peninsula, waveform data have been provided from a variety of stations
431 including BJT, HIA, and MDJ in China; and INCN in South Korea. Data for these GSN stations
432 are archived by the IRIS Data Management Center. In general for discrimination studies we

would wish to work with as much data as reasonably possible, and in particular to use high-sample-rate records if available. We note that for some events in the region there are triggered data sets (notably from MDJ) sampled at 80 sps. However for practical purposes of developing a straightforward strategy for applying the *P/S* discriminant method to signals from the 12 May 2010 event it is necessary to make some choices on what data subsets will be used, and not use all available data. Such choices are guided by the facts that: (a) we have direct access to only a limited number of stations that recorded the event of interest, namely, those of the DBSN, that also recorded signals from nearby known nuclear explosions; (b) at present we do not have access to DBSN data from earthquakes in the region; and (c) our station with the most extensive archive of earthquake and explosion records, namely MDJ, did not provide useful records of the 12 May 2010 event. (At the time of expected arrival at MDJ, of signals from this event of interest, there appears to be an *Lg* arrival from a different event, obscuring the record.) Though the archive for MDJ includes some triggered higher-sample-rate records (80 sps), it is consistently reliable for events recorded in the 40 sps continuous data stream.

In these circumstances, our strategy has been to build up a training set of signals at MDJ for both explosions and earthquakes in the region, and then to find the frequencies at which the observed values of the *P/S* spectral ratio provide the best separation of the two populations (explosions and earthquakes), using standard statistical methods for event classification. We have then used this knowledge to evaluate the *P/S* spectral ratios measured from DBSN stations, noting that their azimuthal direction from the North Korea test site source region is similar to that for station MDJ.

To carry out such a strategy, we have identified 12 earthquakes and 12 explosions in northeastern Korea and northeastern China, mostly within about 200 km from the test site, in

order to carry out multivariate discriminant function analysis. The events we use as training sets are listed in Tables 3 (for explosions) and 4 (for earthquakes), and their locations are indicated in Figure 1. In addition to the UNEs of 2006, 2009, and 2013, and the chemical explosions of 1998, we have included industrial explosions in two separate regions of North Korea (in Table 3, the events numbered 8, 9, and 10; and then numbers 11 and 12).

Three-component waveform data at station MDJ are fairly good for all 24 of the events. Other GSN stations INCN and BJT show weak seismic signals with relatively poor signal-to-noise ratios.

P/S spectral amplitude ratio on vertical-component records

Numerous regional seismic event discrimination studies have shown that the high-frequency (usually higher than 3 – 4 Hz), *P*/*S* spectral amplitude ratios of regional *P* (*Pg* or *Pn*) and *S* (*Sg*, *Sn* or *Lg*) waves are observed over distances ranging from a few kilometers up to about 2,000 kilometers, and provide an efficient method to classify regional earthquakes, quarry blasts and underground nuclear explosions (e.g., Kim *et al.*, 1993; Kim *et al.*, 1997; Hartse *et al.*, 1997, among many others). The high-frequency *P*/*S* spectral ratios on vertical-component regional records, have been used successfully with a misclassification probability of a few per cent (e.g., Kim *et al.*, 1993). Distance corrections have also been applied to *P*/*S* ratios, significantly improving the discrimination power, particularly if the events have data recorded over a wide distance range (e.g., from 3° to 17°). In particular, Kim *et al.*, (1997) and Pasyanos and Walter (2009) have demonstrated improvement in discrimination capability when *P/S* spectral ratios have been corrected for attenuation effects in a region that exhibits significant variation in attenuation (parts of Central and South Asia).

Figure 8 shows vertical-component $\log_{10}(Pg/Lg)$ spectral amplitude ratios at discrete frequency points between 1 and 15 Hz used in discrimination analysis for 12 earthquakes and 12 explosions (five chemical explosions, four shots known to have been conducted in single holes, and three North Korean nuclear tests) recorded at MDJ. To calculate *P/S* spectral amplitude ratios, *Pg* and *Lg* signals are windowed with a Gaussian weighting function centered at group velocities around 5.9 – 6.0 and 3.3 – 3.4 km/s, respectively. A standard deviation, σ, of σ_{ref} = 2.5 s is used for the *Lg* Gaussian window at a reference distance of Δ_{ref} = 100 km. The window includes *Lg* arrivals with group velocities between 3.66 and 3.09 km/sec in its $\pm\sigma$ width centered at 3.35 km/s, and the Gaussian window is truncated at $\pm 1.96\ \sigma$ ($\approx$ 95%), which includes signals within the group velocity range from 4.0 to 2.88 km/s. Window lengths at different distances are scaled by $\sigma = \sigma_{ref} \times \Delta/\Delta_{ref}$. Thus, the time window used for *Lg* has the same group velocity window regardless of distance. The *Pg* window length of σ_{ref} = 2.0 s is used, so that the *Pg* window includes *P* arrivals with group velocities between 6.7 and 5.3 km/s regardless of distance when centered at 5.9 km/s. The *Pg* and *Lg* signals, weighted by the Gaussian functions, are fast Fourier transformed. The resulting amplitude spectra are smoothed with another Gaussian function having σ = 1 Hz and are resampled at every 1-Hz interval from 1 to 15 Hz. 15 to 16 Hz is a conservative upper limit of the seismic data available from MDJ and other GSN-type stations in the region, since stations recorded seismic data continuously at 40 samples/s, and occasionally at 80 or 100 samples/s in trigger mode. For the two training sets of 12 earthquakes and 12 explosions, seismic data with high signal-to-noise were selected for *P/S* spectral ratio analysis. However, *P* or *S* arrivals from two low-magnitude single-hole explosions in the training set at high frequencies (above 10 Hz) fall below the ambient noise on some records, and for this reason the *P/S* ratios at high frequencies were deemed to be less reliable.

The vertical-component *P/S* spectral amplitude ratios from two UNEs and the 12 May 2010 event recorded on Dongbei stations (DBSN) in Northeastern China along the China-North Korea border region are also plotted. Among a total of 17 stations of DBSN, only a few stations—DB08 and DB17—show clear *P*- and *S*-waves from the 12 May 2010 event (Figure 5). Station DB08 also recorded the two UNEs of 2006 and 2009 (see Figure 6a). Station DB17, the other Dongbei station providing good data for the event of interest, recorded the 2009 UNE, but not that of 2006, for which we can turn to a nearby station, DB09, as described in Figure 6b. Hence, stations close together were used to compare spectral characteristcs and *P/S* amplitude ratios (see also Figure 7). Noise analyses of DBSN data, based on use of a pre-*P*-wave-arrival noise window, indicate that signal-to-noise ratios (SNR) are quite high in most of the records, but in some cases, the signals fall to the background noise level above 10 Hz. For example, the vertical record at DB08 from the12 May 2010 event shows that *Lg* wave spectral amplitudes at 11-12 Hz are smaller than the noise amplitude, so the *Pg/Lg* ratio is less reliable in those frequencies (see Figs. 6a, 7 & 8).

Selection of the most effective frequencies for purposes of event discrimination

As mentioned in the previous section, many studies of *P/S* spectral ratios have employed data in the 1 – 5 Hz band, whereas Kim e*t al*. (1993, 1997) and Kim and Richards (2007) have shown the utility of 5 – 15 Hz data in separating earthquakes from explosions when such higher frequencies are available. To explore the discrimination power of *P/S* spectral ratios at various frequencies, we use the Mahalanobis distance-squared measure between the multivariate means of the earthquake and explosion populations, which is related to misclassification probabilities (as shown, e.g., by Kim *et al.*, 1993).

Thus, we are working here with $f_{Eq}(\mathbf{r})$ and $f_{Ex}(\mathbf{r})$ as the probability densities of measurable features of the two types of events, where $\boldsymbol{r}$ is a column vector representing $\log_{10}(P/S)$ values, sampled at d different frequencies. We follow the formalism as described for example in the textbook by Seber (1984). The mean data vector and covariance matrix associated with the training set of 12 earthquakes are $\boldsymbol{\mu}_{Eq} = (1/12) \sum_{i=1}^{12} \mathbf{r}_{Eq}^{i}$ and $\boldsymbol{S}_{Eq} = (1/11) \sum_{i=1}^{12} (\mathbf{r}_{Eq}^{i} - \boldsymbol{\mu}_{Eq})^{T} (\mathbf{r}_{Eq}^{i} - \boldsymbol{\mu}_{Eq})$, and similarly for the training set of explosions to obtain $\boldsymbol{\mu}_{Ex}$ and $\boldsymbol{S}_{Ex}$, where ^{T} denotes a transpose.

We evaluate a linear discriminant function $D(\mathbf{r})$ under the assumptions that sample distributions are normal, the two covariance matrices (describing the scatter of values measured for the two training sets) are effectively the same, and the events in the training sets are correctly classified. This linear function is

$$D(\mathbf{r}) = \boldsymbol{\lambda}^{T} [\boldsymbol{r} - (\boldsymbol{\mu}_{Eq} + \boldsymbol{\mu}_{Ex}) / 2] \qquad \text{where} \quad \boldsymbol{\lambda} = \boldsymbol{S}^{-1} (\boldsymbol{\mu}_{Ex} - \boldsymbol{\mu}_{Eq}) \qquad (1)$$

and $\boldsymbol{S}$ is the average of $\boldsymbol{S}_{Eq}$ and $\boldsymbol{S}_{Ex}$. Conceptually, if $\mathbf{r}$ is a dataset for a particular event, then $D(\mathbf{r})$, which is a scalar, is a dimensionless measure of the distance between $\mathbf{r}$ and the average of the two training sets, expressed in standard deviation units. Further discussion of $D(\mathbf{r})$ is given in a tutorial section of the electronic supplement to this paper. If, *a priori*, we make the simple assumption that a new event is equally likely to be an earthquake or an explosion, then an event with $D(\mathbf{r}) > 0$ is taken to be an explosion, and an event with $D(\mathbf{r}) < 0$ is taken to be an earthquake. Our definition of D is the negative of that used by Kim *et al.* (1993 and 1997). This change is motivated by the tutorial discussion given in the electronic supplement, and noting that typically the log (P/S) values of explosions are higher than those of earthquakes, so components of the vector $\boldsymbol{\mu}_{Ex} - \boldsymbol{\mu}_{Eq}$ are positive.

549 The Mahalanobis distance-squared measure between population means, Δ^2, given by
550 $\Delta^2 = D(\boldsymbol{\mu}_{Ex}) - D(\boldsymbol{\mu}_{Eq})$, is calculated and contoured in Figure 9, for different choices of the LDF
551 (linear discriminant function), using 1, 2, 3, 4 and 5 variables (values of *d*, the number of discrete
552 frequencies), for frequencies in the range from 1 to 15 Hz. Although misclassification
553 probabilities tend to decrease as *d* increases, the computational effort to provide a vector
554 observation goes up with *d*. Both the precision of estimation, and the robustness of the LDF, fall
555 off with increasing *d*. LDF analysis, and related *F*-statistics, work well for two groups
556 (earthquakes and explosions), with parameters less than 5 and number of observations (events)
557 greater than 9 in all groups (Seber, 1984). We have therefore selected $d = 4$ as the optimum
558 number of parameters for our LDF analysis using *P/S* ratios in this study. Furthermore, we have
559 worked within a simple framework in which the multivariate means of the explosion and
560 earthquake populations are found to be significantly different, and the covariance matrices of the
561 two groups (i.e., the scatter of the two populations) are effectively the same. We recognize the
562 merit of working with a subset of the *d* variables in *P/S* ratios that are based upon lower
563 frequencies whenever these can be effective, because low frequency *P/S* ratios may be less
564 subject to frequency-dependent amplitude attenuation during wave propagation, and because
565 data at the lower frequencies may be more readily acquired (for example, at more stations). But
566 it is clear from Figure 8 that the two populations tend to separate better at the higher frequencies,
567 which are available for our use in the present case.
568
569 *Discrimination power of vertical-component P/S spectral ratios at MDJ*
570 For vertical component *P/S* ratios, frequencies between 3 and 6 Hz, and 6 and 12 Hz
571 will carry most of the classification power, as indicated by Figure 9. The shaded contours are for

values of Mahalanobis Δ^2 and they show the classification power of each frequency for chosen sets of 1, 2, 3, 4 and 5 frequency values, resulting in progressively higher Mahalanobis distance-squared values, as expected. From Figure 9, we consider three frequencies in the band from 7 to 11 Hz (i.e. at 7, 9, and 11 Hz), and for four frequencies between 6 and 11 Hz, as the most robust region for discrimination. At lower frequencies, values in the range 3 to 5 Hz and 2 to 5 Hz are most robust, for three and four frequency values, respectively. Ranges including 5 to 6 Hz are not good, because for them the covariance matrices of the two populations are not the same (Seber, 1984; Kim *et al.*, 1993).

For the case of four variables used, and discrete frequencies between 6 and 9 Hz, the analysis is based specifically upon 6, 7, 8 and 9 Hz. In this case all earthquakes and explosions are correctly classified on the basis of MDJ data, and the Mahalanobis distance squared is Δ^2 = 20.7 with misclassification probability of 1.15%. Adjacent frequency bands also provide strong discrimination power: 7 to 10 Hz (Δ^2 = 19.9; misclassification probability = 1.29%), and 8 to 11 Hz (Δ^2 = 19.2; misclassification probability = 1.42%). Broad frequency bands between 6 and 12 Hz provide strong classification power with small misclassification probabilities of about 1.1% to 1.9% (Figure 9).

Our discussions below are based upon use of four frequencies, but here we note that if three variables in the frequency band 8–10 Hz are used (i.e.: 8, 9, and 10 Hz), the Mahalanobis distance is Δ^2 = 19.2 and all earthquakes and explosions are correctly classified with misclassification probability = 1.42%. Using 7, 8, and 9 Hz, we find Δ^2 =18.3 and misclassification probability = 1.62%; using 9, 10, and 11 Hz, Δ^2 =16.5 and misclassification probability = 2.1%. Thus use of three frequencies yields reasonable discrimination power, although we have chosen in the analysis below to work with four frequencies.

Linear discriminant function from vertical-component Pg/Lg ratios

The best LDF from the vertical-component *Pg/Lg* ratio measured from the sample data sets consisting of 12 earthquakes and 12 explosions is obtained for discrete frequencies 6 – 9 Hz. For each event, vertical-component $\log_{10}(Pg/Lg)$ ratios at frequencies of 6, 7, 8 and 9 Hz correspond to the variables r_1, r_2, r_3, and r_4. The linear discriminant function is obtained as

$$D(\mathbf{r}) = -7.46 + 12.88r_1 + 4.28r_2 - 26.81r_3 + 40.19r_4 \qquad (2)$$

for which $\Delta^2 = 20.7$ and the misclassification probability is 1.15%.

Values of $D(\mathbf{r})$ may be called the discriminant score and are plotted in Figure 10 with respect to the mean $\log_{10}(Pg/Lg)$ spectral amplitude ratio of each event. Vertical lines in the Figure denoted as *Eq* and *Ex* are the projection of the multivariate mean of the earthquake and explosion populations, respectively. The vertical line, D_0, is the line $D(\mathbf{r}) = 0$, which serves to classify events when the *a priori* probability of the two populations is the same. The distance between *Eq* and *Ex* is the Mahalanobis *D*-squared measure of distance between two populations. It follows from the logarithmic values in Figure 10 (also, from Figure 8) that all the earthquake records from various paths in northeastern China and North Korea have a mean vertical-component *Pg/Lg* spectral ratio of about 0.6 (and hence a logarithmic value around –0.2), while the explosion records show a mean of about 3.2 (logarithmic value around +0.5) in the frequency band from 6 to 9 Hz.

Classification of known underground nuclear tests and of the claimed event on 12 May 2010

We tested the performance of the high-frequency discriminant function described above, equation (1), by applying it to the two known nuclear tests at the North Korean test site which

were recorded at Dongbei network stations. These two known explosions recorded by this network are correctly classified, as shown in Figure 10, indicating that high frequency, vertical-component *Pg/Lg* spectral ratios are capable of classifying earthquakes from explosions in the northeastern North Korea – northeastern China region. We also tested the claimed event of 12 May 2010 recorded at Dongbei network stations – DB08 and DB17, using the above discriminant function. We found that the claimed event on 12 May 2010 was classified as an earthquake (the red square in Figure 10), based on our study of vertical-component waveforms.

P/S spectral amplitude ratio on three-component records

The high-frequency *P*/*S* spectral ratios of rotated, three-component regional records, have been used successfully with total misclassification probabilities of only a few percent (e.g., Kim et al., 1997). In this study, we apply that previous experience to data collection and analysis for northeastern North Korea and NE China in order to evaluate the *P*/*S* ratios of three-component records from the 12 May 2010 event described by Zhang and Wen (2015a).

The method is simply to rotate the observed NS- and EW-component seismogram pairs to obtain radial (R) and tangential (T) components. The *P*/*S* ratios of three-component records are then formed for each station by defining $P / S = (P_Z^2 + P_R^2)^{½} / (S_Z^2 + S_R^2 + S_T^2)^{½}$ where subscripts indicate the component. For example, P_Z indicates the spectral amplitude of *P* waves on the vertical-component (Z), and S_R indicates the spectral amplitude of *S*-waves on the radial-component. A single three-component *P*/*S* ratio is obtained for each discrete frequency. This approach uses the observation that explosive sources tend to excite relatively strong and impulsive *P* waves and weak shear waves, whereas earthquakes as shear-dislocation sources tend to generate weak *P* waves but relatively energetic *S* waves, particularly on the transverse

component. Figure 11 shows three-component $\log_{10}(Pg/Lg)$ spectral amplitude ratios at discrete frequency points between 1 and 15 Hz used in discrimination analysis for our sample data set of 12 earthquakes and 12 explosions.

Note that when a *P*-wave or *S*-wave is incident upon a seismometer sited near the Earth's free surface, the recorded ground motion consists of a superposition of *P*-waves and *S*-waves (there are reflected *P*- and *S*- waves as well as the incident wave). Kim *et al.* (1997) showed that a slight improvement in discrimination capability can be achieved, if, instead of using rotated components, an additional correction is made, namely, removing the effect of the *P*- and *S*-waves reflected at the free surface. See also Problem 5.7 of Aki and Richards (2009). But such a correction entails making additional assumptions, for example on near-surface structure, and on the horizontal slowness of the recorded waves. The additional capability may not be worth the trouble associated with making the additional assumptions.

The three-component $\log_{10}(Pg/Lg)$ spectral amplitude ratios from the earthquake and explosion populations overlap significantly at frequencies 1 through 4 Hz, but the spectral ratios from the two populations are fairly well separated at frequencies greater than 5 Hz (Figure 11). The overlap and separation of high-frequency spectral ratios from these two populations are also generally observed for other regions such as southern Korea (Kim *et al.*, 1998), and nuclear test sites in China (Hartse *et al.*, 1997) and Nevada (Walter *et al.*, 1995).

Three-component *P/S* spectral amplitude ratios are generally similar to those measured from vertical-components in broad frequency bands: *P/S* ratios of the earthquakes are fairly well separated from those of explosions at frequencies above 5 Hz. The average *P/S* ratio of earthquakes in the 1–15 Hz band is about 0.4, whereas the mean *P/S* ratios of explosions in the same frequency band is about 1.5 (see Figure 11). But the discrimination capabilities when

using three-component data are somewhat improved (over use of vertical components only), as we next discuss.

Discrimination power of three-component P/S spectral amplitude ratios at MDJ

We evaluated the discrimination power of three-component *P/S* ratios using methods similar to those applied to the vertical-component data. The resulting discrimination power is shown at various discrete frequencies, and for different numbers of parameters used, in Figure 12. Broad frequency bands between 6 and 12 Hz, with 3 and 4 parameters used for the LDF, provide strong classification power with very small misclassification probabilities (~ 0.5 – 1.27%). Further details on the discrimination power obtained with different choices of frequency are provided in the electronic supplement. We next go forward using our best choice.

Linear discriminant function based on the three-component Pg/Lg ratio

The best LDF from the three-component *Pg/Lg* ratio for the sample data sets, consisting of 12 earthquakes and 12 explosions recorded at station MDJ, is obtained for four discrete frequencies in the band 6 – 9 Hz. For each event, let three-component $\log_{10}(Pg/Lg)$ ratios at frequencies of 6, 7, 8 and 9 Hz correspond to the variables r_1, r_2, r_3, and r_4. The linear discriminant function we obtain, is

$$D(\mathbf{r}) = -4.33 + 14.43r_1 - 16.77r_2 - 12.04r_3 + 45.91r_4 \quad (3)$$

with $\Delta^2 = 25.6$. All earthquakes and explosions in our training sets are classified correctly (Figure 13), and the misclassification probability is only 0.57%. We can now use this LDF to classify seismic events in northeastern North Korea and northeastern China.

687 *Classification of known underground nuclear tests and the claimed event on 12 May 2010*

688 We tested the performance of the above high-frequency discriminant function, eq. (3), by
689 applying it to the dataset of two known nuclear tests at North Korean test site recorded at
690 Dongbei network stations. We found that two known explosions are correctly classified (Figure
691 13). We also evaluated the seismic event on 12 May 2010 recorded at Dongbei network stations
692 (DB08 and D17), using the above discriminant function, finding that the event is then classified
693 as an earthquake (Figure 13).

694

695 *Lack of effects due to distance and magnitude, on the measured spectral ratio*

696 Figure 14 shows the measured three-component *P/S* ratios plotted against epicentral
697 distance. In cases where high frequencies are significantly attenuated, there can be a distance
698 dependence in the ratio and it may then be useful to make a correction to the observed spectral
699 ratio, as done for example by Kim *et al*. (1997) and Pasyanos and Walter (2009), to account for
700 the fact that the denominator (based on *S*-wave data) may attenuate with distance differently
701 from the numerator (based on *P*). But we have not made a correction for distance, since
702 earthquakes and explosions in the distance range 150 to 550 km show no clear distance-
703 dependence of the ratios, and the training set data have comparable distance ranges and are
704 tightly distributed. The twelve earthquake events have an average epicentral distance of 368±96
705 km, whereas average distance for the twelve explosions is 320±71 km (Figure 14a).

706 Note that *P/S* ratios of the earthquakes and explosions show separation along a horizontal
707 line, namely $\log_{10}(Pg/Lg) = \sim 0.1$ (the data here are for chemical explosions, single-hole shots,
708 UNEs, UNEs recorded at DBSN, and the claimed event on 12 May 2010).

To conclude on this point: in the present case distance corrections can be ignored, since all events are typically within 200 km from each other, and the paths have very low attenuation.

The lack of dependence of *P/S* ratios on magnitude, in the magnitude range we are using, was studied in detail by Pan *et al*. (2007), who wrote that:

"Applicability of regional *P/S* amplitude ratios for the discrimination of low-magnitude seismic events was tested and proved using earthquakes and explosions in central Asia. Results obtained show that regional *P/S* amplitude ratios which may discriminate medium or large magnitude events well are also applicable to low magnitude events…"

The *P/S* ratio of each record is plotted against event magnitude in Figure 14b. Again there is no clear correlation of *P/S* ratios on magnitude, and we have made no correction for a magnitude effect. Note that the low magnitude of the 12 May 2010 event is comparable to magnitudes of the small single-fired chemical explosions in our training set (events numbered 4 to 7 in Table 3); its *P/S* spectral ratio, averaged over the frequencies 6 to 9 Hz, at stations DB08 and DB17, is lower (and hence more earthquake-like) than this ratio for any of these four chemical explosions at the MDJ station.

Event magnitude, and relationships between magnitude, depth, and yield (for explosions)

Zhang and Wen (2015a) assign an *Lg* magnitude of 1.44 to the 12 May 2010 event. Taking the event to be an explosion they interpret this magnitude to obtain a yield estimate on the basis of several assumptions, including:

- use of a relationship between the teleseismic *P*-wave magnitude and log (yield) proposed by Bowers *et al*. (2001) for underground explosions at Novaya Zemlya, $m_b = 4.25 + 0.75 \log Y$ for *Y* in kilotons for Y ≥ 1 kt;

- correcting their magnitude for a depth effect as proposed by Denny and Johnson (1991) and Patton and Taylor (2011), which for a *P*-wave m_b amounts to adding a term $-0.7875 \log \left[h/120\ Y^{1/3}\right]$ to the right-hand-side of the magnitude–yield relation for an explosion conducted at a depth *h* (in meters) rather than at the standard depth of $120\ Y^{1/3}$; and
- taking the source depth *h* to be 230 m on the basis of differencing surface elevations associated with the entry and the end of a presumed adit.

But the numbers 4.25, 0.75, 0.7875, and 230 all have uncertainties in the original studies from which they were derived, with further uncertainties in application to nuclear explosions at the North Korean test site; the application of a *P*-wave magnitude correction for depth to a magnitude based on *Lg*-waves is questionable; and the underlying assumption of a one-to-one correspondence between a magnitude and yield at a particular depth, without regard to coupling conditions or knowledge of local geology or cavity gases, is inappropriate. The overall uncertainty of yield estimation in practice, even accepting the event to be an underground explosion of some sort (chemical or nuclear), is therefore considerable. According to an Office of Technology Assessment (OTA) report of the US Congress (1988), yield estimates are typically associated with a factor of uncertainty. For a yield estimate given as $\hat{Y}$, and *F* as the factor of uncertainty, the actual yield lies within a 95% confidence interval given by $\hat{Y}/F \leq Y(\text{actual}) \leq \hat{Y} * F$); and *F* is about 2 in the case of estimates based upon teleseismic *P* waves. Refinements have been made in the case of well-studied test sites, but even with a unified yield estimate based upon a variety of seismological methods the factor of uncertainty is still around 1.3 according to the same OTA report, and this is in the context of estimating the yield of a large underground explosion (tens of kilotons). For all explosions, yield estimates are subject to

uncertainties associated with the degree of coupling of nuclear energy into seismic waves, and such coupling is potentially easier to modify for small explosions than for multi-kiloton explosions. There is also the fact that just a few regional seismic signals are being used in the present case, with no opportunity to average tens or hundreds of teleseismic signals to obtain a reliable magnitude, as would be the case for a multi-kiloton event.

To summarize here: the yield of tamped explosions with magnitude in the range around 1.5 is on the order of one ton (TNT equivalent), but with considerable uncertainty, perhaps as much as a factor of five.

Discussion

During the late stage of CTBT negotiations in the 1990s, three categories of network were under consideration for treaty verification. These were variously called alpha, beta, and gamma networks; or primary, auxiliary, and supplementary networks. The primary and auxiliary networks have become a reality as components of the International Monitoring System (IMS) of the CTBTO. As noted for example by the US National Academy of Sciences (2012), the IMS and the associated International Data Centre are now providing verification capability that is significantly better than had been anticipated in the 1990s.

The event on which we focus in this paper, at magnitude about 1.5, has seismic signals about 300 times smaller than those of North Korea's first nuclear test, conducted in 2006.

We have been led to write a lengthy paper, analyzing such a very small seismic event, for two main reasons. First, the event has been characterized as a small nuclear explosive test, and we deem this conclusion to be unsupported by the seismological evidence available to us for reasons we have now presented (Figures 10 and 13) and which we augment in this section.

Second, it is surely important to convey a sense of current capabilities when reporting on signals from a region for which extensive supplementary data are available. We advocate for open access to such data, given its potential to help clarify the nature of problem events.

Our analysis has emphasized Mahalanobis methods of event classification, which led us objectively to conclude that *P/S* values measured at frequencies in the range from about 6 to 12 Hz provide the best discrimination. It may be somewhat surprising that higher frequencies (up to the 15 Hz values accessible from 40 sps data) are not more effective; and we are aware of the fact that *P/S* values for the 12 May 2010 event recorded at DBSN stations tend to move upward toward the explosion population at frequencies above 11 Hz in Figures 8 and 11. But at these higher frequencies and for these stations, these two Figures also show that *P/S* values for the UNEs of 2006 and 2009 are notably high within the explosion population. So, comparing *P/S* values just at DBSN stations, the 12 May 2010 event lies well below the two UNEs.

Application of our linear discriminant function to additional data

Although the seismograms used by Zhang and Wen (2015a) are not currently available to us, their paper provides values of the vertical-component *P/S* spectral ratios they have measured from signals recorded at their best station (SMT, in a borehole at a distance of only 120 km from the North Korea test site). We have read these values from their Figure 5, for three earthquakes (events 5, 9, and 10 of our Table 4), for two known nuclear explosions (events 2 and 3 of our Table 3), and for the 12 May 2010 event, at the frequencies needed to evaluate the linear discriminant function we have used in our Figure 10. In Figure 15 we have added these six points, derived from the data of Zhang and Wen, to points shown in Figure 10. Based on measurements made from SMT seismograms, the known earthquakes and explosions fall

appropriately into their respective populations, and the 12 May 2010 event falls among the earthquakes. In Figure 15 we also show a point derived from the NE3C station of the NECESS Array, for the event of interest. It is somewhat of an outlier among the earthquakes, but on the side away from being explosion-like.

The linear discriminant functions we have obtained, in equations (2) and (3), share a characteristic that was exhibited in our earlier studies (Kim *et al*., 1993 and 1997), namely, that these formulas involve a mix of signs, positive and negative, for the various contributing r_i values. The equations define hyperplanes that separate the two populations—earthquakes from explosions. The $D(\mathbf{r})$ measures are quite robust, since we obtain similar results with alternative choices of which frequencies to use and how to weight them. Note that the average values of $\log_{10}(P/S)$ given on the vertical axes of Figs. 10, 13, and 15, represent a quite different linear combination of r_i values, and these average values also lead to separated earthquake and explosion populations, albeit with somewhat greater misclassification potential than use of the horizontal-axis values derived from (2) and (3).

A key assumption of the Mahalanobis analysis in this paper, is that a linear discriminant function developed from training sets of MDJ data is suitable for application to waveform data (recorded at regional distances over low-attenuation paths) from other stations. The NECESS Array stations appear to provide us with an opportunity to evaluate this assumption, and to this end we have examined spectral ratios for 125 examples of three-component *P/S* from six earthquakes and six explosions, using waveforms drawn from 20 NECESS Array stations. The data are far from ideal, in that this was a temporary deployment, too short (from September 2009 to August 2011) to enable study of larger training sets, and it did not acquire data from any one of the four known UNEs in North Korea. Nevertheless it was interesting to carry out a linear

824 discriminant function (LDF) analysis of the training sets for NECESS Array data and to find that
825 *P/S* spectral ratios between 6 and 11 Hz provided the greatest Mahalanobis separation between
826 the two population means. The Δ^2 value was even greater than was the case for MDJ data
827 (discussed in Figure 13).

828 Seber (1994) suggests that LDF analysis should be good for at least nine events in each
829 population when using fewer than five parameters (frequencies, in our case). But we went ahead
830 with our somewhat smaller test data sets and tested for portability of the LDF. The first test, was
831 to classify the 12 explosions and 12 earthquakes of Tables 3 and 4 by applying the NECESS
832 Array LDF to MDJ data. All 12 explosions were correctly classified, but about half of the
833 earthquakes were classified as explosions. Notably, the two populations were simply and
834 completely separated by whether the log *P/S* values for each event, averaged over high
835 frequencies such as 10 – 13 Hz, lay above zero (the 12 explosions) or below zero (the 12
836 earthquakes). A second test was to use the MDJ LDF to classify the six earthquakes and four
837 explosions and two unknown events which provided the NECESS Array LDF. All earthquakes
838 were correctly identified, but in this case two of the explosions were misclassified—due, we
839 suggest, to lower-quality waveforms and the unknown nature of some of the events.

840 To summarize: our efforts to use NECESS Array data were frustrated by the quality of
841 data and lack of training events available for obtaining a reliable linear discriminant function.

842 Our objective criterion for identifying the event has been its Mahalanobis score (negative
843 for an earthquake, positive for an explosion), established on the basis of having just two classes
844 of events, as developed via training sets using MDJ data. But then we might ask if there are
845 other types of seismic event, with characteristics that are not obviously to be associated with the
846 events of our Table 3 (explosions) or Table 4 (earthquakes). For example, on the basis of

observed *P/S* spectral ratios, how would we expect to classify an unusually deep small earthquake—or a small explosion that was ripple-fired rather than single-fired, and/or was conducted at or just above the surface rather than at the contained depth of known nuclear explosions in North Korea? Should we expand the number of source types, or expand the characteristics of the two types we have worked with, in order to include the characteristics of such unusual earthquakes and explosions in our training sets? We raise these questions, to make clear that in this paper we are not stating there is rigorous evidence that the seismic event of May 12, 2010, was an earthquake, but rather that it appears to have been an earthquake on the basis of considerably more seismological data—and objective analysis—than has been provided by Zhang and Wen (2015a).

In the course of our work we have been asked if the 12 May 2010 event could be from a decoupled explosion. Consideration of this possibility seems not to be relevant in the present case, at least for decoupled explosions conducted with the goal of reducing signal amplitudes radiated in all directions and thus with the explosive device not near the cavity walls. Such decoupled explosions are found to generate signals with *P/*S spectral ratios even higher than for signals from tamped explosions (Blandford, 1996), and our observation for the May 2010 event is that its *P/*S spectral amplitudes were lower than for tamped explosions. But then: could the event have been a partially-decoupled explosion conducted asymmetrically in its cavity? There is limited practical experience to show that such an explosive source generates *S*-waves more efficiently (hence appearing more earthquake-like) than for a decoupled explosion at the center of a spherical cavity. Stevens and Baker (2009) describe a 1-ton chemical explosion conducted in Kyrgyzstan in 1960, offset from the center of a spherical cavity with radius 4.92 m. Theory and very limited data from this shot indicate *S*-waves comparable to *P*-waves at around 10 Hz.

The *S*-wave source is effectively a dipole, and therefore would be expected to exhibit amplitudes decreasing significantly at low frequency—a feature we do not observe for DBSN data.

Ultimately, the characterization of a small seismic event by a method such as described here may not be deemed to provide sufficient evidence on the nature of a problem event in the context of the Comprehensive Nuclear-Test-Ban Treaty after entry-into-force, but may serve purely as an indicator. In the presence of conflicting indicators, such as from radionuclide and seismic monitoring, the need for clarification may be a basis for requesting an on-site inspection.

Looking ahead, there is the prospect of being able to analyze additional data, but from the Dongbei network, which was deployed for about seven years. When relevant data become available, we plan to apply our evaluation of *P/S* spectral ratios directly to data from the Dongbei network using training sets from earthquakes and explosions derived from the same stations for which we have signals of the 12 May 2010 event. From the Dongbei waveform archive, we would expect to learn whether there are other small seismic events near the North Korea test site. The LDF method we employed in this study has the merit of allowing new training data to be added to existing data, to update the LDF, thus enabling a process of learning from the new data.

Our analysis of discrimination capability in this paper has emphasized an application to one particular event. But our method potentially has wide application. For example, many data centers that prepare catalogs of seismic events, from waveform data acquired by regional or local networks, are concerned to not include explosions from industrial or construction activities, because the catalog may be intended for studies of tectonic activity—that is, it should contain only natural earthquakes. Our discrimination method based on equations (2) or (3) can be very suitable for such an application.

892 We wish to emphasize the very small size of the event in May 2010 studied here. With
893 signals about 3.5 magnitude units smaller than those of nuclear tests in 2013 and 2016, its
894 seismic signals are about three thousand times smaller. The event is on the margins of what can
895 be detected and identified, and is thus useful as an indication of those margins. It is remarkable
896 that significant data exists, enabling our effort to characterize such a very small event. At this
897 low size, we are stretching current monitoring capability to new low levels, below those
898 described for example in the latest report of the US National Academy of Sciences (2012, Figure
899 2-7 and related discussion of monitoring for CTBT compliance).
900 The need to address problem events, has historically been the way in which the
901 monitoring community has been prompted to develop new methods—which, if demonstrated to
902 be successful, have eventually come into operational use. Over time we can expect monitoring
903 capability to improve, both for broad areas and for particular regions of interest.
904

905 ## Summary and Conclusions

906 First, to locate the 12 May 2010 seismic event, we analyzed *Lg*-wave cross-correlation
907 measurements and phase picks from temporary stations to conclude that the event is more than 1
908 km from the 2009 nuclear test and is most likely located between about 4 and 10 km to the
909 southwest of this known explosion in North Korea.
910 Second, in order to assess the nature of the event, we analyzed waveform data from a
911 Global Seismographic Network (GSN) station, MDJ in northeastern China, for events at
912 distances from about 200 to 600 km, including earthquakes and chemical and nuclear explosions,
913 in and near the northern regions of North Korea, to identify a linear discrimination function
914 (LDF) based on measurements at MDJ of *P*-to-*S* spectral ratios (*P/S*) that could most effectively

915 be used to discriminate between small earthquakes and small explosions. We then applied that
916 experience, acquired from station MDJ, to a limited set of seismic signals recorded by the
917 Dongbei Broadband Seismographic Network which included those from two known nuclear
918 explosions in North Korea, and from a small seismic event (M ~ 1.5) on 12 May 2010 near the
919 North Korea test site, reported by Zhang and Wen (2015a), which was claimed by them to be a
920 small nuclear explosion. When our LDF is applied to DBSN data, and to data from stations
921 SMT and NE3C in China, the LDF values measured from *P/S* ratios from known explosions are
922 explosion-like; but for the event of 12 May 2010 the LDF values are earthquake-like for
923 frequencies between 6 and 12 Hz. In the present case, and taking account of signal-to-noise
924 ratios (which degrade at higher frequencies), the band from 6 to 9 Hz appears to provide the best
925 discrimination capability.

926 Thus we find the event of interest to have the characteristics of an earthquake, on the
927 basis of the best seismological data we have been able to obtain—which includes the *P/S* spectral
928 measurements reported by Zhang and Wen (2015a). This is a preliminary conclusion, which we
929 and/others can examine further if additional data from DBSN, and from other stations in the
930 region, become available.

931 Our results suggest that vertical- and three-component *P*/*S* spectral ratios provide an
932 efficient method for classifying earthquakes and explosions in northeastern North Korea and
933 northeastern China down to only a few tons TNT equivalent.

934 From our work, there is at present still no explosive seismic event to associate with the
935 known radionuclide anomalies reported by De Geer (2012). The event first detected and
936 reported by Zhang and Wen (2015a), being found here to have earthquake-like features, appears
937 not to be an appropriate candidate for consideration as a decoupled explosion.

Data and Resources

Waveform data from the Global Seismograph Network (GSN) stations – MDJ (Mudanjiang, China) and INCN (Incheon, Korea), and stations of the NECESS Array were obtained from the Data Management Center of IRIS (www.iris.edu/SeismiQuery/by_station.html –last accessed January 2016). Waveform data from the DBSN were kindly provided by Dr. Kin-Yip Chun, formerly of the University of Toronto. Earthquake catalog data used are acquired from the International Seismological Centre (via www.isc.ac.uk/iscbulletin/search/catalogue/ – last accessed January 2016). Seismographic networks and data centers in the region include the Korea Meteorological Administration (KMA) (www.kma.go.kr/weather/earthquake/domesticlist.jsp –last accessed January 2016), and the Korea Institute of Geology and Mineral Resources (KIGAM) (quake.kigam.re.kr/pds/db/db.html –last accessed January 2016).

Acknowledgments

We thank Kin-Yip Chun for promptly providing us with requested waveform data recorded by the Dongbei stations he operated in 2004–2011, specifically signals from UNEs in 2006 and 2009, and also from the event that is the main subject of this paper. We also thank

961 Francis Wu and Jiakang Xie for background information, and reviewers of our original
962 submission to BSSA for their detailed and helpful comments. Our work was supported by the
963 Defense Threat Reduction Agency under Award Number HDTRA-1-11-1-00027 to Columbia
964 University, and by the National Nuclear Security Administration via a subcontract between
965 Columbia University and the University of Michigan as part of the Consortium on Verification
966 Technology funded by the National Nuclear Security Administration. Support from the German
967 National Data Center at BGR is further acknowledged. This is Lamont-Doherty Earth
968 Observatory contribution number 8039.
969
970
971 # References
972
973 Aki, K., and P.G. Richards (2009). *Quantitative Seismology*, 2nd ed., University Science Books,
974 Sausalito, CA, 700 pp.
975 Blandford, R.R. (1996). Regional Seismic Event Discrimination, in *Monitoring a*
976 *Comprehensive Test Ban Treaty*, edited by E.S. Husebye and A.M. Dainty, Kluwer
977 Academic Publishers (Dordrecht, The Netherlands), 689–720.
978 Bowers, D., P. D. Marshall, and A. Douglas (2001). The level of deterrence provided by data
979 from the SPITS seismometer array to possible violations of the Comprehensive Test Ban
980 in the Novaya Zemlya region. *Geophysical Journal International* **146**, 425-438.
981 Carmichael, J.D., and H. Hartse (2016). Threshold Magnitudes for a Multichannel Correlation
982 Detector, *Bull. Seismol. Soc. Am.*, **106**, 478–498.
983 Chun, K.-Y., Y. Wan, and G.A. Henderson (2009). *Lg* Attenuation near the North Korean

Border with China, Part I: Model Development from Regional Earthquake Sources, *Bull. Seismol. Soc. Am.*, **99**, 3021–3029.

Chun, K.-Y., and G.A. Henderson (2009). *Lg* Attenuation near the North Korean Border with China, Part II: Model Development from the 2006 Nuclear Explosion in North Korea, *Bull. Seismol. Soc. Am.*, **99**, 3030–3038.

Chun, K.-Y., Y. Wu, and G.A. Henderson (2011). Magnitude Estimation and Source Discrimination: A Close Look at the 2006 and 2009 North Korean Underground Nuclear Explosions, *Bull. Seismol. Soc. Am.*, **101**, 1315–1329, doi: 10.1785/0120100202

De Geer, L.-E. (2012). Radionuclide Evidence for Low-Yield Nuclear Testing in North Korea in April/May 2010, *Science & Global Security*, **20**, 1–29.

De Geer, L.-E. (2013). Reinforced evidence of a low-yield nuclear test in North Korea on 11 May 2010, *J. Radioanal. Nucl. Chem.*, **298**, 2075–2083, doi 10.1007/s10967-013-2678-5.

Denny, M.D. and L.R. Johnson (1991). The explosion seismic source function: Models and scaling laws reviewed, in *Explosion Source Phenomenology, AGU Monograph* 65, edited by S.R. Taylor, H. Patton, and P.G. Richards, 1-24.

Ford, S.R. and W.R. Walter (2015). International Monitoring System correlation detection at the North Korean nuclear test site at Punggye-ri with Insights from the Source Physics Experiment, *Seismol. Res. Lett.*, **86**, 1160–1170, doi: 10.1785/0220150029

Hartse, H., S.R. Taylor, W.S. Phillips, and G.E. Randall (1997). A preliminary study of regional seismic discrimination in Central Asia with an emphasis on Western China, *Bull. Seism. Soc. Am.*, **87**, 551–568.

Hartse, H.E., R.A. Flores, and P.A. Johnson (1998). Correcting regional seismic discriminants for path effects in Western China, *Bull. Seism. Soc. Am.*, **88**, 596–608.

1007 Kim, S.G., Y.C. Park and W.Y. Kim (1998). Discrimination of small earthquakes and artificial
1008 explosions in the Korean Peninsula using *Pg/Lg* ratios, *Geophy. J. Int.*, **134**, 267-276.
1009 Kim, W.-Y., D.W. Simpson, and P.G. Richards (1993). Discrimination of earthquakes and
1010 explosions in the eastern United States using regional high-frequency data, *Geophys. Res.*
1011 *Lett.*, **20**, 1507–1510.
1012 Kim, W.-Y., V. Aharonian, A.L. Lerner-Lamm, and P.G. Richards (1997). Discrimination of
1013 earthquakes and explosions in southern Russia Using regional high-frequency three-
1014 component data from the IRIS/JSP Caucasus Network, *Bull. Seismol. Soc. Am.*, **87**, 569–
1015 588.
1016 Kim, W.-Y., and P. G. Richards (2007). North Korean nuclear test: Seismic discrimination at
1017 low yield, *Eos Trans. AGU* **88**, 158.
1018 Pan, C.-Z., P. Jin, and H.-C. Wang (2007). The applicability of *P/S* amplitude ratios for the
1019 discrimination of low magnitude seismic events. *Acta Seismologica Sinica*, **20**, 553–561.
1020 Pasyanos, M.E., and W.R. Walter (2009). Improvements to regional explosion identification
1021 using attenuation models of the lithosphere, *Geophys. Res. Lett.*, **36**, L14304,
1022 doi:10.1029/2009GL038505
1023 Pasyanos, M.E., S.R. Ford, and W.R. Walter (2014). Testing event discrimination over broad
1024 regions using the historical Borovoye Observatory explosion dataset, *Pure Appl.*
1025 *Geophys*. **171,** 523–535.
1026 Patton, H. J. and S. R. Taylor (2011). The apparent explosion moment: Inferences of
1027 volumetric moment due to source medium damage by underground nuclear
1028 explosions, *J. Geophys. Res.,* **116**, B03310.
1029 Schaff, D.P. (2008). Semi-empirical statistics of correlation detector performance, *Bull. Seismol.*

Soc. Am., **98**, 1495-1507, doi: 10.1785/0120060263.

Schaff, D. P. (2009). Broad-scale applicability of correlation detectors to China seismicity, *Geophys. Res. Lett*., **36**, L11301, doi:10.1029/2009GL038179.

Schaff, D.P., W.-Y. Kim, and P.G. Richards (2012). Seismological Constraints on Proposed Low-Yield Nuclear Testing in Particular Regions and Time Periods in the Past, with Comments on "Radionuclide Evidence for Low-Yield Nuclear Testing in North Korea in April/May 2010" by Lars-Erik De Geer, *Science & Global Security*, **20**, 155–171.

Schaff, D.P., and P.G. Richards (2004). *Lg*-wave cross correlation and double difference location: application to the 1999 Xiuyan, China, sequence, *Bull. Seismol. Soc. Am.*, **94**, 867–879.

Schaff, D. P., and P. G. Richards (2011). On finding and using repeating seismic events in and near China, *J. Geophys. Res*. **116**, doi: 10.1029/2010JB007895.

Schaff, D.P., and F. Waldhauser (2010). One magnitude unit reduction in detection threshold by cross-correlation applied to Parkfield (California) and China seismicity, *Bull. Seismol. Soc. Am.*, **100**, 3224–3238, doi: 10.1785/0120100042

Schaff, D.P., P.G. Richards, M. Slinkard, S. Heck, and C. Young (2016). *Lg*-wave Cross-Correlation and Epicentral Double-Difference Location in and near China, ms. in preparation for submission to the *Bull. Seism. Soc. Amer.*

Schaff, D.P., and F. Waldhauser (2005). Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seismol. Soc. Am.*, **95**, 2446-2461, doi:10.1785/0120040221.

Seber, G. A. F. (1984). *Multivariate Observations*, Wiley, New York, 686 pp.

Song, J., E. A. Hetland, F. T. Wu, X. Zhang, G. Liu, and Z. Yang (2007). P-wave velocity structure under the Changbaishan volcanic region, NE China, from wide-angle reflection

and refraction data, *Tectonophys*., **433**, 127-139.

Stevens, J.L., and G.E. Baker (2009). Seismic wave generation by a nonistropic explosion source, *J. Geophys. Res.*, **114**, B12302, doi:10.1029/2008JBoo5965.

Tao, K., F. Niu, J. Ning, Y. Chen, S. Grand, H. Kawakatsu, S. Tanaka, M. Obayashi, and J. Ni (2014). Crustal structure beneath NE China imaged by NECESS Array receiver function data, *Earth Planet. Sci. Lett.*, **298**, 48–57, doi: 10.1016/j.epsl.2014.04.043

Taylor, S. (1996). Analysis of high-frequency Pg/Lg ratios from NTS explosions and western U.S. earthquakes, *Bull. Seism. Soc. Am*., **86**, 1042–1053.

Taylor, S., A. Velasco, H. Hartse, W.S. Phillips, W.R. Walter, and A. Rodgers (2002). Amplitude corrections for regional discrimination, *Pure. App. Geophys*., **159**, 623–650.

US Congress (1988). *Seismic Verification of Nuclear Testing Treaties*, Office of Technology Assessment, OTA-ISC-361, US Government Printing Office.

US National Academy of Sciences (2002). Technical Issues Related to the Comprehensive Nuclear Test Ban Treaty, Committee on Technical Issues Related to Ratification of the Comprehensive Nuclear Test Ban Treaty, 96 pp. Available from http://www.nap.edu/catalog/10471.html

US National Academy of Sciences (2012). The Comprehensive Nuclear test Ban Treaty: Technical Issues for the United States, Committee on Reviewing and Updating Technical Issues Related to the Comprehensive Nuclear Test Ban Treaty, 204 pp. Available from http://www.nap.edu/catalog.php?record_id=12849

Waldhauser, F., and D.P. Schaff (2008). Large-scale cross-correlation based relocation of two decades of Northern California seismicity, *J. Geophys. Res.*, **113**, B08311, doi:10.1029/2007JB005479

Walter, W.R. and S.R. Taylor (2001). A revised magnitude and distance amplitude correction (MDAC2) procedure for regional seismic discriminants: theory and testing at NTS, LLNL UCRL-ID-146882.

Walter, W. R., K. Mayeda, and H. J. Patton, (1995). Phase and spectral ratio discrimination between NTS earthquakes and explosions Part 1: Empirical observations, *Bull. Seism. Soc. Am.*, **85**, 1050-1067.

Wotawa, G. (2013). Meteorological analysis of the detection of xenon and barium/lanthanum isotopes in May 2010 in eastern Asia, *J. Radioanal. Nucl. Chem.*, **296**, 339-347.

Wright, C.M. (2013). Low-Yield Nuclear Testing by North Korea in May 2010: Assessing the evidence with Atmospheric transport models and Xenon activity calculations, *Science & Global Security*, **21**, 3–52.

Zhang, M., and L. Wen (2015a). Seismological Evidence for a Low-Yield Nuclear Test on 12 May 2010 in North Korea, *Seismol. Res. Lett.*, **86**, 1–8, doi: 10.1785/02201401170.

Zhang, M., and L. Wen (2015b). An effective method for small event detection: match and locate (M & L), *Geoph. J. Int.*, **200**, 1523–1537, doi:10.1093/gji/ggu466.

Zhao, L. F., X. B. Xie, W. M. Wang, and Z. X. Yao (2012). Yield estimation of the 25 May 2009 North Korean nuclear explosion, *Bull. Seismol. Soc. Am.*, **102**, 467–478.

Full Mailing Address for each Author

Won-Young Kim

Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, NY 10960, USA

wykim@LDEO.columbia.edu

Paul G. Richards

Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, NY 10960, USA

richards@LDEO.columbia.edu

David P. Schaff

Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, NY 10960, USA

dschaff@LDEO.columbia.edu

Karl Koch

Federal Institute for Geosciences and Natural Resources (BGR), Stilleweg 2, D-30655 Hannover, Germany

karl.koch@bgr.de

Tables

Table 1. Key detection values and probabilities for a 2010 event on the Dongbei network

Station	CC	t_{peak} (s)	t_{lag} (s)	p
DBN08	0.08	0.18	46.46	0.0039
DBN06	0.06	0.19	10.74	0.017
DBN17	0.16	0.68	450	0.0015
DBN14	0.14	–1.10	450	0.0024
DBN15	0.08	–1.37	24.96	0.055

Note: the combined probability of obtaining these results by chance for all stations is the product of the five separate p values, namely 1.3 E -11.

Table 2. Result of cross-correlation detector using 2009 UNT signals as templates.

P arrival time 2010-05-12	Station	Channel	Magnitude (±1 s.d.)	Template		
				2009-05-25	Station	Magnitude
00:09:10.761	DB08	HHZ†	1.66	00:55:09.121	DB08	4.53
00:09:14.285	DB06	HHZ	1.29	00:55:12.095	DB06	4.53
00:09:16.710	DB17	HH3¶	1.50(0.17)*	00:55:14.550	DB17	4.53
00:09:17.405	DB10	HH3	1.51(0.04)§	00:55:15.105	DB10	4.53

¶ HH3 = Three components used: vertical, NS and EW; †HHZ = only vertical template was available; * Instrument gain was a factor of 4 smaller for the May 2009 explosion than for the event of May 2010; § Instrument gain was a factor of 20 greater for the May 2010 event.

Table 3. Explosions analyzed.

id	Date (year-mo-dy)	Time (hh:mm:s)	Lat (°N)	Long (°E)	Mag (M_L)	Explosion type/ Information source
1	2006-10-09	01:35:28.00	41.287	129.108	3.93	UNE*/Z&W2015§
2	2009-05-25	00:54:43.18	41.294	129.082	4.53	UNE/Z&W2015
3	2013-02-12	02:57:51.33	41.291	129.076	4.89	UNE/Z&W2015
4	1998-08-12	15:00:08.10	42.865	128.223	1.0	SHCE†/Wu
5	1998-08-18	14:00:06.69	42.914	129.324	2.0	SHCE/Wu
6	1998-08-19	15:00:07.79	42.091	128.739	1.9	SHCE/Wu
7	1998-08-25	15:00:07.46	42.427	126.748	1.0	SHCE/Wu
8	2010-01-15	06:18:01.44	41.7488	126.9143	2.9	CE‡/ISC
9	2011-01-05	05:46:05.66	41.7317	126.9674	2.8	CE/ISC, KIGAM
10	2011-02-18	15:25:58.15	41.7345	126.8917	3.5	CE/ISC
11	2011-05-19	09:38:21.58	42.2512	129.3803	2.6	CE/ISC
12	2012-01-21	07:54:45.59	42.2306	129.3680	2.6	CE/ISC
13	2010-05-12	00:08:45.07	41.286	129.079	1.44	Claimed UNE/Z&W2015
14	2016-01-06	01:30:01.4	41.305	129.039	5.1	UNE/PDE

*UNE = underground nuclear explosion; § Z&W2015, Zhang and Wen (2015a);

†SHCE = single hole chemical explosion; ‡CE = chemical explosion.

Table 4. Selected earthquakes analyzed.

id	Date (year-mo-dy)	Time (hh:mm:s)	Lat (°N)	Long (°E)	Depth (km)	Mag (M_L)	Agency
1	1994-01-25	08:51:38.2	42.23	127.12	04	4.0	NK
2	2004-12-16	18:59:14.5	41.79	127.94	10	4.0	PDE
3	2007-12-31	21:33:38.0	40.41	127.25	0	3.2	KMA
4	2009-08-05	12:08:12.6	42.349	127.223	10	3.8	ISC
5	2010-05-18	04:08:10.3	42.83	125.96	10	3.7	Z&W2015
6	2010-10-09	05:45:14.7	42.352	128.388	10	3.4	ISC
7	2010-10-09	06:07:09.2	42.370	128.420	5	3.6	ISC
8	2010-11-07	11:11:39.74	40.062	128.199	18	3.5	KIGAM
9	2010-11-12	02:10:44.8	43.00	125.89	7	2.8	Z&W2015
10	2011-06-09	01:10:35.1	42.44	127.19	6	3.3	Z&W2015
11	2011-12-26	13:34:08.6	42.381	127.246	0	3.6	ISC
12	2014-08-04	21:16:36.0	40.110	127.200	0	2.5	KMA

Earthquakes 5, 9, and 10 were used by Zhang and Wen (2015a)

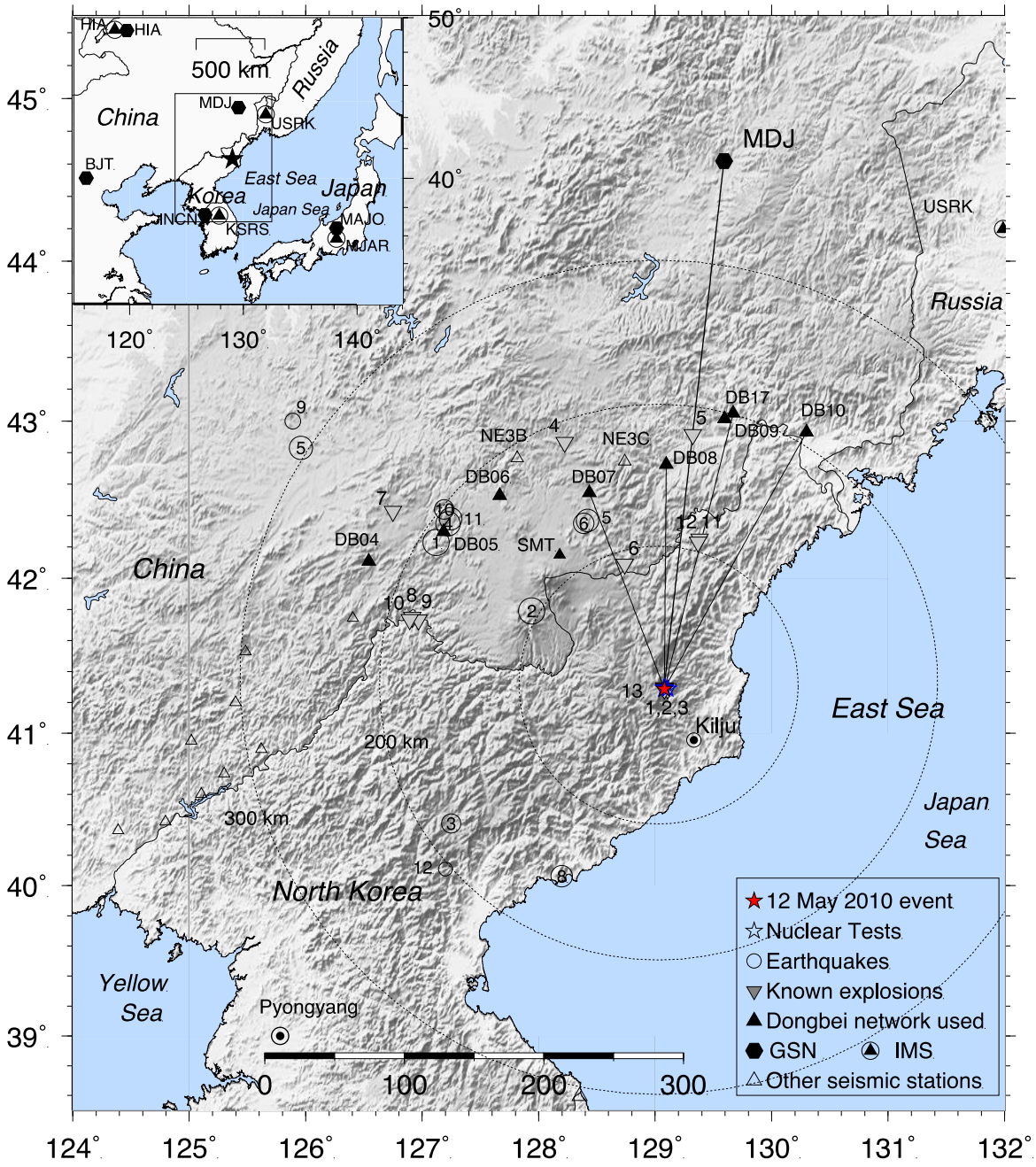

Figure 1. A map showing the location of North Korean nuclear tests (open stars), earthquakes (open circles), single-hole explosions and large industrial explosions (inverted triangles), seismographic stations (triangles), in the northeastern Korean peninsula and in northeast China. Dotted circles concentric with the North Korean test site have radii 100, 200, and 300 km.

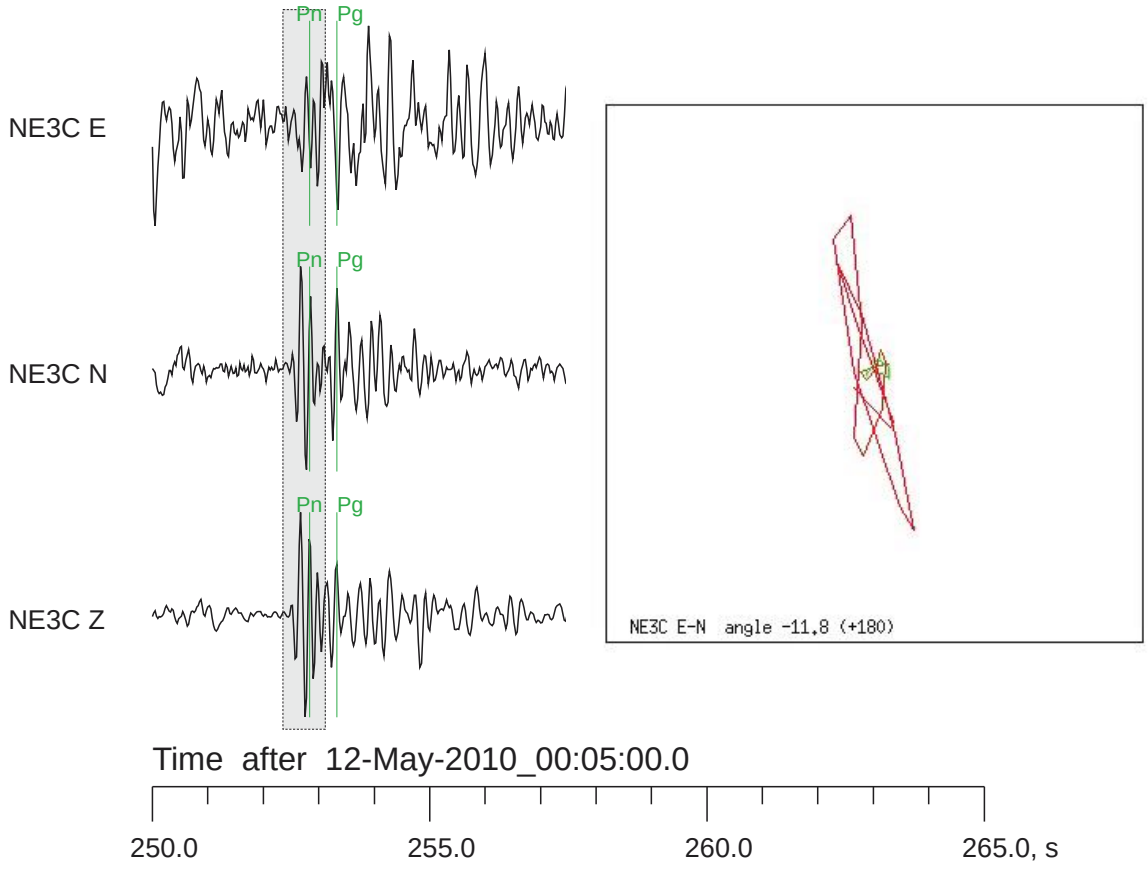

Figure 2. Particle-motion display of the horizontal components of the *P*-wave arriving at station NE3C. The time window used for this analysis is shown on the background seismograms.

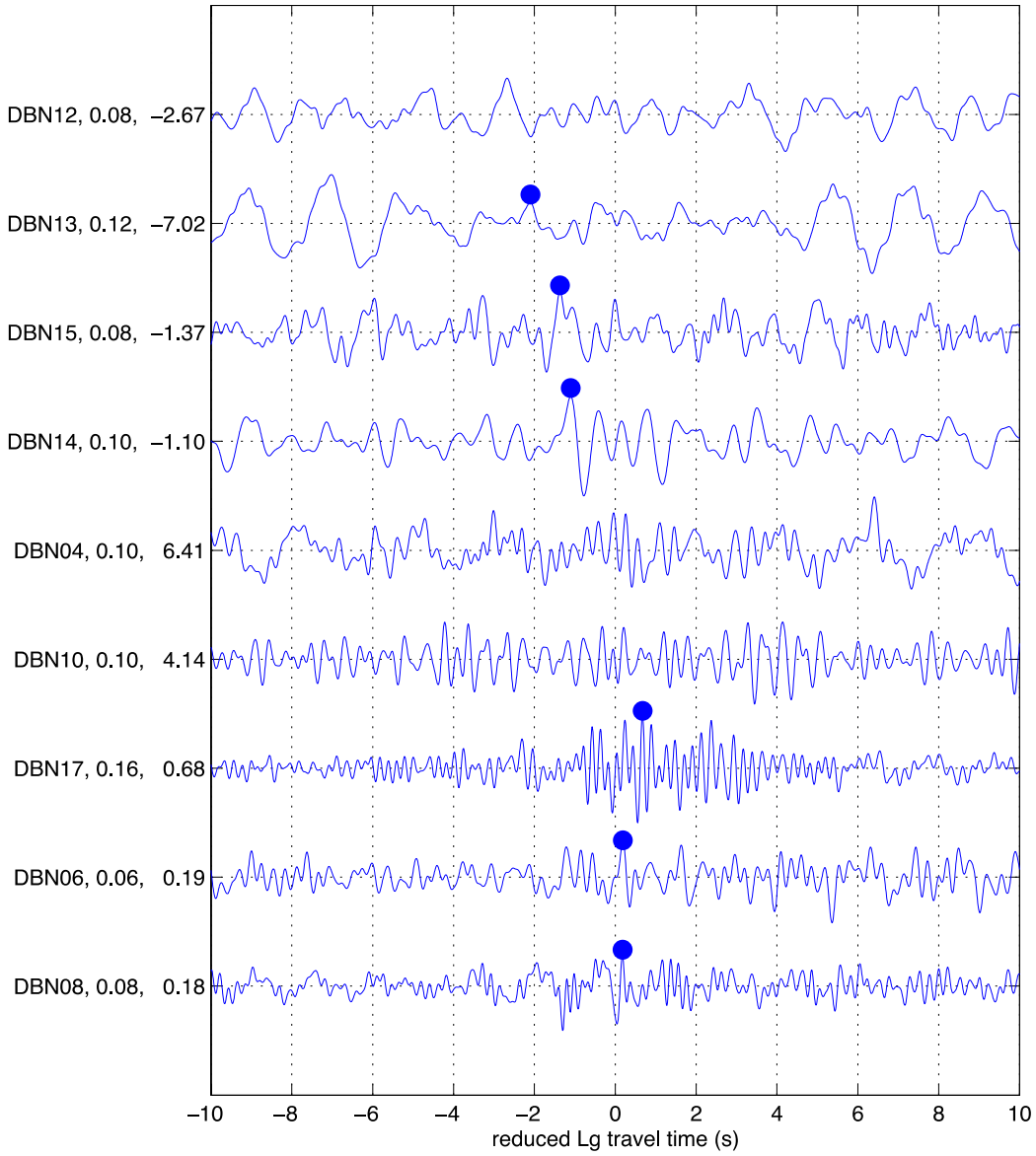

Figure 3

Figure 3. *Lg*-wave cross-correlation traces are shown for DBSN stations, for continuous data from 2010 using master template windows 50 s long from the 2009 nuclear test. These plots show lags searched forwards and backwards 10 s from the expected arrival of *Lg* based on the 2010 hypocenter estimate given by Zhang and Wen (2015a). The data have been filtered from 0.5 to 5 Hz. The correlation traces were averaged for all three components. Stations are in order of increasing distance from bottom to top of the figure. The labels on the vertical axis have the station name, the peak correlation coefficient from the average of three components, and the lag time of the peak in seconds. The peaks of the correlation traces used for the location are shown with large dots.

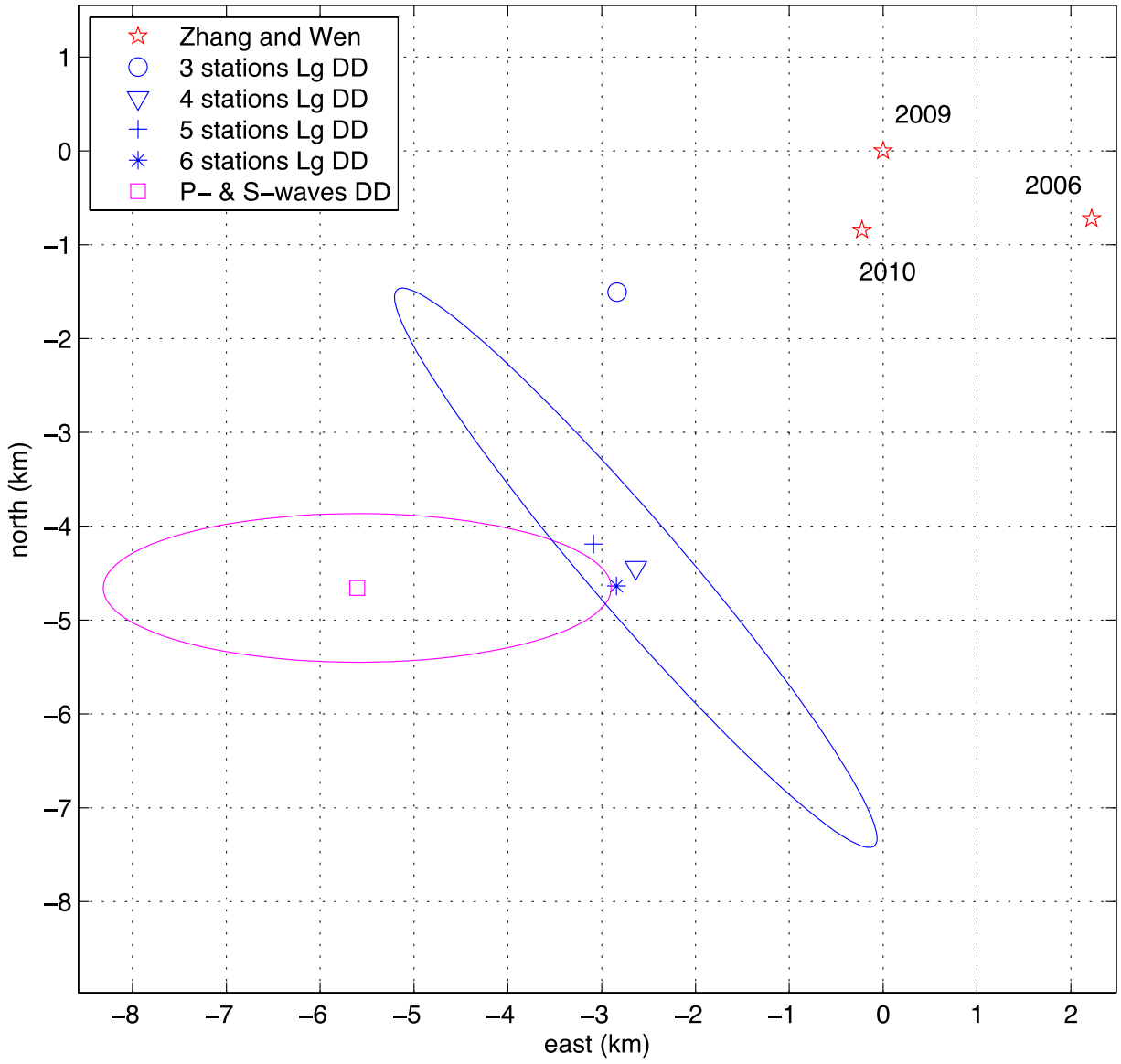

Figure 4

Figure 4. Location estimates for the 2010 seismic event, using different data sets and double-difference (DD) methods based first on *Lg* cross correlation, and second on phase pick data for *P*- and *S*-waves. The local coordinate system centers the origin on the location of the 2009 nuclear test.

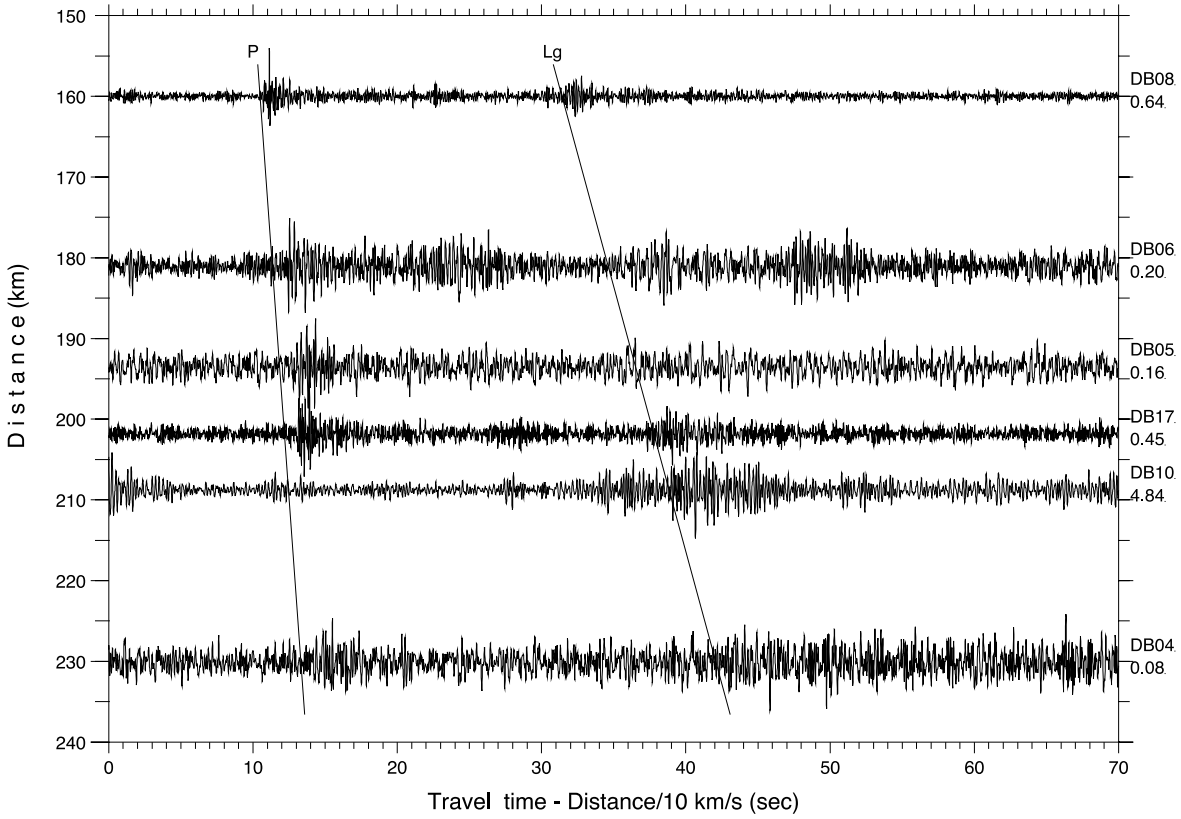

Figure 5. Vertical component seismic record section of DBSN stations within about 230 km from the location of the 12 May 2010 event reported by Zhang and Wen (2015). Records show *P*– and *Lg*–wave arrivals consistent with the reported origin time and location. Signals at stations DB08 and DB17 are the most unambiguous. Peak signal amplitude (in micrometers/s) is given at the end of each trace.

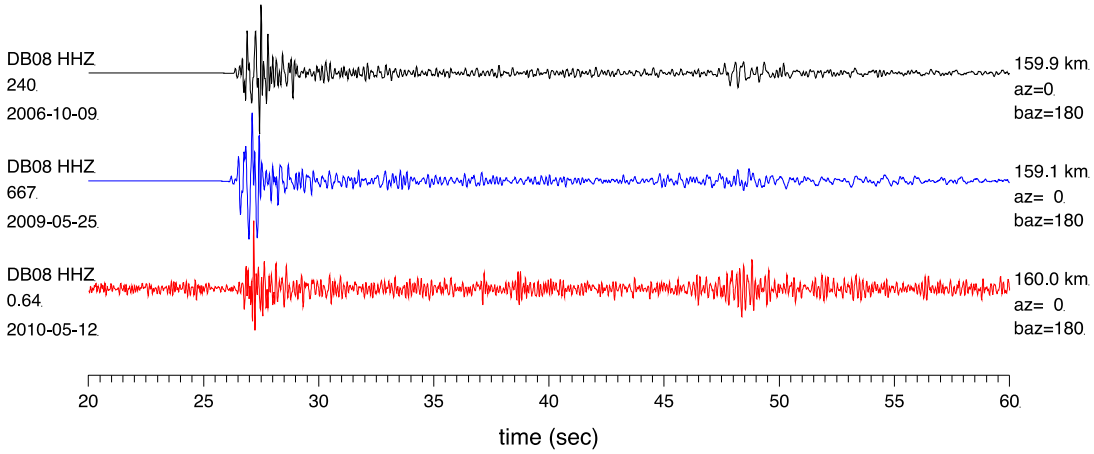

Figure 6a. Vertical records at DBSN station DB08 from the UNE of 9 October 2006 (*black trace*), the UNE of 25 May 2009 (*blue trace*), and from the 12 May 2010 event (*red trace*). Data are passed in the band from 0.8 to 10 Hz, and plotted with equal *P*–wave amplitude. The time-scale origin is the event origin provided by Zhang and Wen (2015a). Note that the relative amplitude of the *Lg*–wave arriving between 46 and 50 s is significantly stronger for the latter event. Peak signal amplitude (in micrometers/s) is given at the beginning of each trace.

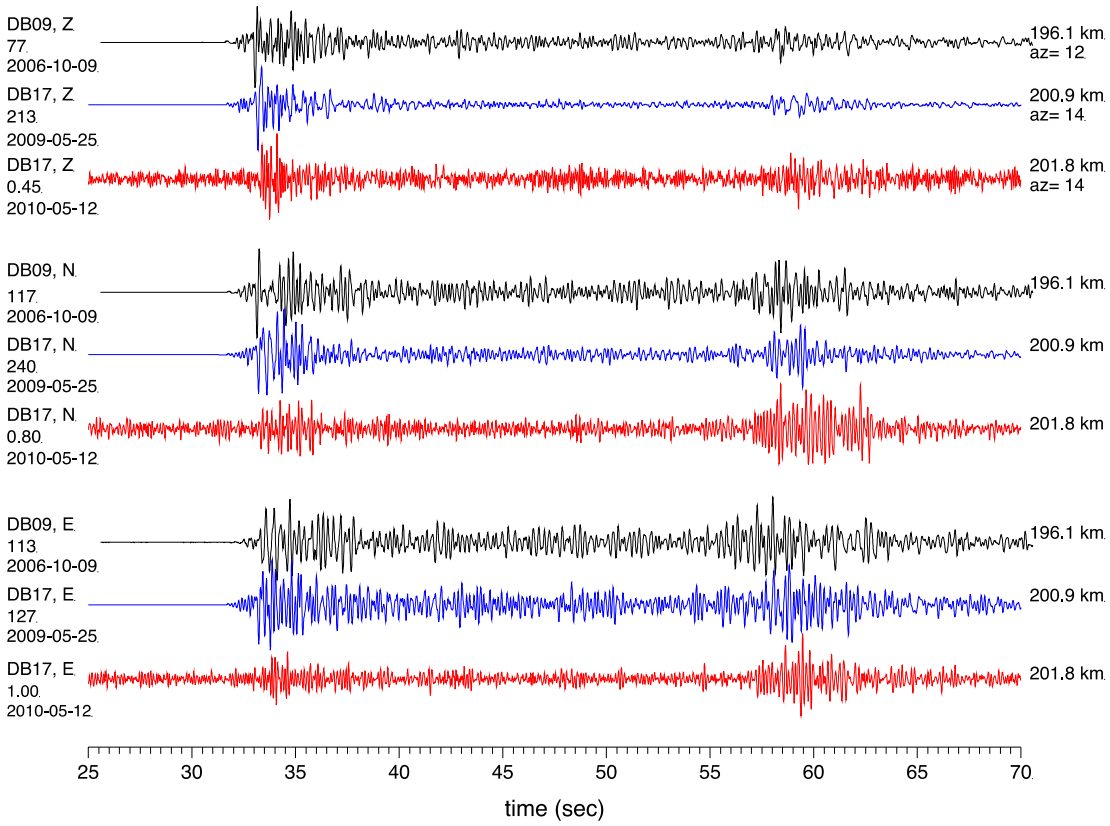

Figure 6b. Three-component records at station DB17 from two events: the UNE of 2009 (*blue traces*), and the event of 12 May 2010 (*red traces*). This station, installed in 2007, had operated earlier as station DB09 at a nearby location, and the three-component records at DB09 are shown from the UNE of 2006 (*black traces*) delayed by about 0.5 s to align with *P*–waves on DB17 records. Each trace is plotted to the same scale for its maximum amplitude (which is given, below each station code, in micrometers/s). Note that the *P* to *Lg* amplitude ratios for these time-domain ground motions from the 2010 event are lower than those from the known UNEs on all three components.

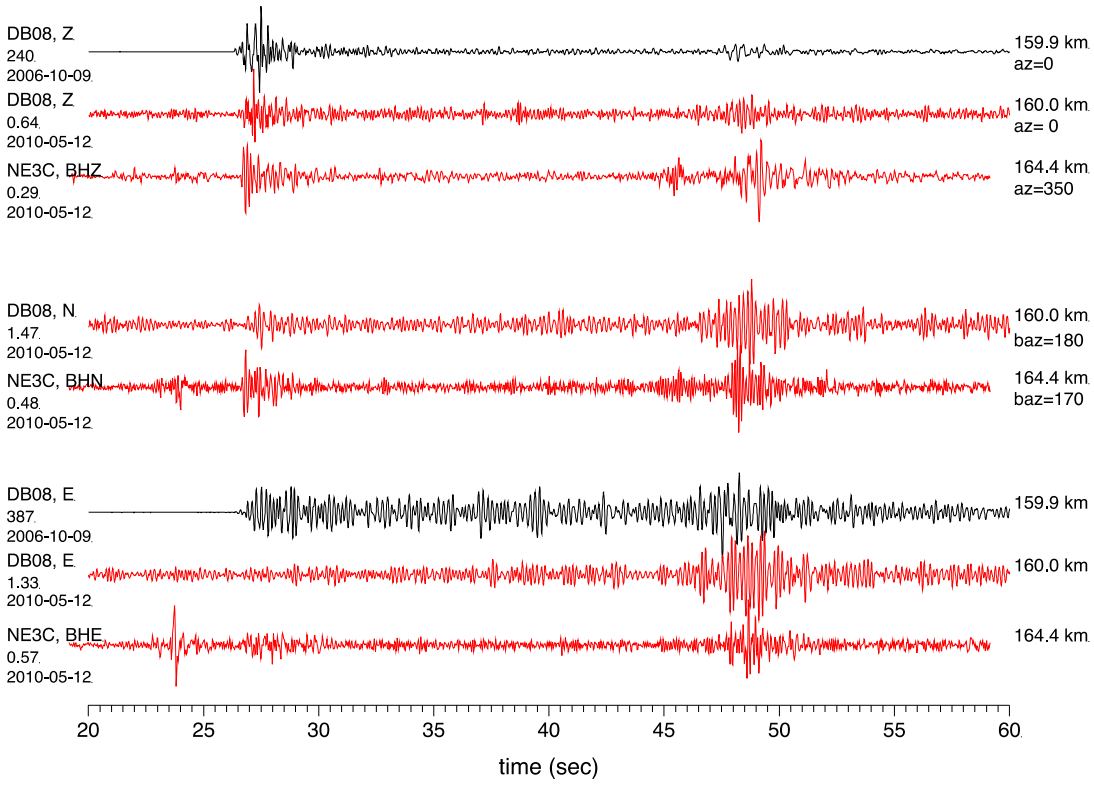

Figure 7. Three-component records at stations DB08 and NE3C from the UNE of 2006, and from the 12 May 2010 event (but the NS-component record is missing on DB08 from this UNE). *P*– and *S*–wave arrival times are generally consistent on records at NE3C and DB08 from May 2010 events; the *P*-wave amplitudes at NE3C are relatively stronger than those on DB08 records. Traces are raw records filtered between 0.8 and 10 Hz, and correspond to ground velocity. Their peak amplitude (in micrometers/s) is given below the station code.

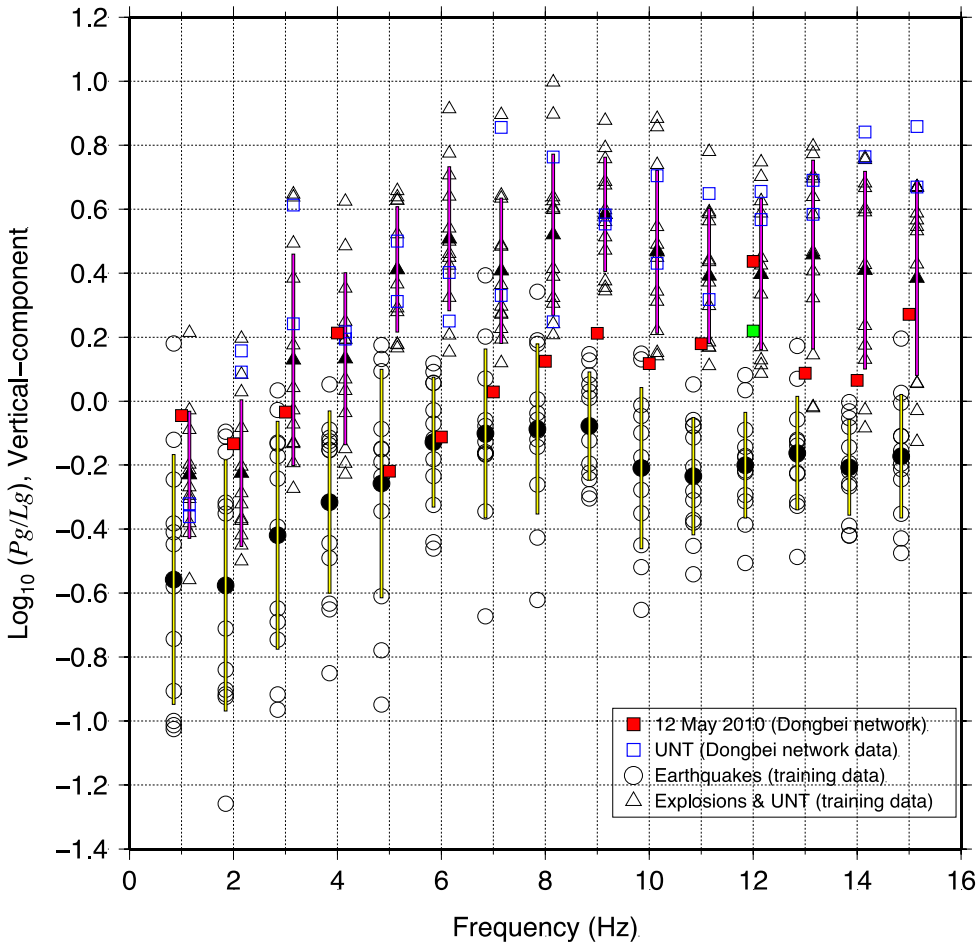

Figure 8. Vertical-component $\log_{10}(Pg/Lg)$ spectral amplitude ratios at discrete frequencies between 1 and 15 Hz used in discrimination analysis are plotted for 12 earthquakes (*circles*), and for 12 explosions (*triangles*) comprised of 4 single-hole explosions, 5 large industrial explosions and 3 known nuclear tests. The data for the earthquakes and chemical explosions are from MDJ, but two underground explosions (2006 and 2009; *blue open squares*), and 12 May 2010 event (*red squares*) are from Dongbei network data (see Figure 1) — stations DB08 and DB17 are used for the analysis. A mean spectral value for each discrete frequency point is plotted both for earthquakes (*black circle*) and for explosions (*black triangle*), with a bar representing the scatter (±1 S.D.). Note that the spectral ratio at 12 Hz of Dongbei network data is less reliable due to very low SNR of *Lg* arrival at DB08, and the *Lg* amplitude at DB17 at 15 Hz is lower than the ambient noise. The *P/S* ratio at 12 Hz excluding DB08 measurement is plotted with a *green square* for comparison.

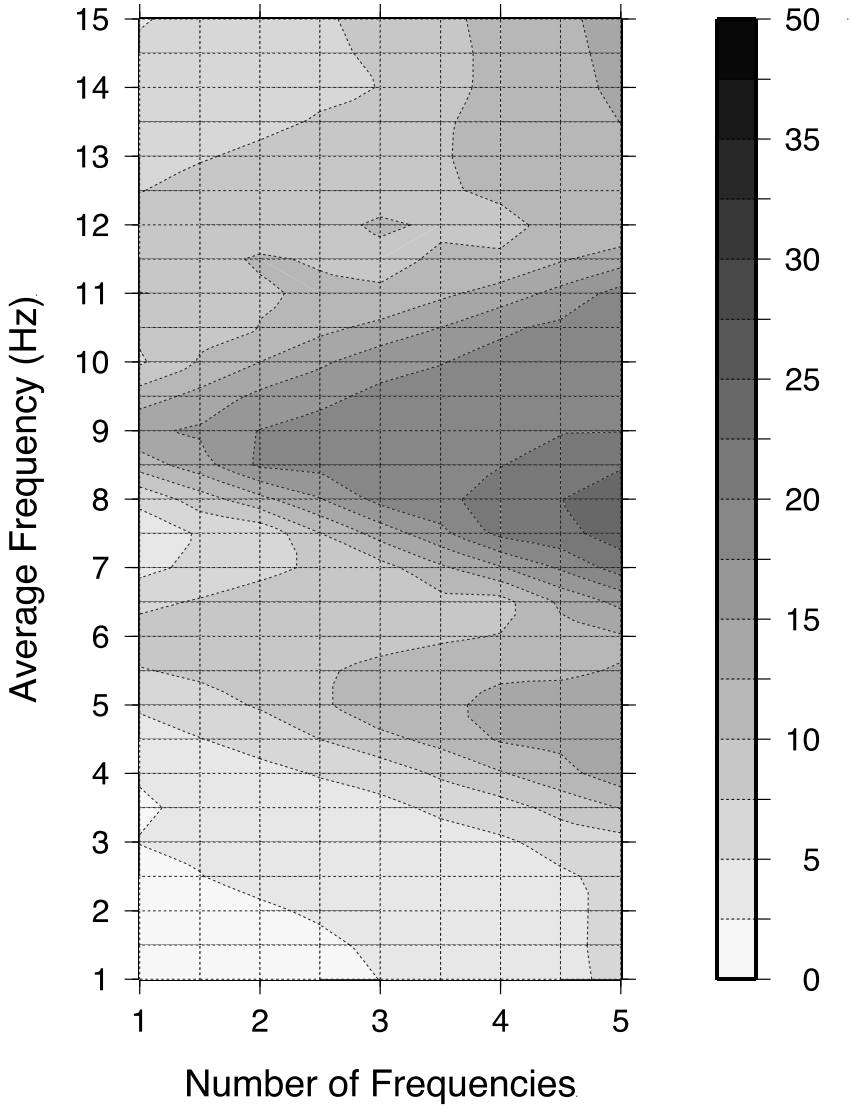

Figure 9. The discrimination power of vertical-component *P/S* spectral ratios—observed at MDJ—at discrete frequencies between 1 and 15 Hz is shown here in terms of the Mahalanobis distance between the multivariate means of the two populations — earthquakes and explosions. The horizontal axis (number of frequencies) represents choices from 1 to 5 discrete frequencies, used for LDF analysis to obtain the Mahalanobis distance and associated misclassification probability. The vertical axis is the average frequency for each run. For example, if choosing four parameters, discrete frequencies 6, 7, 8, 9 Hz are taken for LDF analysis (with average 7.5 Hz), then 7–10 Hz (7, 8, 9 and 10 Hz with average 8.5 Hz)), and 8–11 Hz, and so on.

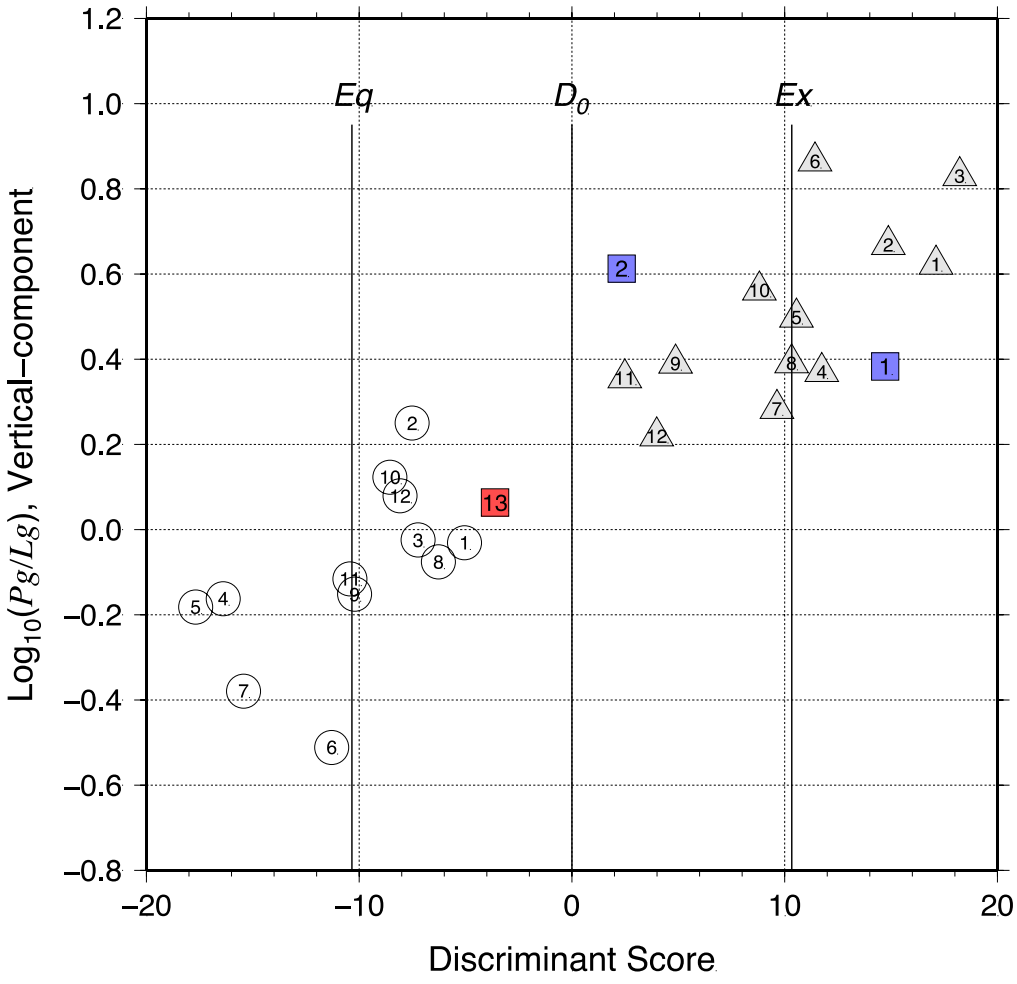

Figure 10. LDF analysis of the vertical *P/S* ratio from 12 earthquakes (*circles*) and 12 explosions (*shaded triangles*) by using four frequencies: 6, 7, 8, and 9 Hz. All the training data—earthquakes and explosions—are correctly classified; the Mahalanobis distance is $\Delta^2 = 20.6$, and the total misclassification probability is 1.15%. Discriminant scores of 12 earthquakes and 12 explosions of the training data are plotted with their mean $\log_{10}(Pg/Lg)$ ratios. Note that the two populations are also separated by a mean vertical-component $\log_{10}(Pg/Lg)$ ratio of about 0.2. Vertical lines denoted as *Eq* and *Ex* are the projection of the multivariate means of the earthquake and explosion population, respectively. The vertical line D_0 is the classification line. Using the linear discriminant function obtained for MDJ, the two UNEs (*blue squares*) for which we have signals recorded at DBSN are correctly classified; and the 12 May 2010 event is classified as an earthquake (*red square*). Event ids are as listed in Tables 3 & 4.

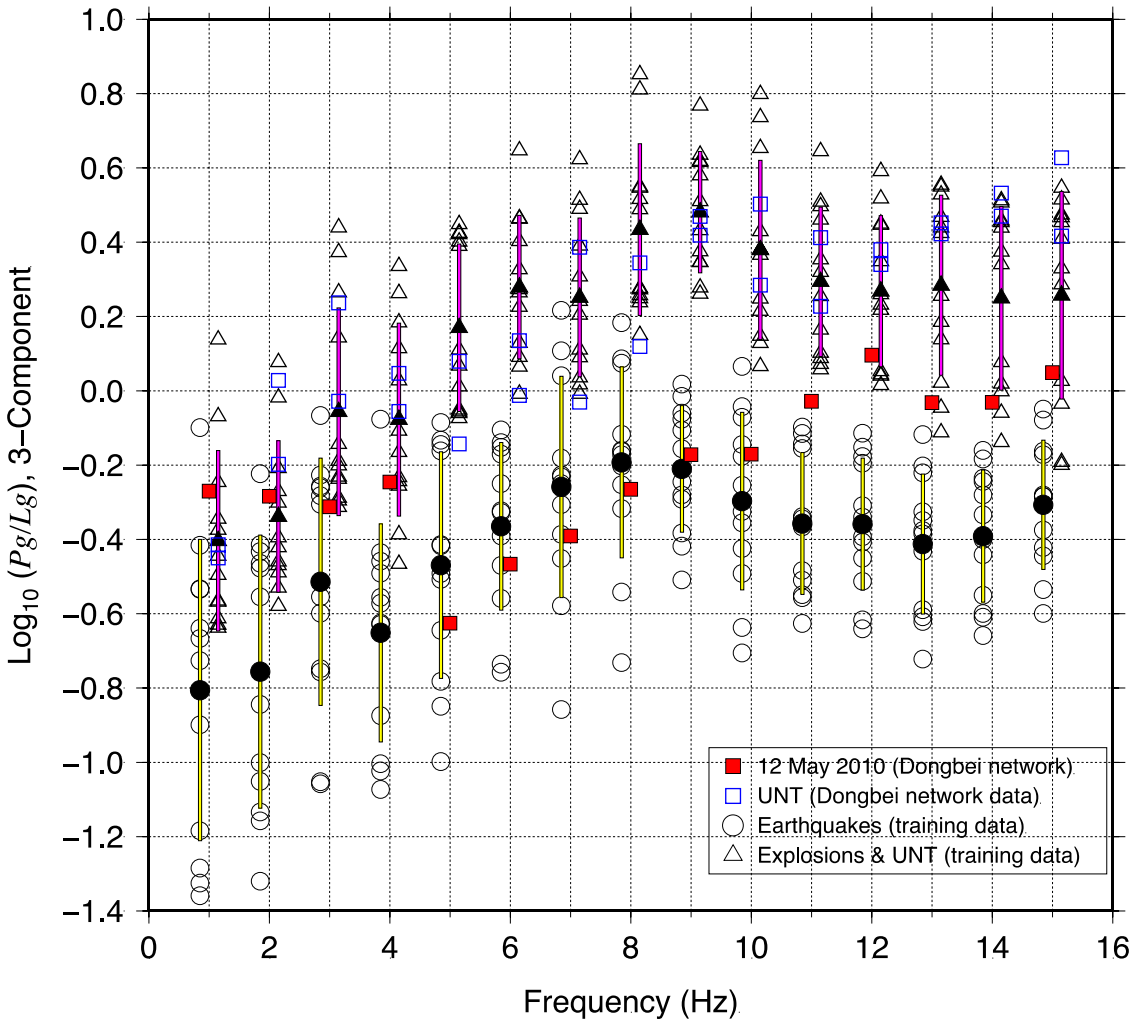

Figure 11. Similar to Figure 8 but now using three-component $\log_{10}(Pg/Lg)$ spectral amplitude ratios. The spectral ratio data for the earthquakes and chemical explosions are from MDJ, but two underground explosions (2006 and 2009; *blue squares*), and the 12 May 2010 event (*red square*) are from Dongbei network data (stations DB07, DB08, DB09 and DB17). A mean spectral value for each discrete frequency point is plotted both for earthquakes (*black circle*) and for explosions (*black triangle*), with a colored bar representing the scatter (±1 S.D.). This use of three-component data achieves better separation of the explosion and earthquake populations than was found in Figure 8, which was based on vertical component data only.

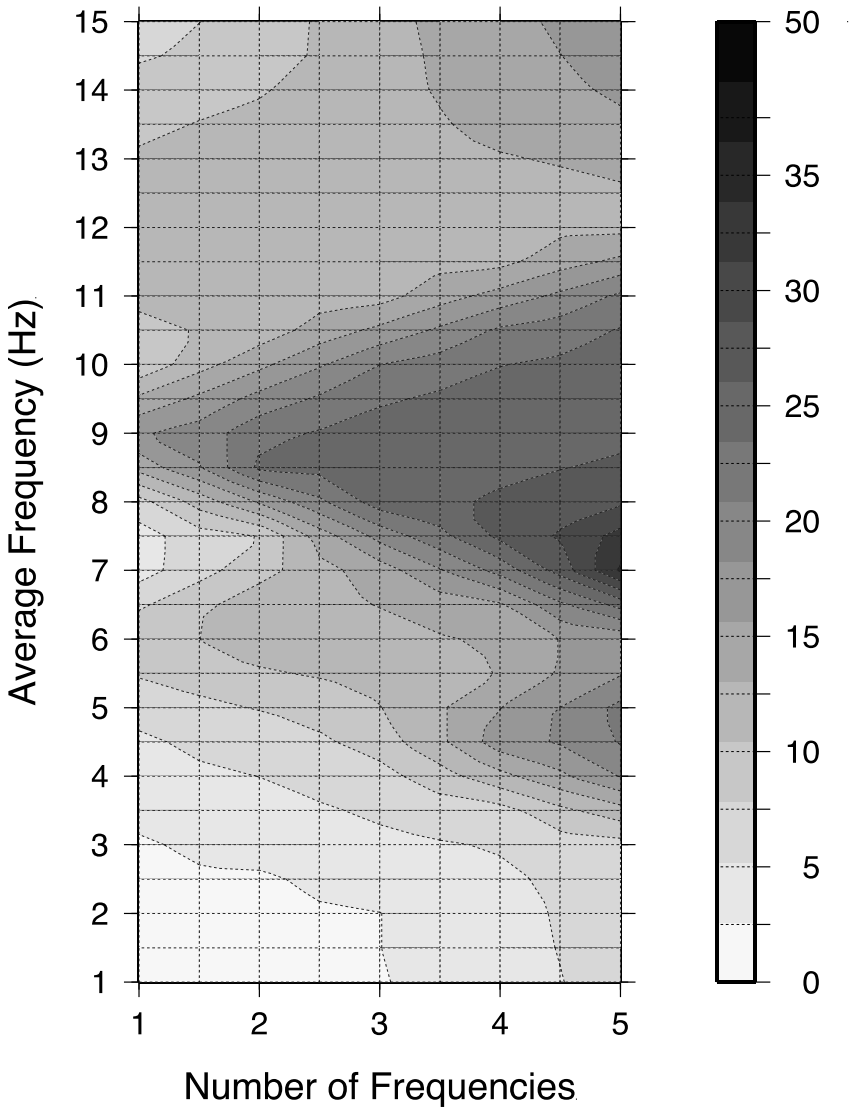

Figure 12. Similar to Figure 9 but now based on three-component data. The goal of maximizing the Mahalanobis distance between the multivariate means of the explosion and earthquake populations, appears to be best achieved by working with frequencies between about 6 and 12 Hz, since this range carries the most discrimination power in *P/S* ratio space for MDJ.

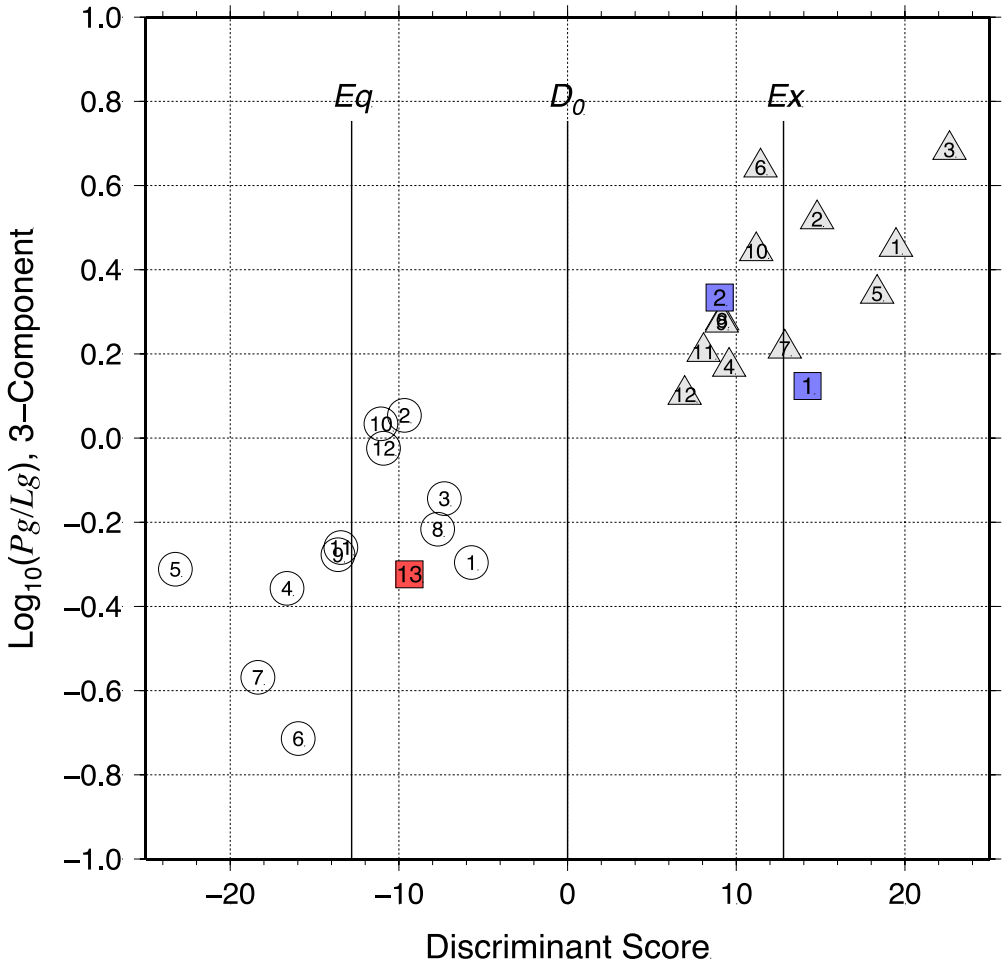

Figure 13. Similar to Figure 10, but now the linear discriminant function analysis is done with 3-component (*Pg/Lg*) spectral ratios at four discrete frequencies between 6 and 9 Hz. The discriminant scores of 12 earthquakes (*circles*) and 12 explosions (*shaded triangles*) of the training data are plotted with their mean $\log_{10}(Pg/Lg)$ ratios. Note that the two populations are also separated by mean 3-component $\log_{10}(Pg/Lg)$ ratio ~ 0.1. [Vertical lines denoted as *Eq* and *Ex* are the projection of the multivariate means of the earthquake and explosion populations, respectively, now separated by a greater Mahalanobis distance, $\Delta^2 = 25.6$, than was the case with vertical component data only. D_0 is the vertical classification line.] The total misclassification probability is 0.57%. Using the linear discriminant function obtained for MDJ, two UNTs (*blue squares*) recorded at DBSN are correctly classified, and the 12 May 2010 event is classified as an earthquake (*red square*).

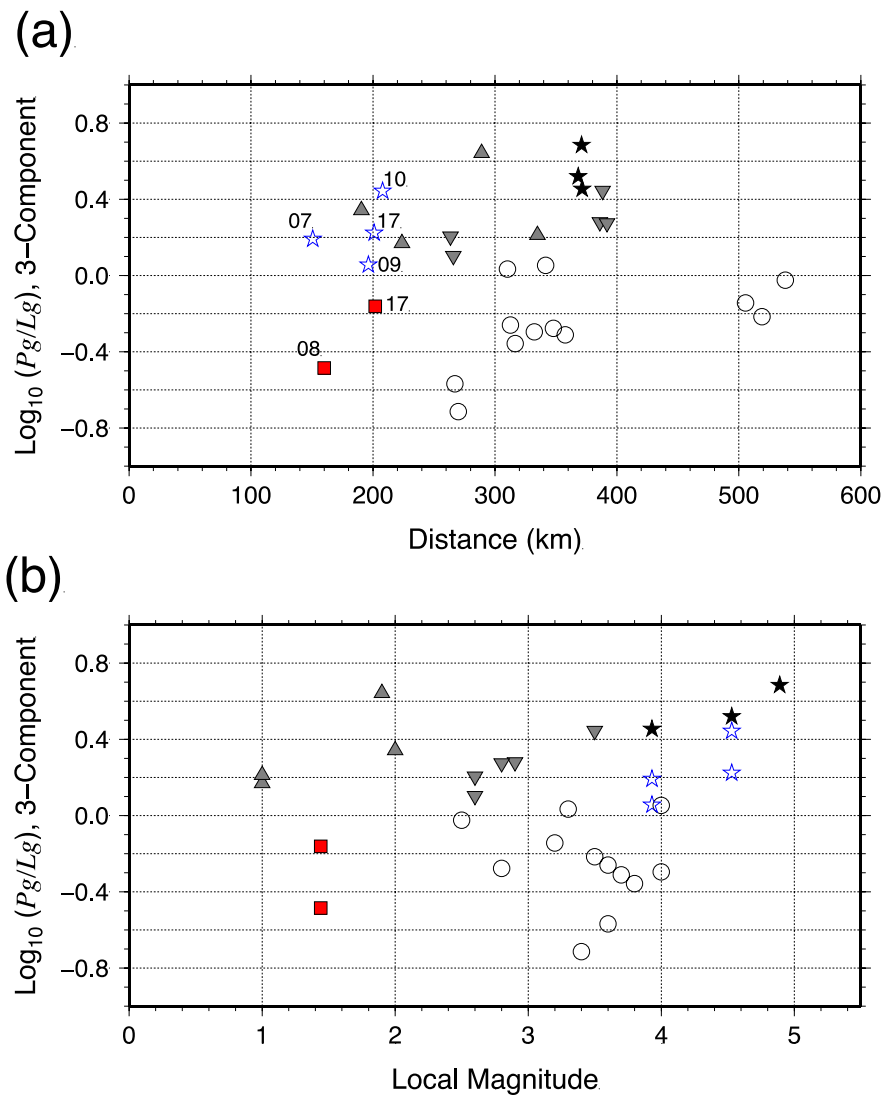

Figure 14. (a) The mean $\log_{10}P/S$ values between 6 and 9 Hz measured from three-component records are plotted against epicentral distance. *P/S* ratios are not corrected for distance. Earthquakes (*circles*) and different types of explosions; chemical explosions (*inverted triangles*), single-hole shots (*triangles*), UNEs (*black star*), UNEs recorded at DBSN (*blue stars* at stations DB07, DB09, DB10 and DB17); and the claimed event on 12 May 2010 (*red squares* at DB08 and DB17). *P/S* ratios of the earthquakes and explosions show separation along $\log_{10}(Pg/Lg)$ = ~ 0.1. (b) The *P/S* ratios from three-component records are plotted against event magnitude. There is no clear correlation of *P/S* ratios with magnitude, within the different types of seismic event, and we have chosen not to make any magnitude correction to the measured ratios prior to applying our methods of event classification.

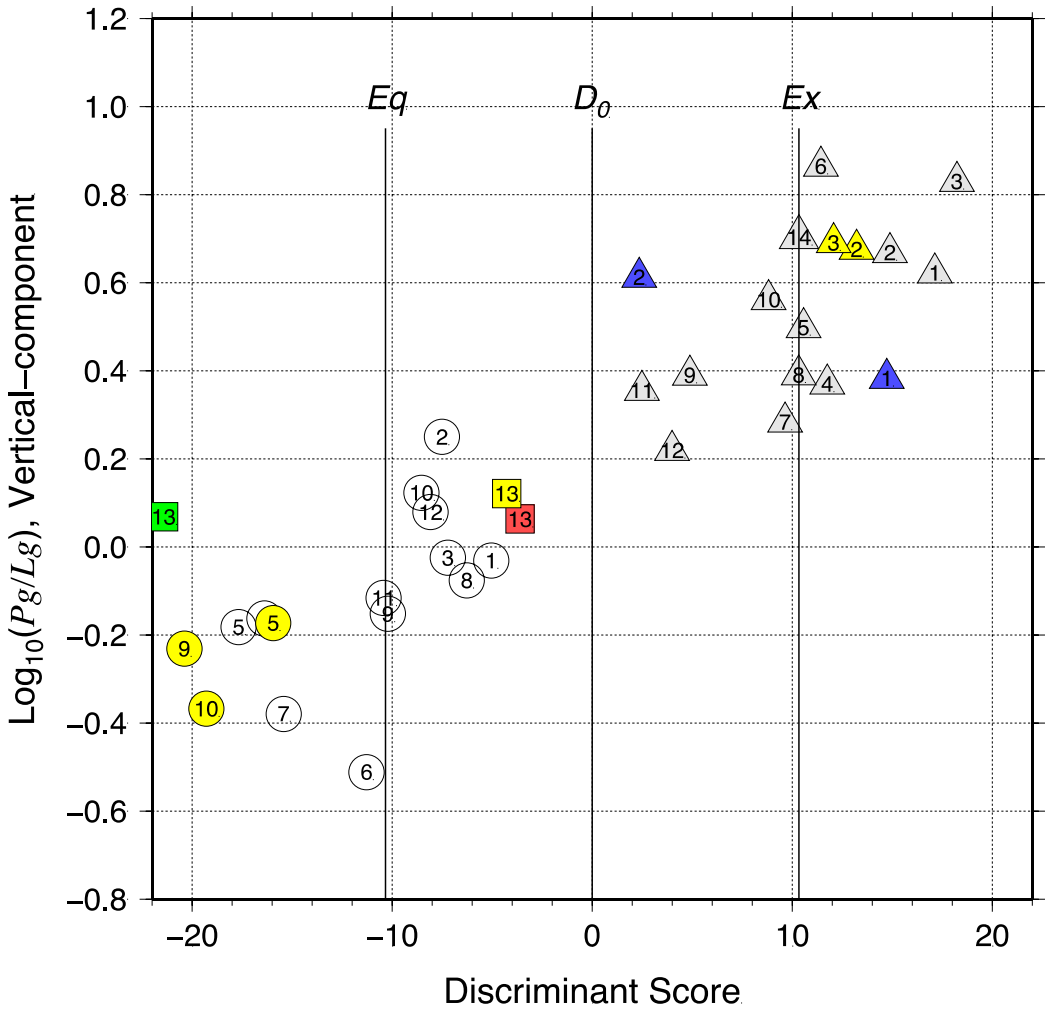

Figure 15. Same as Figure 10 (vertical-component LDF analysis), except that we have added points derived from data recorded by station SMT (shown in Fig. 1) used by Zhang and Wen (2015a, data given in their Figure 5), and from the NECESS Array station NE3C. The six points based on SMT data are shown in yellow, and they correctly place the nuclear explosions of 2009 and 2013 in the explosion population (*yellow triangles*), and the three earthquakes in the training set from Table 4 recorded by SMT are in the earthquake population (*yellow circles*). But the event of 2010 is now with the earthquakes (*yellow square*). The point based on data from station NE3C (*green square*) is for the 2010 event, and it is also with the earthquakes (for this station we do not have explosion data). Finally, a point for the nuclear explosion of January 6, 2016, derived from MDJ data has been added. It lies on the mean for explosions (shaded triangle with event id 14).

The electronic supplement for the paper consists of the following items:

- our measurements of the $\log_{10}$ *P/S* spectral ratios obtained from waveforms recorded at station MDJ for the two training sets (Table_S1.csv);
- a section describing further details of our three-component linear discrimination function analysis, in particular the effectiveness of different choices of the frequency components used to apply *P/S* spectral ratios; and
- a tutorial section on the underlying ideas behind Mahalanobis methods for event classification, with particular reference to Figures 10, 13, and 15 in the main text.

List of Table captions

Table S1. Measurements of the $\log_{10}$ *P/S* spectral ratios obtained from waveforms recorded at station MDJ for the two training sets.

Table S2. Three-Component Discrimination Analysis using four frequencies

List of Figure captions

Figure S1: The log(*P/S*) values measured at 8 hz from vertical component waveforms at station MDJ for two training sets, are shown as circles (earthquakes) and triangles (explosions), together with the normal (Gaussian) probability density functions inferred from these two data sets. Note two length scales; the Gaussian widths, and the distance between the means (explosions, earthquakes).

Figure S2: Probability distributions for the linear discrimination function that best-separates the earthquake and explosion populations, using MDJ data. Shown here, in (a), are the underlying Gaussians for vertical components recorded for our two training sets (earthquakes, explosions) as developed in Figure 10 of the main paper; and in (b) the underlying Gaussians for three-component data as developed in Figure 13. The standard deviation, Δ, is slightly larger in (b); but the distance between the means, which equals Δ^2, is significantly greater in (b) than in (a), providing better classification capability.

Click here to access/download

Supplemental Material (All Other Files, i.e. Movie, Zip, tar)

Table_S1.csv