



A Real-Time Discrimination System of Earthquakes and Explosions for the Mainland Spanish Seismic Network

MARTA GARCÍA VARGAS,¹  JUAN RUEDA,^{1,2,3} ROSA MARÍA GARCÍA BLANCO,³ and JULIO MEZCUA¹

Abstract—Different waveform-based discrimination parameters were tested using multivariate statistical analysis to develop a real-time procedure for discriminating explosions from earthquakes at regional distances in the Iberian Peninsula. This work enabled a purge of the Spanish National Seismic Catalogue for the period 2003–2014. The training data consisted of waveform-based signal properties in the time and frequency domain for events (earthquakes and explosions) recorded during the selected time period by the Spanish Broadband National Network and Sonseca short-period Array of the Instituto Geográfico Nacional (IGN). For each station and its associated training dataset, a discriminant function was defined as a linear combination of the measured variables. All station-specific discriminant functions were then combined with a weighting scheme to test the training events, revealing that 86 % of the events were consistent with the analysts' judgement. The application of this method to the whole of the IGN's seismic database for the studied period gave an 83 % success rate; however, a 91 % success rate is reached if events are classified using at least three stations and 100 % confidence levels.

Key words: Earthquakes, explosions, real-time identification, waveform-based discrimination, multivariate analysis.

1. Introduction

The spatial distribution of earthquakes and the location pattern of quarry blasts in Spain and its vicinity are found to be quite similar. Besides numerous areas known for the occurrence of blasts for a long time, artificial events often occur in areas that are seismically active too. Due to the large number of such explosions routinely detected every

year by the Spanish National Seismic Network (SNSN) of the Instituto Geográfico Nacional (IGN), discriminating between these artificial events and earthquakes poses a great challenge.

During routine data processing, SNSN analysts have the difficult task of classifying by eye analysed events as either earthquakes or artificial events, since only a few of the latter events are ever confirmed by the facility owners. Therefore, seismic discrimination in the SNSN is based simply on analysts' criteria, which, in turn, chiefly depend on their expertise. The detection threshold of the SNSN is very low (at present, around 1.5 mb(Lg) for most of Spain) and, consequently, the number of misclassified events is so large. This contamination of the earthquake catalogue may introduce false interpretations in future hazard studies.

Seismicity in the Iberian Peninsula and the surrounding areas is associated with the contact zone between the Eurasian and African plates along the so-called Azores–Gibraltar fault zone (Fig. 2a), which is purely oceanic. As soon as the collision approaches a zone where the plates are of continental nature (the Ibero–Maghrebian region) (Fig. 1c), the contact becomes more complex, and a single dominant plate boundary cannot be identified (MEZCUA and RUEDA 1997). The seismic activity is confined to the southern and southeastern parts of the Iberian Peninsula, the Iberian cordillera and further north by the Pyrenees.

Blasting in Spain mainly occurs in mines, quarries or within civil engineering projects (e.g., during the construction of roads and dams). Mining in the Iberian Peninsula has been gradually declining since the 1960s, most notably since the 1990s in the coal- and ore-mining sectors. Similarly, the number of active

¹ Fundación José García Siñeriz, Calle Alenza 1, 28003 Madrid, Spain. E-mail: martagvargas@yahoo.es

² Instituto Geográfico Nacional, General Ibáñez de Ibero 3, 28003 Madrid, Spain.

³ Universidad Politécnica de Madrid, ETSI Topografía, Geodesia y Cartografía, Campus Sur, Autovía de Valencia km 7.5, 28031 Madrid, Spain.

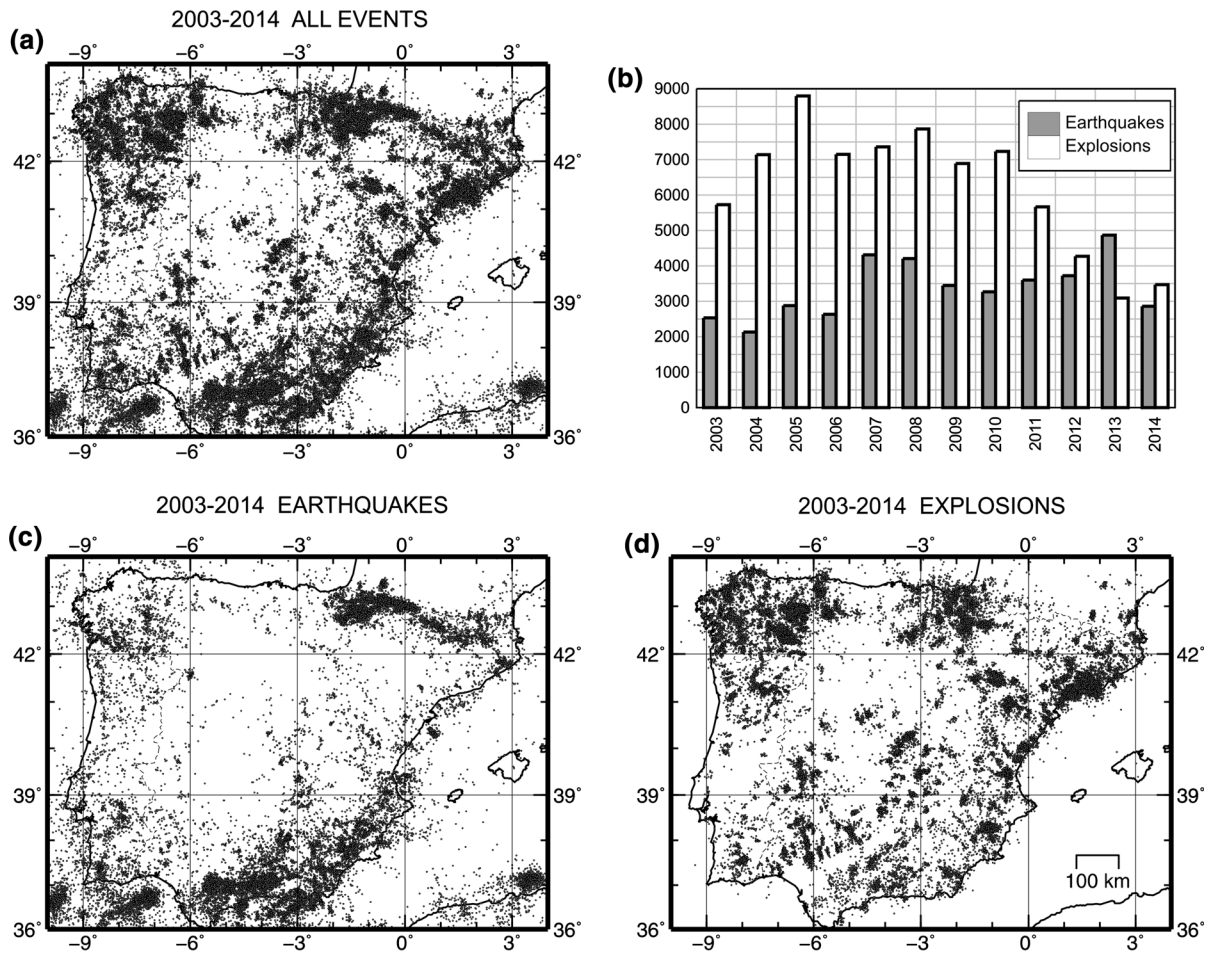


Figure 1

Distribution of events for the period 2003–2014 in the official IGN seismic database according to analysts' judgement: **a** all events (115,126); **b** the time distribution of the annual number of explosions and earthquakes, **c** events classified as earthquakes (40,463) and **d** events classified as explosions (74,663)

quarries and civil engineering projects dropped sharply in the past decade due to the slowing down of economic activity during the economic crisis. Consequently, the number of artificial events in Spain has substantially decreased in this period (Fig. 1b, d). It is clear from Fig. 1 that the distribution of earthquakes and explosions follows similar patterns, and an effective discriminant procedure needs to be implemented to separate these two types of events.

To develop a reliable system for discriminating between explosions and earthquakes, we first selected the most relevant discriminants found in the literature that can be measured automatically from recorded waveforms. We collected short-period data from the

Sonseca Array operating in central Spain and broadband data from the Spanish Broadband National Network (SBNN) for the period 2003–2014. Next, we tested separately training samples from the Sonseca Array data and all the SBNN single-station data, and then performed a statistical multivariate analysis to define a discriminant function for each station's training dataset by a linear combination of the measured parameters. Finally, we combined all the discriminant functions with a weighted discriminant method to test the events in the training data (TRN). The whole dataset for the study period was used to verify the ability of the proposed method to correctly identify fresh seismic events.

2. Review of Waveform-Based Discrimination Methods

The different source properties such as dimensions, time function, focal mechanism and depth, of both explosions and earthquakes are reflected in the observed waveforms. No single method or criterion can yet ensure reliable discrimination between the seismic signals generated by earthquakes and explosions. An effective approach to separate them will have to make use of several discriminants (WÜSTER 1993; WALTER *et al.* 1995; STUMP *et al.* 2002).

We analyse, here, the most effective waveform-based discriminants taken from the literature, which we then evaluate using statistical multivariate analysis. ALLMANN *et al.* (2008) divided these methods into three types: phase amplitude ratios (BENNETT and MURPHY 1986; WÜSTER 1993), spectral methods (TAYLOR *et al.* 1988; WALTER *et al.* 1995) and coda studies (SU *et al.* 1991).

Quarry blasts are commonly rippled-fired and can often be identified by the characteristic time-invariant spectral modulation caused by the time delays between subsequent explosions (ALLMANN *et al.* 2008). BAUMGARDT and ZIEGLER (1988) studied records from the NORESS short-period array and found that the spectra of rippled-fired blasts in Scandinavia have a distinctive modulation that differs from the spectra of earthquakes in the same region. Spectral modulations in the records of quarry blasts were also observed in Kazakhstan (HEDLIN *et al.* 1989) that were absent in spectra from records of single-event calibration shots (CHERNOBAY and GABSATAROVA 1999). However, STUMP *et al.* (2002) provided evidence of accidental, near-simultaneous detonations of a large amount of explosives during standard delay-fired explosions, which further complicates the solving of the discrimination problem. Such events have single-fired characteristics and may prove to be problematic in discrimination analysis.

It is generally believed that both spectral shapes and ratios in regional phases (Pn, Pg, Lg) are useful for distinguishing earthquakes from explosions (POMEROY *et al.* 1982). BENNETT and MURPHY (1986) confirmed the usefulness of measurements of the Lg spectral ratios in the 0.5–1 and 2–4 Hz frequency bands in discriminating between Nevada Test Site

(NTS) (USA) explosions and nearby earthquakes by showing that these measurements were greater in the case of artificial events. Similar results were obtained by TAYLOR *et al.* (1988) testing the performance of spectral ratios for higher frequency bands (1–2 and 6–8 Hz). However, WALTER *et al.* (1995) found that in the case of the NTS, the best discriminant performance is achieved by combining phase and spectral ratios. TAYLOR (1996) used the high-frequency Pg/Lg discriminant between frequencies of 0.5 and 10 Hz that were found to be very effective and displayed improved separation between earthquakes and explosions as the frequency increased. MURPHY *et al.* (2009) found that NTS explosions exhibited a spectral peak in the Lg coda-derived spectra, in sharp contrast to earthquakes, and related it to the result of scattering higher-order Rg near the source region, which scattered off into the coda wavefield. Coda-derived source spectra showed pronounced peaking as a function of depth relative to nearby earthquake source spectra (using the same path and site calibration corrections for both nukes and quakes).

CHERNOBAY and GABSATAROVA (1999) implemented a multi-parameter cluster analysis algorithm for routine classification of seismic events in the northern Caucasus (Russia). They performed event classification for each station separately using the spectrogram method and/or methods based on the P/S spectral ratios. They achieved a 98 % success rate and identified 19 different groups of sources at regional distances in an area of moderate seismicity.

The Vogtland area (Germany and Czech Republic) has been thoroughly studied by WÜSTER (1993, 1995), KOCH and FÄH (2002), and KOCH (2002b). WÜSTER (1993) established a combined discrimination scheme using a simple majority voting technique and obtained a 98 % classification success rate. Events were recorded by GERESS short-period array, and the parameters considered—both individually and in combination with the voting system—were (1) the maximum amplitude ratios (Lg/Pg, Lg/Rg) and (2) the spectral properties of the P- as well as S-wave group. Spectral peaks and slopes were extracted by autoregressive-moving average (ARMA) modelling, master-event correlation and a search for time-independent frequency structures in sonograms. Full discrimination was achieved subsequently by WÜSTER

(1995) using a multivariate quadratic function based on four parameters: the spectral decay and the spectral variance of the P- and S-wave spectra. We considered the spectral variance (KOCH and FÄH 2002) as a measure of the misfit of the regression line fitted to the spectra, that is, the area between the spectrum and the regression line. KOCH and FÄH (2002) pointed out the usefulness of spectral amplitude ratios at two different frequency bands (Lg) and of Pg/Sg amplitude ratios in event identification. Those authors also denoted the possibility of weighting according to station success rates to improve the overall discrimination capability. KOCH (2002b) suggested that spectral variance also provides good performance for event discrimination for other parts of Germany and not only for the Vogtland region.

Using a multivariate statistical analysis, FÄH and KOCH (2002) succeeded in classifying correctly 97 % of events by calculating a discriminant function for a training sample of confirmed explosions and earthquakes in Switzerland. The coefficients of the discrimination function were determined for the whole network and not individually for each station. The parameters analysed were epicentre-station distance, P/S signal energy, S/P ratio for the frequency bands 4–7 and 7–10 Hz, and the relative noise energy. The discrimination of a seismic event was performed using a weighted sum over the discriminant function output of all station data. The weight was the success rate for the training set at the corresponding station. The success rate of the whole training sample (85 events) was 100 % and, with an extended dataset with 194 events added to the training sample (total = 279 events), was 94 %. The introduction of station corrections improved the results of the proposed method.

Using the multivariate discriminant analysis procedure described by WÜSTER (1993), RUEDA (2006) applied multivariate statistics to discriminate between explosions and earthquakes for a limited dataset of 220 events recorded by the Sonseca Array (Spain). This author studied the performance of the spectral variance and decay of the P- and S-waves' spectra using only two parameters at a time and deriving the corresponding linear discriminant function (LDF). To compute all the parameters, RUEDA (2006) developed a weighting scheme according to success rates. This procedure correctly identified nearly 90 % of the

studied events. When the spectral decay and variance of the Lg wave were considered jointly, the performance was better than when they were considered individually. The same applies in the case of the P wave. Also, a sample of 94 events from the SBNN was studied using the same variables as for the Sonseca short-period Array, along with the Lg-wave ratios in two frequency bands (1–2/6–8 Hz and 1–2/7–9 Hz). In this case, the results of the identification agreed with the analysts' criteria in 84 % of cases.

ANDERSON *et al.* (2007) developed a mathematical statistics formulation of the most commonly used teleseismic discriminants, integrating them into a probability model which could be extensible to other discriminants. ANDERSON *et al.* (2009) established a mathematical model to combine the depth from travel time, presence of long-period surface energy, depth from reflective phases and polarity of first motion and compute the standard error of these discriminants at NTS from a multistation approach.

Other approaches such as the application of methods based on artificial neural networks (ANN) (DOWLA *et al.* 1990) have also been successful; these techniques discriminate automatically weak signals at regional distances from low-magnitude events including quarry blasts, underwater explosions and natural sources events (SCARPETTA *et al.* 2005) and landslides, explosion-quakes and volcanic micro-tremor signals (ESPOSITO *et al.* 2006). Recently, FARAHANI *et al.* (2012) have proposed an adaptive neuro-fuzzy inference system (ANFIS) for classifying low-magnitude events (earthquakes and mining blasts) based on a successful application of this method in France (MULLER *et al.* 1999).

In this study, we take advantage of the large amount of data available in the SNSN to test different waveform-based discrimination parameters to develop an automatic procedure using multivariate statistical analysis for discriminating between artificial and natural events at regional distances in the Iberian Peninsula.

3. Data Used in the Analysis

The available dataset for the study consisted of 115,126 events (74,663 explosions and 40,463

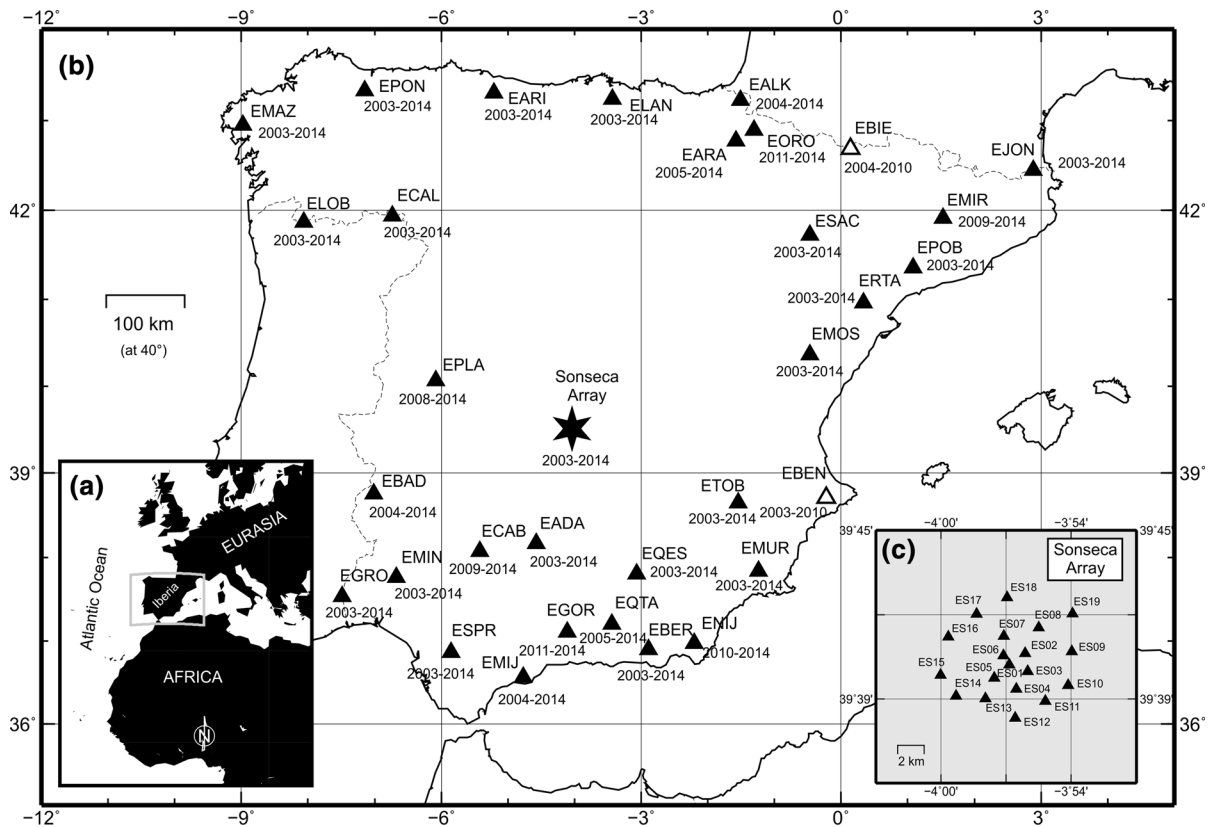


Figure 2

a Location of the study area; **b** distribution of broadband stations (triangles) and the Sonseca Array (star) in the SNSN. Black symbols denote stations currently in operation, while white symbols indicate stations not in operation. The time period for the data used in the analysis is given below the name of the station; **c** distribution of the 19 short-period instruments in the Sonseca Array

earthquakes, as identified by analysts) taken from the IGN seismic database for the period 2003–2014 (Fig. 1a). The analyst decision is based on the morphology of the waveforms and the time of the day the event occurs. Explosions are more likely to occur only during daylight. Besides this, explosions are sometimes confirmed by contacting the local authorities. As continuous digital recording (3C for SBNN and vertical component for Sonseca Array) was required in our study, we limited our analysis to 91,139 events, of which 28,534 were classified as earthquakes ($0.1 \leq m_b(Lg) \leq 5.5$) and 62,605 as explosions ($0.1 \leq m_b(Lg) \leq 3.2$) by the SNSN analysts.

In the Iberian Peninsula, the SBNN consists of 44 broadband stations located in low-noise locations within a distance of 100–150 km of each other. From

these broadband stations, we only used 32 VSAT (very small aperture terminal) satellite communication system stations located on the Spanish mainland, which provided data during the studied period (Fig. 2b). We also used the Sonseca Array located in central Spain formed by 19-element short-period instruments in a circular aperture diameter of 9 km (Fig. 2c). Standard time domain beam forming significantly improves the signal-to-noise-ratio (SNR) of recorded data and is highly effective in suppressing noise and emphasising signals.

To find the most suitable signal properties for separating two populations (explosions and earthquakes) and design a discriminant rule for classifying new events, we first required a sample of confirmed or most likely to be correctly classified events. Separate samples of events referred as training

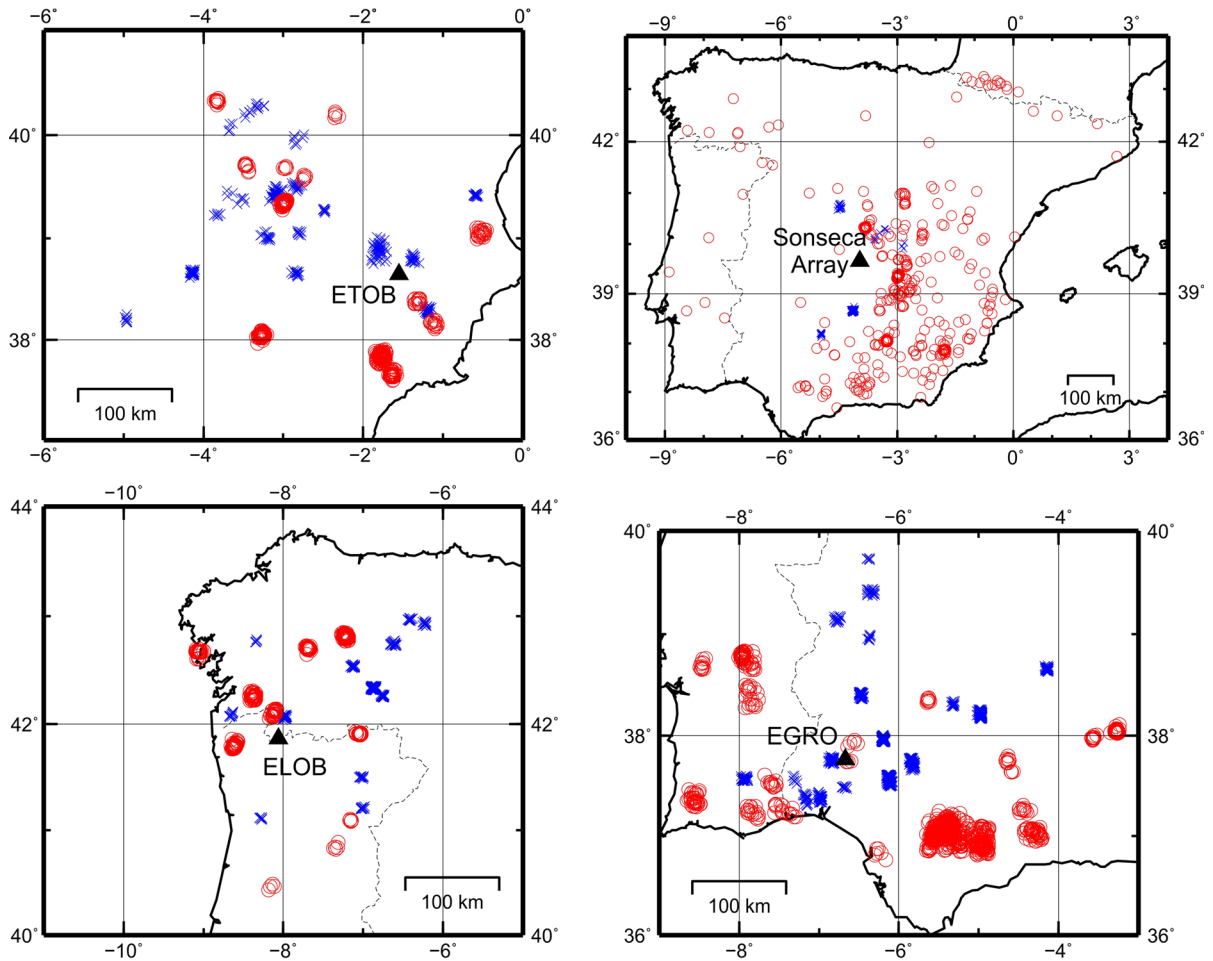


Figure 3

Examples of the training events considered (blue crosses for explosions and red circles for earthquakes) for three broadband stations: ETOB (Albacete), ELOB (Ourense), EGRO (Huelva) and Sonseca Array (Toledo)

samples (TRN) were taken from each station to derive a discriminant function.

We imposed an SNR criterion for broadband stations and Sonseca Array data. We calculated the SNR for the P and Lg by simply dividing the power spectra amplitude of the signal by the power spectra amplitude of the noise in the 7–12 Hz frequency range. For the Sonseca Array beam, we selected from the dataset only events with $\text{SNR} \geq 20$ dB. Given that we had no confirmed sample of explosions and to remove small magnitude natural data that could have been misclassified as artificial, we picked only classified artificial events occurring between 10 AM and 4 PM and excluded earthquake data in the same time

interval. Likewise, artificial events that occurred outside of known blast areas (where most of the explosions occur) were removed from the training event sample. Finally, given that we aimed to select events within the same range of magnitude, we established an 2.5-mb(Lg) upper threshold.

For the broadband station data, we selected events with an $\text{SNR} \geq 23$ dB following the same procedure as for the Sonseca Array training set data. The individual training data sets (TRN) for the three broadband stations – ETOB (Albacete), ELOB (Ourense) and EGRO (Huelva) – and Sonseca Array (Toledo), are shown in Fig. 3 and their corresponding values with the rest of the stations in Table 1.

Table 1

Number of events in the training sample (TRN) for each station and for the whole TRN using the events only one time, and number of explosions and earthquakes. CC stations' canonical correlation coefficients. Success rates of the station training samples with single-station D functions and multi-station weighted D functions (MSWM, where several stations contribute to event classification) applied to each station TRN events

Station	2003–2014 TRN events	EX	EQ	CC	TRN single-station DF success rate (%)	TRN weighted DF success rate (%)
EADA	1077	497	580	0.60	80	86
EALK	1193	518	674	0.70	87	88
EARA	352	166	186	0.84	93	94
EARI	359	181	178	0.65	82	91
EBAD	1181	584	597	0.52	75	88
EBEN	179	92	87	0.38	87	69
EBER	45	22	23	0.67	87	80
EBIE	1788	515	1273	0.73	89	92
ECAB	1053	564	489	0.56	79	89
ECAL	4062	3834	228	0.36	83	85
EGOR	478	50	428	0.54	90	94
EGRO	887	432	455	0.66	85	87
EJON	574	175	399	0.65	86	89
ELAN	740	545	286	0.58	77	82
ELOB	833	415	418	0.71	86	87
EMAZ	385	185	200	0.66	82	88
EMIJ	119	56	63	0.46	67	75
EMIN	2933	1663	1270	0.71	87	89
EMIR	426	141	285	0.51	76	84
EMOS	1528	875	653	0.52	74	81
EMUR	1036	154	882	0.43	80	86
ENIJ	367	93	274	0.63	85	89
EORO	935	261	674	0.68	86	90
EPLA	374	181	193	0.52	75	80
EPOB	3716	2550	1166	0.62	81	84
EPON	3038	2319	719	0.58	81	88
EQES	4050	853	3197	0.68	89	92
EQTA	1860	230	1630	0.50	84	92
ERTA	1349	653	696	0.50	73	84
ESAC	1613	841	772	0.66	84	87
ESPR	420	103	317	0.74	91	91
ETOB	803	372	431	0.78	90	92
SONSECA	974	529	445	0.88	97	98
TRN	27,731	16,172	11,559			86

4. Selection of the Signal Properties Used in the Regression Analysis

To quantify and characterise differences in source type, we selected the following properties that can be automatically calculated from the seismograms recorded by the Sonseca Array and BB stations (Figs. 4, 5):

In the time domain: maximum Lg amplitude/maximum P amplitude ratio (WÜSTER 1993), logarithms of Lg amplitude ratios 1–2/6–8 Hz and 1–2/7–9 Hz (KOCH 2002a; KOCH and FÄH 2002; RUEDA 2006).

Theoretical P and S arrivals are automatically calculated. For each selected event, we used three different time windows containing the preceding noise, the P- and the S-wave trains. Once filtered (0.5 Hz high pass for the P/Lg ratios), the maximum absolute amplitude was measured and stored.

In the frequency domain, we measured spectral peaks and slopes: P and Lg spectral decay and spectral variance (KOCH 2002a; KOCH and FÄH 2002; RUEDA 2006).

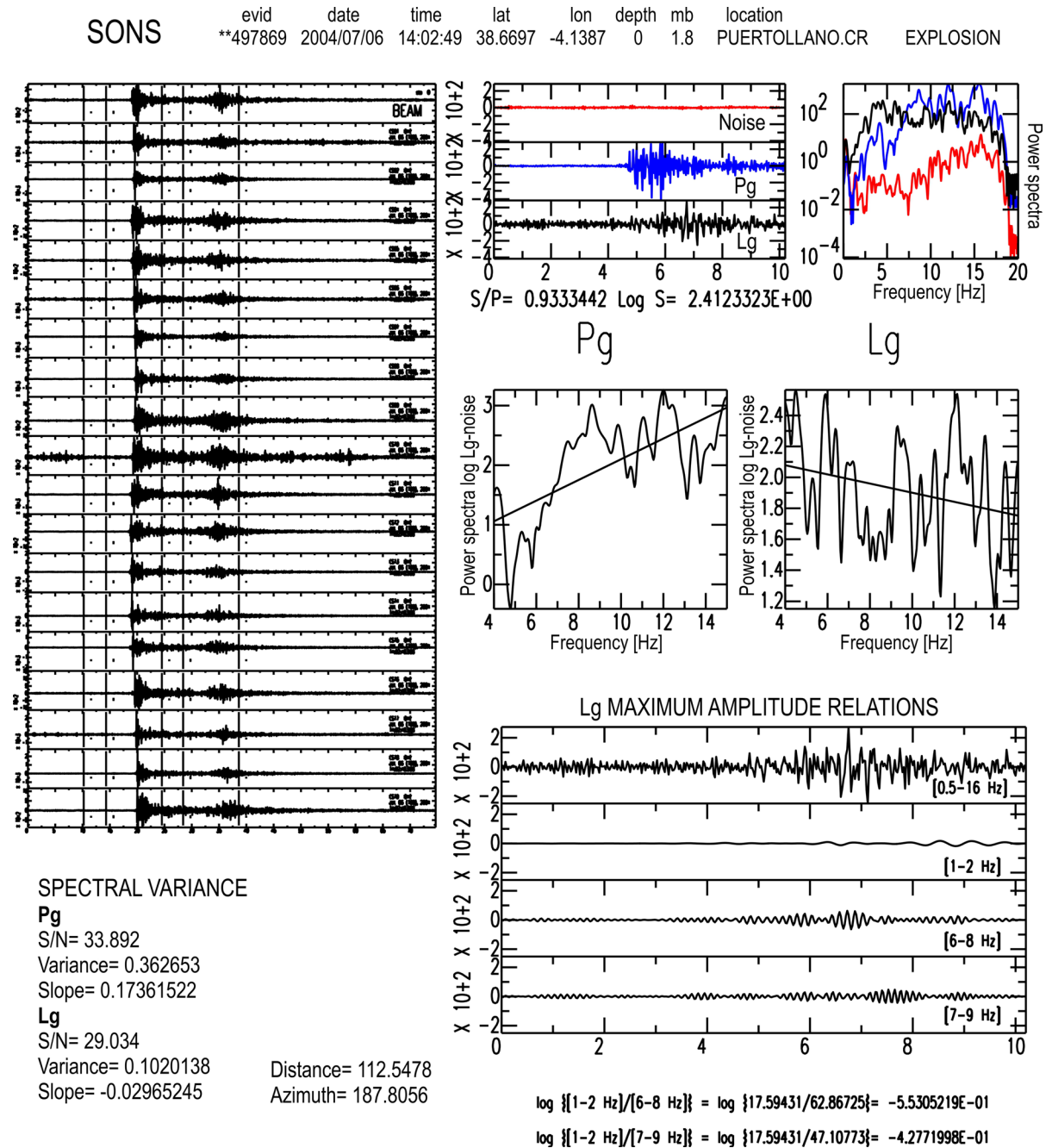


Figure 4

Example of a 1.8-mb(Lg) explosion recorded at the Sonseca Array at a distance of 113 km. *Left* waveforms of 18 short-period Sonseca Array channels and beam. *Top* data segments and associated power spectra for noise, P and Lg waves. *Middle* logarithm of the spectral power of the P and Lg waves (and associated regression line). *Bottom* Lg maximum amplitude relationships

For each selected band pass-filtered (4–15 Hz) window, we calculated the power spectra, which are also corrected for noise by subtracting the noise spectra from the signal spectra, and the spectral power

logarithm is computed. The slope of the regression line fitted for that spectrum shows the spectral decay.

P and Lg amplitudes for the Sonseca Array data were taken from vertical components after

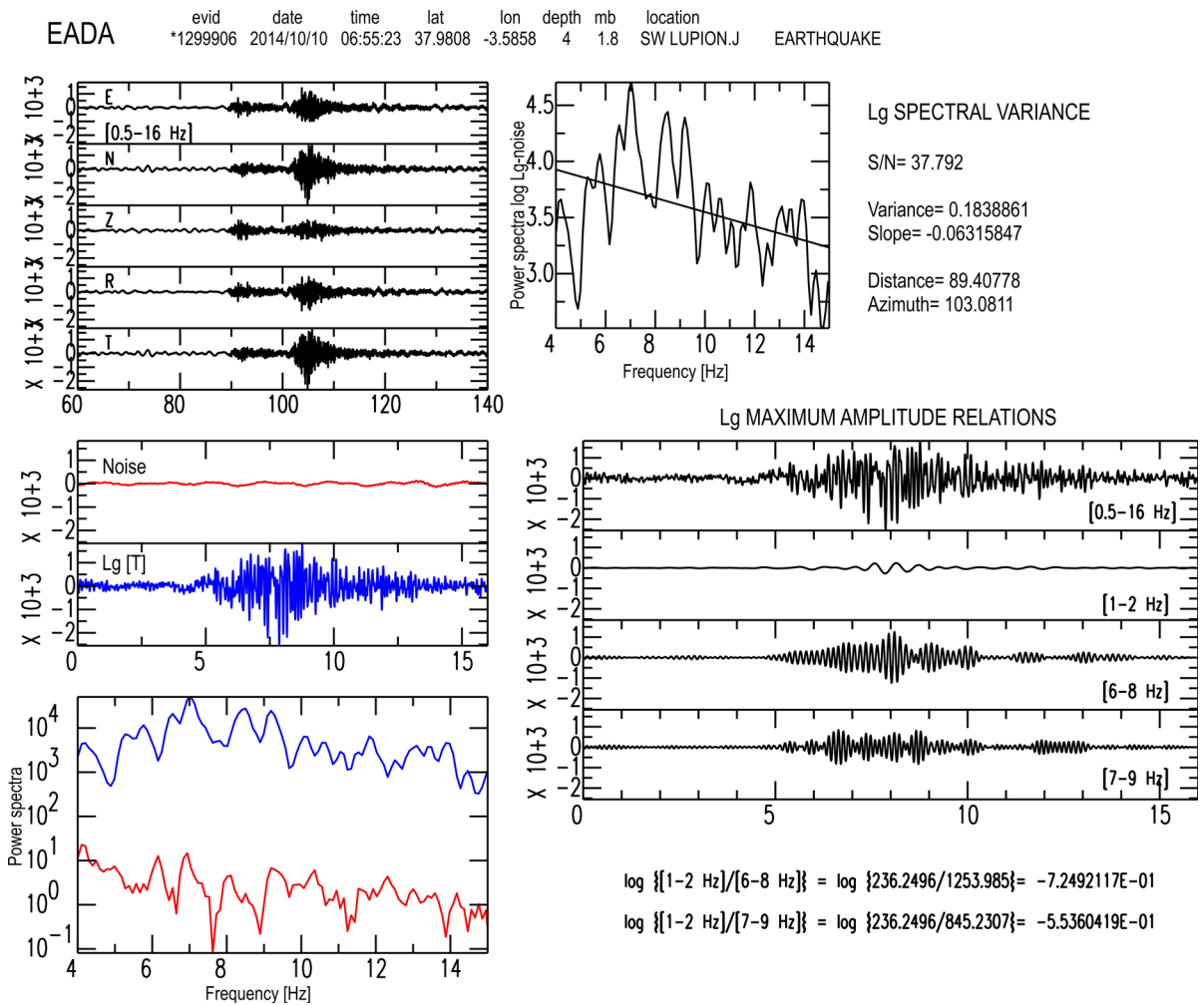


Figure 5

Graphical example of the parameters measured for 3C records from a 1.8-mb(Lg) earthquake at an 89-km epicentral distance from EADA broadband station (Córdoba). *Top left* three-component record (and rotations). *Middle* noise and Lg data segments. *Bottom* the corresponding power Lg and noise spectra. *Top right* the logarithm of the spectral power of the Lg wave and regression line (where spectral variance and spectral decay values are measured). *Bottom right* Lg wave maximum amplitude relationships

beamforming, while for the three-component broadband stations, P was measured in the radial and Lg in the transverse components.

5. Multivariate Statistical Analysis: Single-Station Discriminant Functions

To derive a station' linear discriminant function (D) that encloses those signal properties (X_i)

considered being relevant, based on Fisher's linear discriminant (FISHER 1936) we can write:

$$D = v_1X_1 + v_2X_2 + \dots + v_iX_i + a, \quad (1)$$

where a is a constant and v_i the discriminant coefficients or weights for the variables. The function coefficients (v_i) were determined individually for each station training dataset and chosen to maximise the distance between the two groups (explosions and earthquakes).

To begin with, for each station training dataset, we first examined whether there were any significant differences between the two groups in terms of each independent variable using group means and an analysis of variance (ANOVA). The test of equality of group means gave low values for Wilk's lambda statistic, which provides strong statistical evidence for significant differences between the means of the two groups for all variables and all stations.

We used the stepwise method to introduce the selected parameters into the function to find the best set of predictors and to eliminate those found to be redundant. We first entered the best correlated independent variable and then, one by one, the remaining variables until the subsequent variable adds no significant predictive power to the squared canonical correlation coefficient, which is equivalent to the coefficient of determination (R^2). This procedure resulted in the elimination of the logarithm of the Lg amplitude ratio between 1–2 and 6–8 Hz (L1268) from the Sonseca Array training dataset, as it was found to be redundant. Finally, we tested the role of the predictors in the discrimination: Wilk's lambda criterion reveals that for all the station datasets, all the variables included in the analysis add some predictive power to the discriminant function.

To test the performance of the resulting discriminant function, we used the canonical correlation coefficient (CC), which is obtained from the multiple correlations between the predictors and the discriminant function. This coefficient provides an index of overall model fit, interpreted as the proportion of the variance explained (R^2). The maximum coefficient of 0.88 was obtained for the Sonseca discriminant function, a value that corresponds with a model that explains 77 % of the variation in the grouping variable (Table 1).

For each event (k), we obtained a discriminant score using the station (j) discriminant function (Eq. 1). The discriminant score, $D_{k,j}(X_i)$, is, thus, a linear combination of discriminant variables. The dependent variable in the regression is set as 1 if it is an explosion (EX), or 2 if an earthquake (EQ). To classify new observations, we used 0 as the cut-off value. If for a particular event, the discriminant score of the function was less than or equal to the cut-off

($D_k \leq 0$), the event was classified as an explosion, whereas if it was greater ($D_k > 0$), the event was classified as an earthquake. The closer the group centroids are, the greater the classification error is likely to be, because the overlapping area between the two populations will be larger.

To estimate the success rate of a station (j), we applied the D function (Eq. 1) to each event (k) of the station training sample to obtain a discriminant score $D_{k,j}(X_i)$, which indicates that the event is considered an earthquake (EQ) or an explosion (EX). This process was carried out for all the events in the station dataset, and the station identification success rate was evaluated by comparing the results with analysts' decisions.

The success rates of the discriminant functions applied to single-station training datasets varied from one station to another, with the Sonseca Array reaching 97 % of correctly classified events according to analysts' decisions (Table 1). It is well known that the performance of discriminants depends on parameters such as tectonic region, frequency, phases used (WALTER *et al.* 1995), blasting practices and the complexity of regional wave propagation (STUMP *et al.* 2002). This explains the different success rates we obtained depending on the station dataset (Table 1). The performance of each of the parameters for the Sonseca Array is also shown in Figs. 6 and 7. Therefore, defining the discriminant functions separately for each station compensates for regional wave propagation and local effects.

An example of how taking into account more than one discriminant improves the rate of success is shown in Fig. 6 for the Sonseca Array dataset. Figure 7a depicts two-variable discrimination plots for the Sonseca Array training data. When separating artificial and natural events, three-variable plots are an improvement over two-variable plots (Fig. 7b).

In terms of the signal properties, our results show that Lg spectral variance plays a major role in the single-station discriminant functions, as previously reported by WÜSTER (1995) and KOCH and FÄH (2002). We found that for all stations, explosions had greater values of both Lg and P variance than earthquakes. The analysis of the Pearson correlations between the predictors and the station D function showed that this parameter has a moderate-to-strong

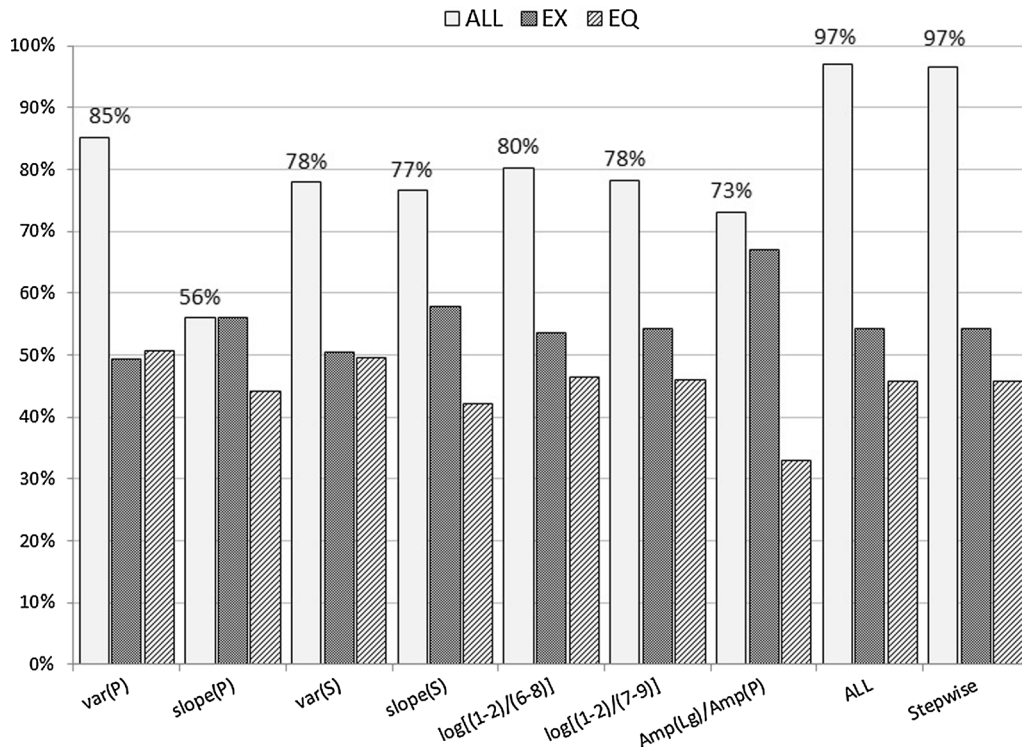


Figure 6

Performance of the discriminant variables tested for the Sonseca Array training dataset: (1) individually, (2) all variables or (3) only variables selected by the stepwise method (excluding L1268). All = rate of success of the variable; EX and EQ = rate of explosions and earthquakes correctly identified, respectively

correlation with the D function in 29 out of 32 stations (91 %). At the Sonseca Array, we observed that the P spectral variance had the best discriminant power.

The second most-correlated variable was the logarithm of the L_g amplitude ratio in 1–2/7–9 Hz (91 % of the stations with moderate-to-strong correlation). The logarithm of the L_g amplitude ratio in 1–2/6–8 Hz (L1268) exhibits very similar results as the logarithm of the L_g amplitude ratio in 1–2/7–9 Hz (L1279). In the end, the L1268 parameter was eliminated from the regression analysis of the Sonseca Array as it was found to be redundant (Fig. 7a). By contrast, the L_g spectral decay gives different results depending on the station, and only in 47 % of the stations did the D functions display medium-to-strong correlation. We generally obtained higher values of L_g spectral decay for earthquakes.

The maximum amplitude ratio L_g/P appears to be only poorly related to group distinctions;

nevertheless, when considered together with the other parameters, these parameters add some predictive power. We found that explosions typically show lower L_g/P ratios than earthquakes but with a considerable degree of scatter, as WÜSTER (1993) had noted.

The L_g spectral variance and L1268 and L1279 seem to be strongly correlated with the D functions in at least 66 % stations. Finally, the amplitude L_g/P ratio was strongly correlated in first station (3 %), whereas L_g spectral decay was only strongly correlated in eight stations (25 %).

6. Application of a Multi-Station Weighted Statistical Discrimination Method to a Training Sample

To classify TRN events occurring during the period 2003–2014, we combined the 33 station discriminant functions (one for each of the 32 broadband

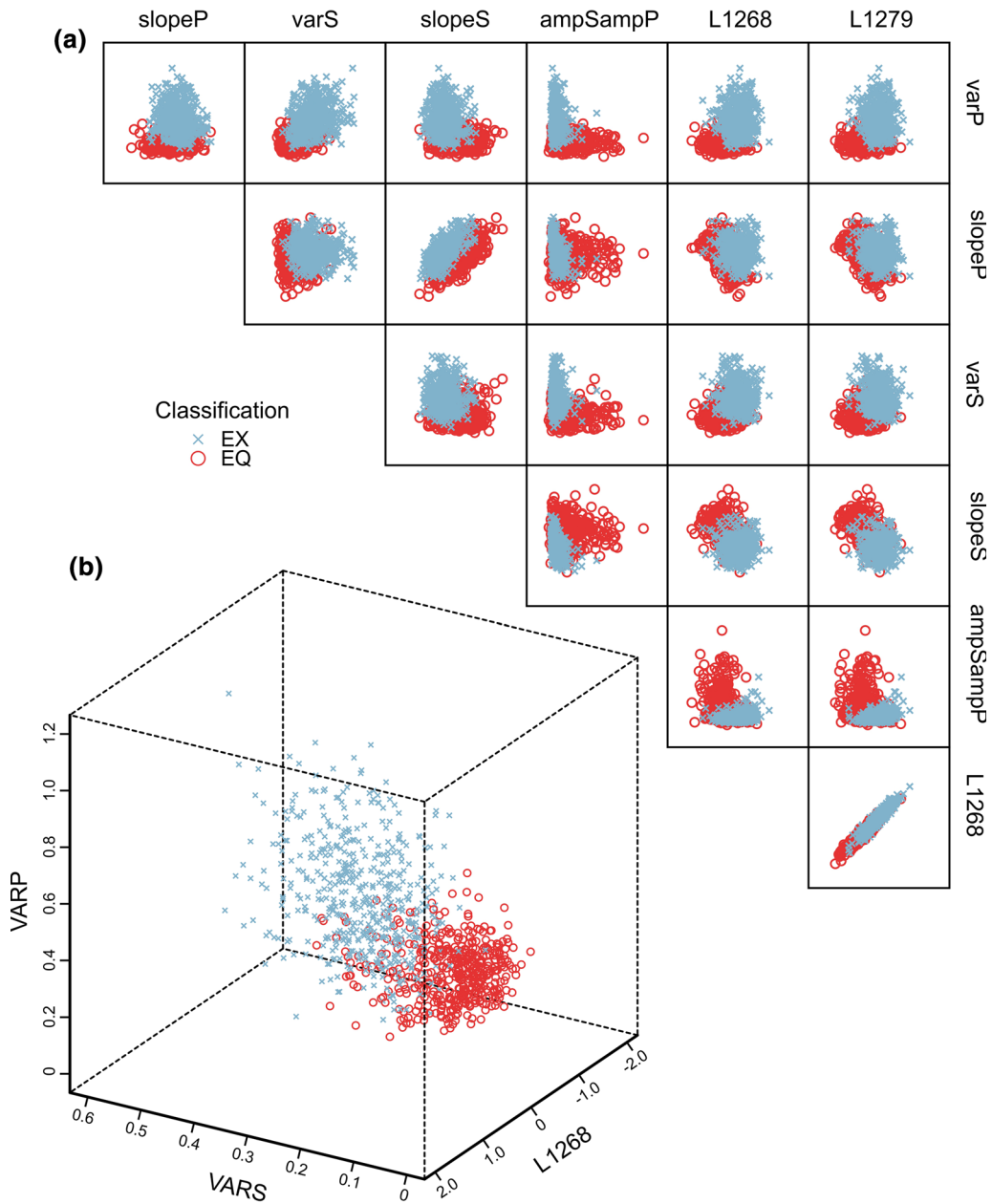


Figure 7

a Separation of explosion and earthquake populations with two-variable discrimination plots and **b** a three-variable plot (log Lg amplitude ratio for 1–2/7–9 Hz, P and Lg spectral variance) for the Sonseca Array training sample as an example of a simple application of the multivariate statistical analysis

stations and one for the Sonseca Array). We employed jointly various stations using the weights based on the canonical correlation (CC) between the predictors and the resulting discriminant function and the position of the discriminant score in the sample

space ($D_{k,j}(X_i)$) (Eq. 2). The use of several stations for classifying new events is called multi-station weighted method (MSWM). Accordingly, for any particular event, the better the performance of the linear function for the training dataset for each station

considered and the higher the discrimination score, the more the station contributes to the final result.

The confidence level assigned for a particular event in the MSWM is the combination of the canonical coefficient of the stations involved and the discriminant score defined before. Thus, the confidence of the results depends on the number of stations used with a reasonable SNR, the performance of each function of the single-station training sample and the discriminant score reached when applied to a particular station discriminant function.

The discriminant score of an event is evaluated by the following expression:

$$D_{\text{MSWM}} = \frac{1}{N} \sum_{j=1}^N [D_{k,j}(X_i) CC_j], \quad (2)$$

where N is the number of stations recording an event with a certain SNR value; X_i represents the signal properties; $D_{k,j}(X_i)$ the discriminant score for an event at a particular station (Eq. 1); CC_j the canonical correlation of the station; and D_{MSWM} the discriminant score for the studied event with the MSWM. When $D_{\text{MSWM}} \leq 0$, the event is identified as an explosion, whereas $D_{\text{MSWM}} > 0$ indicates an earthquake.

We applied this methodology to the TRN used previously as calibration data. The average success rate of the single-station discrimination functions applied individually to the corresponding station training samples was 83 %, while the use of several stations for each event identification (multi-station weighted method) resulted in 87 % of events being correctly classified (Table 1), an increase in the identification success rate of 4 %. The overall success rate for the TRN events using the MSWM reached 86 %. However, this improvement was not confirmed by all the stations since EBEN and EBER showed a significant worsening. We found several reasons for differences in stations' discriminant power, including the goodness-of-fit of the training sample, the blasting practices employed in the area, the propagation path of the waves and station noise.

On the other hand, weighting the D functions according to the canonical coefficient and the discriminant score gave 10 out of 33 (30 %) stations with their training dataset success rates over 90 %,

with only two stations (EBEN and EMIJ) with identification rates below 80 %. The inclusion of the discriminant score as a weight improved the success rate by 1 % (compared with using only the canonical coefficient of the station), although the number of events classified with high confidence levels improved significantly (5 %).

We considered removing from the analysis stations with low canonical coefficients and low identification rates of the single-station D functions throughout the training sample. Nevertheless, we found no change in the success rate of the stations when combined if the contributions of the EBEN, ECAL, EMIJ and EMUR stations were not considered.

7. Results of the Application to the June 2003–December 2014 Dataset

To evaluate the consistency of the MSWM developed here, we applied our discriminant approach to the total dataset using all the weighted discriminant functions. The total sample corresponds to the 91,139 events included in the entire IGN seismic database from the period June 2003 to December 2014. Not all the signals could be used due to the high noise level at several stations. Waveforms with an SNR below 20 dB were removed, which meant that 38 % of events were excluded from the analysis. Accordingly, the method was applied to a sample of 56,857 events, that is, 35,993 explosions and 20,864 earthquakes (according to analysts' classifications), which correspond to all events with at least one station record of $\text{SNR} \geq 20$ dB. The resulting classification of the proposed algorithm agreed with the analysts' criteria in 83 % of cases (47,085 events) and disagreed in 17 % of cases. In total, 54,522 events (96 % of the total dataset considered) had an acceptable level of confidence (over 65 %), 45,701 events (84 %) were well classified by analysts, but 8821 (6728 EX and 2093 EQ) conflicted with analysts' decisions (16 %). The highest percentage (75 %) of events classified differently from analysts' opinions corresponded to artificial events that our method identified as earthquakes.

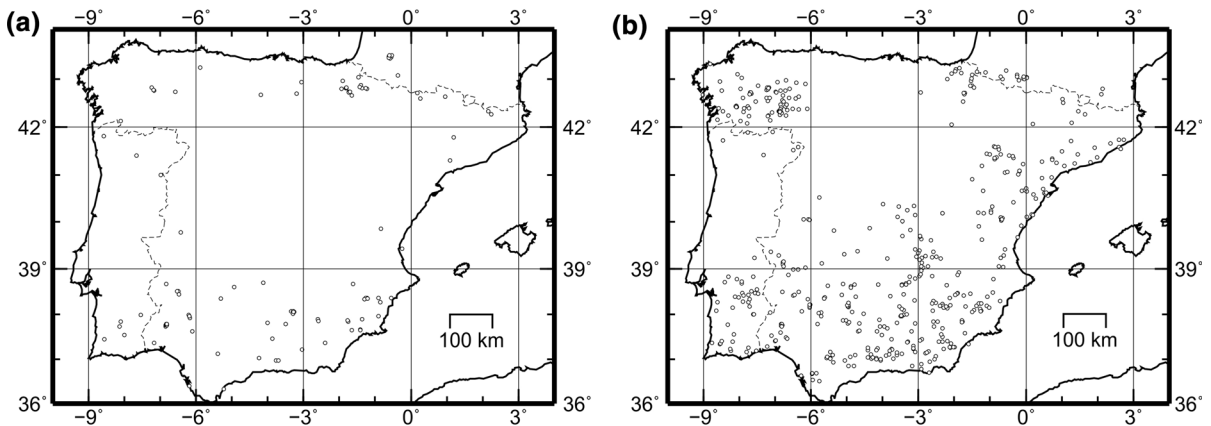


Figure 8

Geographical location of events recorded by at least three stations with $\text{SNR} \geq 20$ dB and 100 % confidence level with classification by the multi-station weighted discrimination method (MSWM) different from analysts' interpretation. **a** Earthquakes from the seismic database now identified as explosions (107 events) and **b** explosions now identified as earthquakes (503 events)

To show the differences between our classification and analysts' decisions, the events recorded by at least three stations with $\text{SNR} \geq 20$ dB and 100 % confidence levels in the results generated by the MSWM method are shown in Fig. 8. This means that all the stations that contributed to this result agree on the diagnosis. The number of such events was 9734; comparison with the analysts' decisions was only conducted for shallow events (at depths less than 10 km) and showed that 90 % of events were correctly identified.

As previously mentioned, the number of explosions in the seismic database that are now classified as earthquakes by the MSWM is greater than earthquakes considered now as explosions. The reason may be related to the fact that analysts tend to wrongly classify as explosions events that occur in areas where explosions are frequent (Figs. 1d, 8b) and misclassify as earthquakes events with epicentres in seismically active areas (Figs. 1c, 8a). However, we found two exceptions, an area near Cáceres (W Spain) and Almodovar (S Portugal), where the algorithm incorrectly identified a series of events with explosion waveforms as earthquakes. These explosions may be near-simultaneously delay-fired detonated with single-fired characteristics that cause the erroneous MSWM discrimination, as noticed by STUMP *et al.* (2002). Consequently, these events are not represented in Fig. 8. Therefore, the events

identified as earthquakes should be considered to be included in the earthquake catalogue, while the explosions should be removed.

8. Conclusions

The method developed in this paper is an objective real-time discrimination tool that does not require human intervention. The results may help the analyst working with a seismic network to finally decide between artificial and natural seismicity. The application of this method to the 2003–2014 Spanish seismic database enabled us to identify new tectonic-related seismic events and to eliminate events once considered to be earthquakes that are now known to be artificial.

We used multivariate statistical analysis to derive linear discriminant functions based on the signal properties of earthquakes and explosions recorded by broadband stations and the Sonseca Array during the period 2003–2014. Our results show that Lg spectral variance plays a major role in the single-station discriminant functions, as well as the logarithm of Lg amplitude ratio in 1–2/7–9 Hz (L1279) and the logarithm of Lg amplitude ratio in 1–2/6–8 Hz (L1268). The maximum amplitude ratio Lg/P appears to be poorly related to group distinctions taken on an individual basis. We observed that the combination of

various discriminant parameters in a discriminant function defining a separating hyperplane improves the separation between explosion and earthquake populations and, consequently, enhances the discriminant power of our analysis.

We obtained a separate discriminant function for each station and for the Sonseca Array, and computed them jointly for new event identification. This improved the separation of the properties of explosions and earthquakes, and enhanced the classification results. When applying MSWM to the total sample, we noticed that disagreements between analysts and the MSWM output were greatest in the case of explosions, as previously noted by RUEDA (2006). We believe that this could be due to the fact that low-magnitude seismic events occurring in daytime in areas where explosions are frequent are typically misclassified as explosions by analysts.

We achieved a 91 % success rate when classifying events using at least three stations with a 100 % confidence level. The summing of the discriminant function results for the different stations can help to reduce the effects of the radiation pattern of the seismic source; therefore, the more stations used, the more efficient will be the subsequent identification of the seismic event be (FÄH and KOCH 2002).

The algorithm proposed here could be used to purge the Spanish National Seismic Catalogue and to classify in real time new incoming events. Nevertheless, the method should only be applied to shallow events (<10 km focus depth) and with stations with an acceptable SNR (at least 20 dB). We observed that the choice of 20 dB as the upper level of SNR did not guarantee a good enough signal for data measurement in all stations. Some stations have systematically high levels of noise and have poor discriminant power. We detected that, when the SNR increases, no improvement is observed in the success rate for events recorded for at least two stations with 100 % confidence levels and remains constant at 90 %. We noted that the overall rate of success of the MSWM remains over 80 % until SNR is lower than 5 dB.

It is important to emphasise that the MSWM is an independent tool that can assist analysts' work and that its results should always be used critically.

We anticipate that the application of the proposed method will introduce a degree of error into

the final results due to a selection of the training events based on factors such as analysts' decisions. Other factors that could affect success rates include possible errors when choosing the training sample and the subsequent weakness of the discriminant function derived from the data, the presence of single-fired detonated explosions, and the consideration of non-shallow events that could hamper the waveform-based discrimination of the type that we used. The method proposed here is based on the main differences detected between delay-fired explosions and tectonic earthquakes. Although the majority of mining explosions are delay-fired, this method will have certain difficulties in classifying contained, single-fired or near-simultaneous delay-fired explosions. As mentioned above, it is important to take into account the lack of a sample of known membership (confirmed explosions or earthquakes), which could result in more classification errors depending on analysts' expertise in the events in question.

We believe that collecting new data from confirmed explosions and earthquakes would enhance the success of the proposed method although, due to the intrinsic nature of the problem, no complete identification of artificial events will be possible.

Acknowledgments

We would like to thank the Instituto Geográfico Nacional of Spain for providing the data for the study. We appreciate the valuable comments of the anonymous reviewers, which helped to improve the manuscript.

REFERENCES

- ALLMANN, B.P., SHEARER, P.M., and HAUSSON, E. (2008), *Spectral discrimination between quarry blasts and earthquakes in southern California*. Bull. Seismol. Soc. Am. 98, 2073–2079.
- ANDERSON, D.N., FAGAN, D.K., TINKER, M.A., KRAFT, G.D., and HUTCHENSON, K.D. (2007), *A mathematical statistics formulation of the teleseismic explosion identification problem with multiple discriminants*. Bull. Seismol. Soc. Am. 97(5), 1730–1741.
- ANDERSON, D.N., WALTER, W.R., FAGAN, D.K., MERCIER, T.M., and TAYLOR, S.R. (2009), *Regional Multistation Discriminants: Magnitude, Distance, and Amplitude Corrections, and Sources of Error*. Bull. Seismol. Soc. Am. 99(2A), 794–808.

- BAUMGARDT, D.R., and ZIEGLER, K.A. (1988), *Spectral evidence for source multiplicity in explosions: application to regional discrimination of earthquakes and explosions*, Bull. Seismol. Soc. Am. 78, 1773–1795.
- BENNETT, T.J., and MURPHY, J.R. (1986), *Analysis of seismic discrimination capabilities using regional data from western United States events*, Bull. Seismol. Soc. Am. 76, 1069–1086.
- CHERNOBAY, I.P., and GABSATAROVA, I.P. (1999), *Source classification in the Northern Caucasus*, Physics Earth and planet. Interiors, 113, 183–201.
- DOWLA, F.D., TAYLOR, S.R., and ANDERSON, R.W. (1990), *Seismic discrimination with artificial neural networks: preliminary results with regional spectral data*, Bull. Seismol. Soc. Am. 80, 1346–1373.
- ESPOSITO, A.M., GIUDICEPIETRO, F., SCARPETTA, S., D'AURIA, L., MARINARO, M., and MARTINI, M. (2006), *Automatic discrimination among landslide, explosion-quake, and microtremor seismic signals at Stromboli Volcano using neural networks*, Bull. Seismol. Soc. Am. 96, 1230–1240.
- FÄH, D., and KOCH, K. (2002), *Discrimination between earthquakes and chemical explosions by multivariate statistical analysis: A case study for Switzerland*, Bull. Seismol. Soc. Am. 92, 1795–1805.
- FARAHANI, J.V., ZARE, M., and LUCAS, C. (2012), *Adaptive neuro-fuzzy interference systems for semiautomatic discrimination between seismic events: a study in Tehran region*, J. of Seismol. 16, 291–303.
- FISHER, R.A. (1936), *The use of multiple measurements in taxonomic problems*, Annals of Eugenics, 7, 179–188.
- HEDLIN, M.A., MINSTER, J. B., and ORCUTT, J.A. (1989), *The time-frequency characteristics of quarry blasts and calibration explosions recorded in Kazakhstan, USSR*, Geophys. J. Int. 99, 109–121.
- KOCH, K. (2002a), *Classification of local and regional events in central Europe based on estimates of S-wave spectral variance*, Geophys. J. Int. 151, 196–208.
- KOCH, K. (2002b), *Seismic Event Identification of Earthquakes and Explosions in Germany Using Spectral Lg Ratios*. Pure appl. Geophys. 159, 759–778.
- KOCH, K., and FÄH, D. (2002), *Identification of earthquakes and explosions using amplitude ratios: the Vogtland area revisited*. Pure appl. Geophys. 159, 735–759.
- MEZCUA, J., RUEDA, J. (1997), *Seismological evidence for a delamination process in the lithosphere under the Alboran Sea*, Geophys. J. Int., 129, F1–F8.
- MULLER, S., GARDA, P., MULLER, J.D., and CANCI, Y. (1999), *Seismic events discrimination by neuro-fuzzy merging of signal and catalogue features*, Physics Chem. Earth. 24, 201–206.
- MURPHY, K.R., MAYEDA, K., and WALTER, W.R. (2009), *Lg-coda methods applied to Nevada Test Site events: Spectral peaking and yield estimation*, Bull. Seismol. Soc. Am. 99(1), 441–448.
- POMEROY, P.W., BEST, W.J., and MCEVILLY, T.V. (1982), *Test Ban Treaty verification with regional data: a review*, Bull. Seismol. Soc. Am. 72, S89–S129.
- RUEDA, J. (2006). *Discriminación sísmica mediante el análisis de las señales generadas por explosiones y terremotos. Aplicación a la región suroeste de Europa-Norte de África*, Ph.D. Thesis, Universidad Politécnica de Madrid. Pp. 450.
- SCARPETTA, S., GIUDICEPIETRO, F., EZIN, E. C., PETROSINO, S., DEL PEZZO, E., MARTINI, M., and MARINARO, M. (2005), *Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks*, Bull. Seismol. Soc. Am. 95, 185–196.
- STUMP, B.W., HEDLIN, M.A., PEARSON, D.C., and Hsu, V. (2002), *Characterization of mining explosions at regional distances: implications with the international monitoring system*, Rev. Geophys. 40, 1011.
- SU, F., AKI, K., and BISWAS, N.N. (1991), *Discriminating quarry blasts from earthquakes using coda waves*, Bull. Seismol. Soc. Am. 81, 162–178.
- TAYLOR, S.R., SHERMAN, N.W., and DENNY, M. (1988), *Spectral discrimination between NTS explosions and western United States earthquakes at regional distances*, Bull. Seismol. Soc. Am. 78, 1563–1579.
- TAYLOR, S.R. (1996), *Analysis of high-frequency Pg/Lg ratios from NTS explosions and western U.S. Earthquakes*, Bull. Seismol. Soc. Am. 86, 1042–1053.
- WALTER, W.R., MAYEDA, K.M., and PATTON, H.J. (1995), *Phase and spectral ratio discrimination between NTS earthquakes and explosions. Part I: Empirical observations*, Bull. Seismol. Soc. Am. 85, 1050–1067.
- WÜSTER, J. (1993), *Discrimination of chemical explosions and earthquakes in central Europe—a case study*, Bull. Seismol. Soc. Am. 83, 1184–1212.
- WÜSTER, J. (1995), *Discrimination of earthquakes and industrial explosions in the Vogtland Region and NW-Bohemia*, Ph.D. Thesis, Ruhr-University Bochum, Series A, No.42.