DEVELOPMENT OF A NETWORK DATA SET FOR EVALUATING DETECTION AND NETWORK PROCESSING PERFORMANCE

Benjamin Kohl, Theron J. Bennett, István Bondár, Brian Barker, Walter Nagy, Colin Reasoner, Hans Israelsson, and Paul Piraino

Science Applications International Corporation

Sponsored by Army Space and Missile Defense Command

Contract No. DASG60-03-C-0009

ABSTRACT

Practical implementations of new or improved monitoring technologies, such as signal detectors, network phase association algorithms, location and event identification methods, rely on quantitative assessments of performance such as detection probabilities and false alarm rates. These types of performance metrics are typically obtained through experiments using data sets constructed from archival data records. However, such experimental data sets implicitly contain signal and event recordings from numerous unknown sources (e.g., small earthquakes not reported in published local, regional, or teleseismic bulletins) potentially contaminating the data set and complicating the interpretation of processing results. Furthermore, they are only representative of events and station network characteristics contained during the time interval of the archival data. Our objective is to develop an experimental network data set in which all the target signal and event detections are known and ultimately to extend those results to represent expected network data from potential surrogate events and stations, which may not be included in the historical archive. To achieve this objective, we have been developing the framework for synthesizing a database including continuous waveform data for a network of seismic and infrasound stations relevant to nuclear explosion monitoring which contains signals from actual events, scaled to various sizes, and embedded in a variety of background noise.

Our initial focus for this study has been a large region in southern Asia (15°-45°N 50°-115°E). We have identified a network of 51 core seismic and infrasound stations, most useful for monitoring this region; and we have been collecting waveform data from those stations to represent background noise and signals from historical nuclear explosions as well as earthquakes and seismo-acoustic sources. In constructing the data for background noise, we are seeking to form long, continuous waveforms of detection-free clean noise spanning several days into which we can then embed real event signals and signals which have been scaled down on the basis of source scaling predictions to magnitudes representing lower levels. Formation of clean noise waveforms has required meticulous analysis to exclude time-windows with phase arrivals predicted from global and regional seismic bulletins as well as phases picked by standard signal detectors. Resulting noise segments have been carefully merged together to produce several days of continuous clean noise waveforms while maintaining basic noise attributes with respect to overall level and seasonal, weekly, and diurnal variations. From our effort to date, we have generated clean noise waveforms of two-days duration, as well as reversed noise waveforms of similar duration, for 42 of the seismic stations.

We have assembled the seismic signal waveforms from 6 underground nuclear explosions and approximately 100 well-recorded earthquakes with high signal-to-noise ratio (SNR) which occurred in southern Asia along with seismo-acoustic signals from 23 mine blasts and one bolide recorded by infrasound stations in Mongolia and Kazakhstan. We have been testing and employing frequency-dependent explosion (e.g., Mueller/Murphy) and earthquake (e.g. Brune with both inverse cube- and quad-root corner frequency dependence on moment) source scaling models to scale down the large, high-SNR events to small events covering a range of yields/magnitudes approaching the monitoring thresholds. In addition to describing target events for analyzing monitoring performance, the scaling/embedding process is also being used to represent potential sources of regional and teleseismic clutter signals, which increases processing complexity (while continuing to maintain control of the contributing sources) and provides a more realistic background condition than the clean noise scenario. Preliminary event detection experiments are quantifying the systematic time, amplitude and azimuth measurement biases that can be expected from low-SNR detections. Methodologies for analyzing the performance of the infrasound stations for monitoring seismo-acoustic events from the southern Asia source region are also being assessed.

OBJECTIVES

Our objective is to develop an experimental network data set in which target signal and event detections are known, as well as having realistic distributions of false or "clutter" detections and background noise characteristics. We are utilizing a variety of actual nuclear explosion, earthquake, mine blast and infrasonic event recordings and developing scaling and embedding algorithms to yield continuous waveforms with numerous target events at or near the detection threshold in southern Asia. This will allow for detection, location and identification experiments utilizing the known characteristics of small events under realistic background noise and seismicity conditions. The background noise, scaled signals, embedded waveforms and relevant meta-data from this effort are available to the monitoring research community via the mechanisms of the Research and Development Support Services (RDSS) web site (http://www.rdss.info/, Woodward et al. 2005).

RESEARCH ACCOMPLISHED

The basic framework for the comprehensive network data set was developed previously (Kohl et al., 2004) and is depicted in Figure 1. It involves taking well-recorded, high signal-to-noise ratio (SNR) signals, scaling them down to various sizes based on source theory and embedding them in a variety of background noise conditions. We are scaling and embedding nuclear explosion, earthquake, seismo-acoustic (e.g., mine blast) and infrasonic event recordings at levels spanning the detection threshold.

We assembled a background noise library of two days of continuous detection-free clean noise, two days of reversed clean noise and fours days of whole background noise for a core network of 42 stations. We constructed a signal library from 6 nuclear explosions, over 100 earthquakes, 23 mine blasts and one bolide. We developed source scaling models for nuclear explosions and earthquakes and scaled the nuclear explosion records and earthquakes to equivalent m_b ranging from 1.8 to 4.5. We embedded the scaled signals several hundred times in varying noise conditions and conducted a number of signal detection experiments yielding Receiver Operating Characteristic (ROC) curves and quantitative assessments of arrival time, azimuth and amplitude biases as a function of SNR for selected stations. We continue to develop scaling models for infrasound signals and plan on conducting experiments to demonstrate the utility of this approach to assess network processing performance.

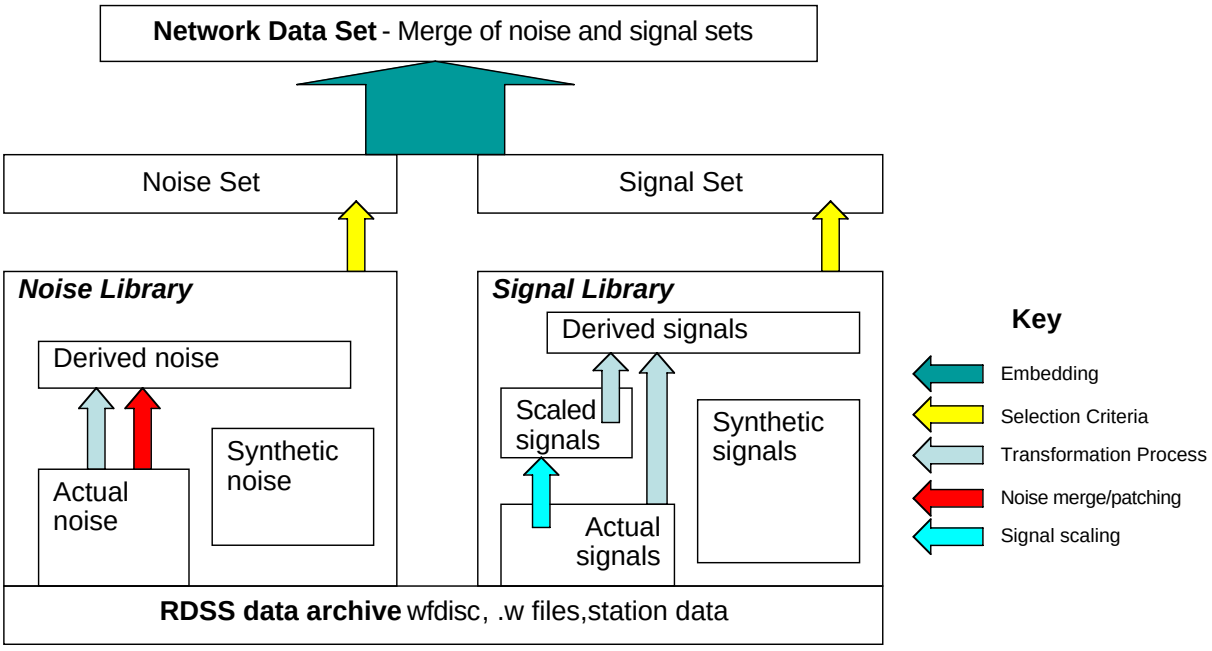


Figure 1. Basic framework for constructing a network data set for systematic testing and evaluation of new or improved monitoring technologies. Scaled signals from a variety of sources are embedded in a variety of noise conditions to construct a data set in which target detections have been characterized.

Noise and Signal Library

In the past year we completed the construction of three background noise datasets for a network of 42 stations useful for building experimental data sets for testing monitoring capabilities in southern Asia (15°-45°N 50°-115°E). The first of these data sets (clean-noise, two days) was constructed by stitching together detection free segments. To insure that the stitched clean-noise realistically represented actual noise, including known seasonal and diurnal variability, we first assembled a set of reference spectra for each station and channel spanning the range of variability for each station (Figure 2). For example, for stations that exhibited strong diurnal variability in the noise levels, we computed a separate reference spectrum for every hour of the day. Only those detection-free segments derived from the same hourly span, *and* whose spectra matched the reference spectra were used. To minimize the effects of the merging process we used 10 to 120 second tapers at the ends and overlapped neighboring segments. All the channels of stations and arrays were merged consistently in time to retain the noise coherency characteristics originally present in the data.

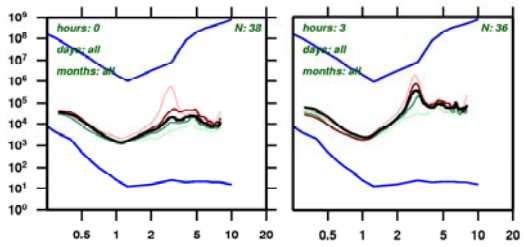


Figure 2. Example reference spectra for hours 0 and 3 for the CM16 element of CMAR.

The second noise dataset (reversed, two days) was constructed by simply time-reversing the clean background noise, thus also yielding two days of continuous noise. The third noise dataset (whole-background) was constructed by simply extracting four days of raw continuous data (June 1998). This third dataset is analogous to what is normally used in signal processing experiments. As a quality control measure, and to establish a baseline background detection rate, we ran standard signal processing (DFX) against the clean noise. Despite the fact that the clean noise was constructed from detection-free segments, low-level detections were still made against the clean and reversed noise. On average the detection rate against the clean-noise was 20% of that against whole-background noise and reversed noise had a detection rate of about 30% of the whole-background noise. Figure 3 shows the waveform and the spectrogram of a waveform stitched from detection-free segments.

In the past year we assembled a signal library from the waveforms of 6 historical nuclear explosions, more than 100 earthquakes, 23 mine blasts and one bolide that occurred in central and southern Asia. We are computing a wide variety of signal characteristics (e.g. Figure 4) on the original event records, and the scaled signals.

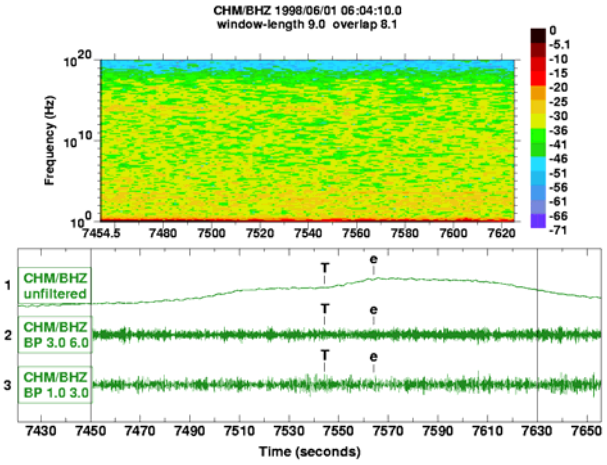


Figure 3. Spectrogram of merged clean-noise around a stitching boundary (marked start – "T" and end – "e" markers).

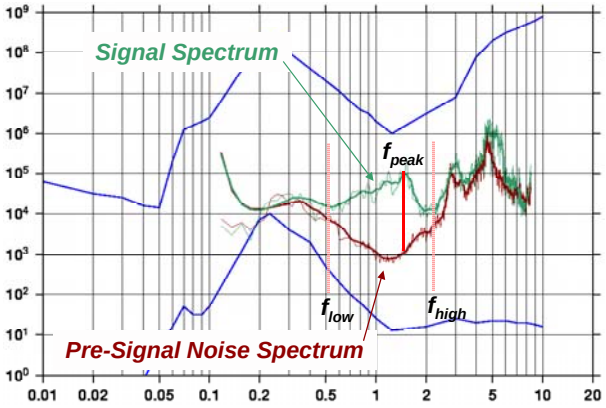


Figure 4. Schematic showing the signal bandwidth and peak-frequency measurements routinely computed for the signal library.

Explosion and Earthquake Scaling

One of the principal objectives of this project is to provide realistic assessments of monitoring performance for smaller events for which the signal detections from station networks are often incomplete. To better understand the factors affecting this performance for our southern Asia study area, we have been using source scaling theory to scale the signals from larger events. This scaling introduces a frequency-dependent change in signal amplitudes.

For explosion scaling we have been using the Mueller-Murphy (MM) model which has been validated over the years for a range of explosion observations (Mueller and Murphy, 1971; Murphy 1977). Alternative models (e.g., vonSeggern and Blandford, 1972) would be expected to produce very similar predicted behavior. The MM model is formulated in terms of explosion yield (W) and provides an expression of the P-wave spectrum as a function of source media properties, explosion yield, depth, and empirical constants corresponding to different geologic emplacement media. For southern Asia we use the explosion scaling relations for granite, which worked well for nuclear explosions at Semipalatinsk and Lop Nor test sites. For scaling the explosions in terms of body-wave magnitude m_b, we use the relation $m_b = 4.45 + 0.75 \log W$, which has been previously validated for these test sites, to convert to yields. Our MM model results were verified by comparing observations of Pn spectral ratios from nearly co-located nuclear explosions with the predictions based on the source scaling theory. Some of these comparisons showed very good matches, although in other cases there appeared to be corner frequency differences, which may require future modifications to some of the model parameters.

To test the explosion scaling model, we applied the MM scaling procedures to scale the signals recorded on a network of regional and teleseismic stations from 6 southern Asia underground nuclear explosions, down from their original magnitudes ($4.5 \le m_b(\text{REB}) \le 6.0$) to a range of lower magnitudes ($m_b(\text{REB}) = 4.5$ and in 0.1 magnitude unit steps from 4.0 m_b(REB) to 1.8 m_b(REB)). We then applied standard IDC processing to the scaled signals to measure initial P amplitudes and associated periods, which would be used for computing station magnitudes. The results are presented in Figure 5a, in which we plot the observed magnitude differences at each stations, $\log A_i/T_i$(original unscaled) – $\log A_i/T_i$(scaled) where i is a station index, versus the target m_b difference, m_b(original for the network) – m_b(target for the network). Obviously the ideal result would be for the observed magnitude difference at each station to equal the target magnitude difference, and this is achieved quite well in the explosion scaling measurements in Figure 5a. The observations are scattered around a line with a slope approximately equal to 1.0. With the exception of a few outliers related to data quality issues, the scatter in the observations is less than half of a magnitude unit, and the least-squares linear fit to the observations is 1.02 with only a slight bias indicated at the largest target magnitude differences. Thus we conclude that the MM explosion scaling procedure appears to be performing as expected over a fairly large range of magnitudes.

For earthquake scaling, we began with a Brune ω^2-source model. For this model the corner frequency is proportional to velocity of the source medium and inversely proportional to a source dimension term, which scales with moment. In scaling the earthquake signals, we consider both cube-root (as indicated in original models by Brune 1970, Hanks and Bakun 2002) and quad-root (as suggested by Mayeda and Walter 1996 amongst others) for

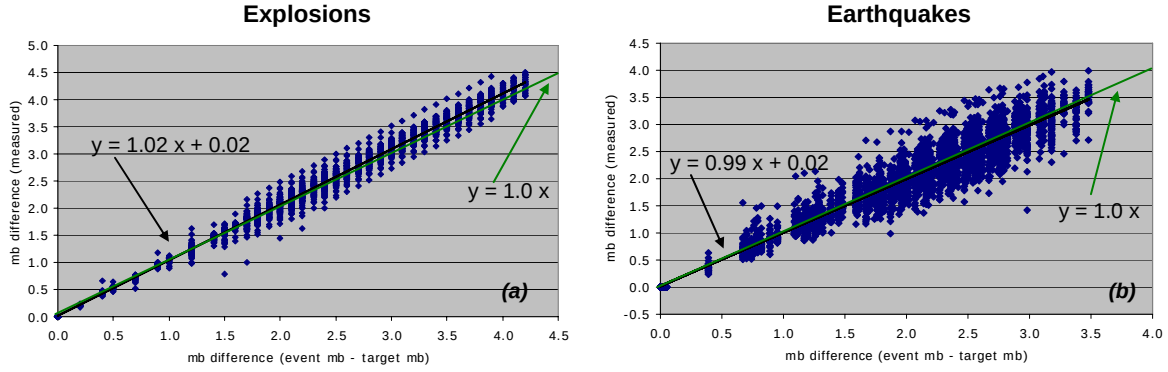


Figure 5. Results of magnitude differences for P and Pn signals measured from routine processing of the observations from 6 southern Asia nuclear explosions scaled using MM explosion source scaling (a) and from 17 southern Asia earthquakes scaled using our preferred cube-root earthquake model (b).

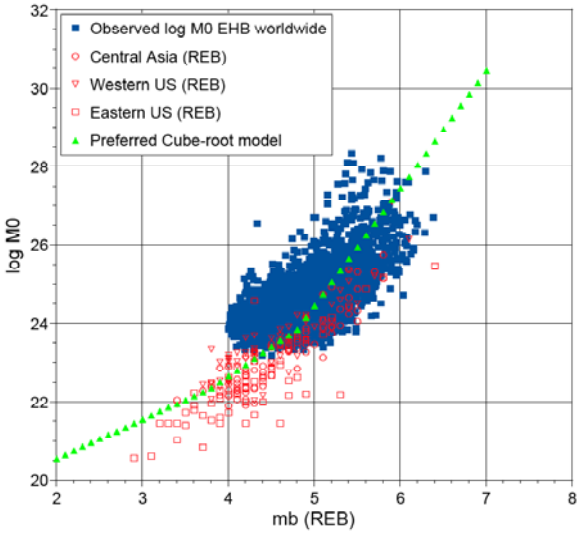


Figure 6. Log M0 versus mb(REB) for a large global earthquake sample reported by EHB and elsewhere.

corner frequency dependence on seismic moment. Our analyses to date focus on the model with cube-root corner frequency dependence. In the earthquake model we followed the approach of Hanks and Bakun (2002), using the definition of moment magnitude, M_w, to establish the relationship to source moment – i.e., $\log M_0 \equiv 1.5\ M_w + 16.05$. We began with a simple linear model for relating M_w to m_b. We drew upon observations from the global EHB (Engdahl et al. 1998) earthquake sample for large events (m_b(REB) > 4) along with observations from smaller events for selected areas reported by Patton (2001), which were adjusted to equivalent REB m_b's (assuming $m_b \approx m_b$(REB) + 0.3). Although the data scatter (Figure 6) would appear to permit a linear M_w-vs-m_b model (i.e. essentially a straight-line relationship between $\log M_0$ and m_b), the implied effects on spectral behavior are not realistic. In particular, for small events with magnitudes measured from signals with frequencies below the corner frequency, a one unit change in m_b should correspond to a factor of ten change in M_0. To meet this objective we would need to have $M_w \propto 2/3 m_b$ – i.e. $\log M_0 \propto 1.0 m_b$. The observations in Figure 6 cannot support such a slope over magnitudes $3 \le m_b(\text{REB}) \le 6$. We, therefore, decided to investigate a model for which the M_w-vs-m_b relationship is nonlinear; Taylor et al. (2002) reached a similar conclusion and developed a model only slightly different from those described below. In particular, we sought to determine an earthquake model for which $\log M_0$ scales directly as the m_b difference for small events, produces $M_w \approx m_b$ over some intermediate range in magnitude, and has M_w greater than m_b for larger events (where we know m_b is saturated).

As a preliminary validation test to further constrain our earthquake model, the earthquake scaling relations were used to scale down the signals from selected samples of large southern Asia earthquakes which were well recorded at regional and teleseismic stations. Just as for the scaled explosion signals above, we compared the A/T measurements from the scaled P and Pn signals to the target values for the earthquakes, m_b(REB) = 2.5 – 4.5. When this scaling model was applied to the signals from 17 southern Asia earthquakes, the A/T measurements were in very good agreement with the expectation (Figure 5b). There appears to be very little indication of bias in the results; the slope of the least-squares straight line fit to the observations is 0.99 and the intercept is 0.02. Scatter about the line is about 0.7 magnitude units, only slightly greater than the scatter seen in the corresponding explosion measurements. We conclude that the cube-root model combined with our $\log M_0 - m_b$ relation (Table 1), provides a reasonable procedure for scaling the southern Asia earthquakes. We are continuing to look at additional validation with spectral ratios (e.g., Figure 7) from nearly co-located events and to evaluate alternatives, including an earthquake model with quad-root corner frequency dependence.

Table 1: Preferred moment-vs-m_b relationship

$\log M_0 - m_b$ relation for cube-root model	
$\log M_0 = m_b$(REB) + 18.55	m_b (REB) ≤ 3.8
$\log M_0 = 1.5\ m_b$ (REB) + 16.65	3.8 ≤ m_b (REB) ≤ 4.8
$\log M_0 = 3.0\ m_b$ (REB) + 9.45	m_b (REB) ≥ 4.8

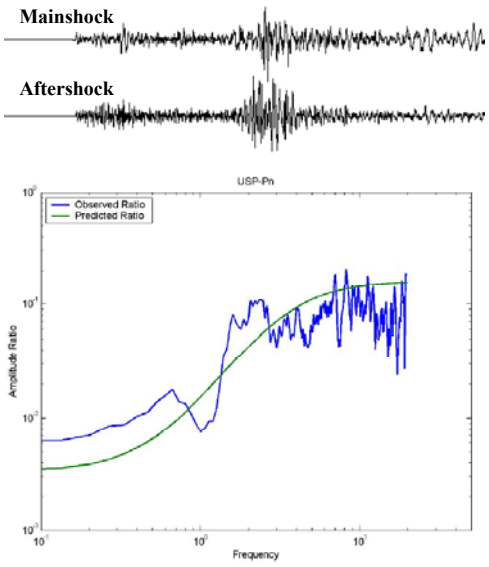


Figure 7. Spectral ratio of a mainshock/aftershock pair in southern Asia, showing good agreement with the cube-root model (green).

Signal Processing Experiments

Controlled experiments with the embedded data offer several opportunities for evaluating algorithms for station signal processing with regard to signal detection as well as estimation of signal parameters. We give examples below with results from initial seismic signal processing experiments. Test data sets were constructed from background noise and scaled signals for three stations: CMAR, FINES, NIL. Signals scaled to magnitudes between m_b =1.8 - 4.5 in increments of 0.1 magnitude units (m.u.) from three large and similar (m_b ~ 5.5) underground nuclear explosions at the Lop Nor test site were embedded in noise of different types – whole background, clean, and reversed clean. The test data sets were processed with DFX signal detection and parameter extraction programs with configurations currently employed at the IDC.

Signal Detection Probabilities

Detection probabilities as a function of m_b were calculated as the ratio of the number of detected signals/total number of embedded signals of magnitude m_b. Figure 8 compares the probabilities for the three stations as a function of m_b. The probability curves are in reasonable agreement with a Gaussian cumulative distribution functions with mean values corresponding to the 50% detection probability threshold, and the standard deviation is a measure of noise amplitude variation. The data in Figure 8 thus support the common assumption of network detection simulations that incremental detection probabilities as a function of m_b can be approximated by cumulative Gaussian distribution functions (Kvaerna and Ringdal, 1999).

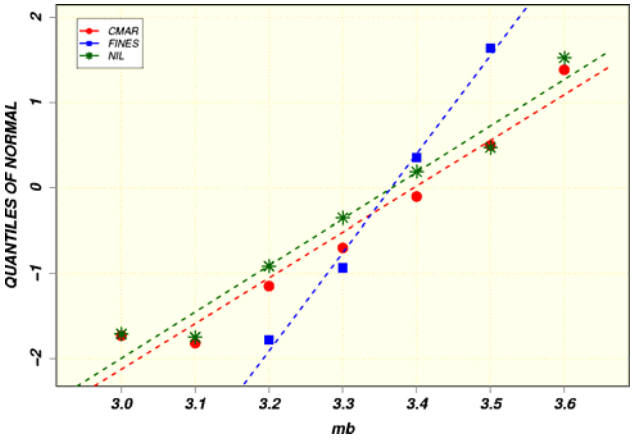


Figure 8. Detection probabilities as a function of m_b for the stations CMAR, FINES, and NIL. A Gaussian scaling is used for the detection probabilities on the vertical axis so that true Gaussian data would follow a straight line. The data points for the detection probabilities follow the fitted lines closely suggesting a that the detection probability as a function of m_b can be approximated with a cumulative Gaussian distribution.

False alarm rates

Estimates of false alarm rates from the rates of unassociated detections become uncertain when based on observations of real data; such unassociated detections include detections of signals from events of unknown origin as well as detections triggered by signal coda. Clean noise, free of signals, can provide a more robust estimate of the false alarm rate, that is unbiased both due to signals of unknown origin and due to coda detections.

In Table 2 we compare the false alarm rates of the three types of noise samples. The false alarm rates for clean and clean-reversed noise with no embedded signals represent genuine false alarms of the detector configuration, whereas for the real noise data some portion of the detections are from signals of seismic events. The number of detections of the whole background noise data that could be associated (using a 4 second allowance for initial P) with seismic events in the ISC database was small (< 10%). The ISC reported close to 200 worldwide events/day for the time period analyzed here, but its event catalog is incomplete for small events. Furthermore, secondary phases or coda detections were not considered in the association, so total detection rates, with ISC associations subtracted, still overestimate the actual false alarm rate. The difference is most striking for NIL, for which the unassociated rate of the clean noise is about a factor of eight lower than that for the real data, while this factor is around 5 for FINES and a little more than 2 for CMAR.

Table 2. Daily Unassociated Detection Rates for Different Noise Types

Station	Clean[1]	Reversed Clean	Whole Background
CMAR	190 | 37	160 | 29	411
FINES	20 | 6	30 | 6	165
NIL	62 | 7	85 | 3	507

1. The two numbers represent the number of unassociated detections in clean noise with no embedded signals/number of additional unassociated detections when signals were embedded.

With signals embedded in the clean noise the false alarm or unassociated rate goes up slightly due to additional coda detections that normally would have been assumed to be false alarms. The rates of such additional unassociated detections are also given in Table 2. The median of this increase in false alarm or unassociated rates when signals are embedded for the clean and reversed clean noise is about 20%.

ROC curves

The performance of a signal detector is often defined by the so-called Receiver Operating Characteristic, ROC, which describe the trade-off between signal detection probability and false alarm probability (Van Trees, 1968). Using scaled signals embedded in clean noise affords the opportunity to construct ROC curves, which is difficult, if not impossible, to do from real data. An example of an ROC curve constructed from detection experiments with the array FINES is shown in Figure 9. The histograms to the left show the distributions of the logarithm of the short term/long term averages (sta/lta), or SNR, for noise (red) and for signals (plus noise) (in blue) of the same size (m_b=3.30). Note that the histograms for the signals include data for signals that were not detected. SNR values for undetected signals could be calculated because of their known embedding times. Gaussian curves draped on the histograms show reasonable agreement with the empirical SNR distributions for both noise and signal data. The detector triggers if the SNR is above a preset threshold and the default threshold used in the processing experiment is marked as a vertical dashed line (green) to the right. A significant portion of the SNR distribution for the explosion signals is below the threshold and hence went undetected. If the threshold were lowered to, e.g., the mean of the SNR distribution (black dashed line to the left marked "NEW THRESHOLD"), the probability to detect an explosion signal would increase significantly (to 50%) without significantly increasing the probability of triggering on noise, or of a false alarm.

In the right diagram the two Gaussian distributions for noise and signal detection probabilities were combined to a standard ROC curve. The “default” and the "new" thresholds are marked showing that by lowering the threshold the false alarm probability would not change much, whereas the detection probability would go up from less than 20% to 50%. It should be noted that the data in Figure 9 represent beam forming with steering of a single beam that is optimum for FINES and the Lop Nor test site.

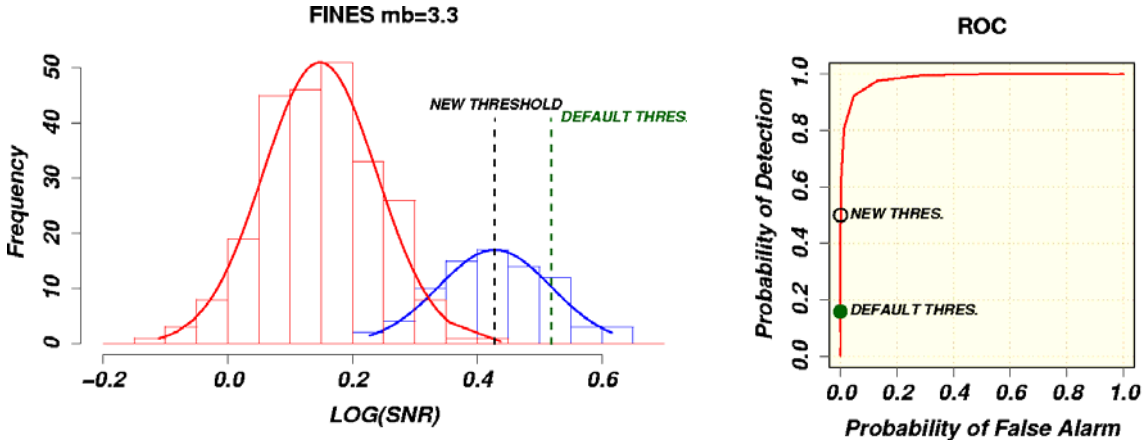


Figure 9. The histograms in the left diagrams for SNR (log scale) of noise (red) and of signals (blue) are used to construct the ROC curve in the right diagram. The default threshold used in the detection experiment is marked as a green dashed line in the left diagram and as a filled dot in the ROC curve. The ROC curve shows that lowering the default threshold (marked as “NEW THRES” in the diagrams) has little effect on the probability of false alarm, but would increase the probability of detection from less than 20% to 50%.

Estimation of Signal Parameters

Apart from detection the processing algorithm that we use (DFX) estimates signal parameters that characterize a detection, such as time of arrival, SNR, amplitude, and for arrays, slowness vector. With the ground truth of the characteristics of embedded signals, the distributions of errors in parameter estimates can be estimated with accuracy. Figure 10 summarizes some statistics of estimation errors in amplitude/period ratios (to the left) and

azimuths of slowness vectors (to the right) for the DFX algorithm for experiments with FINES. The boxplot to the left shows that the automatically estimated amplitude/period ratio becomes increasingly positively biased with decreasing SNR of the detected signal. The bias at the threshold of the detector (log SNR=0.5) is, on average, about 0.3 m.u. The bias is probably caused by low frequency noise that the automatic algorithm cannot account for as the amplitude/period ratios for the downscaled signals were unbiased prior to embedding in the noise (see Figure 5). The bias raises the question whether amplitude/period measurements at low SNR should be used uncorrected for magnitude estimation.

The errors in azimuth estimates (right diagram in Figure 10) grow fast with decreasing SNR. A bias is not as clearly defined as for the amplitude/period ratios, but the scatter increases drastically as indicated by the widening of the boxes which represent 50% of the data around the median. The dashed red lines outline the 50% limits of the azimuth uncertainty (measurement errors) assigned by the automatic algorithm, which clearly underestimates the spread of the actual azimuth errors at low SNRs around 4 and below. An analysis of the empirical distributions of the amplitude and azimuths errors revealed that they are well represented by a Gaussian for all SNRs.

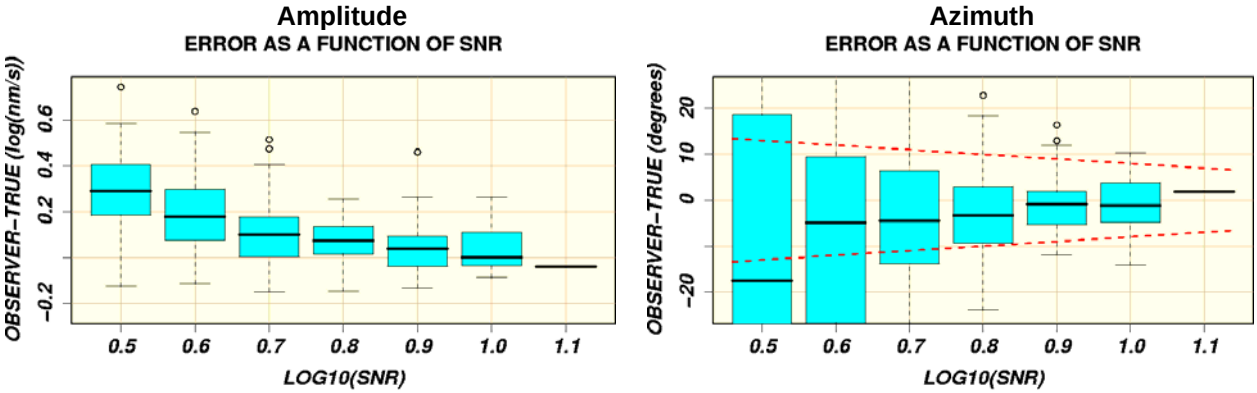
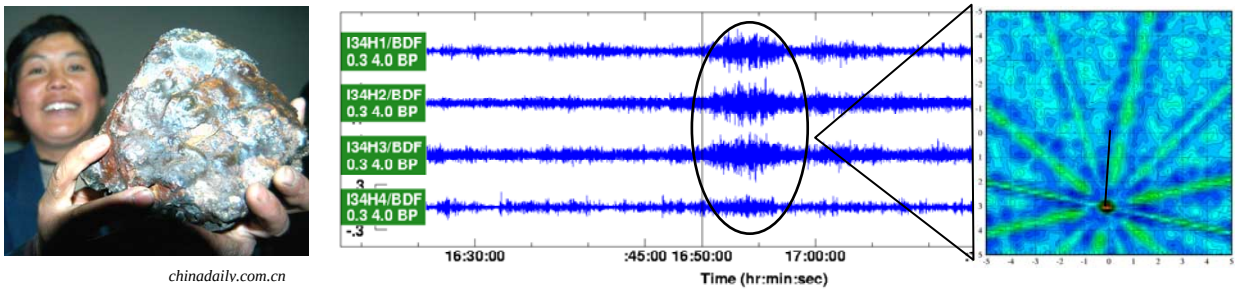


Figure 10. Errors in amplitude/period ratios (left) and azimuth (right) of the automatic DFX algorithm for signals embedded in clean noise of the array FINES. The boxplots show the errors as a function of log SNR; the dashed lines for azimuth in the boxplot to the right outline the 50% limits of errors assigned by the automatic algorithm (measurement errors).

Seismo-Acoustic Event Database and Scaling

Detection of infrasound signals from seismo-acoustic events (e.g., mine blasts) can provide valuable insights into a variety of research topics, such as infrasound propagation, seismic vs. acoustic coupling and the effects of wind-generated noise on infrasound detection. It is our objective to provide to the nuclear monitoring research community a controlled data set where scaled infrasound signals are embedded in a variety of noise conditions. This will be particularly useful for evaluations of detection algorithms. Currently evaluations with respect to variations in ambient noise are dependent on finding data sets where signals from known sources are detected in a variety of noise conditions. Given the current dearth of such infrasound ground truth, a comprehensive evaluation of infrasound detection capabilities using actual recordings is difficult.

We are following the same basic methodology for infrasound that we used for seismic data (described above). Namely, we are building a background noise database from historical recordings, collecting high-SNR infrasound signals from known sources, applying a source scaling function to the recorded data, and embedding the scaled signals in the background noise. In contrast to nuclear explosion or earthquake source theory, general theoretical models cannot be used to predict the scaling of infrasound signals from mine blasts. This is in large measure due to the fact that differing mine practices, acoustic coupling and other local conditions vary immensely from mine to mine and region to region. Therefore we are following an empirical approach to assess the scaling of infrasound signals and limiting the scope of the scaled and embedded data. We are building the data set such that each scaled/embedded event represents effectively the same source and meteorological conditions as the original event, with only the signal amplitude (by no more than one order of magnitude) and the ambient noise conditions varying.



chinadailv.com.cn

Figure 11. Bolide near Lanzhou, China on December 11, 2004 recorded on the IMS infrasound array I34MN and reported in the local media.

We obtained seismo-acoustic recordings (seismic and infrasonic signals from the same source) from 23 mine blasts (c) and 1 bolide (Figure 11) in central Asia for inclusion in our signal library. The mine-blast data offer an opportunity to look for evidence of non-linear source scaling in the infrasound data. b shows a section of 9 infrasound signals recorded at I31KZ for mines at ranges between 50 and 400 km. Traces 1 and 3 originating from the same mine, however, had signal amplitudes differing by a factor of about 10. Similarly traces 2 and 4, again originating from the same mine, had signal amplitudes differing by a factor of about 3. In both cases, spectra ratio comparisons showed that the signals scaled approximately linearly with frequency, consistent with the visual similarity of the waveforms. Thus, based on this admittedly limited data set we are proceeding with a linear scaling function and embedding these in a variety of noise conditions.

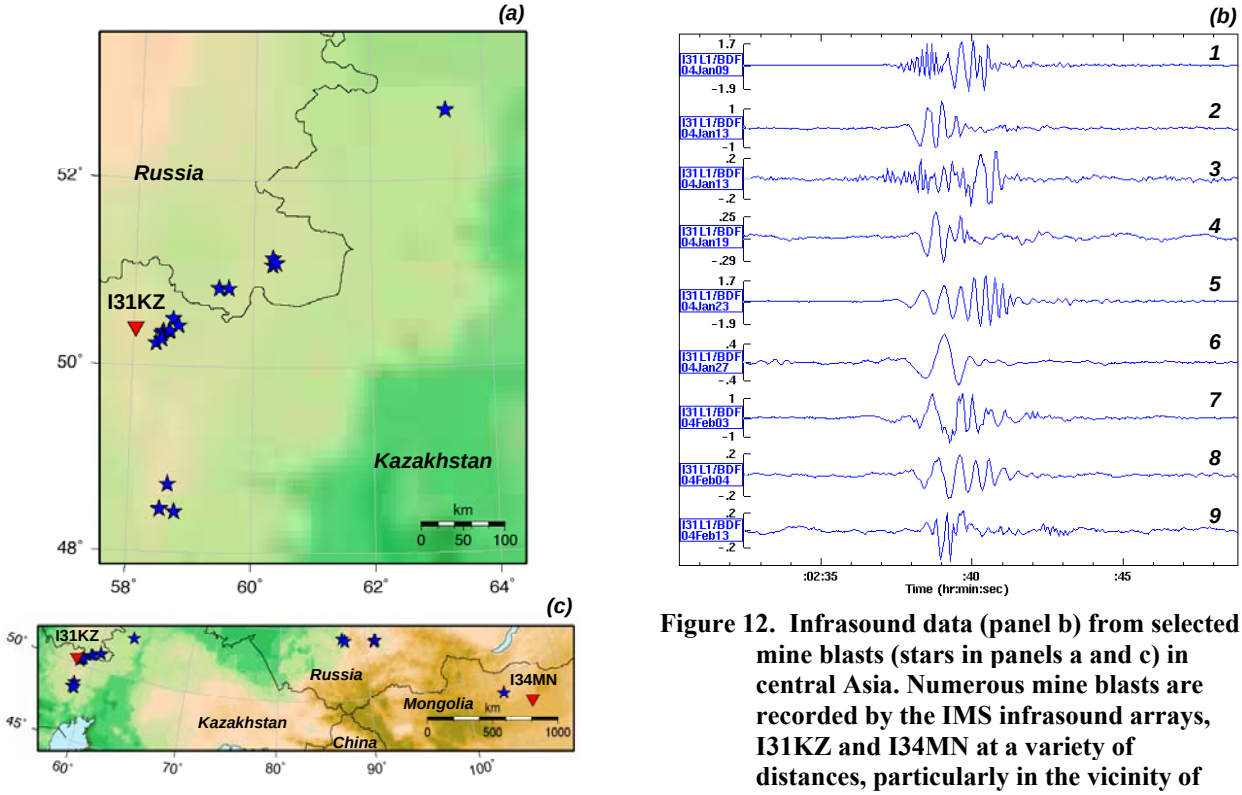


Figure 12. Infrasound data (panel b) from selected mine blasts (stars in panels a and c) in central Asia. Numerous mine blasts are recorded by the IMS infrasound arrays, I31KZ and I34MN at a variety of distances, particularly in the vicinity of I31KZ (panel a).

CONCLUSIONS AND RECOMMENDATIONS

We developed a network data set where scaled signals were embedded in varying background noise conditions, including carefully constructed clean noise. Analysis of reprocessed scaled data show that both a Mueller-Murphy model for nuclear explosions, and modified Brune model with cube-root corner frequency scaling and an empirical $\log M_0$-m_b relation for earthquakes, give consistent amplitude scaling for events in southern Asia.

The detection experiments we executed illustrate some of the benefits that using an embedded data set can provide for assessment of signal detection and characterization algorithms. We found clear evidence for an amplitude bias for low SNR signals and that measurement uncertainties for azimuth probably underestimate the error at low SNR. In general we conclude that experiments with an embedded data set can:

- Validate assumptions for detection probabilities underlying network simulations of real data
- Provide accurate estimates of station detection probability as a function of source strength
- Provide unbiased estimates of false alarm rates
- Accurately map ROC curves for relative assessment of signal detectors and detector configurations
- Estimate the statistical characteristics of errors of signal parameters such as arrival times, slowness vectors, and signal amplitudes.

The background noise, scaled signals, embedded waveforms and relevant meta-data from this effort are available to the monitoring research community via the RDSS web site (http://www.rdss.info/, Woodward et al. 2005).

REFERENCES

Brune, J. N. (1970), Tectonic stress and spectra of seismic shear waves from earthquakes, *J. Geophys. Res.* 75: 4997-5009.

Engdahl, E.R., R. van der Hilst & R. Buland (1998), Global Teleseismic Earthquake Relocation with improved Travel Times and Procedures for Depth Determination, *Bull. Seism. Soc. Am.*, 88: 722-743.

Hanks, T. C., and W. H. Bakun (2002). A bilinear source-scaling model for M-logA observations of continental earthquakes, *Bull. Seism. Soc. Am.* 92: 1841-1846.

Kohl, B., T. Bennett, I. Bondár, B. Barker, W. Nagy, C. Reasoner, and J. Hanson (2004), Development of a Network Data Set for Evaluating Detection and Network Processing Performance, in *Proceedings of the 26th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, LA-UR-04-5801: Vol. 2, pp. 725-734.

Kværna, T. and F. Ringdal (1999), Seismic Threshold Monitoring for Continuous Assessment of Global Detection Probability, *Bull. Seism. Soc. Am.*, 89: 946-959.

Mayeda, K., and W. R. Walter (1996), Moment, energy, stress drop, and source spectra of western United States earthquakes from regional coda envelopes, *J. Geophys. Res.* 101: 11,195-11,208.

Mueller, R. A., and J. R. Murphy (1971), Seismic characteristics of underground nuclear detonations, part 1: seismic spectrum scaling, *Bull. Seism. Soc. Am.* 61: 1675-1692.

Murphy, J. R. (1977), Seismic source functions and magnitude determinations for underground nuclear detonations, *Bull. Seism. Soc. Am.* 67: 135-158.

Patton, H. J. (2001), Regional magnitude scaling, transportability, and MS:mb discrimination at small magnitudes, *PAGEOPH* 158: 1951-2015.

Taylor, S. R., A. A. Velasco, H. E. Hartse, W. S. Phillips, W. R. Walter, and A. J. Rodgers (2002), Amplitude corrections for regional seismic discriminants, *PAGEOPH* 159: 623-650.

Van Trees, H. L., (1968), *Detection, Estimation and Modulation Theory*, John Wiley and Sons.

vonSeggern, D. H. and Blandford, R. (1972), Source time functions and spectra for underground nuclear explosions, *Geophys. J.* 31: 8-97.

Woodward, R., M. Skov, M. Bahavar, G. Davis and Y.-L. Kung (2005), Data and tools to support nuclear explosion monitoring research and development, in *Proceedings of the 27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies,* in current Proceedings.