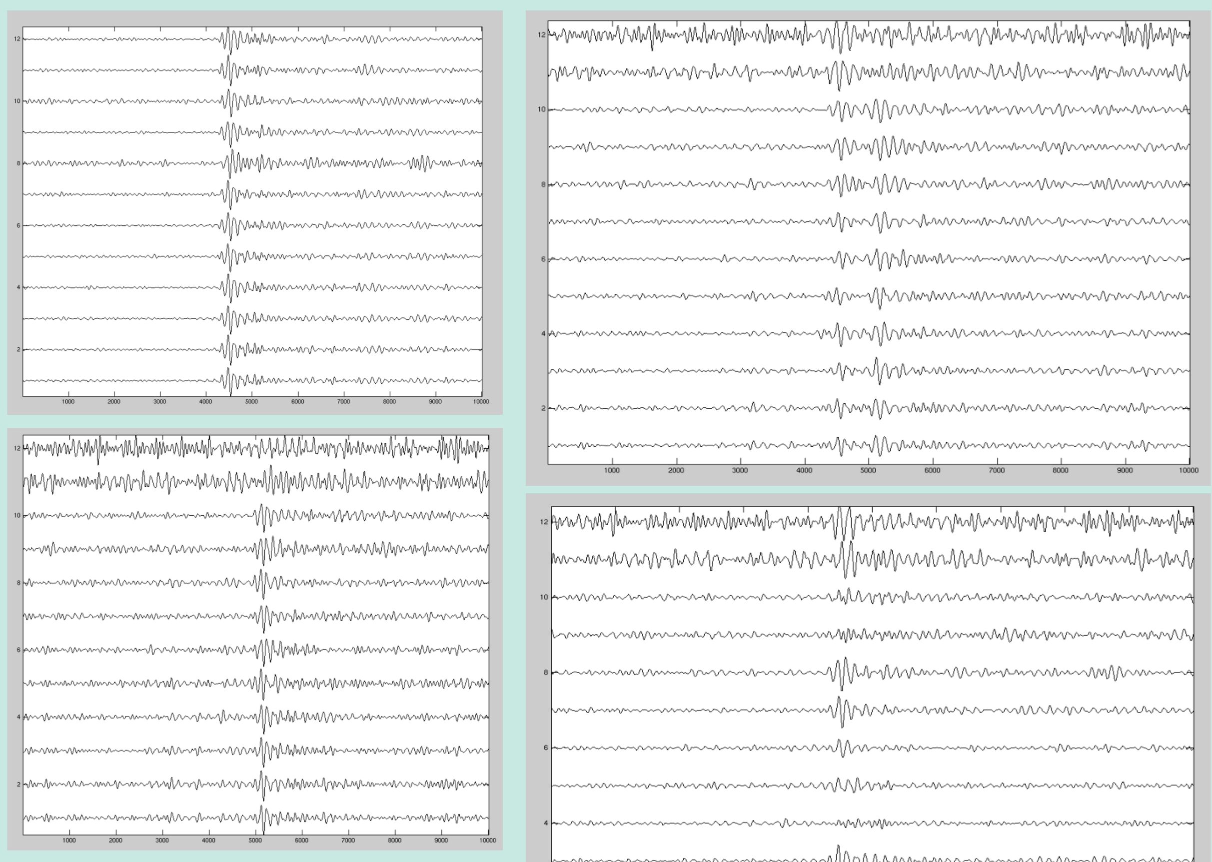
Blind Source Separation of Seismic Events with Independent Component Analysis: CTBT related exercise

Mikhail Rozhkov[1], and Ivan Kitov[2] (1 - International Data Centre, CTBTO, 2 - Institute of Geosphere Dynamics, RAS)

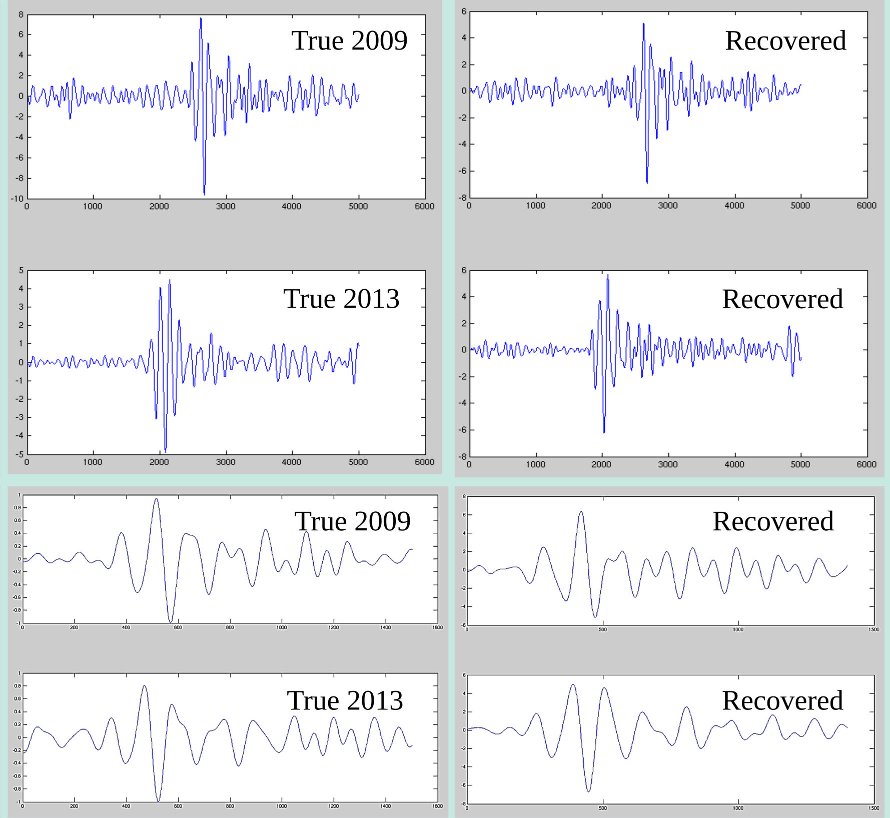
CTBTO PREPARATORY COMMISSION — preparatory commission for the comprehensive nuclear-test-ban treaty organization

Blind Source Separation (BSS) methods used in signal recovery applications are attractive for they use minimal a priori information about the signals they are dealing with. Homomorphic deconvolution and cepstrum estimation are probably the only methods used in certain extent in CTBT applications that can be attributed to the given branch of technology. However Expert Technical Analysis (ETA) conducted in CTBTO to improve the estimated values for the standard signal and event parameters according to the Protocol to the CTBT may face problems which cannot be resolved with certified CTBTO applications and may demand specific techniques not presently used. The problem to be considered within the ETA framework is the unambiguous separation of signals with close arrival times. Here, we examine two scenarios of interest: (1) separation of two almost co-located explosions conducted within fractions of seconds, and (2) extraction of explosion signals merged with wave-trains from strong earthquake. The importance of resolving the problem related to case 1 is connected with the correct explosion yield estimation. Case 2 is a well-known scenario of conducting clandestine nuclear tests. While the first case can be approached somehow with the means of cepstral methods, the second case can hardly be resolved with the conventional methods implemented at the International Data Centre, especially if the signals have close slowness and azimuth. Independent Component Analysis (in its FastICA implementation) implying non-Gaussianity of the underlying processes signal's mixture is a blind source separation method that we apply to resolve the mentioned above problems. We have tested this technique with synthetic waveforms, seismic data from DPRK explosions and mining blasts conducted within East-European platform as well as with signals from strong teleseismic events (Sumatra, April 2012 Mw=8.6, and Tohoku, March 2011 Mw=9.0 earthquakes). The data was recorded by seismic arrays of the International Monitoring System of CTBTO and by small-aperture seismic array Mikhnevo (MHVAR) operated by the Institute of Geosphere Dynamics, Russian Academy of Sciences. Our approach demonstrated a good ability of separation of seismic sources with very close origin times and locations (hundreds of meters), and/or having close arrival times (fractions of seconds), and recovering their waveforms from the mixture. Perspectives and limitations of the method are discussed.

DPRK 2006 and 2013 nuclear tests mixtures, recorded at the MHVAR array. Sampling rate is 200 Hz.



DPRK 2013 (top) and 2006 records at MHVAR array, Moscow region. Mixtures on the right: 3 (top) and 0.3 seconds signal separation



Original (left) and Recovered (right) signals for different time separation cases. Correlation coefficients (CC) is between restored and original signals are in a range of (0.84, 0.94). CC's variability is due to instability and ambiguity of the ICA procedures.

Independent Component Analysis

A Problem ("cocktail party"):

- ✓ Several Sources
- ✓ Several Sensors

Recover source signals given only mixtures

- ✓ No prior knowledge of sources or mixing matrix

- typical Blind Source Separation (BSS) case

Assumptions:

Source signals are statistically independent

- ✓ Knowing the value of one of the components does not give any information about the others

ICs have non-gaussian distributions

- ✓ Initial distributions unknown.
- ✓ At most one Gaussian source

Recovered sources can be permutated and scaled

Independence

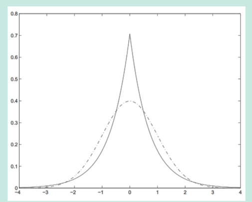
- *If variables are independent, they are uncorrelated*
- *Uncorrelation doesn't equal to independence*

ICA has to prove independence

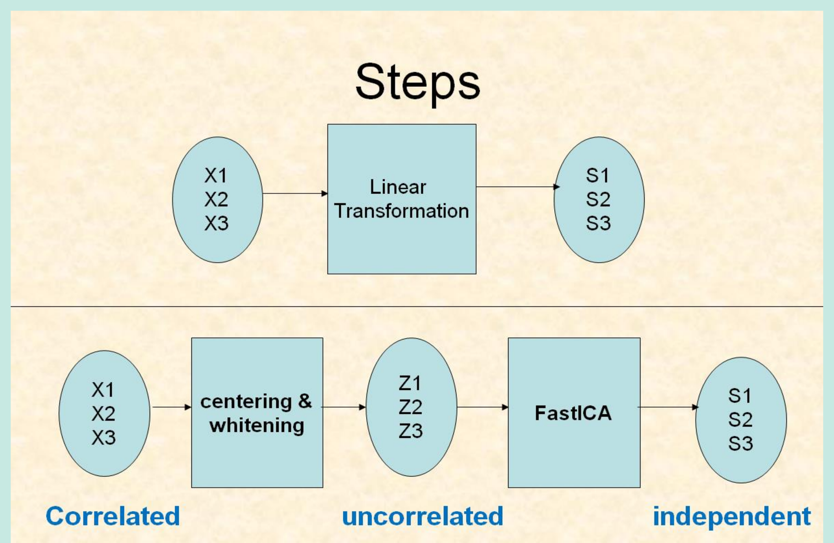
Non-gaussianity estimates independence

Maximizing non-gaussianity gives independent components

A key to understanding the ICA is a Central Limit Theorem: *The distribution of a sum of independent random variables tends toward a Gaussian distribution.* **Finding Independent Components through maximization of non-gaussianity.**



The density function of the Laplace distribution, which is a typical supergaussian distribution. For comparison, the Gaussian density is given by a dashed line. Both densities are normalized to unit variance.



Hyvärinen, A. (1999). Survey on independent component analysis. Neural Computing Surveys, 2:94–128.

Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. Neural Computation, 9(7):1483–1492.

Math model:

Given; *x1(t),x2(t),x3(t)* ↓ *Find;* *s1(t) , s2(t), s3(t)*

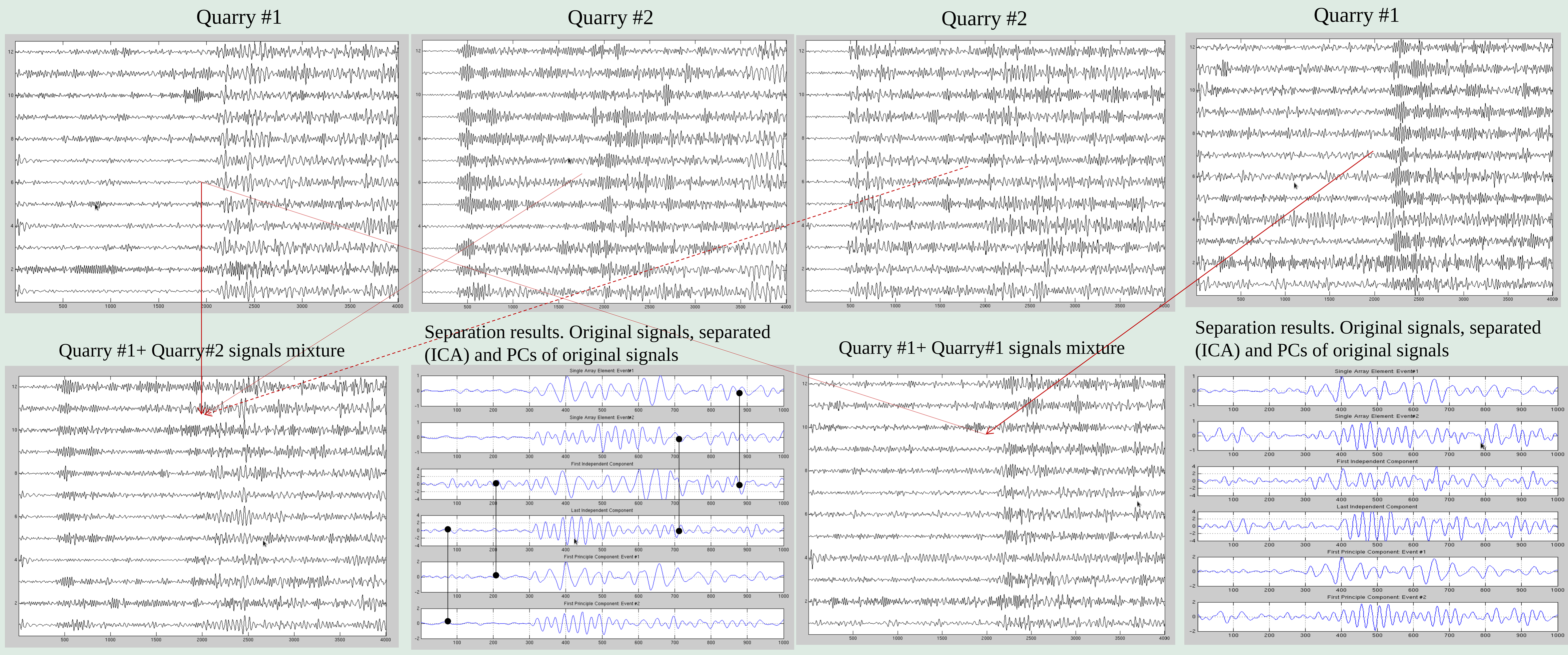
$$x_1(t)=a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t)$$
$$x_2(t)=a_{21}s_1(t) + a_{22}s_2(t) + a_{12}s_3(t)$$
$$x_3(t)=a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)$$

Find a **W** such that **S=WX, W** = $\mathbf{A}^{-1}$

The FastICA learning rule finds a direction, i.e. a unit vector **W** such that the projection $\mathbf{W}^T$X **maximizes nongaussianity**. Nongaussianity is here measured by the approximation of **negentropy** J($\mathbf{W}^T$X). The FastICA is based on a fixed-point iteration scheme for finding a maximum of the non-gaussianity of $\mathbf{W}^T$X, see (Hyvärinen and Oja, 1997; Hyvärinen, 1999). It can be also derived as an approximative Newton iteration (Hyvärinen, 1999).

Conclusion. We have investigated a few feasible scenarios for avoiding comprehensive explosion test monitoring with hiding specific signals from one seismic event on the background of another. Independent Component Analysis appears to be a promising way in seismic signals separation taking into account some limitations of this approach. ICA in seismology was used mainly for the seismic exploration data, separating a non-gaussian reflectivity sequence from the minimum phase sounding signal. In many seismic applications ICA showed better separating performance then SVD–based methods and classical PCA. We have successfully applied ICA methods to separation of signals from earthquakes, quarry blasts, and nuclear explosions conducted at different ranges from recording stations. Data from teleseismic (IMS array MKAR) and regional (small aperture array MHVAR operated by Russian Academy of Sciences) seismic stations was processed showing applicability of the method for different types of equipment and sampling rate (40 Hz at MKAR vs 200 Hz at MHVAR). Also, good results were achieved for recovery of nuclear explosion buried in a wave-train of strong earthquakes (MKAR array example). ICA shows good results when other approaches fail, for example, separation with FK-analysis based on different slowness and azimuths, or cepstral methods for echo. However ICA performance is proportional to the azimuthal gap between the composing signals, slowness and sampling rate. Due to the ICA's ambiguity (polarity, energy and order of components (permutation)) we consider this approach so far as a tool for extended interactive analysis, and more work to be done to introduce it in automatic processing pipeline. Further plans of using the ICA in seismic monitoring is improving the robustness of this approach, enforcing homomorphic deconvolution with the ICA applied in quefrency domain, and introducing the multi-station association independent components association for improved seismic event location. We also work on development of formal criteria exposing the presence of several composing signals in the observed data.

Quarry blast signals separation

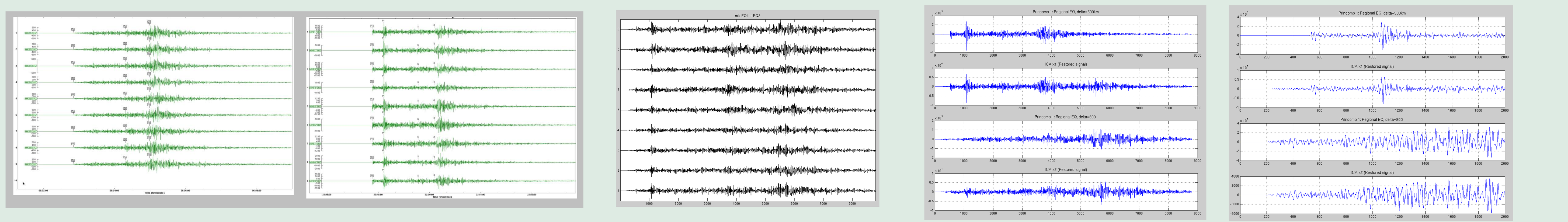


Comparing: ICA components (restored signals) with first principle component of the array seismogram

Separation of signals from different quarries. Solid lines indicate input waveforms for the actual example (mixture, and decomposition on the right). Dash line indicates typical waveform used to produce similar signal separation result. In first case primary phase of Q#1 event fall onto the secondary phase of Q#2 event making them visually inseparable. In second case all arrivals of both events fall on the same time producing visually a new event. A figure with separation results consists of: (1) a pair of input signals (from same channels) composing the mixture, (2) separated signals, and (3) first principle components of the array seismograms corresponding to each signal.

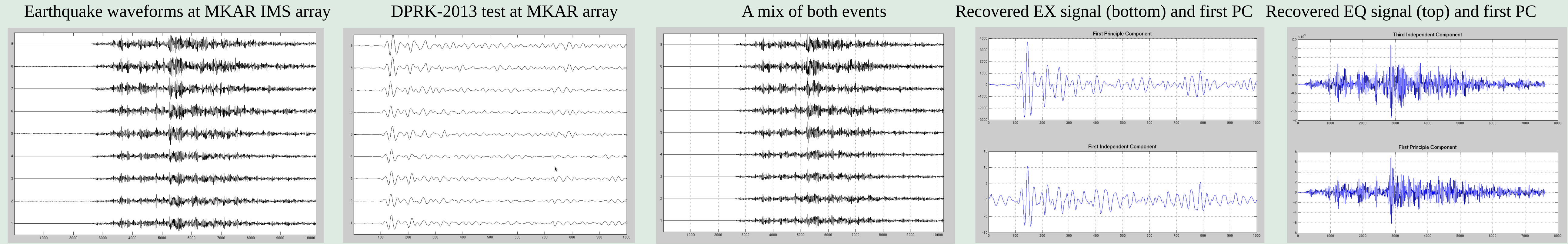
We present the PCs to emphasize that the restored signals inherit the property of the whole array, not just some single component, since the best similarity is provided between the first PCs of the input array and the best Independent Component. The black vertical rounded lines indicate corresponding signals (omitted on right figure). Correlation coefficients between the original and restored signals are in a range (0.84, 0.92).

Separating two regional earthquakes (left to right): EQ1, Δ=8º; EQ2, Δ=5º; Mixture; Decomposed, 400 sec; Decomposed, 50 sec.



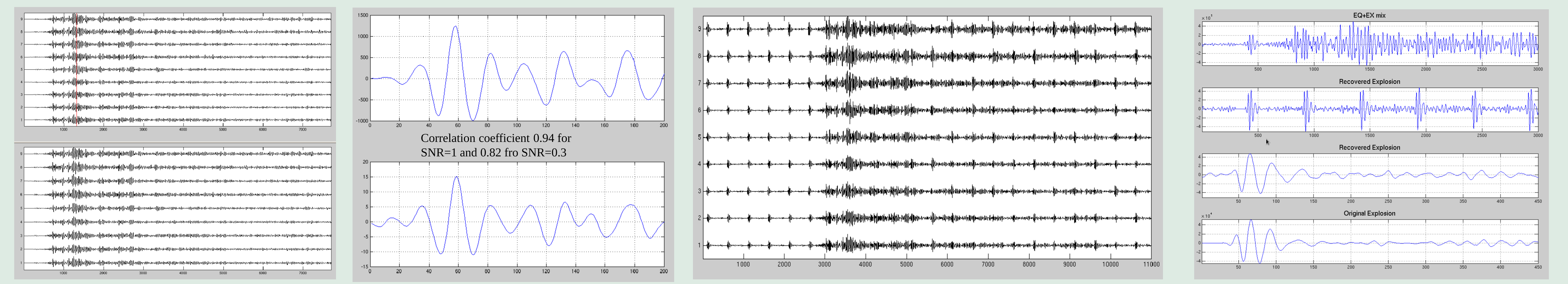
Separating events with close azimuths from the mix: DPRK-2013, Feb-2013 Mb=4.9 (scaled up to Mb=7) test and Tohoku, March 2011 Mw=9.0 EQ.

Earthquake waveforms at MKAR IMS array — DPRK-2013 test at MKAR array — A mix of both events — Recovered EX signal (bottom) and first PC — Recovered EQ signal (top) and first PC

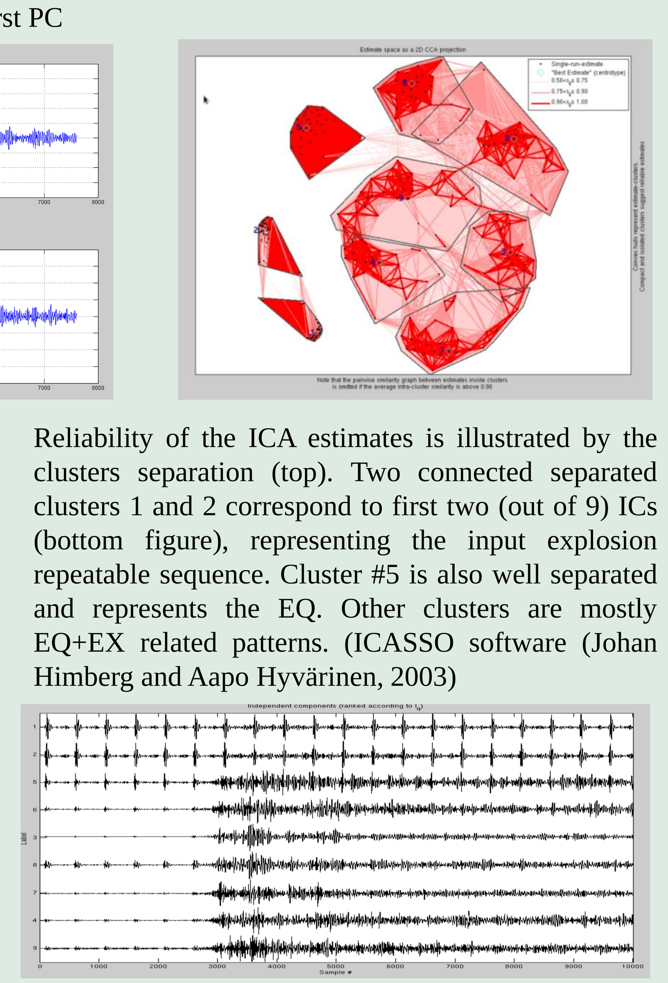


A case of separation of events with different back-azimuths from the mix: DPRK-2013, Feb-2013 Mb=4.9 (scaled up to Mb=7) test and Sumatra, April 2012 Mw=8.6 earthquake

Earthquake waveforms at MKAR array (top) and mix with DPRK-2013 test — Recovered test signal (top) and first PC — Repeatable mix of the synthetic DPRK-2013 array seismogram with the EQ array seismogram, MKAR IMS array — EQ and synth-DPRK-2013 mix, restored explosion in different time scales, and 1st PC of the original



Reliability of the ICA estimates is illustrated by the clusters separation (top). Two connected separated clusters 1 and 2 correspond to first two (out of 9) ICs (bottom figure), representing the input explosion repeatable sequence. Cluster #5 is also well separated and represents the EQ. Other clusters are mostly EQ+EX related patterns. (ICASSO software (Johan Himberg and Aapo Hyvärinen, 2003)



Preparatory Commission for the Comprehensive Nuclear-Test Ban Treaty Organization, Provisional Technical Secretariat, Vienna International Centre, P.O. Box 1200, A-1400 Vienna, Austria. E-mail: mikhail.rozhkov@ctbto.org

IDG RAS
http://idg.chph.ras.ru/en

International Data Centre
http://www.ctbto.org