DEBLENDING SEISMIC SIGNALS
PROJECT REPORT

August 7, 2019

Team Members: Yanhua Yuan (Mentor, Exxon Mobil),
Dilek Erkmen, Michigan Technological University
Anthony Nguyen, UC Davis
Samantha Pinella, University of Michigan
Jose Pedro Rodriguez Ayllon, Houston University
Nandita Sahajpal, University of Kentucky

Contents

0.1 Executive Summary 2
0.2 Project Roadmap 3
0.3 The Problem and Assumptions 4
0.4 The Approaches 5
0.5 Independent Component Analysis (ICA) 10
0.6 Looking Forward 13

0.1 Executive Summary

Fossil fuels are very intricately woven into the fabric of our day to day lives in both very obvious and subtle ways. Starting from the hot water and electricity we use to the clothes we wear, the houses we live in, the vehicles we drive and most of the consumer goods we enjoy, fossil fuels are building blocks of modern society and life as we know it. The widespread use of these resources has immensely improved the standard and quality of life across the globe. As such, life the way humans are accustomed to today would be simply impossible without them.

Based on the World Energy Council (WEC, 2013), today, fossil energy accounts for about 80 % of the total primary energy supply, . It is expected that in 2020 76% of the total primary energy supply will still be covered by fossil energy. However, the fossil energy sources that are needed to continue to sustain human life on earth, are not discovered yet. Therefore, the survival of humankind on earth depends a great deal on the exploration of new fossil fuel reservoirs. It is not only important to find new sources of fossil fuels but it is also very important to do so in a way that is efficient and reliable.

An advanced exploration technique called the Seismic Surveying or Acquisition is one of the ways of finding new reservoirs of fossil fuels below the surface of earth underneath oceans. Since 70 % of the earth is covered by the ocean, seismic surveying is the key to understanding the geology of the surfaces underneath ocean.

It consists in sending signals through a source and recording the echoes to get an image of the earth's surface.

In seismic surveying, there is usually a streamer which has multiple seismic sources(like airguns) and a line of sensors or receivers (hydrophones). The sensors are located on either side of the streamers.

The seismic sources are used to direct low frequency pulses towards the ocean floor, which are reflected by various geological structures underneath the ocean. These reflected pulses are recorded by the sensors/receivers. The data collected by the sensors/receivers is analyzed and interpreted to create a map of the resources that lie below.

Acquisition of seismic data, especially in a marine environment, is a very expensive process. Using just one source to send pulses costs about 1 million dollars, and hence it is not very cost effective to use just one source. Since the majority of the survey costs is heavily dependent on the time for which the data is recorded, it makes sense to fire shots from multiple seismic sources. This not only lowers the cost of data acquisition, but also increases the quality of data obtained. However, the data obtained by using multiple simultaneous sources is highly blended, and the current processing techniques cannot deal with blended data. Therefore, the data must be de-blended before it can be analyzed and interpreted. However, de-blending blended seismic data is an extremely challenging and a difficult problem to solve.

Based on our analysis so far, we recommend using Independent Component Analysis(ICA) with neural networks for de-belending. We are in no way claiming that this is the best approach but based on our work we believe that this is a very promising approach and therefore requires further investigation.

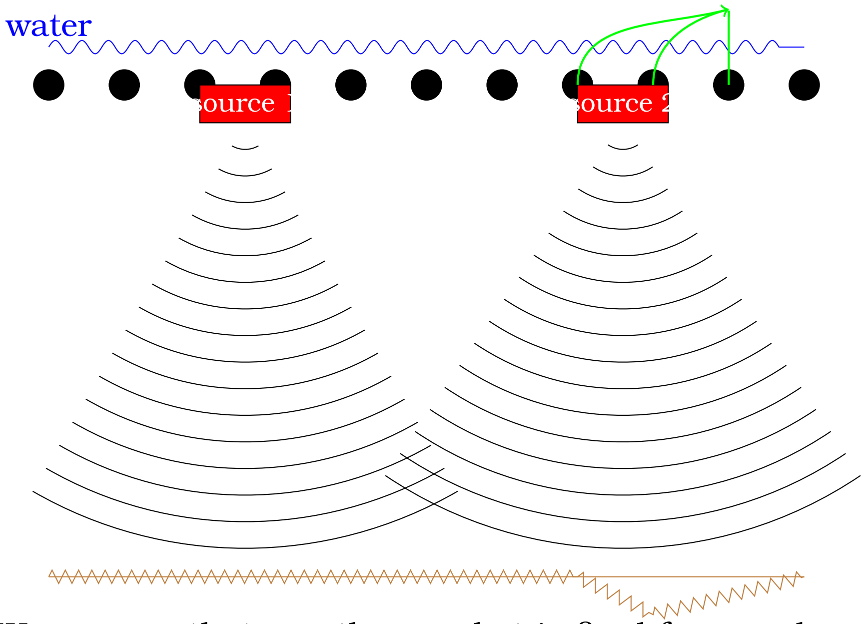
0.2 Project Roadmap

Keeping in mind the complexity of the problem, our main goals for in this project were

1. Understand the problem thoroughly and generate our own data for a system of 2 simultaneous sources using the Marmousi Model.
2. Detailed study of some of the approaches that have been made used in the past to approach this problem.
3. Understand why certain approaches do not work for the our data set to get more insight into the complexity of the data generated.
4. After completing the first 3 steps, focus on one approach that seems to work for our data set.

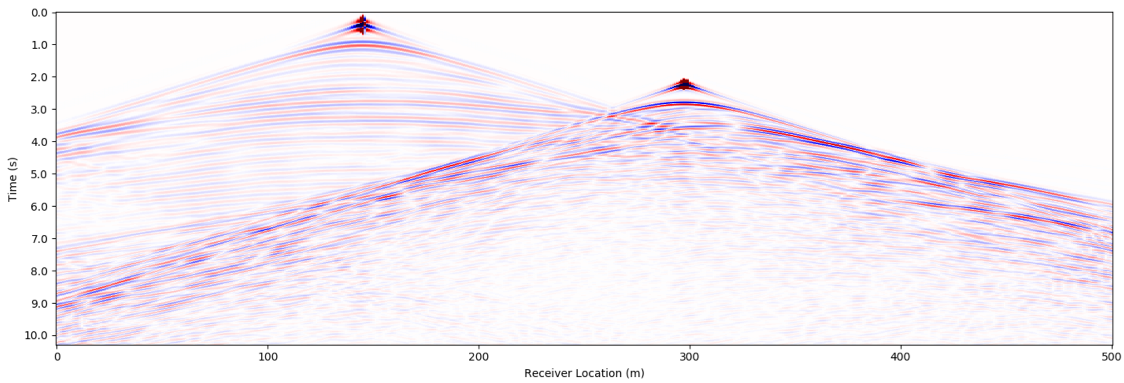
0.3 The Problem and Assumptions

Our model for generating data consists of two independent seismic sources and 501 sensors. The data is recorded for a duration of 10 seconds with 8573 time steps. To make it the problem slightly easier, the two sources are fired at different times, and with different frequencies.

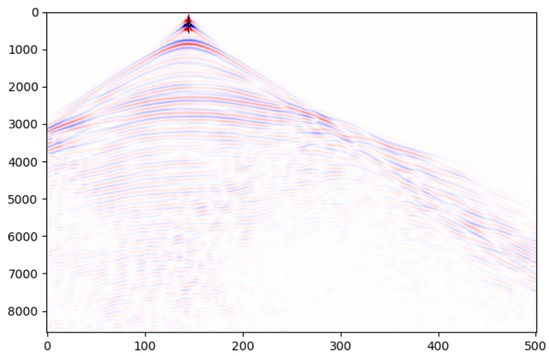


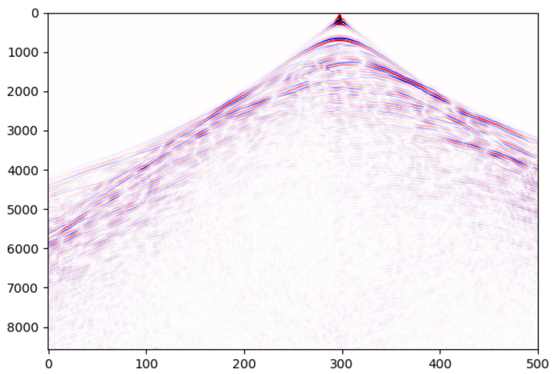
We assume that exactly one shot is fired from each source and the time interval between the first and the second shot is 2 seconds. The frequency of the pulse from the first and the second source is 2 Hz and 8 Hz, respectively.

The data obtained is highly blended as demonstrated in the image below:



Our problem is to find a robust algorithm than can be used to de-blend the data into the data from two different sources as shown in the image below:





0.4 The Approaches

1. **SINGULAR VALUE DECOMPOSITION (SVD):** This was the first approach that was used to understand this problem. For this approach we first converted the data into an 8573 x 501 matrix 'A', where each column recorded the data corresponding to one sensors.

   Then we computed the singular value decomposition of A

   $$A = U\Sigma V^H$$

   where $\Sigma \in \mathbb{R}^{8573\times 501}$ is a rectangular matrix whose main diagonal consists of the singular values σ_i of A arranged in decreasing order, $U \in \mathbb{C}^{8573\times 8573}$ and $V \in \mathbb{C}^{501\times 501}$ are both unitary matrices. Equivalently, we can decompose the image A as the sum of rank one eigen-image:

   $$A = I_1 + I_2 + \cdots + I_r$$

   where the eigen-images $I_i = \sigma_i u_i v_i^H$ $r = rank(A)$.

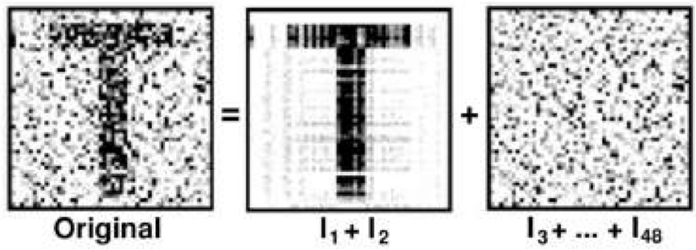


Figure 3.5: A 48x48 pixel picture showing a noisy T and its eigenimages. One can see that laterally coherent signals will be captured in the first two eigenimages and the noise is in the following eigenimages. Taken from Trickett (2003).

The singular values σ_i are measures of the lateral correlation, i.e., how correlated events are from sensor to sensor. Hence, a laterally coherent event will have a high singular value. Since the singular values are sorted in decreasing order, the first eigen images consist of laterally coherent signals while the latter ones are associated to random noise as the degree of correlation is low.

The intuition behind using this method was eliminate one of the signals as noise but this method did not work for our data set, since there was no random noise in our data, and none of signals could be eliminated as random noise.

2. **FAST FOURIER TRANSFORM (FFT):** The fast Fourier transform is a mathematical method for transforming a signal of time into a signal of frequency. It is very useful for analysis of time-dependent phenomena. We used the FFT method to convert our image A from time-domain signal to the frequency domain, used Numpy's FFT shift, which shift the zero-frequency components to the center of the spectrum, and plotted the magnitude of the signal in the frequency domain.

   The intuition on implementing the FFT here is that our signal in the time domain can be decoupled in the frequency domain via the Fourier Transform, but our efforts were fruitless because the signals in the frequency domain were visually inseparable due to the similar amplitudes of the two sources in our problem. Through this approach, we have gained a better understanding of the problem.

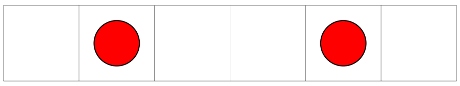
3. **SVD + FFT:** This approach was implemented around the same time we

tried the SVD and FFT. The idea was instead of implementing the SVD on the image A, we would apply the FFT on the image A and apply the SVD to that. However, this approach was unsuccessful due to the absence of noise in the image, so none of the signals in the frequency domain could have been eliminated.

4. **INVERSE PROBLEM:** In this approach, we set up the forward version of the de-blending problem. This means that we will assume that we already have the un-blended data. We can start with the un-blended data (in a matrix), blend it, and then use inverse problem techniques to go back to the un-blended data. We follow the techniques from [5] although the de-blending process did not work for us.

   In order to discretize the problem, we assume that the sources and detectors can only be placed along a line at l equally spaced locations. Let n_s be the number of sources and let n_d be the number of detectors.

   First, we set up the source matrix S. This is a $l \times n_s$ matrix, i.e. each column is a source and each row is a position. Each entry at (i,j) will have a 1 if source j is at position i, and a 0 otherwise. For example, if $l = 6$ and there are two sources arranged like



   then

$$S = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

   The detector matrix D is a $n_d \times l$ matrix and it is set up analogously to the source matrix, except as the transpose (i.e. swap row and column). Finally, the $l \times l$ matrix X represents measurements that come from the experiment, [5] calls it the *multidimensional transfer function of the earth.*

   However, it is not our goal to find X. We are looking for the $n_d \times n_s$ matrix

P where

$$P = DXS.$$

This is the unblended data matrix. For example, if there are six detectors and two sources,

$$P = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{pmatrix}$$

This means that detector 1 observed a pressure of a_{11} from source 1 and a pressure of a_{12} from source 2, etc., at some fixed time.

Our data will be given to us not as the unblended matrix P, but as the blended matrix P'. Using the previous example, our detectors would not know which measurement came from which source, so $P' = \Big(a_{11} + a_{12} \quad a_{21} + a_{22} \quad a_{31} + a_{32}$
P and P' are related by the blending matrix Γ:

$$P' = P\Gamma$$

where Γ is a $n_s \times$ (number of blended sources) matrix. In the previous example $\Gamma = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ because the pressures are added together. More typically, Γ would include a change in amplitude or a time shift for the source waves. Since Γ is not an invertible matrix, we have to use inverse problem techniques to find P. The thesis [5] provides an algorithm for finding P although we were unsuccessful in using it to find a reasonable solution. The main struggle was how to handle the time dimension.

5. INDEPENDENT COMPONENT ANALYSIS (ICA)

Based on Wikipedia, "In signal processing, ICA is a computational method for separating a multivariate signal into additive sub-components. This is done by assuming that the sub-components are non-Gaussian signals and that they are statistically independent from each other. ICA is a special case of blind source separation. A common example application is the "cocktail party problem" of listening in on one person's speech in a noisy room."

ICA works in two steps:

(a) The first step in an ICA algorithms is to whiten the data. This means that we remove any correlations in the data. The whitening process is simply a linear change of coordinate of the mixed data.

(b) Once the ICA solution is found in the new "whitened" coordinate frame, it is projected back into the original coordinate frame.

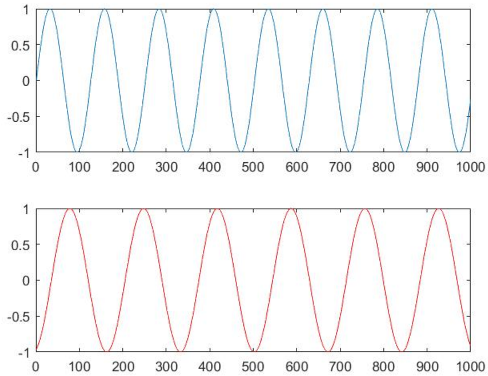
This approach proved to be the most fruitful based on the research that our team has done so far. More information about this is available in the next section.

0.5 Independent Component Analysis (ICA)

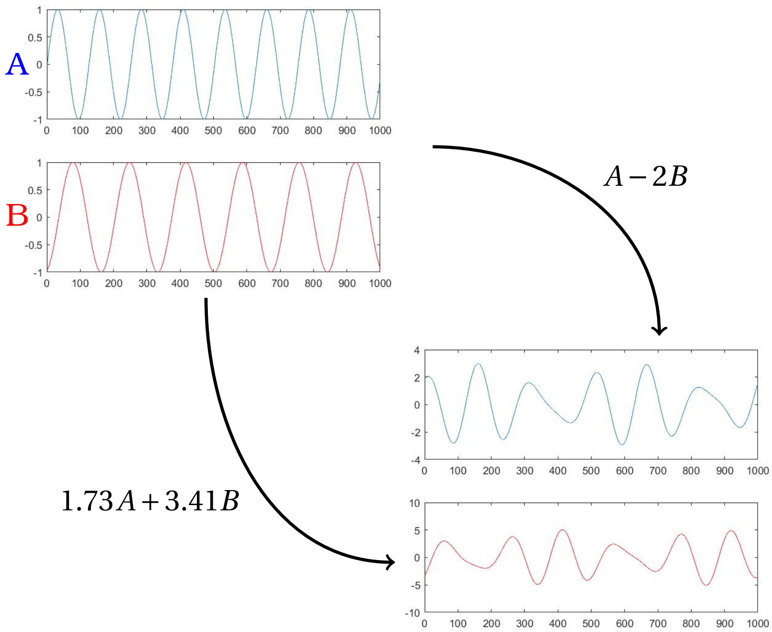
Independent Component Analysis is a signal processing method to separate independent sources linearly mixed in several sensors.

For example, let's try to mix two sound waves from two different sources, and then separate the two sources.

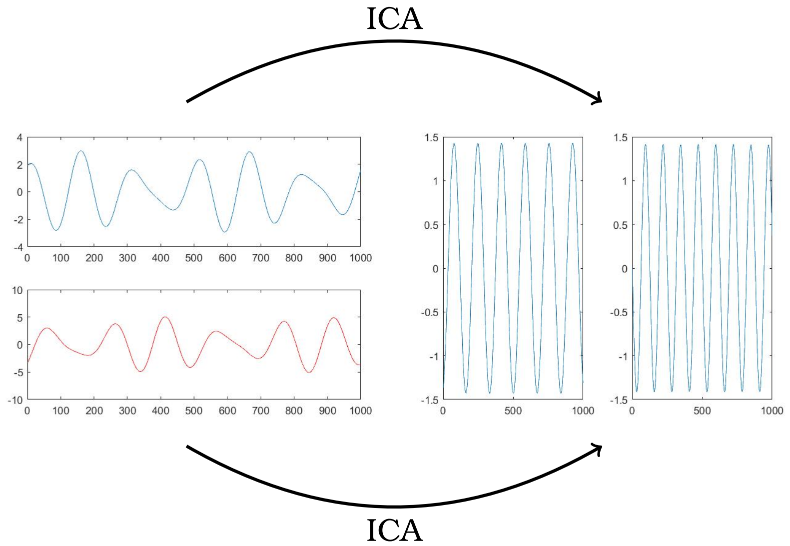
Let A (top) and B (bottom) be the two independent sound waves.



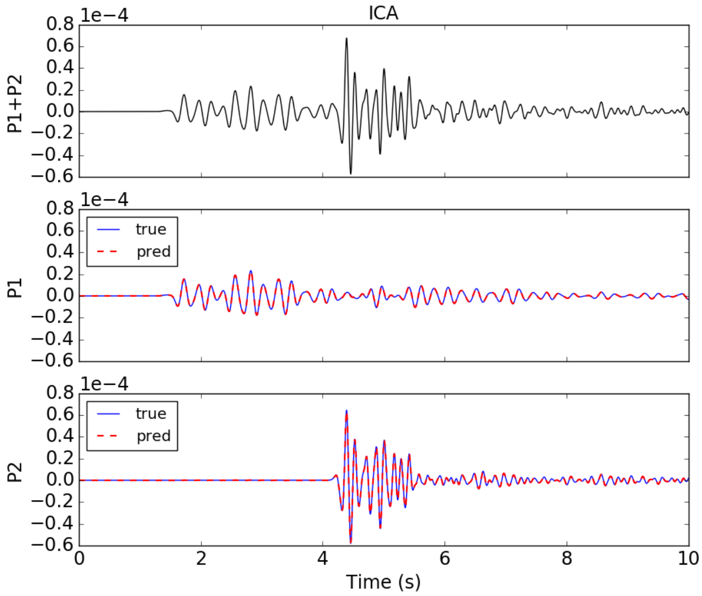
Let us blend these two independent waves in two different ways.



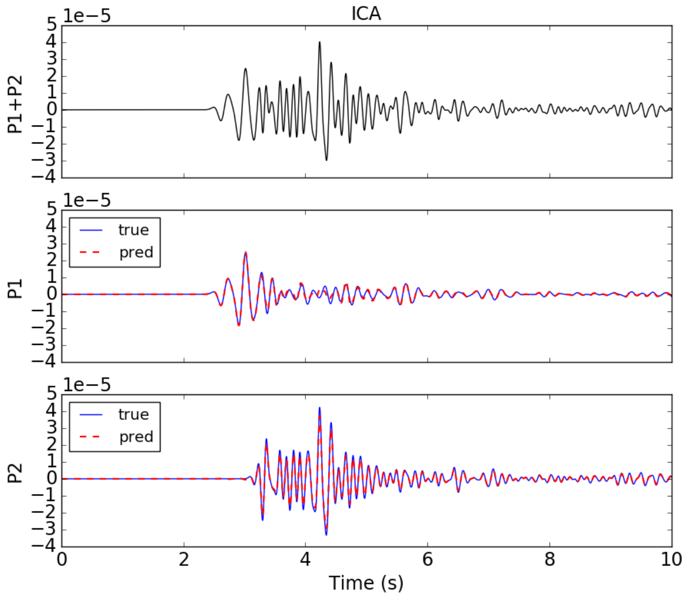
Now we apply ICA to these two blended waves to get the original waves. It is important to note that ICA does not recover the exact amplitude of the source waves.



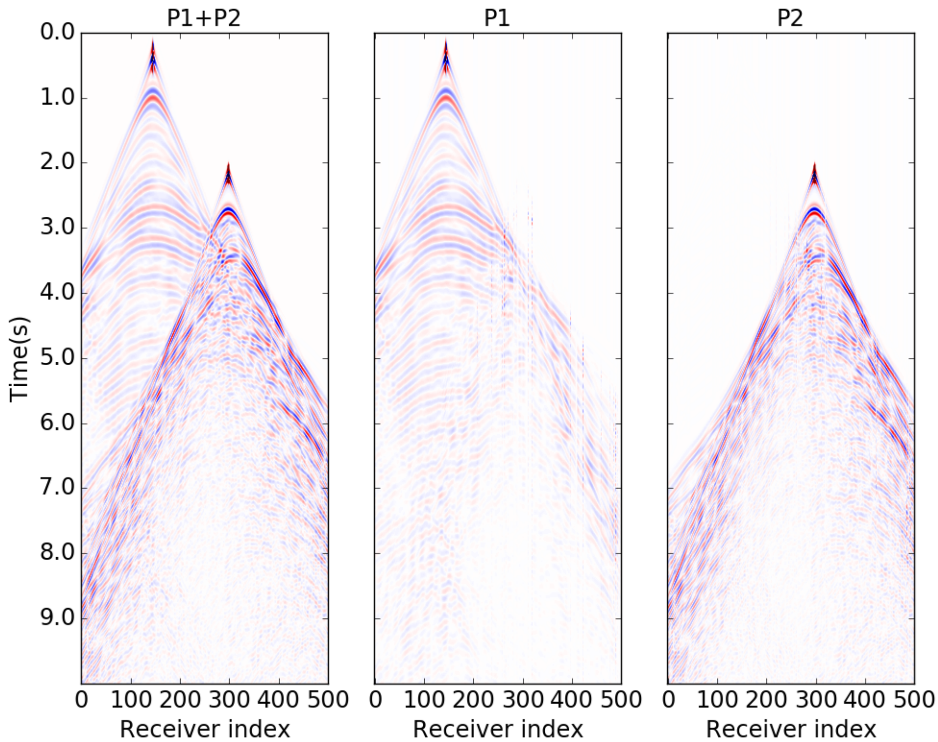
Based on this example, we decided to apply ICA to the blended waves received by a single sensor. But note that, to be able to apply ICA, we need to have two independent waves blended two different ways. So we used the un-blended data from the two independent sources and blended it two different ways such that the resulting waves were still independent. On applying ICA to the 200th sensors, we got really promising results as shown in the picture below:



We also applied ICA to the data from the 250th sensor, and again we got some really good results:



After this we decided to apply ICA to each of the 501 sensors one at a time one at a time and combined the results to get the following image.



Note that in the image of P1(center), the parts which belong to P2 are suppressed just enough so that we can see P1 more clearly.

0.6 Looking Forward

The results from using ICA just by itself have been really promising. Going forward, our hope is to apply ICA to the outputs of a deep neural network trained on the data generated by blending the un-blended data in two independent ways.

Bibliography

[1] Mahdad, Araz. 13 Nov. 2012, “Deblending of Seismic Data" master’s thesis, Technische Universiteit Delft.

[2] Wikipedia contributors. 2019, “Independent component analysis — Wikipedia, The Free Encyclopedia", "https://en.wikipedia.org/w/index.php?title=Independent_component_analysis&oldid=906212528."

[3] Magnussen, Fredrik. June. 2015, “De-blending of Seismic hydrophone and multicomponent data " master’s thesis, University of Oslo.

[4] Richardson,A. & Feller,C. *Seismic data denoising and deblending using deep learning* (arXiv:1907.01497v1 [eess.IV] 2 Jul 2019)

[5] Delorme, Arnaud. *Independent Component Analysis for Dummies*, "http://arnauddelorme.com/ica_for_dummies/."