# Application of Independent Component Analysis and Gaussian Mixture Models in micro-seismic signal detection

*Amir Shamsa and Mehdi Paydayesh*
*Schlumberger*

## Summary

With the emergence of new algorithms in machine learning, advances in hardware, computation power, and abundance of big data, a renewed investment is dedicated to research on artificial intelligence and machine learning in the oil and gas industry. The main attractions of using these modern techniques are cost-efficiency, accuracy, and reducing turnaround time on contrast to current workflows and practices. In this paper, we present a simple and effective machine learning workflow to automate signal recognition in microseismic applications. This workflow can be plugged into a real-time monitoring task for a fast, robust, and cost-efficient solution.

## Method and workflow

We propose a machine learning technique that is an unsupervised clustering method known as Gaussian mixture models (GMM) or expectation maximization learning (Guoshen, 2012). The two main clustering features are fracture signal and background noise. The objective is to enhance the signal related to the fractures and suppress the background noise by acquiring more data. To condition the data for GMM and enhance feature detection we have applied, the independent component analysis (ICA), as explained by Ranjan (2008). ICA is applied to raw measurements to transform the data into a domain where features are more separable. Figures 1 a and b shows the result of this workflow applied to the data for 250 samples and 6250 samples, respectively. A significant improvement is observed by adding more data as acquisition progresses in real time. The more data, the better the ICA algorithm can discriminate the signals from noise. This, in turn, will help the GMM process produce a better split of the noise and signal. It would be reasonable to think that feeding the algorithms with more data always gives better results; however, our tests show that using excessive data can lead to the noise dominating and masking the signal due to an increased chance of including inconsistency in the data. A remedy is to split the entire data space to smaller portions and run ICA on each portion before applying GMM on entire data sets.

To validate our result, apart from looking and the well data, the silhouette score was used.  silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighbouring clusters. As demonstrated in Figure 1 c, the more data the better score.

## Conclusions

A simple and robust machine learning technique is applied to automate signal detection and analyze recorded microseismic data. The method's performance is tested and evaluated on real data. The fracture signals were well detected using the proposed workflow and techniques when more data were introduced.

In contrast to conventional methods, the techniques implemented herein described work on training the model prediction with additional data without restarting from the beginning, making them viable for continuous online learning.

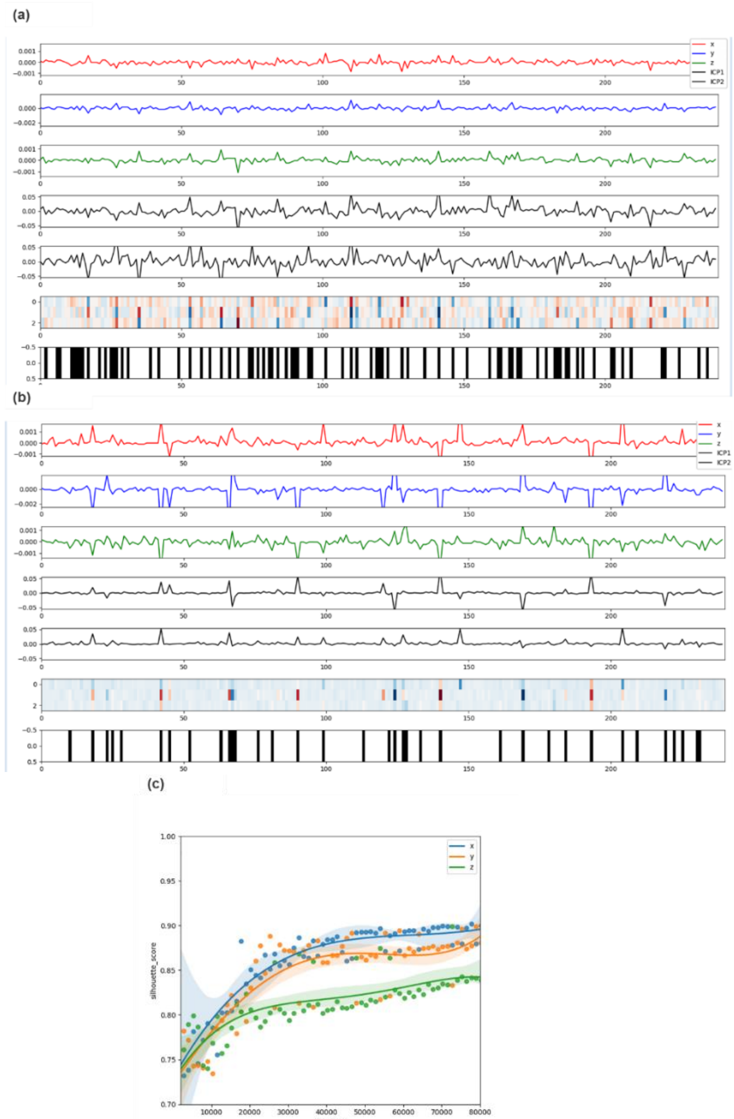


*Figure 1:a) Initial clustering result based on 250 sample points equivalent to 12 ms. The tracks from top to bottom are: X, Y, and Z components of geophone, IC1, IC2, (first and second independent components, independent components visualized together and the detected signals; b) Final clustering result after feeding the algorithm with 6250 sample points.; c) Silhouette score*

**References**

Ranjan, A. [2008] A New Approach for Blind Source Separation of Convolutive Sources. Wavelet Based Separation Using Shrinkage Function ISBN 3-639-07797-0 ISBN 978-3639077971

Guoshen, Y. [2012] Solving Inverse Problems with Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity. *IEEE Transactions on Image Processing*, 21 (5): 2481–2499