**Seismic noise labeling in pre-stack data: the initial step for automatic identification of noises affecting towed-streamer**

Glaucia Sakai¹, Pablo Carvalho¹, Anderson Santos², Stephanie Tavares¹, Eduardo Gaspar¹; (1) Fundação Gorceix, (2) Petrobras-GDGEOF

Copyright 2023, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 18th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 October 2023.

Contents of this paper were reviewed by the Technical Committee of the 18th International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

# Abstract

Supervised machine learning technique is defined by its use of labeled dataset to train algorithms to classify data or predict outcomes precisely. In this work, it is proposed a workflow to label different types of seismic noise to fed into a supervised machine learning model. The initial stages of seismic data processing involve a large volume of data. The seismic data is subject to various types of noise and according to its characteristics, the geophysicist will decide on the most appropriate filters, processing flows and parameters. In many cases a certain type of noise will not affect the entire survey, only a few lines. For example, the attenuation of linear noises such as seismic interference can only affect some lines, so knowing which ones are affected the geophysicist will be able to apply the linear noise attenuation processing only in the shots in which the noise occurs, saving computational time.

# Introduction

A machine learning dataset is a collection of data used to teach the algorithm how to make predictions. The types of data can be text data, audio data, video data, etc. Choosing, preconditioning, and labeling the most appropriate dataset is one of most crucial steps in training an Artificial intelligence model.

During machine learning algorithm test are necessary to split the dataset into three subsets: training dataset, validation dataset and test dataset. Training dataset can be categorized into two categories: labeled data and unlabeled data. Labeled data is a set of data flagged with one or more significant labels.

Labeled dataset is applied in supervised learning. It allows Machine Learning algorithm to learn the attributes associated with specific labels. Flagged data is useful during training of neural network and later to check the accuracy of Machine learning model.

Supervised method has been applied in reflection data focus on detection of seismic facies analysis (Wrona et al., 2019), identification of faults (Huang et al., 2017), seismic data anomalous detection (Almeida et al., 2019; Carvalho et al., 2022; Santos et al., 2023) and in early stages of seismic processing (Jia et al. 2018; You et al., 2020; Dong et al, 2022).

Towed-streamer marine acquisition is extensively applied to mapping the subsurface. Recorded traces during the seismic survey consist of reflected signal (wanted signal) and noise (unwanted signal). One of Achilles ‘heel of this method is the contamination by different types of noise including swell noise, spike-like noise, seismic interference noise, passing vessel noise, etc. Each noise has a pattern and machine learning algorithms are trained to recognize the pattern, being applied to different types of problems, such as classification, regression and prediction.

Seismic data processors should be familiar with the origin and nature of these specific noise types, to appropriately determine the data-processing steps seeking for the optimal attenuation achievement. In a dataset a seismic noise may not affect all lines. For instance, while a seismic interference is present in a line, this may not occur in another line.

# Method

1. *Data description*

Field marine streamer data is recorded in shot gather domain (Figure 1), it means the traces of the gather are originated from a single shot and many receivers. The initial quality assessment tools in field are performed in shot domain.

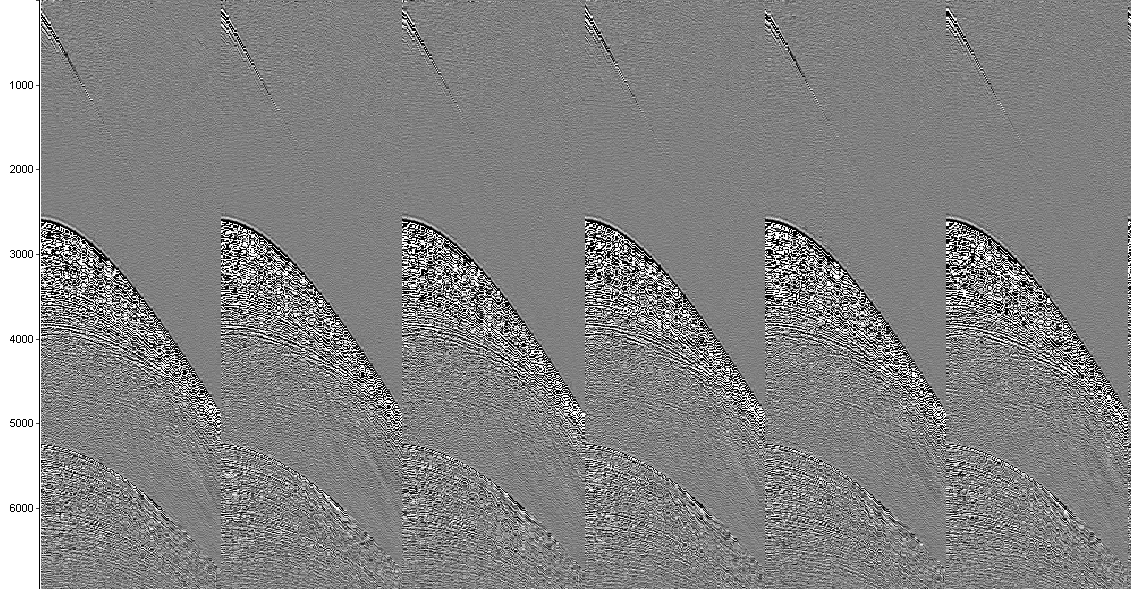


Figure 1: Shot gather.

Mostly seismic data are recorded and stored in SEGY. SEGY format is one of several standards developed by the Society of Exploration Geophysicists (SEG) for storing geophysical data. In this standard each seismic trace has a trace header associated with information on geographic coordinates, CMP, shotpoint, receiver number, among other information. A suggestion of the present work is the use of some free byte for the labeling of the noise that affects that trace or gather.

1. *Classification and identification of noise*

The noise classification applied is based on the work of Hlebnikov et al. (2021) and Elboth and Hermansen (2009). According to these authors, noise can be classified with the following aspects:

1) The origin of the noise: background (e.g., swell noise) or instrument noise (e.g., autofire);

2) If coherent or noncoherent from trace to trace in the observation data domain (Figure 2 and Figure 3);

3) The final feature is the characteristic frequency range of the noise.

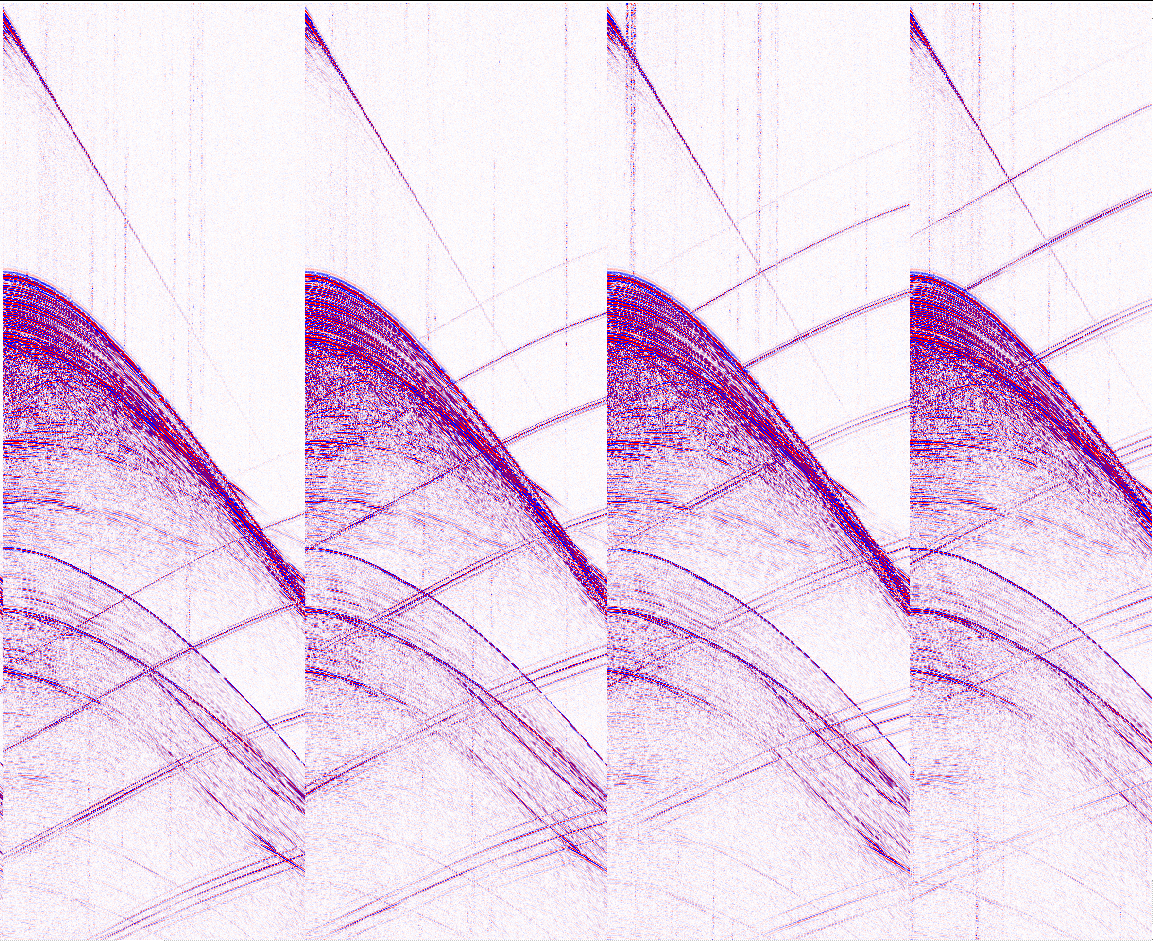


Figure 2: Example of coherent noise in shot domain.

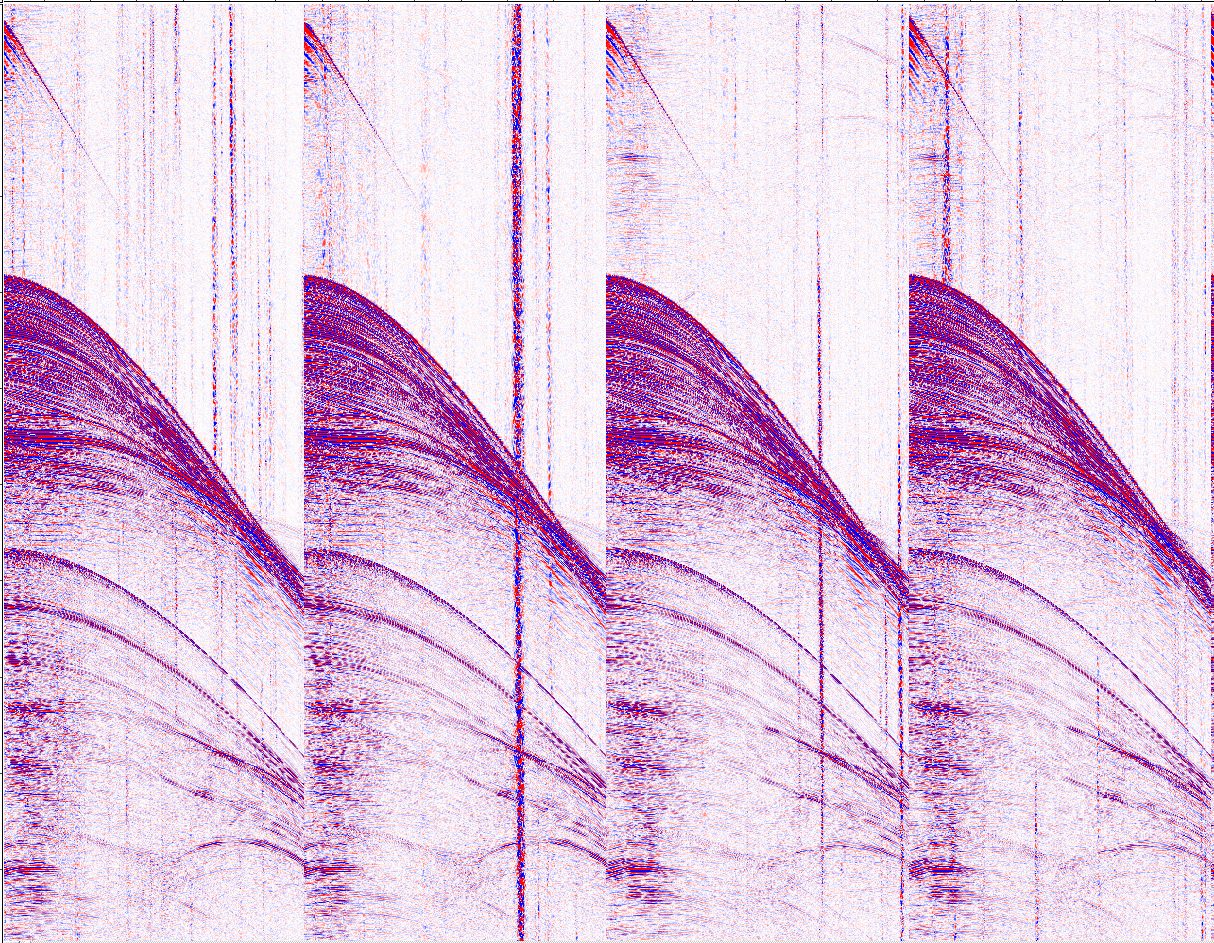


Figure 3: Example of random noise in shot domain.

The table below presents the main characteristics of the noises that affect towed streamer survey. The figure 2 shows shot gathers contaminated by coherent noise (seismic interference) and figure 3 shows a shot gather contaminated by random noise.

Table 1: Classification of noise, after Hlebnikov et al. (2021)

|  |  |  |  |
| --- | --- | --- | --- |
| **Type of noise** | **Origin of the noise** | **Random or Coherent?** | **Frequency** |
| Swell | Background | Random | 0-15 Hz |
| Spike | Instrument | Random | Broadband |
| Turn noise | Background | Random | Low frequency |
| Barnacle | Background | Random | Low frequency |
| Cable hit | Background | Random | Broadband |
| SI | Background | Coherent | Broadband |
| Rig Noise | Background | Coherent | Medium to high |
| Tail buoy | Background | Coherent | Low frequency |
| Autofire | Instrument | Coherent | Broadband |
| Crossfeed | Instrument | Coherent | Broadband |
| Passing vessel | Background | Coherent | Medium to high |

1. *Manual interpretation and labeling*

In this step seismic shot-gather are manually classified by experiment geophysicists. The identification of the noises occurred through visual inspection in the shot domain. Although it consumes a lot of time, the benefit will be to teach the algorithm to automatically identify the features of the main noises found in streamer survey.

Coherent noise has a trace-by-trace consistency and can be tracked over several traces (e.g., seismic interference). While random noise, does not have a systematic consistency from one trace to another (e.g., swell noise). Based on this characteristic we can elaborate a strategy to label the seismic noise.

**Results**

The data examples have been collected throughout recent 15 years on various seismic survey vessels. All examples are acquired by streamers with a group spacing of 12.5 m.

The following shot gathers represent the main noises found in surveys conducted in offshore Brazil in the last 15 years (Figure 4 and Figure 5).

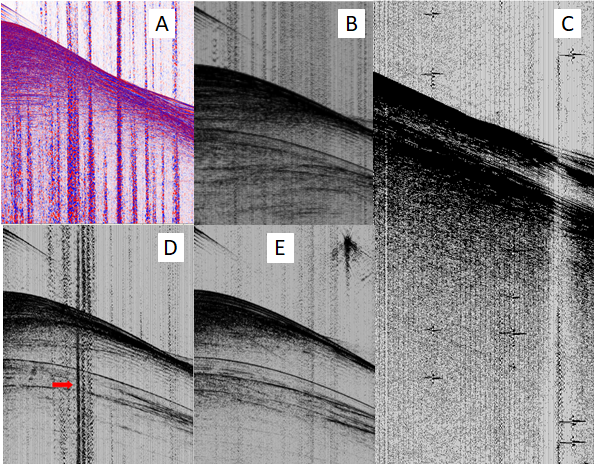


Figure 4: Shot gather contaminated by random noise: A) Swell noise; b) barnacle noise; c) Spikes; d) Turn noise and e) Cable strike.

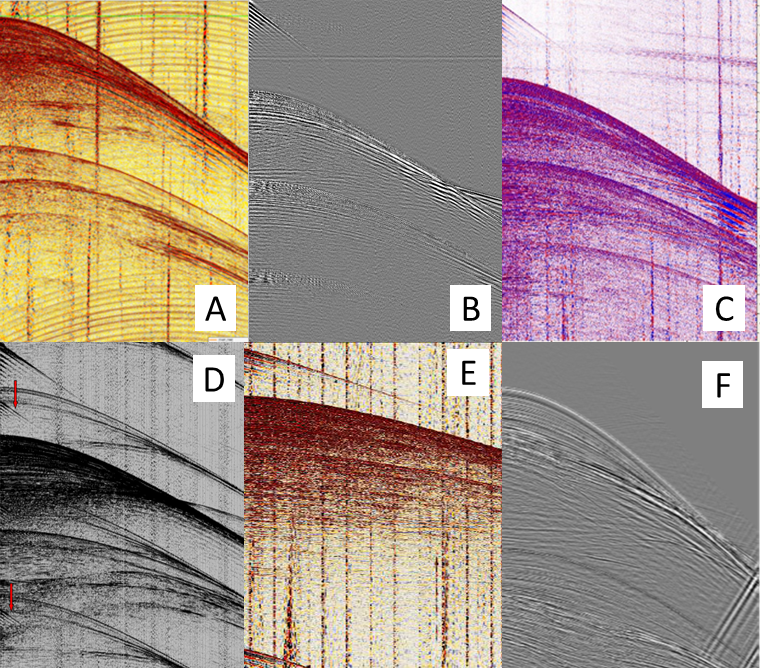


Figure 5: Shot gather contaminated by coherent noise: a) Ship noise, b) crossfed, c) Rig noise, d) Autofire, e) Wrap around and f) Tailbuoy noise.

**Examples of labeling**

The labeling strategy for random noise is to perform it trace by trace according to its occurrence. The suggested example is for the labeling of a spike, identified with the number 3 while the unaffected strokes were labeled with the number 1 (Figure 6).

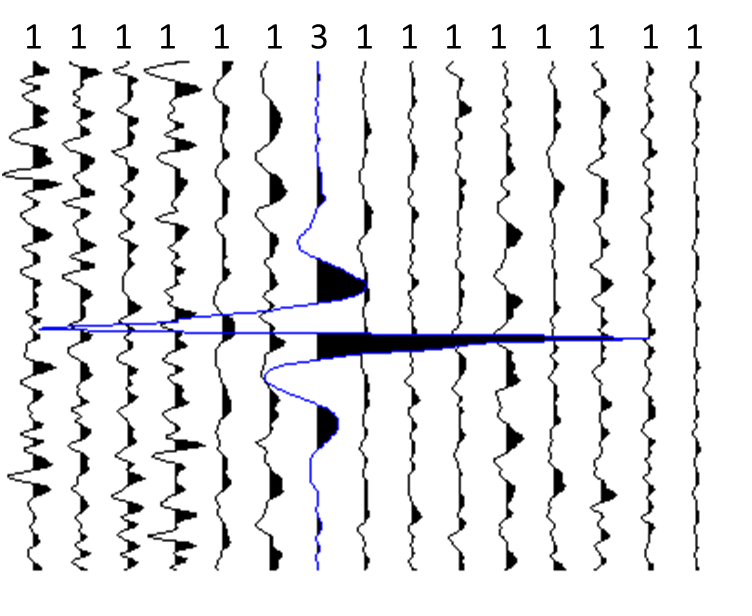


Figure 6: Labeling traces individually. The number 1 represents traces without spike and 3 is the trace with spike.

Below we have the examples of the labeling performed for a shot gather contaminated by seismic interference, because it is a coherent noise it is chosen to label by gather (Figure 7).

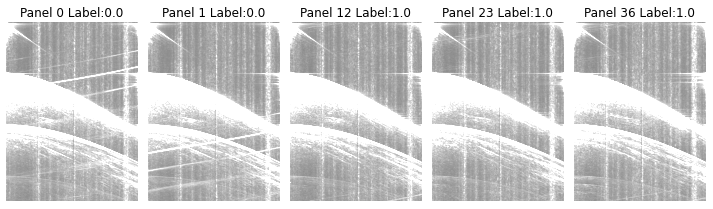


Figure 7: Labeled shot gather with seismic interference.

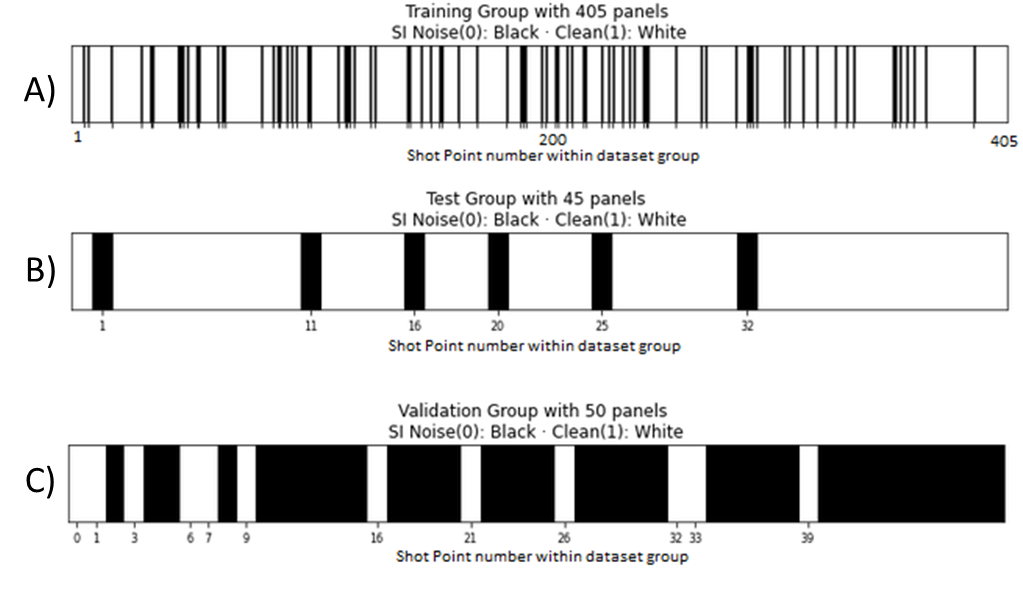


Figure 8: Input dataset to machine learning model.

With a data set labeled for a given problem, the next step is to divide this set into data for training (learning the algorithm), data for testing the learning of the algorithm and data for validating the learning. There are several ways to perform this division, such as using the Scikit-Learn library. The figure above shows a visualization of a data set for training, testing and validation randomly generated by a function created by the authors of this work.

# Conclusions

We demonstrate a methodology to label pre stack data to be used in supervised machine learning noise identification. Initially, the process of manually identifying and labeling seismic data can take up to weeks. However, in the long run, the identification process will be automated and will assist the processor in developing the best noise attenuation technique.

# Acknowledgments

The authors would like to thank Petrobras and Fundação Gorceix for the data and permission to publish this work.

**References**

ALMEIDA, I.N.; VARGAS, M.; NETO, C.S.; NOBRE, L.; SILVA, B.; BULCÃO, A.; DIAS, B.P.; LANDAU, L., EVSUKOFF, A. **Machine learning applied in swell noise classification**. *In*: **16th International Congress of the Brazilian Geophysical Society**, Rio de Janeiro, [s.n.], 2019.

CARVALHO, P., SANTOS, A. S., AZEVEDO, N., MACIEL, G., PORTUGAL, R., BARROS, P., BULCÃO, A. **Controle de qualidade de dados geofísicos usando uma rede neural artificial: identificação e classificação de swell noise**. *In*: **Simpósio** **Brasileiro** **de** **Geofísica**, nº 9, Curitiba, [s.n.], 2022.

DONG, X.; LIN, J.; LU, S.; HUANG, X.; WANG, H.; LI, Y. Seismic Shot Gather Denoising by Using a Supervised-Deep-Learning Method with Weak Dependence on Real Noise Data: A Solution to the Lack of Real Noise Data. **Survey in Geophysics**, v. 43, p. 1363-1394, 2022.

ELBOTH, T., HERMANSEN, D**. Attenuation of noise in marine seismic data**: *In* 79th Annual International Meeting, SEG, Expanded Abstracts,  
3312-3316, 2009

HLEBNIKOV, V.; ELBOTH, T.; VINJE, V.; GELIUS, L, J. Noise types and their attenuation in towed marine seismic: A tutorial. **GEOPHYSICS**, V.: 86, n.: 2, 2021.

HUANG, L.; DONG, X.; CLEE, T.E. A scalable deep learning platform for identifying geologic features from seismic attributes: **The Leading Edge,** n. 36, p. 249–256, 2017.

JIA, Z.; LU, W.; ZHANG, M.; MIAO, Y. Separating ground-roll from land seismic record via convolutional neural network, *In*: **Society of Exploration Geophysicists 2018 Workshop:** SEG MaximizingAsset Value Through Artificial Intelligence and MachineLearning, Beijing, China, 60–63, 2018.

SANTOS, A. S.; AZEVEDO, N.; CARVALHO, P.; MACIEL, G.; PORTUGAL, R. **Identificação de traços duros usando uma simples rede neural perceptron e recursos de multiprocessos de python**. *In*: **Simpósio** **Brasileiro** **de** **Geofísica**, nº9, Curitiba, [s.n.], 2022.

YOU, J.; XUE, Y.; CAO, J.; Li, C. 2020. Attenuation ofseismic swell noise using convolutional neural networks in frequency domain and transfer learning, **Society of Exploration Geophysicists and American Association of Petroleum**, v. 8, p. 941-952, 2018.

WRONA, T.; PAN, I.; GAWTHORPE, R.L.; FOSSEN, H. Seismic facies analysis using machine learning. **Geophysics**, n. 83, p. 83-101, 2018.