

Th\_R11\_13

## Leveraging a Supervised Machine Learning Toolkit For Better Seismic Processing Quality Control

M. Chambefort<sup>1,2\*</sup>, N. Salaun<sup>1</sup>, E. Chautru<sup>2</sup>, S. Clémençon<sup>3</sup>

<sup>1</sup> CGG; <sup>2</sup> MinesParisTech; <sup>3</sup> Université Paris Saclay

### Summary

---

Geophysicists are facing great challenges to process increasingly large volumes of seismic data in a faster, better and more reliable way. In order to achieve this, one key element is to be able to perform quality control of processing steps in an effective and efficient manners. In this paper we present two cases where machine learning techniques are used to help speed up the quality control process. The first case describes an application of supervised learning with K-nearest neighbor to identify areas that may potentially cause cycle-skipping in full waveform inversion. The second case uses a logistic regression machine learning approach to detect and classify the presence of rig noise on shot points. Furthermore, we also tested an automated data reduction method using a long short-term memory auto-encoder, that could speed up the algorithm significantly without losing key information in the original data.

## Introduction

Machine learning has been widely used in reservoir characterization for facies recognition or classification (Coléou et al., 2003). Recent developments in deep learning with advanced and sophisticated neural network topologies have gained significant attention in our industry, especially for opportunities to automate some of the time consuming seismic interpretation tasks such as picking fault planes and salt bodies, and facies identification (Zao et al., 2018; Ma et al., 2018). A natural way forward is to use such techniques for quality control (QC) during pre-imaging of seismic data, where the amount of data to consider is large (Martin et al., 2015). For many classification methods, data reduction followed by manual attribute extraction (also known as feature engineering), is needed. This step requires significant expertise in seismic processing, as the geophysicist needs to know precisely which feature(s) to check and visualize. It can also be time-consuming depending on the amount and the type of extracted features. To reduce the feature engineering step, we present two methods tested on real data examples where the classification has been performed on data in its original T-X domain (i.e. shot gathers). First, we show a supervised learning with K-nearest neighbor (KNN) for full waveform inversion (FWI) cycle-skipping QC. Then we show a noise detection case study, first using logistic regression (LR) and then we propose a semi-supervised method based on a long short-term memory (LSTM) auto-encoder, which allows us to use a reduced number of features for the QC.

### Cycle skipping QC case study

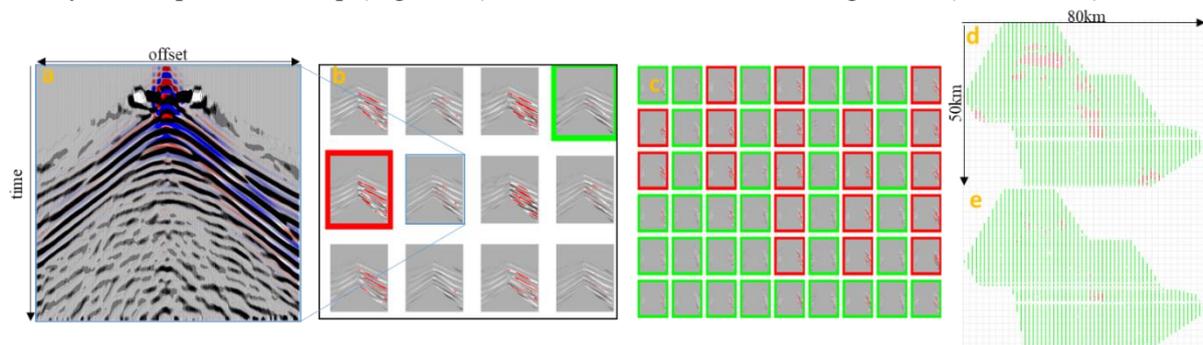
Our first example is cycle-skip QC for FWI. During velocity model building with FWI, initial QC is done by observing the mismatch between the generated synthetics using the FWI input model and the real data, which can be observed visually by overlaying these two datasets (Figure 1a).

Traditionally, time lag values between these two datasets, obtained by either warping or other fitting methods, are calculated for each trace, and then the average time-lag for each trace is used as the cycle-skip QC metric.

Here we use a supervised classification approach on the superimposed image directly (Figure 1a), meaning that both the seismic data character and the time-variant time-lag information will be included in the QC. The shot gathers are then muted to get only what will be used in the FWI.

First a training set (Figure 1b) containing 4% of the shot gathers for the entire dataset for the entire survey was selected. Gathers were sporadically selected across the survey to capture the diversity of the area and to avoid any geological bias in our prediction. Second, we labelled all of the superimposed shot point images into either a low or a high risk group (Figure 1c). In our case, the low and high risk labels represent 79% and 21% of the training dataset, respectively.

Finally for the prediction step (Figure 1c), we chose to use the K-NN algorithm (Harts, 1968).



**Figure 1** (a) Shot overlay between the model (blue-red) and the observed data (wiggle). These training set shots are then labelled, (b), for cycle-skip risk as high (red) and low (green) and automatic classification is run for all survey shots (c). (d) and (e) show a cycle-skip risk map before and after 6 iterations of FWI. Areas affected by gas pockets are identified and the remaining high-risk areas need further manual investigation.

This method is the most commonly used in shallow learning as it can capture the non-linearity in the data while remains computation efficient. Using the accuracy of the classifiers and a receiver operating characteristic (ROC) curve analysis (Powers, 2011), we chose a 3-NN classifier based on the  $l_2$  distance

as the similarity metric. The cross-validation score, using k-fold with k=10, gave us a good accuracy score of 92% ( $\sigma = 3.7 \times 10^{-3}$ ).

We applied this to the whole dataset and generated a map view of the classification result (Figure 1d) with FWI input model. The red points indicate gathers possibly affected by cycle-skipping and these areas are well correlated with the presence of large gas pockets. This classification QC map view of helps us to quickly design specific mutes for the affected shots to remove severely cycle-skipped data so FWI can update the velocity successfully. This is shown in Figure 1e where we show the results of the cycle-skipping QC classification after FWI. Inspection of the results indicated that the application of supervised learning directly on shot gathers provided effective and automated cycle skipping QC. Note that it could be interesting to extend this binary classification to larger group of classes to study different cycle-skipping patterns, so that the mute design can be optimized and automated for each individual class.

### Rig noise detection case study

In our second example we propose a classification scheme for rig noise detection, the goal being to identify data contaminated by rig noise for further processing. Supervised classification methods often provide alternatives when conventional seismic attributes (e.g. absolute amplitude, dominant frequency) do not provide a clear distinction between signal and noise. This is especially the case for rig noise where amplitude, frequency and dips are quite similar to the recorded signal; however, despite similarities in terms of conventional attributes, rig noise can be easily identified visually.

The direct use of original full shot points was not practical for hardware memory limitation, so data reduction was needed, keeping in mind that noise information should be preserved. We first calculated the amplitude envelope, and then reduced both the spatial and temporal sampling to reduce the data size by a factor of 20.

A comparison of several classification methods was performed on the training set (17% of the survey) prior to applying classification on the full survey. Of the methods compared, LR and support vector machine (SVM) (see Hastie et al., 2001 for both methods) were both ranked first with a cross-validation score of 98%. We chose LR because it has the advantage of outputting the probability of belonging to a class, and therefore provides an uncertainty estimate for the classification.

We used a LR to classify the full 60,000 shot points into clean and noise-polluted classes (Figure 2). This classification shows the area where rig noise is expected (in red), forming a circle around the platform location and the clean lines (shot after the drilling) passing through the rig location (in green). An inner circle of unpolluted data is also noticeable where rig noise is appearing on the very shallow part of the section, above our defined detection window, as rig noise in the very shallow part of the data is not an issue.

Having obtained a satisfactory classification on shot gathers of 50,000 samples each, we wanted to define a strategy to further reduce the data dimension to speed up the classification process and to study the level of compression that can be applied to the data without losing important information.

We chose a bi-directional LSTM auto-encoder (Hochreiter and Schmidhuber, 1997; Graves and Schmidhuber, 2005; Zhao et al., 2017) and treated the data as a time series (as we would in signal processing) rather than an image. An auto-encoder's purpose is to reduce the information to a small number of features via convolutional layers (Bengio, 2013) and to decode it based on a cost function minimization.

Our auto-encoder is unsupervised, since it has to reconstruct a given input image under the constraint of data reduction without any labelled data. However, the classification itself (this time applied on the encoded part of our auto-encoder) is supervised, and needs labelled data.

Based on this, we define two "losses", one for the auto-encoder and one for the classifier (Figure 3):

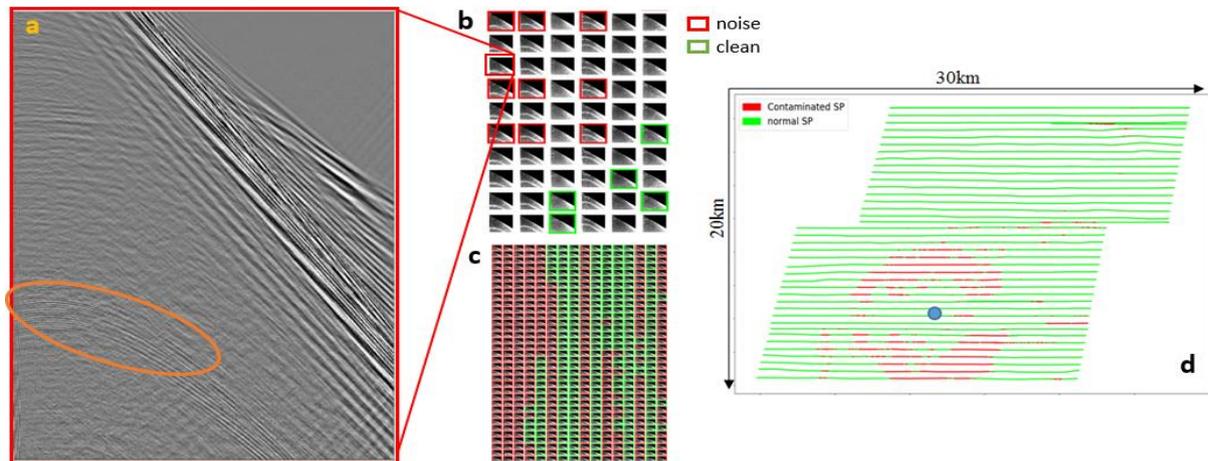
- The loss of the auto-encoder  $l_{CCE}(X, X^*) = \sum_{i=1}^m X_{j,i} \log(p_{j,i})$
- The loss of the classification  $l_{BCE}(Y, \bar{X}) = -(Y \log(p) + (1 - Y) \log(1 - p))$

with  $p$  the probability of the observation  $j$  to be in the  $i^{th}$  class and  $X_{j,i}$  the true label. The auto-encoder loss is the categorical cross-entropy loss and the classification is the binary entropy loss. We kept the same classifier as before, i.e. logistic regression with a  $l1$  regularization term.

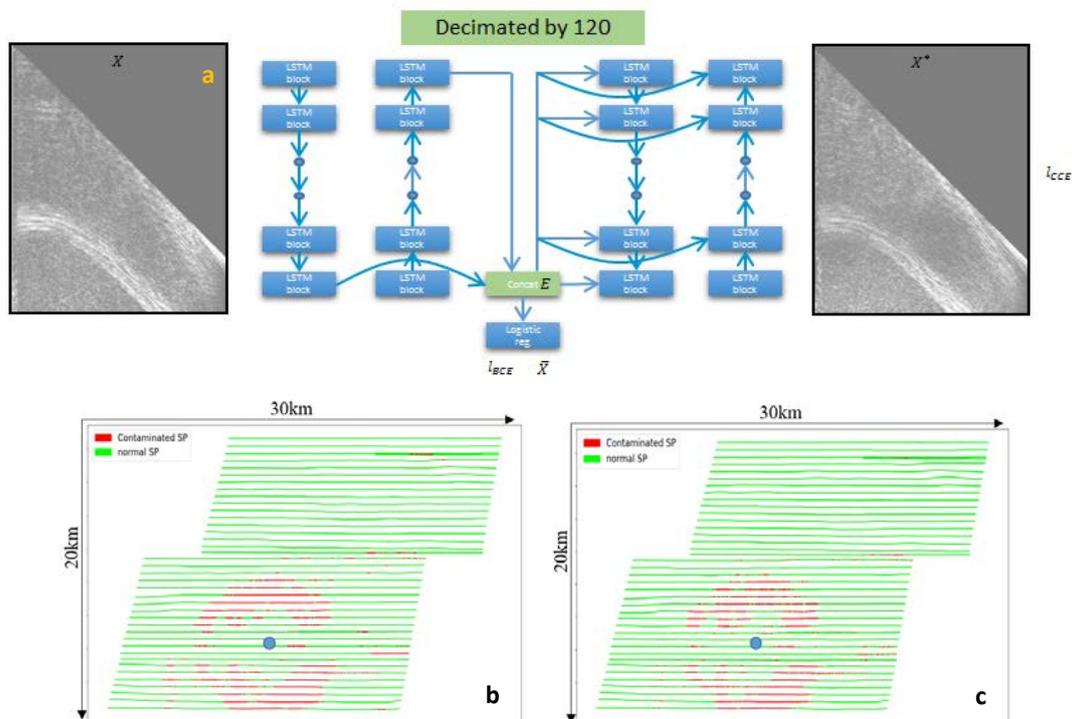
To train the network, we have to minimize  $l_{net}(X, \bar{X})$ , which represents the loss of the classification considering the input image of the auto-encoder. To achieve this, we tested three methods:

- Minimize  $l_{CCE}(X, X^*)$  using all the data (labelled and unlabelled), then minimize  $l_{BCE}(Y, \bar{X})$  using the labelled data only
- Minimize  $l_{net}(X, \bar{X}) = l_{BCE}(Y, \bar{X})$ , with no constraint from the auto-encoder
- Minimize  $l_{net}(X, \bar{X}) = l_{CCE}(X, X^*) + \beta l_{BCE}(Y, \bar{X})$ , with  $\beta = 1$  when the input gather is labeled, 0 elsewhere.

By testing on a large set of synthetic data, we obtained a classification accuracy of 63%, 66% and 92% for the first, second and third methods respectively.



**Figure 2** (a) shows a recorded shot contaminated with rig noise (orange ellipse), (b) which is then labelled and predicted with logistic regression (c) prior to building a rig noise map (d). Noise (red circle) is detected around the rig (blue dot). However there are some new lines acquired through the rig location after drilling finished so they are seen as free of noise in green.



**Figure 3** Illustration of LSTM auto-encoder used for data reduction (a), original classification (b) and encoded classification (c). Recorded data are accurate enough to show that reduced vectors contain all needed information required to perform a good prediction. LR used from this encoded data as inputs gives a very similar prediction and is less sensitive to noise (c) to the original one using 120 times more information (b)

We applied the third method to the data from the previous example using the same input, *i.e.* the resampled and trace dropped amplitude envelope of the signal. By comparing results in Figure 3c with previous results duplicated in Figure 3b, we first observe that the rig noise is correctly classified. We also observe that the classification from the encoded information is less sensitive to other noise, as the Figure 3c contains fewer misclassified points.

This unsupervised data reduction (by a factor of 120 by design) may solve memory limitation issues (a single shot point can contain several million samples). There are also multiple advantages in not relying on manually extracted attributes: no need to define which attribute to use; speeding up the process; potentially avoiding the creation of unbiased further predictions.

## Conclusions

Supervised statistical learning is becoming a common tool in the industry for assisting geophysicists during seismic data processing. Its effectiveness in defining regions, which necessitate a specific additional process or QC, as demonstrated with two examples using field data, is valuable in ensuring high quality seismic data for interpretation. Currently, data reduction is still a task which requires human intervention and feature engineering and is therefore inefficient; so we have proposed methods based on the LSTM auto-encoder to automate this step.

## Acknowledgements

We wish to thank Equinor, Petoro, and Lundin for the data examples. This paper gives the views of CGG, and may not necessarily be supported by Equinor, Petoro and Lundin. Furthermore, we thank Celine Lacombe for her help and advice, and we also thank Song Hou and Vivien Massart for their feedback.

## References

- Bengio, Y. [2013]. Deep learning of representations: Looking forward. In *International Conference on Statistical Language and Speech Processing*, 1-37
- Coléou, T., Poupon, M., & Azbel, K. [2003]. Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *The Leading Edge*, **22**(10), 942-953.
- Graves, A., & Schmidhuber, J. [2005] Framewise phoneme classification with bidirectional LSTM networks. In *Neural Networks International Joint conference on* (Vol. 4) 2047-2052
- Hart, P. [1968] The condensed nearest neighbor rule. *IEEE Transactions on Information Theory*, vol. 14, no. 3, 515-516
- Hastie, T., Tibshirani, R., Friedman, J. [2001]. *The Elements of Statistical Learning*. New York, NY, USA: Springer New York Inc.
- Hochreiter, S., & Schmidhuber, J. [1997] Long short-term memory. *Neural computation*, **9**(8), 1735-1780
- Ma, Y., Ji, X., BenHassan, N. M., & Luo, Y. [2018] A deep learning method for automatic fault detection. *SEG Technical Program Expanded Abstracts*, 1941-1945
- Martin, T., Saturni, C., & Ashby, P. [2015] Using machine learning to produce a global automated quantitative QC for noise attenuation. *SEG Technical Program Expanded Abstracts*, 4790-4794
- Powers, D.M.W [2011] Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies Volume 2, Issue 1*, 37-63
- Wong, T., & Luo, Z. [2018] Recurrent Auto-Encoder Model for Multidimensional Time Series Representation.
- Zhang, Z., & Wang, P. [2015] Seismic interference noise attenuation based on sparse inversion. *SEG Technical Program Expanded Abstracts*, 4662-4666
- Zhao, R., Yan, R., Wang, J., & Mao, K. [2017] Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors*, **17**(2), 273
- Zhao, T., & Mukhopadhyay, P. [2018] A fault detection workflow using deep learning and image processing. *SEG Technical Program Expanded Abstracts*, 1966-1970