

# Determining the Added Value of Surface Distributed Acoustic Sensors in Sparse Geophone Arrays using Transfer Learning with a Convolutional Neural Network

Samir F. Jreij, Whitney J. Trainor-Guitton, James L. Simmons

Colorado School of Mines - Department of Geophysics

## SUMMARY

Distributed sensors have widely been used in boreholes and their added value is apparent in these environments. Surface acquisitions with distributed sensors have not been quite as successful due to the limited understanding of the types of waves that the instrument records. This paper discusses experiments to identify if there is any added value to using distributed acoustic sensors with sparse geophone arrays in 2-D surface acquisition. The results qualitatively show that 2-D surface DAS arrays are able to recover migrated images similar to sparse, multi-component geophone arrays. Quantitative analysis was also performed using transfer learning in a convolutional neural network. The quantitative analysis shows that adding distributed sensors for this experiment only helped in decreasing false negatives and increasing the true negatives in identifying reflectors. This paper provides the framework for future quantitative analysis in the geophysics field using machine learning.

## INTRODUCTION

Distributed acoustic sensing (DAS) is a technology that uses Rayleigh scattering in a fiber-optic cable to detect elastic signals when the particle motion is parallel to the sensing fiber (Hornman et al., 2013). The two main components used in distributed sensing are the interrogator unit and the fiber-optic cable. The interrogator unit works as a light source and receiver. It sends a known pulse of light down the fiber. Nearly any type of existing fiber-optic cable can be used in conjunction with an interrogator unit. Small imperfections within the fiber causes backscattering of light. The fiber undergoes a strain when a seismic wavefield approaches and a scattering of light is produced that is different from base conditions. The interrogator unit is able to measure the Rayleigh backscattering that is produced and relate it to the local strain along the fiber (Parker et al., 2014).

Although DAS is utilized mainly in borehole acquisition, it has been shown repeatedly that the technology is capable of recording seismic data in surface acquisition (Daley et al., 2013; Yavuz et al., 2016; Jreij et al., 2018). Both Daley et al. (2013); Yavuz et al. (2016) concluded that observing P-wave reflections in a single surface DAS fiber is difficult. The issue that is apparent in their experiments is that short offset P-waves arrive perpendicular to the DAS fiber. DAS, however, is limited by its broadside insensitivity, or in other words, the

fiber is most sensitive to waves that have particle motion parallel to the orientation of the fiber. In a geophone, this directional insensitivity is observed as well, but by recording multiple components, the insensitivity can be minimized.

DAS has many advantages in various industries. One of the most notable for this paper is that DAS enables seismic surveys to be acquired with dense sampling (as small as 10-centimeter receiver spacing) at large distances (tens of kilometers long). Achieving even 1-meter sampling with conventional geophone is expensive and logistically difficult.

This paper is inspired by the PoroTomo survey geometry shown in Figure 1. The geophone acquisition has 238 geophones spaced nonuniformly anywhere from 50 meters to 150 meters apart. This receiver spacing is much greater than a conventional reflection seismic surveys which leads to aliased data as there is not enough spatial sampling. Fortunately, the PoroTomo survey also included almost nine kilometers of trenched, horizontal DAS cable. The experiments presented in this paper utilize the dense spatial sampling of DAS fiber and the broadside sensitivity of sparse, multi-component geophones. These experiments translate the source and receiver spacings from the 3-D survey into 2-D.

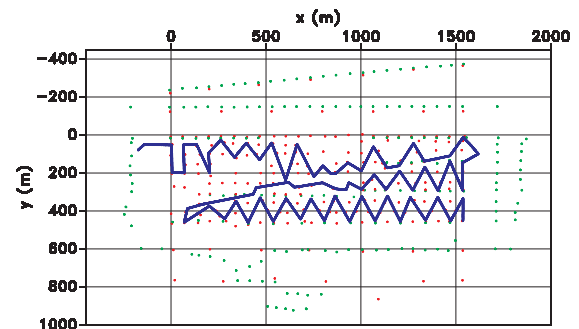


Figure 1: Sources (shown in green), multicomponent receiver (shown in red), and distributed acoustic sensing (shown in blue) geometry used in the PoroTomo survey.

## Fiber Attributes

Understanding how DAS fiber works is essential prior to modeling data. As discussed in the introduction, fiber records data by monitoring phase lags between injected pulses. These phase lags are produced when a seismic wave disrupts the original conditions by stretching and squeezing the fiber. The particle motion of the incoming seismic wave must be polarized properly to stretch and squeeze the fiber. Fiber directionality can

## Determining the Added Value of Surface DAS in Sparse Geophone Arrays

be visualized with a simple 2-D example (Figure 2): consider wave propagation in the x-z plane in a homogeneous, flat-layered, isotropic or vertical transverse isotropic medium.

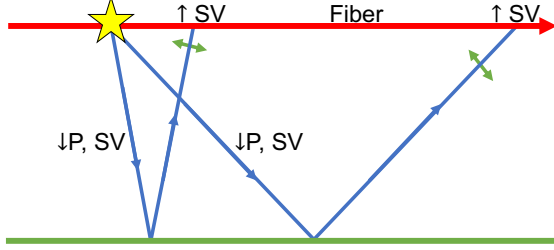


Figure 2: 2-D homogeneous, flat-layered, isotropic ray path example. The blue lines represent ray paths of the labeled waves, the blue arrow represents the propagation direction of the wave, the green line represents a horizontal reflector in the subsurface, and the yellow star represents the source. This figure shows a demonstration of P-SV or SV-SV wave effect on fiber using ray paths. Particle motion perpendicular with the propagation direction (green arrows). The fiber will only record data at short offsets.

It was shown in Jreij et al. (2018) that only certain reflected waves are recorded by the DAS fiber. These depend on the particle motion of the wave. Jreij et al. (2018) concludes that, for the model presented in Figure 2, only long-offset P-waves and short-offset SV-waves are able to be recorded by the surface DAS fiber in a 2-D line. SH-waves will not be recorded for a surface DAS fiber in a 2-D line.

### MODEL DESIGN

Siler and Faults (2013) mapped the faults beneath the PoroTomo survey. It is important to image these faults in detail as they are driving factors behind the recharge of the geothermal reservoir (Feigl, 2017; Folsom et al., 2018). For this reason, a slice from the fault model of the PoroTomo survey shown in Figure 3 is used as a reflection velocity model.

A modified version of the conventional elastic finite difference modeling (FDM) code (*ewefdm*) present in Madagascar is utilized (Equation 1),

$$\rho \frac{\partial^2 U}{\partial t^2} = \nabla \cdot \vec{t} + \vec{f} \quad (1)$$

where  $U$  represents displacement,  $t$  represents time, and  $\vec{f}$  represents the source function. The conventional elastic FDM code extracts displacement data at receiver locations. Although this can be used to extract the geophone data, it is not the right solution for DAS data as

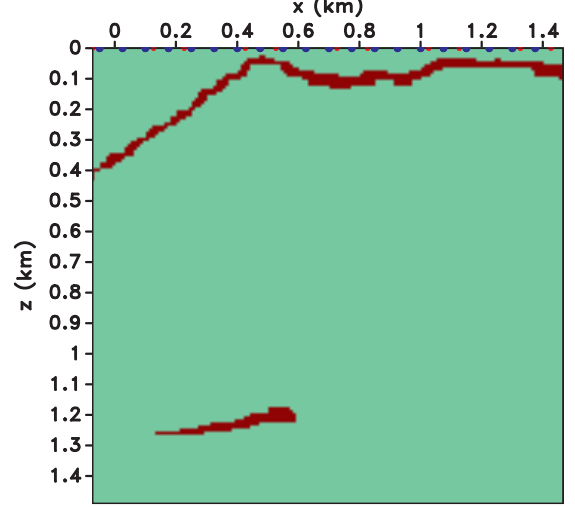


Figure 3: Reflectivity model from (Siler and Faults, 2013) extracted at the inline slice location of a borehole DAS acquisition within the PoroTomo survey.

fiber records strain along the fiber. Since strain is the spatial derivative of displacement, the code was modified to output strain and displacement by taking the spatial derivative of the displacement wavefield. After taking the spatial derivative of the wavefield, the strain can be extracted at all of the locations where DAS exists. In the 2-D examples shown, a single surface fiber only records the x-component of strain, the contributions from ZZ- and XZ-data are muted when modeling the DAS strain data.

Seismic sources in the PoroTomo survey are not on a uniform grid. In fact, some consecutive sources are as far as 150-meters from each other. The aim of this paper is to discuss how DAS and multi-component geophones can be used together to create a more detailed image. Seismic illumination describes how much of the subsurface can be imaged given a source-receiver geometry. Illumination in seismic surveys is highly influenced by source-receiver spacing. For the purpose of this paper, a uniform source spacing of 75 meters (which is about the average source spacing in the PoroTomo survey) was utilized to minimize migration artifact effects from poor illumination.

As discussed in the previous section, S-waves are more useful using surface DAS fiber. SV-sources are utilized for the numerical experiment in this paper. The resulting wavefield is recorded by multi-component geophones at 100 meter spacing. The average geophone spacing is about 70 meters originally in the PoroTomo survey, but to analyze the geophone spacing closer to the extremes of this experiment, a geophone spacing of 100 meters is utilized. The wavefield is also recorded by a DAS fiber in the x-direction at 1-meter spacing, consistent with the actual PoroTomo survey. The data are then imaged

# Determining the Added Value of Surface DAS in Sparse Geophone Arrays

using the energy norm imaging condition.

## Energy Norm Imaging Condition

An imaging condition is needed to produce images from the wavefields. The energy norm imaging condition is utilized for this experiment (Rocha et al., 2016). Conventional imaging conditions for reverse time migration rely on taking the inner product of source and receiver wavefields at each spatial location (Claerbout, 1985). The energy norm imaging condition relies on the energy conservation principle. The imaging condition is shown in Equation 2,

$$I_E = \sum_{e,t} [\rho \dot{U} \cdot \dot{V} + (\lambda + \mu)(\nabla \cdot U)(\nabla \cdot V) + \mu(\nabla U) : \nabla V] \quad (2)$$

where  $U$  is the source wavefield oriented forward in time,  $V$  is the receiver wavefield oriented backwards in time, and  $\lambda$  and  $\mu$  are the Lamé parameters. This imaging condition outputs an image that represents the total reflection energy from extrapolated wavefields (Rocha et al., 2016). There are various benefits to using this imaging condition, but the most important for this experiment is the reduction of images from the conventional four images (PP, PS, SP, SS) to one finalized elastic image. The resulting migrated image of the model presented in Figure 3 is shown in Figure 4a and Figure 4b for the combined sparse multi-component geophone data and the DAS data with the multi-component geophone data, respectively.

The images shown in Figure 4 are 2-D numerical examples which allow for quantitative analysis. All of the experiments presented in the paper can be qualitatively analyzed and discussed, but qualitative analysis is always different between people due to different biases and perspectives. A method to quantitatively analyze the experiments is needed to do effective comparisons. Calculating the posterior reliability allows for quantitative image comparisons. A quantitative image comparison requires a technology to compare two images on a one-to-one scale. This paper discusses using a technology from the machine learning field called the convolutional neural network.

## Convolutional Neural Network

Machine learning is a field within computer science that focuses on the ability of computer systems to learn weak or higher-dimensional patterns within data without being explicitly programmed for these patterns (Samuel, 1959). One of the most powerful machine learning algorithms is the neural network. They consist of many layers in parallel and every layer consists of a number of nodes. All neural networks consists of at least two layers: the input layer and output layer. All of the extra layers in between the input and output layers are the hidden layers. Each layer receives the output from the previous layer based on if the previous neuron is fired or not.

Convolutional Neural Networks (CNN's) in particular are at the core of most state-of-the-art computer vision solutions for a wide variety of tasks. One of the most accurate CNN image classifiers is the Inception-v3 model (Szegedy et al., 2015). The original Inception-v3 model is trained and tested on the 2012 ImageNet Visual Recognition Challenge (Russakovsky et al., 2015). The training dataset consisted of 1,000 object categories. The Inception-v3 model was able to perform with a 17.3% top-1 error. Top-1 error means that the class with the highest probability classification is the same as the target label. A top-1 error of 17.3% means the Inception-v3 model is able to perform with high accuracy, making it a top contender for a geophysics image classification problem. The Inception-v3 model utilizes "convolutions" to perform at a high accuracy which means the network separates the image into subsets to train the network for different features.

The Inception-v3 model's ability to identify features within an image can be leveraged within the geophysics field using transfer learning (the use of pretrained models on other data). Since the reverse time migration (RTM) is a synthetic example, the Inception-v3 model can be trained on "reflectors" and "not reflectors" that are produced by the RTM because the true model is known. There are about 500 other slices along both the X and Y axes of the PoroTomo grid. We take a number of these slices, migrate them, and produce training data for identifying reflectors. The training data is based on 100 meter by 100 meter (10 grid cell by 10 grid cell) windows of the migrated image. Examples of the reflectors are shown in Figure 5a and examples of not reflectors are shown in Figure 5b. The CNN is trained on 600 images of reflectors and 1900 images of not reflectors, giving a 25% to 75% split for reflectors to not reflectors.

## RESULTS

The quantitative measure of how accurate the CNN is in classifying the reflectors can be calculated using the posterior reliability of information shown in Equation 3,

$$Pr(\theta = \theta_i | \theta^{int} = \theta_j) = \frac{(Pr(\theta = \theta_i))Pr(\theta^{int} = \theta_j | \theta = \theta_i)}{Pr(\theta^{int} = \theta_i)}; \quad \forall i, j = R, NR \quad (3)$$

where  $\theta$  represents a true value of reflector or not reflector,  $\theta^{int}$  represents an interpreted reflector or Not Reflector.

A 100 meter by 100 meter image subset of testing data is created the same way the training data was created. The testing data is kept separate from the training data. The testing data is specifically from the images produced by the RTMs (Figures 4a-4b). The RTM is decomposed

## Determining the Added Value of Surface DAS in Sparse Geophone Arrays

Table 1: Confusion matrix for CNN created from the multi-component geophones only.

	$\theta_R^{int}$	$\theta_{NR}^{int}$
$\theta_R$	386 (94.14 %)	24 (5.86 %)
$\theta_{NR}$	2800 (87.10 %)	415 (12.91 %)

Table 2: Confusion matrix for CNN created from both data types together.

	$\theta_R^{int}$	$\theta_{NR}^{int}$
$\theta_R$	377 (92.95 %)	33 (7.05 %)
$\theta_{NR}$	2871 (89.31 %)	344 (10.69 %)

into 3,625 (100 meter by 100 meter) images with labels of “reflectors” and “not reflectors” so testing validation can be performed after classification. The results are shown in confusion matrix form in Table 1 and Table 2 with both the number count and the posterior reliability percentage.

## CONCLUSIONS

Conclusions can be made regarding the accuracy of the classifier and the added value of DAS with these tables. The classifier was only able to identify about 13% true negatives and 87% false negatives meaning it is over-trained on identifying true reflectors. The geophone image shown in Figure 4a is already quite a sharp image which is why a high posterior reliability is obtained for true positives. When stacking and normalizing with the DAS image (Figure 4b), mostly the artifacts are added in. This is why the true positives decrease when the DAS data are added. Although the classifier was over-trained on true reflectors, the CNN gives a general idea of what would happen if DAS fiber were added to sparse geophone arrays. The false positives and false negatives will decrease with more training iterations.

Conclusions can also be made on the use of machine learning in the geophysics field for quantitative analysis. It was shown that transfer learning with convolutional neural networks has value in calculating the posterior reliability of information. We have provided a framework for future quantitative analysis studies.

## ACKNOWLEDGMENTS

We would like to thank the PoroTomo team, whose financial support made this research possible. The PoroTomo project was funded by the Office of Energy Efficiency and Renewable Energy (EERE), U.S. Department of Energy, under Award Numbers DE-EE0006760 and DE-EE0005510. The reproducible numeric examples in this paper use the Madagascar open-source software package (Fomel et al., 2013).

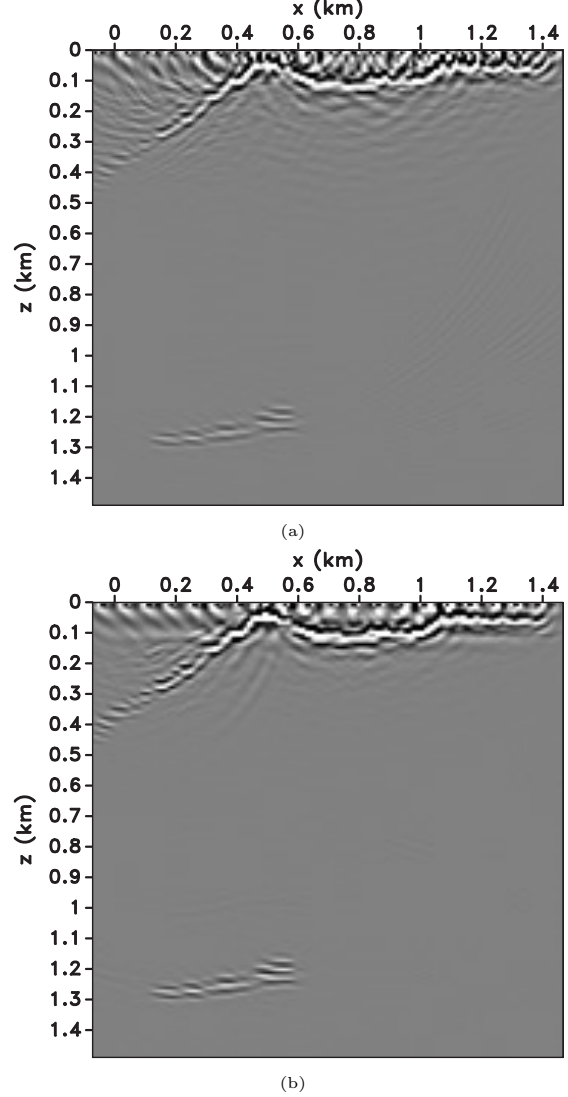


Figure 4: (a) Resulting image from migrating the sparse multi-component geophone data. (b) Combined image from migrating a single DAS fiber and the multi-component geophone synthetic data.

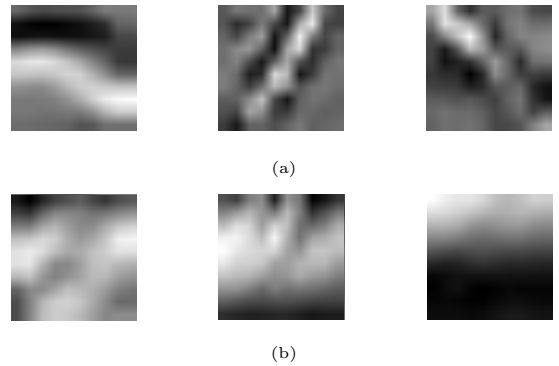


Figure 5: (a) Examples of reflectors fed to the CNN (b) Examples of not reflectors fed to the CNN.