

An Empirical Analysis of the Influence of Seismic Data Modeling for Estimating Velocity Models with Fully Convolutional Networks

Luan Rios Campos

Manufacturing and Technology Integrated Campus – SENAI CIMATEC
Salvador, Bahia 41650-010, Brazil

Peterson Nogueira

Manufacturing and Technology Integrated Campus – SENAI CIMATEC
National Institute of Science and Technology for Geophysics of Petroleum - UFBA
Salvador, Bahia 41650-010, Brazil

Davidson Moreira

Manufacturing and Technology Integrated Campus – SENAI CIMATEC
Salvador, Bahia 41650-010, Brazil

Erick Giovanni Sperandio Nascimento

Manufacturing and Technology Integrated Campus – SENAI CIMATEC
Salvador, Bahia 41650-010, Brazil

ABSTRACT

Seismic modeling is the process of simulating wave propagations in a medium to represent underlying structures of a subsurface area of the earth. This modeling is based on a set of parameters that determine how the data is produced. Recent studies have demonstrated that deep learning methods can be trained with seismic data to estimate velocity models that give a representation of the subsurface where the seismic data was generated. Thus, an analysis is made on the impact that different sets of parameters have on the estimation of velocity models by a fully convolutional network (FCN). The experiments varied the number of sources among four options (1, 10, 25 or 50 shots) and used three different ranges of peak frequencies: 4, 8 and 16 Hz. The results demonstrated that, although the number of sources have more influence on the computational time needed to train the FCN than the peak frequency, both changes have significant impact on the quality of the estimation. The best estimations were obtained with the experiment of 25 sources with 4 Hz and increasing the peak frequency to 8 Hz improved even more the results, especially regarding the FCN's loss function.

Keywords: Deep Learning, Geophysics, Velocity Model Estimation, Seismic Data Analysis, Fully Convolutional Networks.

1. INTRODUCTION

The exploration of subsurfaces of the earth is an expensive process. The first step is to place sources and receivers along a certain area and then propagate waves from one equipment to be recorded by the other. This process generates seismograms that

have much information of the structures underneath the region where the acquisition was made. The understanding of these seismic data may lead oil and gas companies to more assertively drill an area that may contain, for example, petroleum. However, two problems arise: the raw data by itself does not provide such kind of detailed information and they are too big and complex to be analyzed by humans.

In this scenario, computer simulations aim to replicate the process of seismic data modeling so certain methods, such as the Reverse Time Migration (RTM) or Full-Waveform Inversion (FWI), can be used. These techniques try to alleviate the aforementioned problems when a subsurface section is investigated. The RTM is a method that outputs an image where it is possible to identify the underlying structures of a subsurface, whilst the FWI is an iterative method that tries to solve a nonlinear inversion problem to output a high-resolution model of velocities of the subsurface. The latter complements the former, since its output is an input for the other, and both methods require a signal representation of the subsurface in order to operate, i.e., the seismic data. Moreover, the FWI also needs an initial velocity model in order to produce a new one with higher resolution. Offering an optimal initial velocity model to the FWI can diminish the computational power required to perform the method, increase its convergence rate by avoiding local minima and produce a high-resolution velocity model.

In the geophysics literature there are methods that help the production of initial velocity models. Authors such as [2] and [3] have, respectively, studied the use of reflection tomography and migration-based velocity analysis for such tasks. There are also approaches that consider the use of global methods such as genetic algorithms [4] and simulated annealing [5]. However,

the first two methods pose as a high time-consuming task and the last two demand more computational resources as the subsurface being analyzed increases in size, since they will require a larger population and, consequently, more modeling steps to carry the search on. More recently, researchers have been experimenting the use of deep learning techniques to solve geophysics problems [6], including seismic inversion [7] [8] [9] [10]. As far as it is of our concern, the first use of a fully convolutional network (FCN) for the velocity model estimation problem was addressed by [11], on which the FCN is trained with the seismic data of 1000 velocity models and tested with 20 examples of seismic data not seen during training. The seismic data was generated with sources and receivers placed on both the top and the bottom layers of the subsurface, which characterizes a well log seismic acquisition. The work of [12] shows how the same network used by [11] can be applied to a more conventional seismic acquisition, where the sources and receivers are positioned only on the top of the subsurface.

None of the works previously mentioned address the consequences of changing the number of sources or the peak frequency when training a deep learning method, except for [11], which compares only the case of seismic data with 1 and 10 shots. This comparison led the authors to conclude that training the FCN with 10 shots not only offers better results than using only 1 shot, but also contributes to reduce overfitting.

The goal of this study is to empirically analyze how the seismic data generated from synthesized velocity models can influence the estimation of such models using a FCN. This can contribute to the oil and gas industry by either demonstrating that deep learning methods may not necessarily require a high number of seismic shots, as it happens with other techniques, in order to be able to estimate a comprehensible velocity model, which can lead to reduce the expense to simulate, store and process non-synthetic seismic data, or offering a technique that perhaps is less sensitive to higher frequencies. Moreover, a brief comparison with the results of the experiments of different shots made by [11] is carried in this work. However, it is important to state that, because the scheme to generate the seismic data and the velocity models used in this work differ from the ones used by [11], their results are only discussed, not reproduced.

The experiments discussed here consider a finite-differences approach for the seismic modeling and alterations on some of its parameters, such as the number of sources and peak frequency, with the former varying from one central shot to 10, 25 and 50 equally spaced shots and the latter varying from 4 to 8 and 16 Hz. The seismic data is generated with basis on the same dataset of velocity models independently of changes on the modeling parameters, which consequently yields the same training and testing dataset throughout the entire analysis with modifications only on the resolution of the seismic data due to the differences of parameters.

The analysis is twofold: to compare the graphical results of the estimated velocity models of each experiment made as well as their metrics obtained after the FCN is completely trained. Analyzing the metrics can offer a statistical and more precise evaluation of the results obtained after training the neural network, since only a graphical analysis can mislead the interpretation of how changing the modeling parameters effects on the neural network training.

This study is organized as follows: the following section presents the mathematical and physical theory behind the seismic modeling; section three briefly presents the importance of velocity models; section four overviews fully convolutional networks applications and theory; section five describes the methodology and experiments; in section six a discussion of the results obtained with the experiments is made; and section seven concludes this work and points new directions of research based on the results obtained.

2. SEISMIC MODELING

Seismic modeling simulates the process of propagating waves on a subsurface area. This is done so researches can advance on processes that aid the understanding of subsurface areas prior to going into expeditions to them. This section is dedicated to briefly present some of the equations considering the modeling via the acoustic wave equation.

$$\frac{1}{v(x)^2} \frac{\partial^2 P_s(x,t)}{\partial t^2} - \nabla^2 P_s(x,t) = s(x,t) \quad (1)$$

The acoustic wave equation [13] [14] is described by Eq. (1), of which $x = (x', z')$ is the position on the subsurface for a 2D representation, $v(x)$ is the velocity at a given position, $P_s(x, t)$ is the source wave field and $s(x, t)$ defines the seismic source of the acoustic wave. Eq. (2) denotes the second spatial derivatives, i.e., the Laplacian operator (∇^2), for the two-dimensional case as:

$$\nabla^2 = \frac{\partial^2}{\partial x'^2} + \frac{\partial^2}{\partial z'^2} \quad (2)$$

One way to perform the seismic modeling is with the finite-differences method [14], which offers a simple and easy implementation [15] through the Taylor series, consequently leading to a discretization of the equations. Both Eq. (1) and Eq. (2) can be expanded by a Taylor series, but some conditions must be met in order to avoid the numerical dispersion and instability that may arise when discretizing a continuous-time equation [15].

On one hand, Eq. (3) [15] denotes the conditions to avoid the numerical instability of a 2D model, on which Δt is the time sampling interval, $\max(v)$ is the maximum velocity of the model, $\Delta x'$ and $\Delta z'$ are the spatial sampling interval respectively on the x and z axes.

$$\Delta t \leq \frac{1}{\max(v) \sqrt{\frac{1}{\Delta x'^2} + \frac{1}{\Delta z'^2}}} \quad (3)$$

On the other hand, Eq. (4) [15] illustrates the conditions to avoid the numerical dispersion problem of a bi-dimensional model: f_{max} is the maximum value of frequency allowed so the dispersion does not occur considering a given model, i.e., its maximum spatial sampling interval ($\max(\Delta x', \Delta z')$) and its minimum velocity ($\min(v)$). The parameter F is constant according to the order used for the Taylor series and it decreases as the order increases.

$$f_{max} = \frac{1}{F} \frac{\min(v)}{\max(\Delta x', \Delta z')} \quad (4)$$

The peak frequency (f_{peak}) is defined as approximately half of the max frequency (Eq. (5)) and represent the point of the spectrum of frequency with maximum amplitude.

$$f_{peak} = \frac{f_{max}}{2.3} \quad (5)$$

The information generated by the simulated wave propagation is translated into the seismic data, which corresponds to the values of transit time of the wave, the amplitudes and the phase of the events. The seismic data varies and respects undulation phenomes such as reflection, refraction and transmission

3. VELOCITY MODEL

A velocity model offers a representation of the structures present in a subsurface based on the velocity of propagation of the waves emitted from the sources and recorded by the receivers that are placed on the surface when the seismic data is being modeled. This is because the velocity of propagation directly depends on the type of medium through which a wave travels. Therefore, it is possible to determine a structure, i.e., rock, water, salt body, etc., according to its velocity.

As said before, there are different approaches to handle the initial velocity model problem in the geophysics literature and an optimal model can help when applying the full-waveform inversion. These models, however, are said to be smoothed (Figure 1a) and, although they can display an initial guess of the velocities of the subsurface, they lack details on its structural composition. In that sense, estimated models that have their structures clearly identified by their velocity values and are highly correlated to their ground-truth (Figure 1b) counterpart are known as high resolution models.

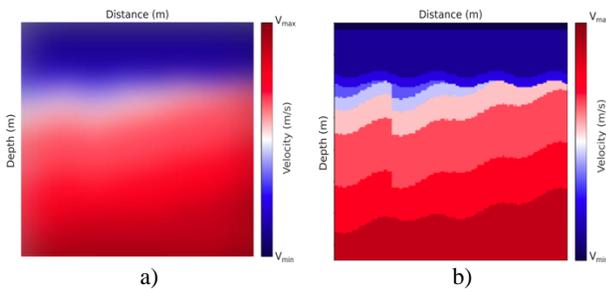


Figure 1 - An example of a) smoothed and b) ground-truth velocity models

4. FULLY CONVOLUTIONAL NETWORKS

Convolutional Neural Networks (CNN) were firstly introduced by [16] as an option for recognizing handwritten digits from the U.S. Postal Service. Later it was proved that CNNs can handle, besides images, speech and time-series problems [17]. In the recent years, deep learning has gained even more importance, especially after the Imagenet contest in 2012 and the development of AlexNet [18]. Since then, different proposals of deep learning methods with CNNs have been made, including the fully convolutional networks (FCNs).

The first proposition of use of an FCN was for handling semantic segmentation problems [19], which is the task of segmenting an

image into parts and classifying those parts into one of the predetermined classes.

Eq. (6) demonstrates the operation of the basic components of CNNs as [19] point out. In this case, x_{ij} is the data vector, y_{ij} is the next layer, k is the size of the kernel, s is the subsampling factor and f_{ks} defines the type of the layer (convolution, pooling or activation function). Therefore, [19] nominate CNNs that contain only layers ruled by Eq. (6) as fully convolutional or deep filter, since, differently from conventional approaches that use CNNs, the FCN does not contain fully connected (dense) layers, producing with its operations a nonlinear filter instead of a nonlinear function and reducing the number of parameters, computational time and dependency of the size of the image.

$$y_{ij} = f_{ks}(\{x_{si+\Delta i, sj+\Delta j}\}_{0 \leq \Delta i, \Delta j \leq k}) \quad (6)$$

5. METHODOLOGY AND EXPERIMENTS

In this section we describe the methodology and experiments of this work. Firstly, the synthetic velocity models are presented with details regarding their construction such as number of layers, minimum and maximum velocities and other characteristics. Then, the description moves on how the seismic data is generated and how the FCN is configured to handle it as inputs and estimate velocity models.

The velocity models and seismic data are both synthetic and they are built in different occasions. We first generate 1020 random velocity models and then we apply the finite-differences seismic modeling on each one of the recently-generated velocity models to create its corresponding seismic data.

The subsurface area being represented by the synthetic velocity models is a marine region of 3000 m in length by 3000 m in depth. The models are two-dimensional grids of 150 samples on both x (nx') and z (nz') axes and their number of layers vary from 8 to 12 layers, of which the first layer represents a water blade of 100 m deep and velocity of 1500 m/s. Subsequent layers have their depth randomly defined and their velocity is incremented (V_{incr}) in a crescent order, from the first layer velocity onwards, depending on how many layers (n) the model has and on its maximum ($V_{max} = 3500$ m/s) and minimum ($V_{min} = 1500$ m/s) velocities (Eq. 7), e.g., if the model has 12 layers, then the velocity will be incremented in 166,66 m/s at each layer. Furthermore, the models can have their layers inclined, undulated or containing fault structures. Figure 2 displays an example of such model.

$$V_{incr} = \frac{V_{max} - V_{min}}{n} \quad (7)$$

The seismic modeling is conducted on two fronts and it considers an arrangement of sources and receives as used by [12], i.e., they are simulated as they were placed on the top of the subsurface. The first front is to make different modeling changing only the number of sources and fixing a low frequency of 4 Hz. Since the sources are positioned on points (x', z') of the subsurface, by decreasing their quantity we might inflict on the acquisition of information belonging to certain regions of the given subsurface. Hence, the goal is to analyze how the changes on the number of sources will affect and how much of the velocity model the FCN can estimate.

The experiments of the second front consider modifications of the frequency using three different bands: 4 Hz, 8 Hz and 16 Hz. As said before, the frequency is important to avoid the numerical dispersion that may occur when calculating the acoustic wave equation through the Taylor series. The frequency of 16 Hz is the frequency of peak obtained from Eq. (5) after calculating the maximum frequency needed to avoid such dispersion when applying the parameters depicted in Table 1 in Eq. (4) using $F = 2$ as we considered a 32-order finite-differences. It is safe to say that any value below this threshold does not disperse the wave equation modeling, whereas frequencies above it disperse.

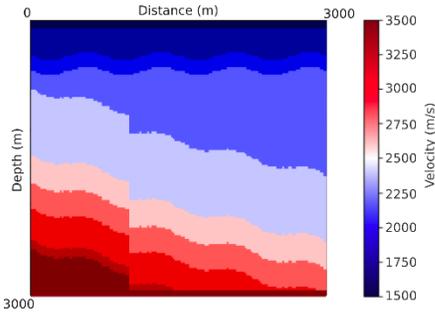


Figure 2 - A synthetic velocity model containing 10 layers, undulations, inclinations and fault structures

The frequency influences on how much of detail of the structures the modeling will be able to capture. It is expected that by lowering the frequency, the seismic data becomes smoother and consequently the non-linearity of the problem is decreased. Hence, the values picked for the experiments represent low-band (4 Hz), medium-band (8 Hz) and high-band (16 Hz) frequencies and aim to aid the understanding of how different bands can determine the level of details of the estimated models.

Table 1 - Fixed parameters considered when modeling the synthetic velocity models

Parameter	Value
nx'	150 samples
$\Delta x'$	20 m
nz'	150 samples
$\Delta z'$	20 m
nt	1500 samples
Δt	0.002 s

The FCN implementation takes the seismic data previously described as input and tries to estimate the velocity model corresponding to the input by minimizing the error between the estimated model and the ground-truth that generated the seismic data.

The work of [11] proposes the use of a U-Net [20] to perform the inversion of a seismic data into a velocity model. This FCN consists of two parts: an encoder and a decoder. The encoder is composed of convolution and max-pooling layers, which

gradually reduces the size of the image at the same time it determines what are the features of the input data. The decoder also has convolution layers, but the max-pooling are replaced by up sampling layers. This results in an increasing of the image size, to match the original image, and consequent localization of the features identified during encoding.

This study relies on the same U-Net proposed by [11], having the same quantity of layers and the same number of filters on each convolutional layer. However, two major changes were made in order to improve the results. Firstly, the stochastic gradient descent (SGD) optimization function was replaced by Adamax [21], which computes adaptive learning rates for each parameter and offers a more robust solution than the SGD's fixed learning rate when training a neural network model. Secondly, the rectified linear unit (ReLU) activation function was replaced by the parametric rectified linear unit (PReLU).

By using the ReLU activation function one can avoid the vanishing gradient problem that might occur in neural networks trained with gradient-based optimizers, such as Adamax. However, such function has another issue called the Dead ReLU, which might compromise a network from learning since the output of some of its neurons can be zero due to this function's nature (Figure 3a). The PReLU is an alternative to avoid this issue as it learns to parameterize the negative inputs of the neurons instead of assigning zeros to them as ReLU does (Figure 3). A more detailed study demonstrating how these changes on the activation function and optimizer leverage better velocity model estimations can be seen in the work of [22].

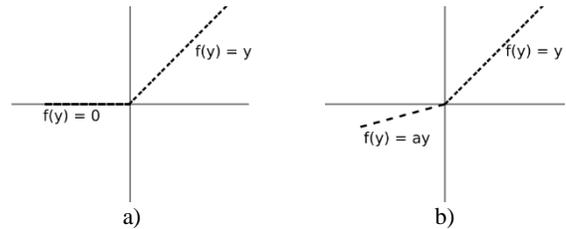


Figure 3 - Plot showing how the a) ReLU and b) PReLU activation functions work

The FCN is trained for 200 epochs with a batch size of 2 on 80% of the total of seismic data generated, saving 20% for the testing stage. The testing dataset is a portion of the original dataset unknown to the FCN, i.e., that has never been presented to it during the training phase, so it can offer an unbiased analysis of the model's performance. The batch size is small due to the size of the input and, although it could increase as the number of sources used during modeling decreases, since less sources means a reduction of size of the seismic data, it was kept unchanged throughout all experiments.

The evaluation of the FCN is made based on five different metrics with respect to the testing dataset: mean squared error (MSE), which is also the loss function, mean absolute error (MAE), coefficient of determination (R^2), Pearson's coefficient of correlation (r) and factor of two (fac2).

In this context, the MSE (Eq. (8)) measures how far an estimated model is from its respective ground-truth model. The bigger the

differences between one output and its corresponding target, the greater the penalization and, consequently, the associated error.

$$MSE = \frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (8)$$

The MAE (Eq. (9)) have lower values when compared to MSE's and indicates how much the difference of velocities between an estimated and its ground-truth model vary, i.e., if the MAE is of, say, 100, it means the output have 100 m/s of average error compared to the target.

$$MAE = \frac{1}{N} \sum_{k=1}^N |y_k - \hat{y}_k| \quad (9)$$

The coefficient of determination (Eq. (10)) indicates how better the estimation is when compared to a baseline model - either \bar{y} or $\bar{\hat{y}}$ variables of Eq. (10).

$$R^2 = \frac{[\sum_{k=1}^N (y_k - \bar{y})(\hat{y}_k - \bar{\hat{y}})]^2}{\sum_{k=1}^N (y_k - \bar{y})^2 \sum_{k=1}^N (\hat{y}_k - \bar{\hat{y}})^2} \quad (10)$$

The Pearson's coefficient (Eq. (11)) quantifies the linear relationship between an estimated model and its ground-truth counterpart, of which the value of -1 means opposite correlations, 0 means no correlation at all and 1 means total correlation

$$r_{y\hat{y}} = \frac{\sum_{k=1}^N (y_k - \bar{y})(\hat{y}_k - \bar{\hat{y}})}{\sqrt{\sum_{k=1}^N (y_k - \bar{y})^2} \sqrt{\sum_{k=1}^N (\hat{y}_k - \bar{\hat{y}})^2}} \quad (11)$$

The factor of two (Eq. (12)) determines how much of the estimation can be considered an outlier.

$$fac2 = 0.5 \leq \frac{\hat{y}_k}{y_k} \leq 2 \quad (12)$$

The parameters from Eq. (8) to Eq. (12) are as follows: N is the size of the velocity model grid, y_k is the k^{th} velocity of the ground-truth model (target), \hat{y}_k is the k^{th} velocity of the FCN's model (estimated output), \bar{y} is the mean of velocities of the target output and $\bar{\hat{y}}$ is the mean of velocities of the estimated output

6. RESULTS

An analysis and comparison of the experiments discussed previously are carried along this section. Both the experiments with variations on the number of shots and peak frequency are discussed and compared amongst them. It is important to restate that the experiments with adjustments of the number of shots were made with a peak frequency of 4 Hz, whereas when changes on the peak frequency occurs the number of shots is fixed at 25. The comparison is made both graphically and with regards to the metrics presented beforehand that are calculated after the deep learning model is fully trained.

The statistical comparison is to give a more reliable analysis, since considering only the estimated image of the velocity model can mislead the interpretation of the results. In this case, the goal is to minimize both the loss (MSE) and MAE metrics at the same time it maximizes R^2 , r and $fac2$ to values as close to 1 as possible. Besides the metrics, the time (in hours) taken to train the model also composes the analysis.

The graphical results of the estimation of one ground-truth model from the testing dataset can be seen in Figure 3. The ground-truth model contains undulated and inclined layers, and a simple fault structure that is identified by the yellow ellipsis in Figure 4a. Analyzing only the images leads to pointing out that Figure 4c, Figure 4f and Figure 4g obtained the best representation of the ground-truth model because they contain not only well-positioned layers, with identification of their undulation and inclination, and a high precision of the velocities on each layer, as the other estimations do, but also a fair depiction of the fault structures.

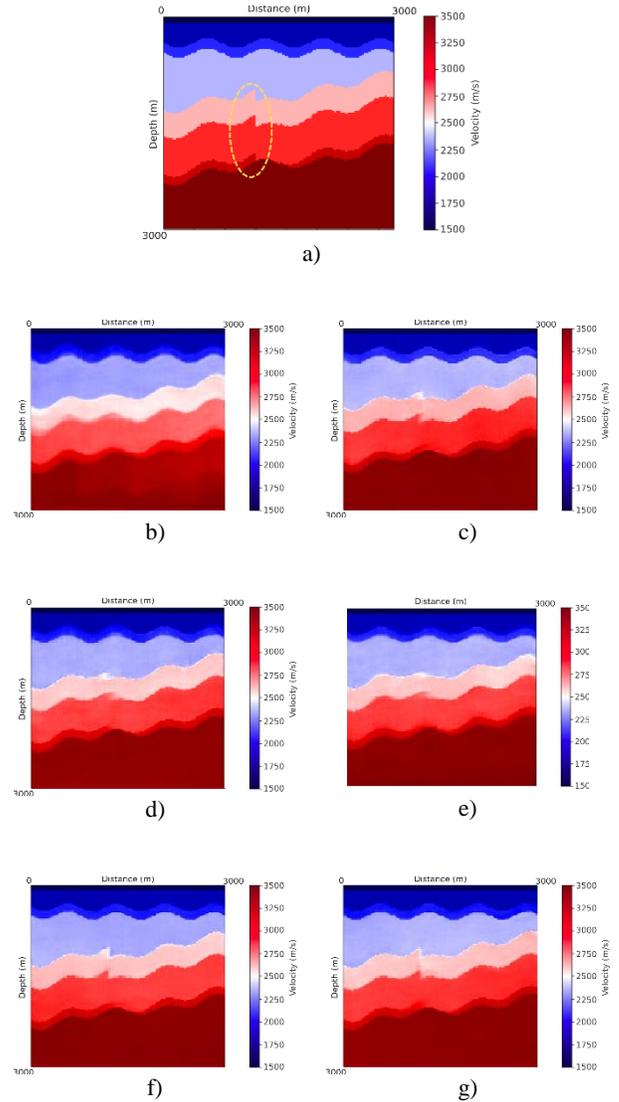


Figure 4 - a) Ground-truth velocity model and graphical results obtained with the experiments of b) 1 source, c) 10 sources, d) 25 sources and e) 50 sources all modeled with $f_{peak} = 4$ Hz and f) 8 Hz and g) 16 Hz both having 25 shots

On one hand, it is not safe to infer so straightforwardly that these representations are the best because this velocity model represents only one example of the entire testing dataset. This may indeed be a case where the FCN models estimated an optimal velocity model from the seismic data they were trained by, but there may also exist cases that the estimations greatly

differ from their ground-truth models. On the other hand, this analysis indeed validates the use of FCNs to produce velocity models from unknown seismic data.

Once the graphical investigation of many examples is imprecise and impractical, a quantitative evaluation of the statistical indicators belonging to each one of the experiments is conducted. These metrics are measured after the training phase using the entire testing dataset. Table 2 displays the metrics, the corresponding time it took for the models to be trained and the peak frequency for each experiment.

Before venturing into the comparison of the metrics, an association between the experiments and the computational time is conducted. It is possible to see from Table 2 that as the number of shots increases, so it increases the computational time taken to train the FCN. This happens because the number of shots have a direct influence on the size of the seismic data as additional shots mean adding matrices of size $nt \times nx'$ to the seismic data.

On the other hand, the peak frequency does not seem to have much importance to the computational time. Considering the experiment of 25 shots in Table 2, since the modeling had a peak frequency of 4 Hz, and comparing it with the time of the experiments of 8 Hz and 16 Hz, as both have 25 shots, there is no clear relation of computational time and higher or lower frequencies. In fact, the result that achieved the lowest time is the one with the highest frequency and the experiment with medium frequency took the longest to train.

Table 2 - Results of the evaluation metrics and the time for training (in hours) for each one of the experiments with changes on the number of sources (shots) with fixed peak frequency of 4 Hz and on the frequency with a fixed 25 number of shots

	1 shot	10 shots	25 shots	25 shots	25 shots	50 shots
f_{peak} (Hz)	4	4	4	8	16	4
Time (h)	7.19	7.43	8.10	8.14	8.09	9.07
MSE	14172	7313	6837	6126	7578	7207
MAE	75.39	45.41	44.19	46.79	54.69	49.72
R^2	0.954	0.975	0.977	0.980	0.974	0.976
R	0.983	0.989	0.990	0.991	0.990	0.990
fac2	0.999	1.0	1.0	0.999	1.0	0.999

The evaluation metrics of each experiment, in general, demonstrated close values, but it is possible to notice that training the FCN with seismic data that have more shots does not necessarily indicate a better estimation. Even though the experiment with 50 shots demonstrates valuable results, i.e., it accomplished values close to 1 for the r , R^2 and $fac2$, and relatively low values for MAE and MSE, other experiments were able to surpass it. In this case, both experiments with 10 and 25

shots obtained better values in all metrics, of which the latter bested the former. Moreover, the experiment with 25 shots could be further improved when the modeling was made with 8 Hz, reaching the lowest value with the loss function (MSE) and the highest with R^2 and r metrics for all experiments. This, however, happened at the expense of slightly decreasing the MAE and $fac2$ metrics to values below the experiment of 25 shots and 4 Hz.

On the other hand, neither reducing much the number of shots nor increasing even more the peak frequency mean improvement on the estimation either. The worst results belong to the experiment with the central shot. In this case, the values of the metrics R^2 , r and $fac2$, though show little differences from the same metrics of the other experiments, were the lowest and the MSE and MAE were the highest amongst all. Moreover, the experiment with 16 Hz resulted in worst metrics than the one with 50 shots.

Although having the worst metrics, the FCN successfully inverted a seismogram of one shot into a velocity model. This possibly happened due to the size of the subsurface and the velocity model, which are considered small from a geophysics perspective, but this cannot be confirmed to happen as the subsurface becomes larger considering only the analysis made in this work.

Hence, considering the extent of the experiments conducted in this work, it is possible to conclude that the FCN not only can produce velocity models from unknown seismic data, but it can also deliver high-resolution models. Furthermore, the parameters used to generate the seismic data, combined with the size of the subsurface area and the size of its velocity model representation, play an important role in determining how high the model's resolution is going to be.

7. CONCLUSIONS

This work demonstrated how changing the number of sources and peak frequency of the seismic modeling can affect the training and evaluation of an FCN model that takes seismic data as input to estimate 2D velocity models.

The experiments firstly fixed the peak frequency at 4 Hz and varied the number of shots amongst one central shot, 10, 25 and 50 shots and then fixed 25 shots and varied the peak frequency to 8 and 16 Hz. The results showed that the best metrics for the FCN were obtained with the experiments of 10 and 25 sources and increasing the peak frequency from 4 to 8 Hz improved even more the estimation, especially regarding the FCN's loss. When the peak frequency was increased once again, the FCN reached lower metrics than the experiment with 50 shots. Nevertheless, the worst results amongst all were obtained with the seismic data produced by a single central shot. These results partially contradict the affirmation made by [11], since the results were indeed improved when increasing the number of shots from 1 to 10 and from 10 to 25 but they worsened when considering 50 shots. Additionally, there is no clear evidence whether the number of shots influences on the model overfitting or not. This might have happened due to the size of the dataset used in their work.

Initial conclusions for the experiments addressed in this work indicate that, depending on the size of the subsurface, training

the FCN with seismic data that have few shots is enough to estimate a velocity model. However, as the size of the model and subsurface increases, more shots may give a better representation of the area. Furthermore, the results imply that the FCN is, up to a certain point, less sensitive to higher peak frequencies as the results improved when the modeling was changed from 4 to 8 Hz, but they worsened when 16 Hz was considered.

In general, the results demonstrated to be valuable, since they show the possibility of training deep learning models with seismograms containing few shots and high frequencies to estimate optimal velocity models.

Further studies point to the need of analyzing whether few shots are indeed enough to estimate velocity models of larger and more complex subsurfaces. Furthermore, improvements on the training stage, such as mixing the dataset with low, medium and high frequencies or substituting the max-pooling layers for convolutional layers, can be made, and other deep learning methods, such as generative adversarial networks (GAN), may be studied to determine whether they behave differently from the FCN for the seismic inversion problem.

8. REFERENCES

- [1] CARCIONE, Jose M.; HERMAN, Gérard C.; TEN KROODE, A. P. E. Seismic modeling. *Geophysics*, Vol. 67, No. 4, 2002, pp. 1304-1325.
- [2] STORK, Christof. Reflection tomography in the postmigrated domain. *Geophysics*, Vol. 57, No. 5, 1992, pp. 680-692.
- [3] AL-YAHYA, Kamal. Velocity analysis by iterative profile migration. *Geophysics*, Vol. 54, No. 6, 1989, pp. 718-729.
- [4] SAJEVA, Angelo et al. Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the Marmousi model Genetic FWI for acoustic macro models. *Geophysics*, Vol. 81, No. 4, 2016, pp. R173-R184.
- [5] DATTA, Debanjan; SEN, Mrinal K. Estimating a starting model for full-waveform inversion using a global optimization method. *Geophysics*, Vol. 81, No. 4, 2016, pp. R211-R223.
- [6] WANG, Wenlong; YANG, Fangshu; MA, Jianwei. Automatic salt detection with machine learning. In: **80th EAGE Conference and Exhibition 2018**. European Association of Geoscientists and Engineers, 2018.
- [7] RÖTH, Gunter; TARANTOLA, Albert. Neural networks and inversion of seismic data. *Journal of Geophysical Research: Solid Earth*, Vol. 99, No. B4, 1994, pp. 6753-6768.
- [8] LEWIS, Winston; VIGH, Denes. Deep learning prior models from seismic images for full-waveform inversion. In: **SEG Technical Program Expanded Abstracts 2017**. Society of Exploration Geophysicists, 2017, pp. 1512-1517.
- [9] ARAYA-POLO, Mauricio et al. Deep-learning tomography. *The Leading Edge*, Vol. 37, No. 1, 2018, pp. 58-66.
- [10] WU, Yue; LIN, Youzuo; ZHOU, Zheng. InversionNet: Accurate and efficient seismic waveform inversion with convolutional neural networks. In: **SEG Technical Program Expanded Abstracts 2018**. Society of Exploration Geophysicists, 2018, pp. 2096-2100.
- [11] WANG, Wenlong; YANG, Fangshu; MA, Jianwei. Velocity model building with a modified fully convolutional network. In: **SEG Technical Program Expanded Abstracts 2018**. Society of Exploration Geophysicists, 2018, pp. 2086-2090.
- [12] Campos, Luan R.; Nogueira, Peterson; Nascimento, Erick. Estimating Initial Velocity Models for the FWI Using Deep Learning. In: **Proceedings of the 16th International Congress of the Brazilian Geophysical Society**. Brazilian Geophysical Society.
- [13] ALFORD, R. M.; KELLY, K. R.; BOORE, D. Mt. Accuracy of finite-difference modeling of the acoustic wave equation. *Geophysics*, Vol. 39, No. 6, 1974, pp. 834-842.
- [14] BAYSAL, Edip; KOSLOFF, Dan D.; SHERWOOD, John WC. Reverse time migration. *Geophysics*, Vol. 48, No. 11, 1983, pp. 1514-1524.
- [15] DOS SANTOS, A. W. G. **Waveform inversion applied to the analysis of seismic velocities using a multi-scale approach**. Master's thesis, Universidade Federal da Bahia, 2013
- [16] LECUN, Yann et al. Handwritten digit recognition with a back-propagation network. In: **Advances in neural information processing systems**, 1990, pp. 396-404.
- [17] LECUN, Yann et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, Vol. 3361, No. 10, 1995, pp. 1995.
- [18] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**, 2012, pp. 1097-1105.
- [19] LONG, Jonathan; SHELHAMER, Evan; DARRELL, Trevor. Fully convolutional networks for semantic segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**, 2015, pp. 3431-3440.
- [20] RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. In: **International Conference on Medical image computing and computer-assisted intervention**. Springer, Cham, 2015, pp. 234-241.
- [21] KINGMA, Diederik; BA, Jimmy. Adam: A Method for Stochastic Optimization. **International Conference on Learning Representations**, 2014.
- [22] Campos, Luan R., Nogueira, Peterson, & Nascimento, Erick. Tuning a Fully Convolutional Network for Velocity Model Estimation. In: **Offshore Technology Conference Brasil**. Offshore Technology Conference.