

# Validated artificial neural networks in determining petrophysical properties: A case study from Colombia

Ursula Iturrarán-Viveros<sup>1</sup>, Andrés M. Muñoz-García<sup>2</sup>, Jorge O. Parra<sup>3</sup>, and Josué Tago<sup>4</sup>

## Abstract

We have applied instantaneous seismic attributes to a stacked P-wave reflected seismic section in the Tenerife field located in the Middle Magdalena Valley Basin in Colombia to estimate the volume of clay  $V_{\text{clay}}$  and the density  $\rho$  at seismic scale. The well logs and the seismic attributes associated to the seismic trace closer to one of the available wells (Tenerife-2) is the information used to train some multilayered artificial neural networks (ANN). We perform data analysis via the gamma test, a mathematically nonparametric nonlinear smooth modeling tool, to choose the best input combination of seismic attributes to train ANNs to estimate  $V_{\text{clay}}$  and  $\rho$ . Once the ANNs are trained, they are applied to predict these parameters along the seismic line. From the continuous estimations of  $V_{\text{clay}}$ , we distinguish two facies: sands for  $V_{\text{clay}} < 0.5$  and shales when  $V_{\text{clay}} \geq 0.5$ . These estimations confirm the production of the Mugrosa C-Sands zone, and we draw the brown shale that correlates with the high-amplitude attributes and the yellow sand that correlates with the low-amplitude attributes. Using the well-log information for  $V_P$  and the facies classification (also in the well log), two cubic polynomials that depend on time (or depth) are obtained, one for sands and the other for shales, to fit the  $V_P$ . These two cubic polynomials and the facies classification obtained from the  $V_{\text{clay}}$  at the seismic scale enable us to estimate  $V_P$  at the seismic scale. To validate the 2D  $V_P$  and  $\rho$  predicted data, a forward-modeling software (the Kennett reflectivity algorithm) is used. This model calculates synthetic seismograms that are compared with the real seismograms. This comparison indicates a small misfit that suggests that the  $V_P$  and  $\rho$  images are representing the reservoir description characteristics and the ANN method is accurate to map these parameters.

## Introduction

A main goal of the oil industry worldwide is the determination of accurate reservoir models. Fundamental challenges for geologists and geophysicists to predict these properties are reservoir particularity and heterogeneity, which affect the performance of the reservoir and their well productivity. To this end, the integration of well-log and seismic data (with their seismic attributes) for reservoir characterization has been a constant aim of geoscientists and has grown exponentially. This fact has been extensively documented in the literature, to mention but a few authors (see e.g., Gastaldi et al., 1997; Russell et al., 1997; Hampson et al., 2001). In Roy et al. (2014), authors explored one of the recent techniques known as generative topographic mapping, which takes care of the shortcomings of Kohonen self-organizing maps and helps in data classification. In Ma et al. (2017), a combination of principal compo-

nent analysis and the artificial neural network (ANN) for facies classifications and porosity prediction is obtained by combining seismic attributes and well-log information. The study by Faraji et al. (2017) integrates core data, rock elastic properties, and 3D seismic attributes to delineate tight and low-reservoir-quality zones of the South Pars gas field, and they estimate rock elastic properties such as Young's modulus and Poisson's ratio along with porosity using ANN. Machine learning has recently gained a great deal of attention in the oil and gas industry. Some examples of this recent work are deep learning neural networks, autoencoders, deep convolutional networks for seismic facies segmentation, seismic salt body delineation, and seismic interpretation (see Chevitarese et al., 2018; Di et al., 2018; Shafiq et al., 2018; Veillard et al., 2018). Random forest and ANN have also been applied to estimate densities (see Prietzhev and Stanislav, 2018).

<sup>1</sup>Universidad Nacional Autónoma de México, Circuito Escolar S/N, Facultad de Ciencias, México and Coyoacán, Ciudad de México, C.P. 04510, México. E-mail: ursula@ciencias.unam.mx.

<sup>2</sup>Instituto Tecnológico Metropolitano de Medellín, Colombia. E-mail: andmmunozgar@unal.edu.co.

<sup>3</sup>JPGeosciences, Helotes, Texas, USA. E-mail: joparram@yahoo.com.

<sup>4</sup>Universidad Nacional Autónoma de México, Circuito Escolar S/N, Facultad de Ingeniería, México and Coyoacán, Ciudad de México, C.P. 04510, México. E-mail: tago@unam.mx.

Manuscript received by the Editor 16 January 2018; revised manuscript received 6 July 2018; published ahead of production 22 August 2018; published online 24 October 2018. This paper appears in *Interpretation*, Vol. 6, No. 4 (November 2018); p. T1067–T1080, 16 FIGS., 2 TABLES. <http://dx.doi.org/10.1190/INT-2018-0011.1>. © 2018 Society of Exploration Geophysicists and American Association of Petroleum Geologists. All rights reserved.

We present an algorithm that integrates well logs and seismic data aimed at predicting velocity and density distributions from a 2D seismic section. For this purpose, we use several well logs and a 2D seismic data set from the Tenerife field oil reservoir. The data consist of one well log (Tenerife-2) with conventional well logs that tie to the 2D seismic section. We perform depth-to-time conversion of well logs and up scale the well logs to seismic wavelengths. We sought a relationship between log-derived physical properties and seismic attributes. We applied the gamma test to obtain the best combination of seismic attributes that better estimates the density  $\rho$  and the volume of clay  $V_{\text{clay}}$ .

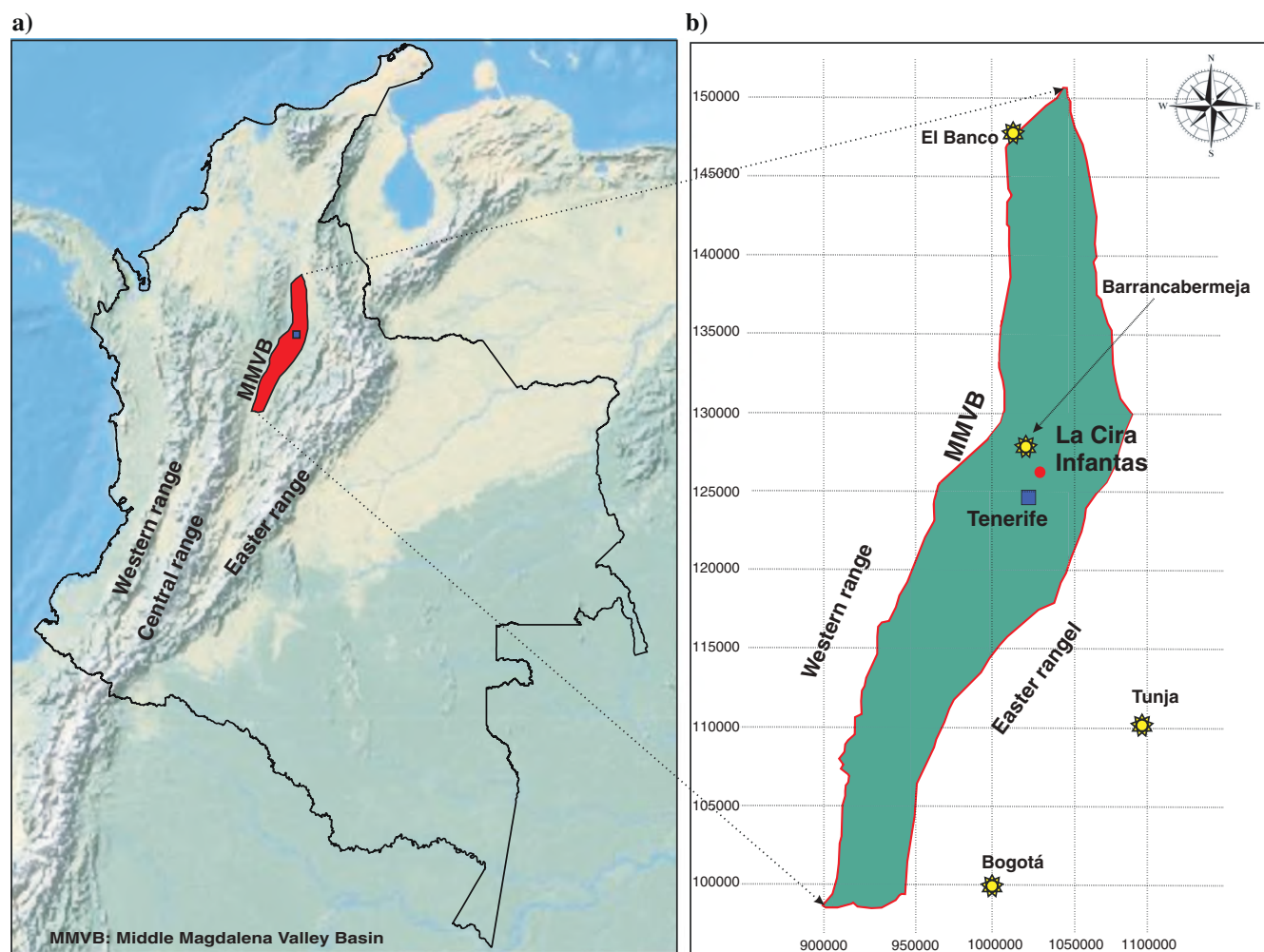
In this paper, we begin describing the basin evolution and the stratigraphy of the geologic site under study. We continue with the interpretation of the available seismic and well-log data to establish the framework of the lithologic features of interest. We follow with the implementation of a flowchart to validate the numerical algorithms for the data analysis and processing. Next, we applied the ANN design and training to obtain the optimal results that include the P-wave velocity and density images. Finally, we present the validation of the ANN method

that predicts rock physical properties from seismic data, using forward modeling.

### Basin evolution and stratigraphy

Exploration in Sub-Andean basins in Colombia has been traditionally challenging due to the complex geology, rough topography, and inaccessible areas. Within the Northern Andes mountain range is the central range (CR) and the eastern range (ER); they form the topographic flanks of the north-trending Middle Magdalena Valley Basin (MMVB) (see Figure 1). The MMVB takes its name from the Magdalena River that flows between these two mountain ranges (ER and CR) from the south to the Caribbean sea. This valley overlies the continental basement composed of granulite-grade metamorphic rocks overlaid by low-pressure meta sedimentary rock that also constitutes the crystalline basement of the CR and ER (see Restrepo, 1995).

The MMVB is one the most prolific petroleum basins of Colombia with a long history of hydrocarbon exploration that started with the discovery of a giant oil field called La Cira-Infantas in 1918. The oil fields in the basin occur mainly as either structural or stratigraphic traps



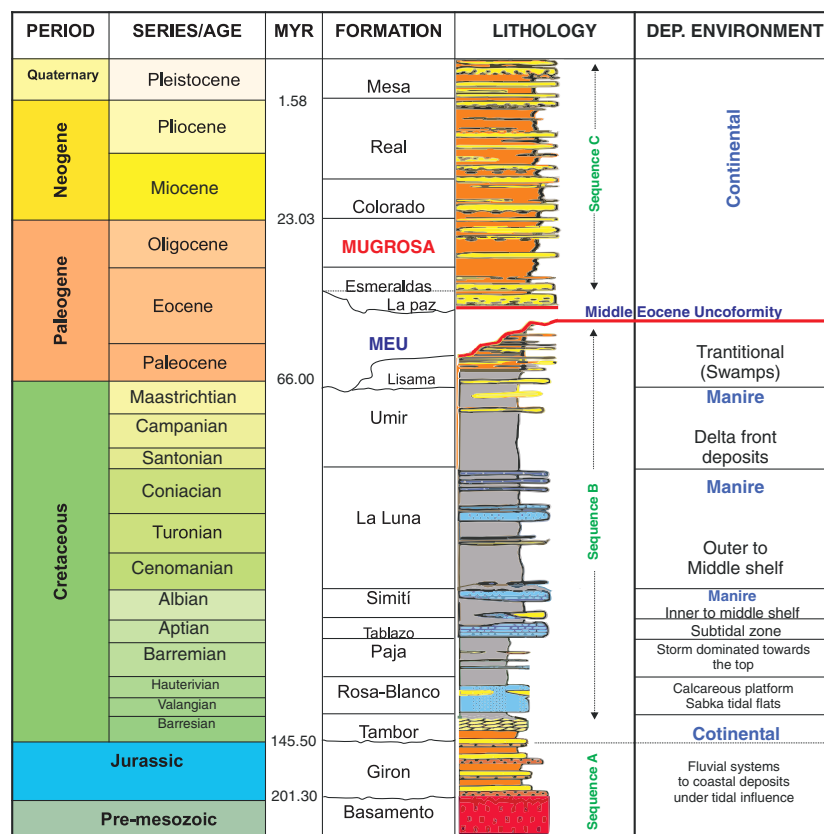
**Figure 1.** The MMVB. (a) Location and (b) fields La-Cifra-Infantas and Tenerife; the blue square is the study area.

in Paleogene clastic reservoirs. The Tenerife field (see Figure 1), operated by Ecopetrol S.A., is located in the central MMVB, approximately 20 km southwest of the giant La Cira-Infantas field. This field was discovered in 1971 after positive results of drilling the Tenerife-1 well. The appraisal strategy was followed by the drilling of wells Tenerife-2 and Tenerife-3. The development of the field was stopped later due to failed results in Tenerife-3. Tenerife-1 and Tenerife-2 had production of approximately 100 STBO/day of 22.8 American Petroleum Institute (API) crude oil; however, they are almost totally depleted today (see [Ecopetrol, 1983](#); [Sandoval, 2010](#)).

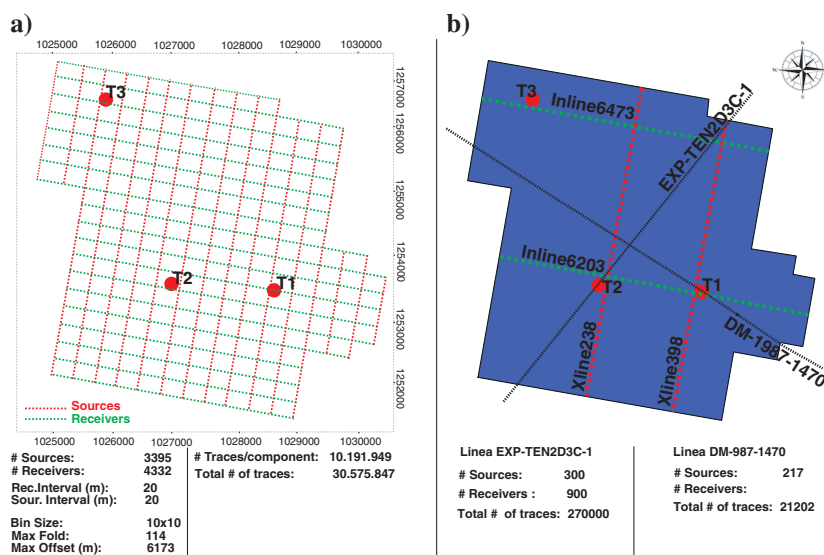
The continental deposits in the MMVB are composed mainly of red beds and volcanic effusive and pyroclastic deposits, although some marine facies appear locally. For the time of the Cretaceous, the basin sedimentation continued in a back-arc setting east of the Andean subduction zone (see [Cooper et al., 1995](#)). Throughout the Aptian time, the pure extension terminated and a postrift sedimentation phase began, controlled by thermal subsidence. At the same time, the following formations were deposited (see Figure 2): Tablazo, which consists of a set of limestones interbedded with black shales; Simiti, which is mainly made of shales deposited in an inner to middle shelf; La Luna, which consists of interbedding of limestones and black organic-rich shales; Umir, which marks a transition into a more clastic environment and is composed of gray shales and shaly sandstones with coal intervals. The major deformation phase of the CR affecting the MMVB marks a regional unconformity called the Middle Eocene unconformity (MEU). It is the most important boundary within the Cenozoic record and can be observed throughout the basin (see Figure 2). The MEU separates a premiddle Eocene deformation with thrust-controlled exhumation. For a comprehensive study in this area, see [Velasquez-Espejo \(2012\)](#).

### Interpretation and available data

The main goal of this work is to produce a petrophysical characterization for the MMVB at seismic scale and validate it. To achieve this goal, we combine information of the well logs from the well Tenerife-2 and the seismic line



**Figure 2.** Stratigraphic column of the MMVB illustrating the main petroleum system elements. The highlighted red curved lines are the Cenozoic unconformities, being the MEU the most important boundary that separated the transition between environments of marine and continental regimes. Figure modified from International Commission on Stratigraphy.



**Figure 3.** Map of Tenerife 3D-3C seismic acquisition. (a) The geometry is a regular square with sources oriented north-northeast/south-southwest, and receptors run nominally east-west in an orthogonal array. The total area of the survey is approximately 29 km<sup>2</sup>. The sources are in red, and the receivers are in green. The red dots denote the location of the three available wells in the field: Tenerife-1, Tenerife-2, and Tenerife-3. (b) Location of line: EXP-TENE-2D-3C and other lines close to one or two wells are marked. This figure is modified from [Velásquez-Espejo \(2012\)](#).



[illegible]

EXP-TENE-2D-3C. ANNs are trained to estimate the well-log measurements for  $\rho$  and  $V_{\text{clay}}$  using seismic attributes. The trained ANN is then applied to the section to produce these estimations at seismic scale.

In 2009 for the first time in Colombian exploration history, a 3D-3C survey was acquired in the MMVB over an area of approximately 29 km<sup>2</sup> (see Figure 3). An explosive type of source was used to generate a P-wave impulse, and 3C receivers measured the upgoing energy in three perpendicular directions at the surface. The vertical component primarily records the reflected P-waves, and the horizontal components record the displacement associated with converted S-waves. The survey was designed with high lateral resolution (bin size = 10 x 10 m) and long offsets (more than 6000 m) to guarantee high coverage (fold) and sufficient migration aperture. The acquisition geometry is an orthogonal survey. Receiver lines were east-west at 100° azimuth at a spacing of 280 m. The shot lines were nominally perpendicular to the receivers, running north-south (see Figure 3a).

Previous to this acquisition, two experimental lines were acquired and these are shown in Figure 3b. In this work, we use well-log and seismic data from the Tenerife field; in particular, we focus our attention in the line EXP-TENE-2D-3C P-wave reflected stacked seismic section (see Figure 4) with an extension of 9 km with 300 sources and 903 receivers. This requires seismic processing that includes time poststacked migration favoring data control and a structural and stratigraphic interpretation. The line EXP-TENE-2D-3C crosses the borehole Tenerife-2, which has the following well-log information: spontaneous potential (SP), resistivity and induction (deep induction log [ILD]), and sonic log of the P-wave. Some of these logs are more recent, such as the gamma ray (GR), dipolar sonic logs, check shots, and one offset vertical seismic profile. Well Tenerife-2 has a depth of 7513 ft, and the information available from the core analysis shows that Tenerife-2 crosses some formation units such as the Colorado (depth 2462.32 ft), the top of the Mugrosa zone (depth 5045.83 ft), the unit Mugrosa B (depth 5004.17 ft), and, finally, the Esmeraldas formation zone D (depth approximately 7500 ft). The formation unit Mugrosa C is divided into Mugrosa C-Sands and Mugrosa C-Shales. The first one has more oil sands; this fact defines Mugrosa C-Sands as the main reservoir of the basin (see Figure 5). It is a sandstone filled with oil; it has a porosity between 20% and 25%, a water saturation  $S_w$  between 40% and 50%, an oil saturation

tion between 30% and 40%, gas saturation between 10% and 30%; and the approximate permeability is 40.5 cp. With respect to well Tenerife-2, the zone of the Mugrosa C-Sands is located between depths 6857 and 7255 ft at the top of this formation and there are thin layers with depths no greater than 10 ft with a high clay content greater than 80%, 10% of sandy clay, pyrite, and other minerals. From this top to the bottom, we have large sand bodies greater than 100 ft depth with high porosities greater than 15% and small intercalations of clay with small depth widths (approximately 5 ft). On the other hand, the Mugrosa C-Shales define the bottom of the Mugrosa C-Sands; it has a depth of 244 ft, it is made of an 80% clay-type shale, with a porosity less than 15%; and it has low or null hydrocarbon saturation and very low permeability. The time window of the seismic section 574–1438 ms considers the formation units from the top of Colorado to the base of the MugrosaC-Shale. These units of continental origin are conformed by intercalations of sands and shales (see Figure 2). Structural (faults and cracks) and stratigraphic interpretation has been carried out in the seismic section (see Figure 4) based on the well-log information, cores, and formation markers, by picking the amplitudes along the offset. The core information shows an inverse fault that crosses well Tenerife-2 at depth of approximately 4710 ft on the top of the Mugrosa Formation. The petrophysical model is build based on this information. We compute pseudologs for the volume of clay  $V_{\text{clay}}$  using SP and the GR well logs. This methodology is based on [Mavko et al. \(2009\)](#) and [Tiab and Donaldson \(2004\)](#).

### Flowchart

To build a  $V_P$  and a  $\rho$  sections using ANN and validate the results (the estimated sections are verified with the forward-modeling software), we implement a full procedure that is shown in Figure 6 and it is described in the following steps.

### Seismic survey and well-log data

We use the seismic line EXP-TENE-2D-3C and the well-log data that correspond to the well Tenerife-2, which is the closest to this line. The reservoir characterization has been defined based on the seismic and geologic interpretations, petrophysical, and reservoir engineering data. These data were used to define the geologic model of the area for predicting sandy units of the reservoir. We upscale the well-log information using a low-pass filter and subsampling the results to match the seismic sampling. To estimate the time-depth (T-Z) curve, we have used available VSP information.

### Computing seismic attributes

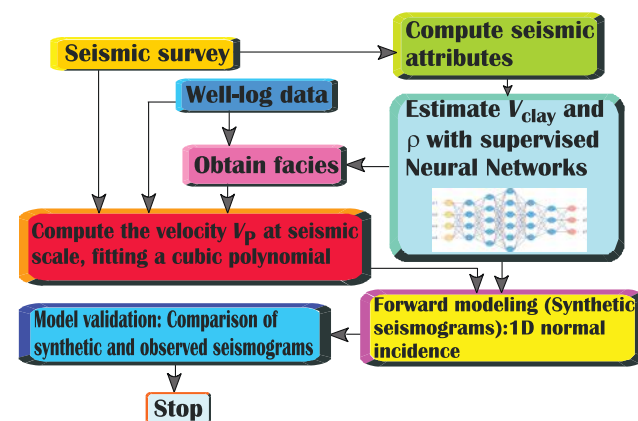
From the seismic line EXP-TENE-2D-3C, we take the seismic trace closer to the well Tenerife-2 and we compute a set of seismic attributes. From this set of seismic attributes, we will choose a subset to train ANNs to estimate each one of the desired petrophysical parameters.

### Analyze the data by means of the gamma test and train ANNs

We upscaled the well logs. Then, we select the combination of seismic attributes that produced the smallest gamma statistic  $|\Gamma|$  (when  $|\Gamma|$  is small, then there is a strong predictive relationship between the input variables and the output. On the contrary, if  $|\Gamma|$  is large, then there is no smooth predictive relationship between inputs and output) and trained an ANNs to learn how to estimate  $V_{\text{clay}}$  for the selected attributes. The same procedure is accomplished to estimate  $\rho$ .

### Testing the trained neural network and applied them to obtain predictions

We test the ANN via cross validation. This is a systematic removal of some data and re-estimating their values based on the model selected. In our case, the data set was divided into two halves: one for training and the other for testing. The areal distribution of the variable of interest is mapped by integrating the well data and seismic attributes. We use two trained ANNs to finally estimate the  $V_{\text{clay}}$  and  $\rho$  at seismic scale using the inputs drawn by the gamma test, given in Table 1. From the continuous volume of clay, we obtain two facies: sands and shales for the entire 2D seismic section.



**Figure 6.** Procedure to validate petrophysical properties estimated using ANNs generating synthetic seismograms and compare them with the real acquired seismic traces.

**Table 1.** Best combination of input attributes for  $\rho$  and  $V_{\text{clay}}$  drawn by the gamma test.

Property to predict	Best input combination of seismic attributes	$ \Gamma $
$\rho$	Attenuation, structural smoothing, $\omega_B$ , isofrequency, dip deviation, amplitude, edge evidence, envelope	$3.3151 \times 10^{-7}$
$V_{\text{clay}}$	$s(t)$ , attenuation, sweetness, $\omega_B$ , structural azimuth dip, iso-frequency, dip illumination, amplitude, $ds(t)/dt$	$3.699 \times 10^{-6}$

## Computing $V_P$ at seismic scale

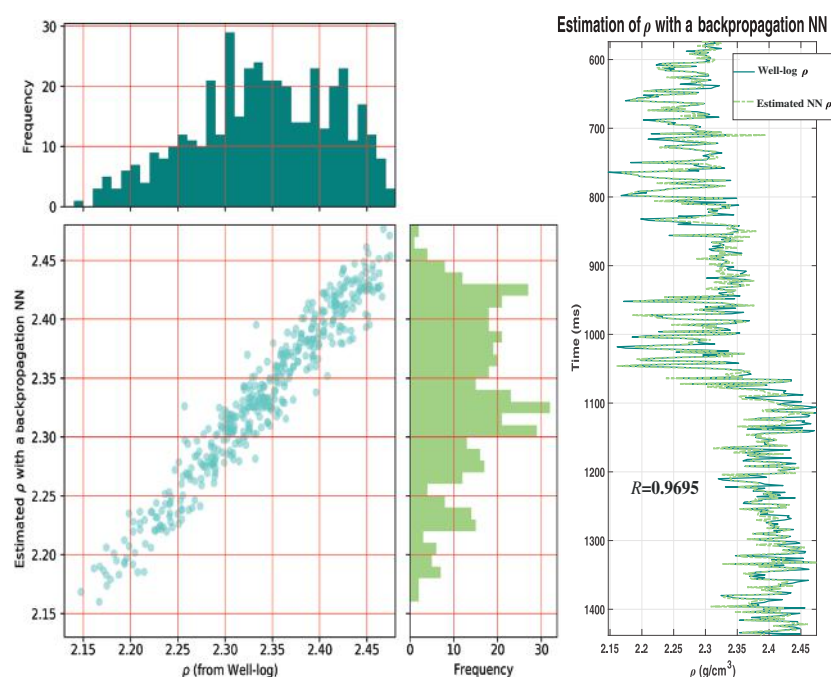
From the well-log information of  $V_P$  and the facies discrimination at well-log scale, we build (in the least-squares sense) two cubic polynomials: one to fit the  $V_P$  of sands and one for shales, given the knowledge of their respective depths. For any given point in the seismic section, we have its classification: Either is a sand or a shale. Using this information, we apply the corresponding cubic polynomial to all points in the seismic line EXP-TENE-2D-3C to obtain an estimation of  $V_P$  at seismic scale.

## Model validation

Each seismic trace within the 2D seismic section has properties only varying with the vertical axis

**Table 2. Architecture and training paradigms for the best ANNs to estimate  $\rho$  and  $V_{\text{clay}}$ .** The architecture description is given as the number of inputs imposed by the gamma test (the seismic attributes in Table 1), the number of neurons in each hidden layer (in both cases, the ANNs have two hidden layers), and the last number is the number of outputs.

Property to predict	Architecture of the best ANN	Training algorithm
$\rho$	8-10-10-1	Backpropagation
$V_{\text{clay}}$	9-10-10-1	CG



**Figure 7.** Estimation of density  $\rho$  (at well-log scale) using a backpropagation ANN. The scatterplot shows the fit between the ANN estimated  $\rho$  and the measured  $\rho$ . The histograms show the distribution of each of these two values for  $\rho$ , and we have a correlation factor of  $R = 0.9695$  between these two parameters.

(i.e., a layered medium), and this 1D medium has a profile determined by  $\rho$ ,  $V_P$ , and the thickness of each layer. The thickness of the layer depends on or might vary according to the T-Z curve. Using the ANNs estimated  $\rho$  and  $V_P$  seismic sections, we input these parameters in a 1D analytic method by Kennett (1983), which allows us to build synthetic seismograms that result from our estimations of reservoir properties. This is a simple way to validate the ANN estimations of these reservoir parameters. In the following sections, the steps involved in this methodology are explained in detail.

## Neural networks: Design, training, and results

The strength of ANNs derives from their capability to infer complex, nonlinear, underlying relationships without any a priori knowledge of the model. Multilayer feed-forward networks are a class of universal approximators (see Cybenko, 1989; Hornik et al., 1989). Multilayer perceptron networks with backpropagation learning can solve a wide range of classification and estimation problems. When the performance of this network is unsatisfactory, either because of speed or accuracy, we have also considered to use conjugate gradient (CG) networks, which is a variation and improvement on two-layer vanilla backpropagation; this is generally more effective (faster convergence) but it requires more memory (see Charalambous, 1992; Towsey et al., 1995). Then, we describe the results obtained after the data analysis to choose the best input combination of seismic attributes to construct a nonlinear model, in this case ANNs, to estimate  $V_{\text{clay}}$  and  $\rho$  at seismic scale.

## Data analysis, input selection, and training

The enormous diversity of possible modeling techniques and the difficulty of assessing the quality of the data has made the construction of nonlinear models from sampled data a hard process. The gamma test is a nonlinear modeling and analysis tool that allows us to examine the nature of a hypothetical input/output relationship in a numerical data set. First reported in Stefánsson et al. (1997) with the conjecture that a very simple algorithm (the gamma test) could be used to estimate directly from a given input/output set of data the extent to which the data derived from an underlying smooth model, even though this model was unknown. The gamma test works on the supposition that if two points  $\mathbf{x}'$  and  $\mathbf{x}$  are close together in input space, then their corresponding outputs  $\mathbf{y}'$  and  $\mathbf{y}$  should be close in output space. If the outputs are not close together, then we consider that this



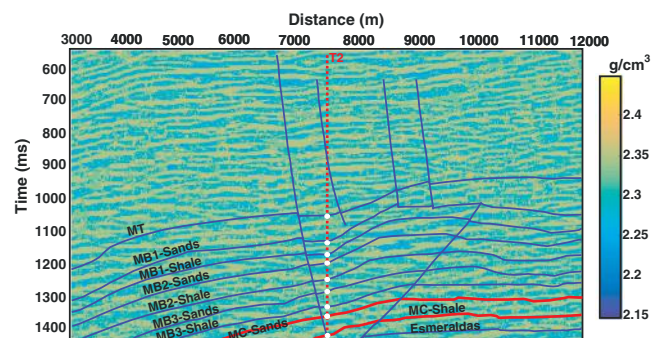
difference is because of noise. The input vectors are the seismic attributes, and the output is either  $V_{\text{clay}}$  or  $\rho$ . We assume that the input vectors contain factors that are useful for influencing the output. A second assumption is that the underlying relationship of the system under investigation is of the following form:

$$[V_{\text{clay}}, \rho] = f(\text{attributes}) + r, \quad (1)$$

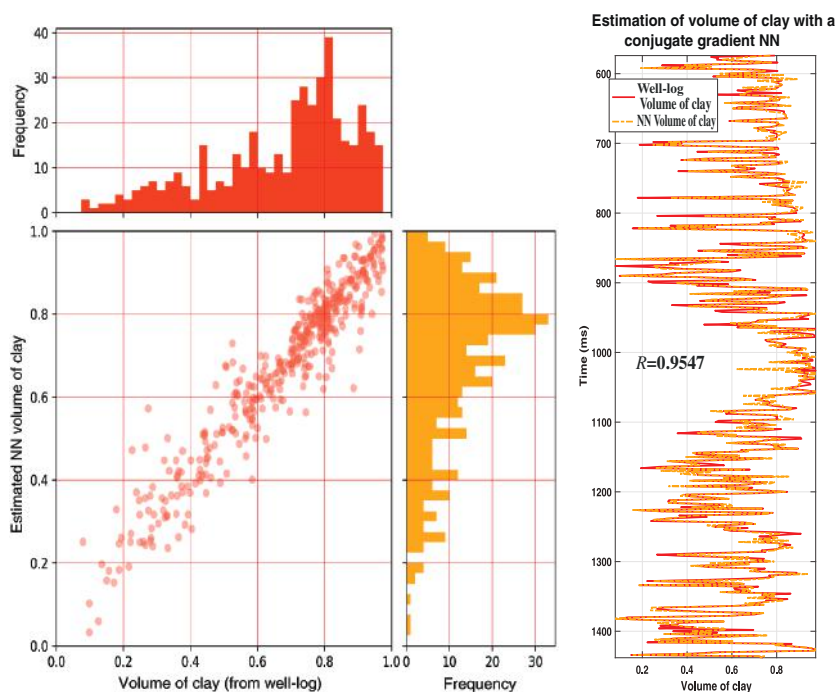
where  $f$  is a smooth function and  $r$  is a random variable that represents noise. The domain of possible models is restricted to the class of smooth functions that have bounded the first partial derivatives. This conjecture was proved for a wide class of situations (see Evans and Jones, 2002, 2008; Evans et al., 2002). The data analysis phase consists of taking all possible combinations of seismic attributes (we take the number of seismic attributes  $n \leq 20$  because the number of possible combinations is  $2^n - 1$  and it becomes computationally prohibited as  $n$  increases) for each output and perform the gamma test that delivers (for each possible combination of attributes) a parameter called the gamma statistic ( $\Gamma$ ), which is the estimate of that part of the variance of the output that cannot be accounted for by a smooth data model. If  $|\Gamma|$  is small, then there is a strong predictive relationship between the input variables and the output. On the contrary, if  $|\Gamma|$  is large, then there is no smooth predictive relationship between inputs and output — the inputs are irrelevant to the output. From this analysis, we keep that combination of attributes that produces the smallest gamma statistic and we use it for training the ANNs.

The aim in this study is to establish and validate models to predict  $V_{\text{clay}}$  ( $V_P$  is derived indirectly by the  $V_{\text{clay}}$ ) and  $\rho$  logs at seismic scale combining the well-log information and a seismic section in the depth interval 2615–7414 ft that corresponds to the time interval 574–1438 ms. In this interval, we have well-log information from the bore-hole Tenerife-2 as the training data set. We follow the same approach given in Parra et al. (2015), Iturrarán-Viveros and Parra (2014), and Iturrarán-Viveros (2012) and perform the gamma test analysis; see Jones (2004) to select the best combination of inputs to train an ANN. This is a critical step to devise a systematic feature selection scheme that provides guidance on choosing the most representative features for estimation of petrophysical parameters. This also saves time when designing the ANN architecture and provides the least mean square error (MSE) that any smooth data model can achieve on the given data without over-training the ANN. The number of units required

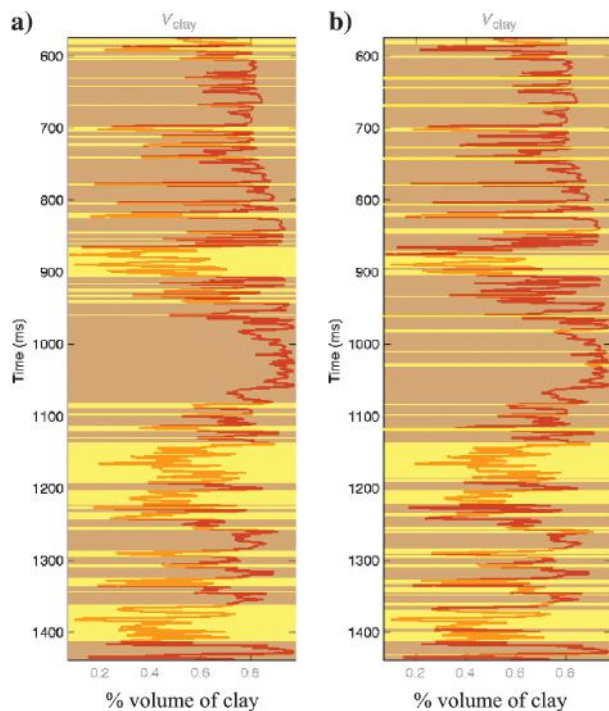
to achieve a good model will depend on the complexity of the unknown function that we are trying to approximate. Unfortunately, here, there are few rules to guide us; the best approach is to try to train using relatively few units (we have started with five units per layer) and if training fails to converge to the target MSE, progressively increase the number of hidden units in each layer. We have used two hidden layers, and these neighboring layers are fully connected; each neuron uses a sigmoid as the activation function. We have used the winGamma software; see Durrant (2001) for the gamma test analysis and also for training. Currently, there is powerful



**Figure 8.** Estimation of density  $\rho$  (at seismic scale) using a backpropagation ANN. We replicated the structural and stratigraphic interpretation showing the intercalation of sands and shales in the Mugrosa Formation as shown in Figure 4. The well Tenerife-2 (T2) is in the red.

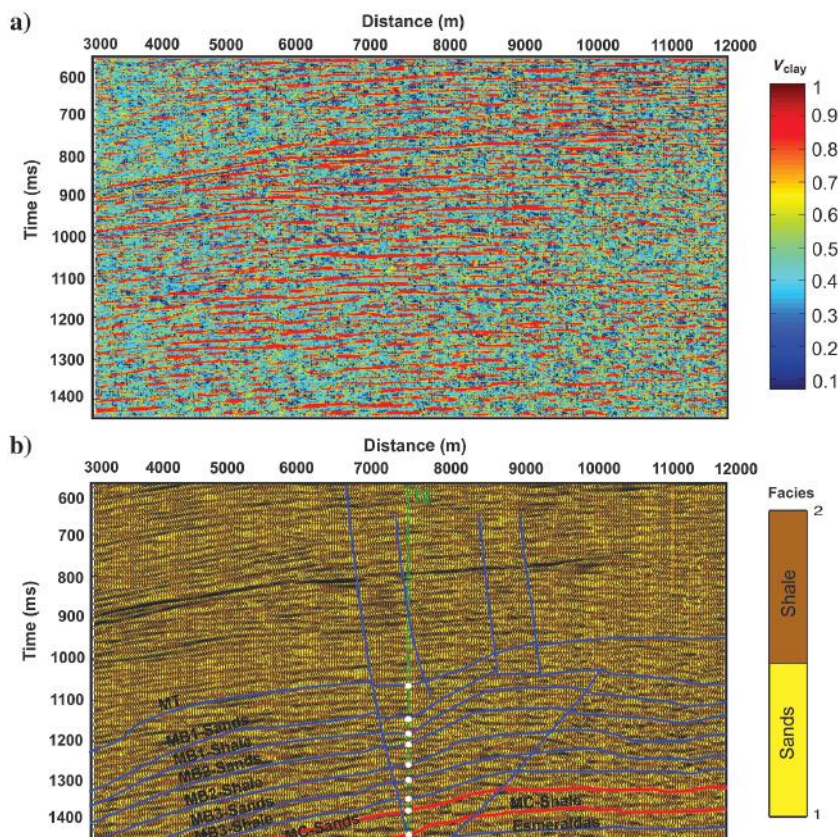


**Figure 9.** Training an ANN to estimate the volume of clay. The scatterplot shows the fit between the ANN estimated  $V_{\text{clay}}$  and the measured  $V_{\text{clay}}$ . The histograms show the distribution of each of these two values for  $V_{\text{clay}}$ , and we have a correlation factor of  $R = 0.9547$  between these two parameters.



**Figure 10.** The red curve in both figures represents the measured volume of clay in the well log. (a) Volume of clay from the well log and its associated facies. (b) Volume of clay from the neural network estimation and the two facies obtained. This shows that the facies estimated with the ANN method at seismic scale correlate with the petrophysical units identified in the well log.

**Figure 11.** (a) Continuous estimation of volume of clay using a CG ANN (b) Estimation of two facies: yellow sands and brown shales based on the continuous  $V_{\text{clay}}$  estimation. We overlaid the seismic traces and observe a very good correlation between high amplitudes and shale layers.



open software based on Julia and/or Python, such as TensorFlow or Scikit-learn, that could be used for training. The training process is expensive; in some cases we trained the ANNs for two or three days to let the MSE to reach a level close to the target MSE dictated by the gamma statistic. Nevertheless, the prediction is more efficient and accurate than using ANN trained faster with higher MSE.

Poststack attributes are derived from the stacked data and these constitute the candidates as input data to the ANN. The set of 19 seismic attributes considered in this work is given in Appendix A. For this data set, the best combination of attributes to estimate  $\rho$  and  $V_{\text{clay}}$  drawn by the gamma test are given in Table 1. Next, we describe the results obtained by these ANNs.

## Results

Once we have defined the best combination of inputs (seismic attributes) to estimate  $V_{\text{clay}}$  and  $\rho$ , we trained ANNs and test them via cross validation. (Cross validation is a standard machine-learning technique for tuning parameters. For  $k$ -folder cross validation, the training samples are randomly split into  $k$  groups with one reserved for validation and the rest  $k - 1$  for training. We randomly split the training set into two halves, a two-folder cross-validation, and the training shall be repeated twice. This allows one to monitor the ability of the trained network to generalize those inputs not seen during the training



phase). When testing the best (i.e., the one with the lowest MSE) ANN to estimate  $\rho$ , we obtain a correlation factor of  $R = 0.7183$  between the target curve and the results obtained by the ANN. When using the full set of points in the well log, the correlation factor is  $R = 0.9695$ . The training algorithms and the architectures of the best ANNs to estimate  $\rho$  and  $V_{\text{clay}}$  are given in Table 2. Figure 7 shows the best estimation of  $\rho$  (using all data points in the area of interest) and a scatterplot, in which a perfect fit between the well-log data and the produced ANN  $\rho$  is a straight line. We do not have a perfect straight line; however, we have a good fit, and the distributions (shown by the histograms) are also very similar. The  $\rho$  from the well log and the  $\rho$  estimated by the ANN is also depicted on the right side of Figure 7.

The test set used to evaluate the performance of the network trained to estimate  $V_{\text{clay}}$  gives a correlation factor of  $R = 0.7853$  between the target curve and the response obtained by the ANN. When using all the points in the well log, the correlation factor is  $R = 0.9547$ . This gives us confidence that the trained ANNs (to estimate  $\rho$  and  $V_{\text{clay}}$ ) have shown their capability to generalize. Figure 9 is an analog to Figure 7 but for the  $V_{\text{clay}}$ .

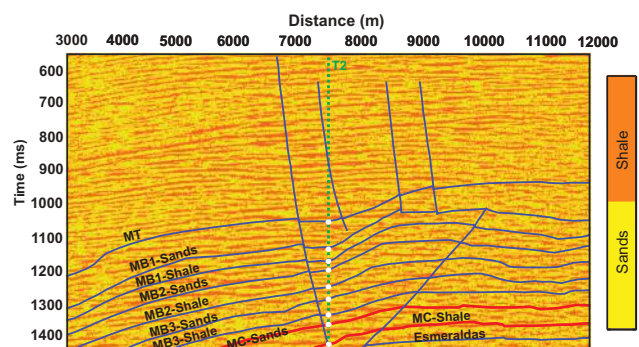
The estimations for  $V_{\text{clay}}$  and  $\rho$  at the seismic scale are obtained by applying the trained ANNs to the 903 seismic traces, and Figure 8 shows the estimations of  $\rho$ . In Figure 11a, we have the continuous estimation of the volume of clay  $V_{\text{clay}}$ ; from this estimation we assign two facies: sands to the values of  $V_{\text{clay}} < 0.5$  and shales to those values that satisfy  $V_{\text{clay}} \geq 0.5$  — this classification is shown in Figure 11b. The results far away from the borehole allow us to identify continuous sand bodies with shally intercalations, and in many cases, it is possible to distinguish between the limits of the sand zone. As shown in Figure 11b, some brown shales correlate with the high-amplitude attributes and the yellow sands correlate with the low-amplitude attributes. Figure 12 shows the two facies: sands and shales with the structural and stratigraphic interpretation including the zones described earlier and the faults in blue. There is a remarkable coincidence between the sands obtained with the ANN estimation and the reported sand areas.

### P-wave velocity model validation

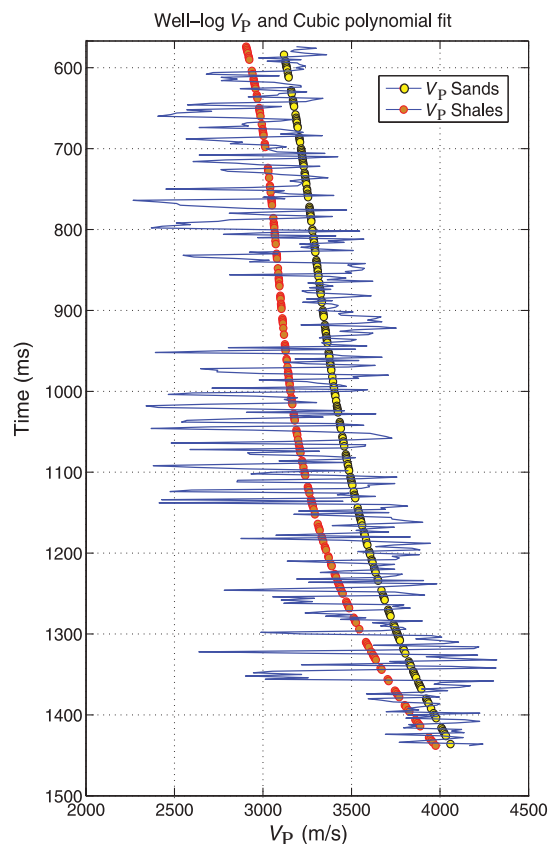
To test our ANN method to estimate the rock physical properties from the surface seismic, we compute synthetic seismograms and compare them with the observed seismic data. We need to estimate  $V_P$  at the seismic scale and then construct a layer model at each receiver location, with  $V_P$  and  $\rho$  predicted at that location. In this section, we discuss how to compute the  $V_P$  at seismic scale; we use it together with the estimation for  $\rho$  to compute synthetic seismograms in a layered medium applying a method of propagator matrices for the 1D forward simulation.

### Computing $V_P$

We have the well log for the velocity  $V_P$  in well Tene-rife-2. The reason for us not to use ANNs to estimate this parameter combined with the seismic attributes as we have done for the  $V_{\text{clay}}$  and  $\rho$  is that when analyzing the data using the gamma test taking  $V_P$  as the output, we have obtained a high (with respect to the other values of  $\Gamma$  for  $\rho$  and  $V_{\text{clay}}$ ) gamma statistic  $|\Gamma|$  indicating that we cannot achieve a good model for  $V_P$ .



**Figure 12.** Interpretation with the structure on top (faults, lines for the horizons are in blue, and the red curves delineate the main productive zone Mugrosa C-Sands) of the facies classification: yellow sands and brown shales.



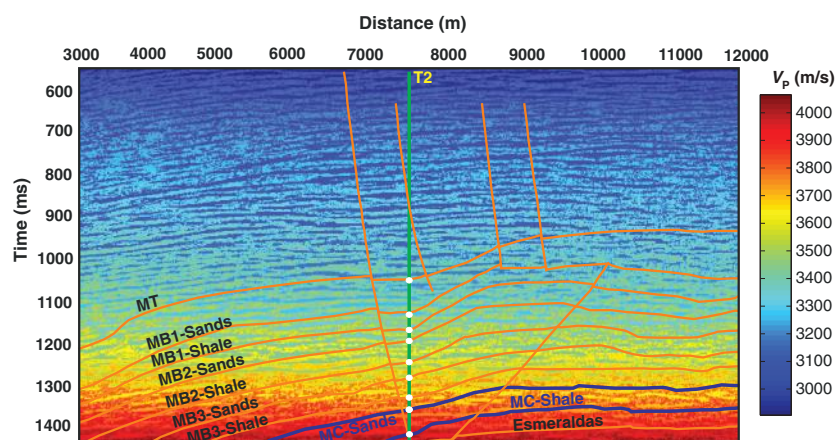
**Figure 13.** We fit the data that correspond to sands or shales with two different cubic polynomials. Both depend on time (or depth) and the  $V_P$  from the well log for sands or shales.

with the given data set — at least not as good as the models obtained for  $\rho$  and  $V_{\text{clay}}$ . Instead, we chose the method of least squares, fitting a cubic polynomial to the data. Using the facies classification (obtained from the  $V_{\text{clay}}$ ), we choose the times (or depths) along the borehole that correspond to sands, and using their values for  $V_P$ , we fit a cubic polynomial to this data set. We apply the same procedure to shales. Once we obtained the two cubic polynomials, we apply them to the entire section. The following polynomial is for sands:

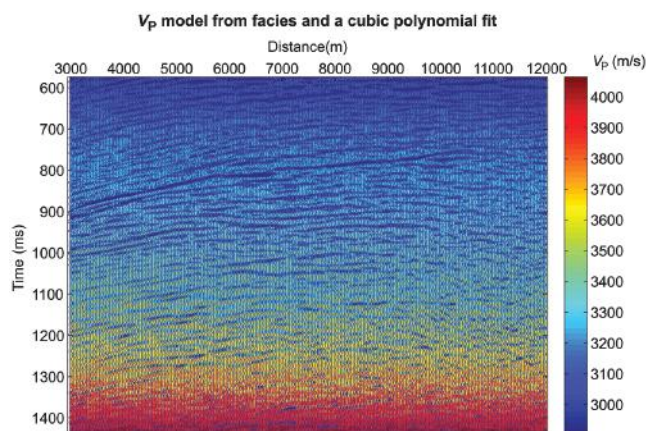
$$f_{\text{sand}}(t) = 0.000006389t^3 - 0.01615t^2 + 15.53t + 5398. \quad (2)$$

For shales, we have

$$f_{\text{shale}}(t) = 0.00001066t^3 - 0.02724t^2 + 24.51t + 2421 \quad (3)$$



**Figure 14.** With the facies and the well-log  $V_P$  velocity, we build the following velocity model as a supply for the 1D layered forward modeling. The structural and stratigraphic interpretation is included.



**Figure 15.** With the facies and the well-log  $V_P$  velocity, we build this velocity model at seismic scale. Note that the seismic on top of the velocity model helps to distinguish that the structure driven by the seismic data is preserved by the proposed velocity model.

Both polynomials depend on time  $t$ . We fit an average trend for each one of the facies in the well log; the result is shown in Figure 13. From equations 2 and 3, we note that the overall increase in seismic wave speed with respect to time (or depth) is mostly linear because the quadratic and cubic coefficients in these polynomials are small. These polynomials are applied to the entire section; as a result, we have  $V_P$  at seismic scale given in Figure 14. Using this approach, we obtain a velocity model that clearly distinguishes between sands or shales, allowing lateral variations imposed by each facies. The sand bodies with high porosities (i.e., the paying zones Mugrosa-C sands and Mugrosa-C shales) bend, and we note lateral and vertical changes in velocity in the 2D seismic section (Figure 14); these facts are in agreement with the structural and stratigraphic interpretation included in this figure. In the paying zone, we can distinguish velocity anomalies with ranges that finely adjust the oil sands identified by the well logs.

Figure 15 shows the same  $V_P$  but with the wiggle traces of the seismic section on top. The next section continues with the validation of this  $V_P$  model and the neural network estimation of  $\rho$  (at seismic scale). These two parameters are the ingredients for the 1D layered reflection method to generate synthetic seismograms and compare them with the acquired seismic traces. We have used this velocity model as a first guess of an inversion algorithm, obtaining satisfactory results (see Iturrarán-Viveros et al., 2018).

### Validation: 1D modeling

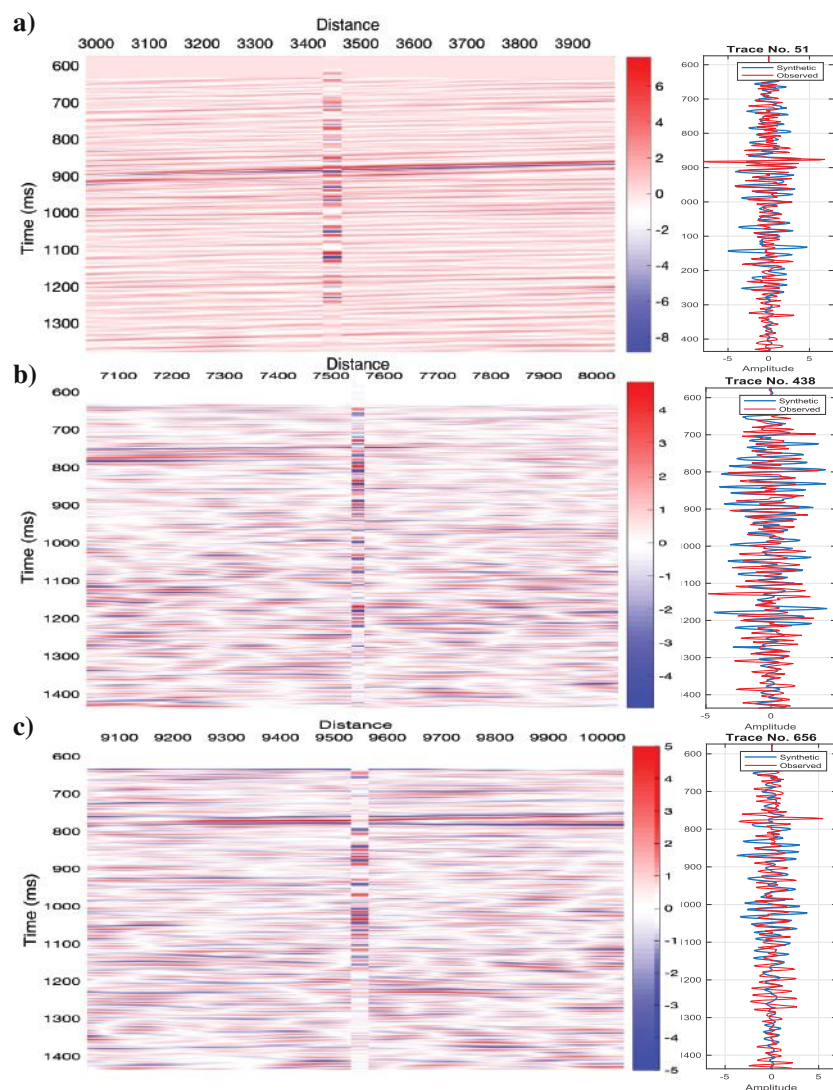
We used a multilayered earth model algorithm that assumes a normal incident plane wave as a source (see Mavko et al., 2009). The source wavelet extracted from the seismic data is approximated by the following expression (the integral of the Ricker wavelet  $r(t)$  is used to produce the synthetic full waveforms):

$$\begin{aligned} \int r(t)dt &= \int \left[ b^2 - \frac{1}{2} \right] \exp[-b^2] dt \\ &= -\left( \frac{t-t_0}{2} \right) \exp\left( -\left[ \frac{(t-t_0)f_c\pi}{2} \right]^2 \right), \end{aligned} \quad (4)$$

where  $b = [((t-t_0)f_c\pi)/2]$ , with  $f_c$  is the dominant frequency and  $t_0$  is the time delay. In this case, we use  $f_c = 25$  Hz; applying this method to each trace of the seismic section, we obtain synthetic seismograms that we compare with real acquire seismic traces. The comparison of some traces is shown in Figure 16; this selection of seismic traces is made up roughly to consider one representative trace in each part of the 2D seismic sec-



tion. Figure 16a corresponds to trace number 51. To better illustrate the agreement between the synthetic traces and the real seismic traces, a magnification of the complete seismic line, given in Figure 4, is taken in each plot. Figure 16b and 16c corresponds to trace numbers 438 and 656, respectively, on the top of 2D surface reflection seismograms. Although there are differences in amplitude, it is remarkable that the phases are in agreement. Even though we are modeling each trace as a 1D layered medium, the agreement is acceptable. Indeed, a perfect match is not possible due to the 3D effects. The 3D reflection seismic response is associated with the Fresnel zone, rather than with a single point used in the idealized 1D convolution model (see Gao, 2012; Gao and Di, 2015).



**Figure 16.** Comparison between real and synthetic traces. We have taken three traces to exemplify the results for different locations within the seismic section (on the left, one in the middle, and one on the right). To better illustrate the agreement between the synthetic traces and the real seismic traces, a magnification of the complete seismic line, given in Figure 4, is taken in each plot. (a) Corresponds to trace number 51. (b) Corresponds to trace number 438 on top of the 2D surface reflection seismograms. (c) Corresponds to trace number 656. There are some differences and some remarkable coincidences, especially in the phase agreement.

## Conclusion

We have shown the the facies estimated with the ANN method at seismic scale correlate with the petrophysical units identified in the well logs. The results of the comparison between the observed and calculated waveforms show a reasonable fit, validating the estimations of  $V_{\text{clay}}$  and  $\rho$  obtained from the ANN method. The model for  $V_p$  is indirectly obtained via the ANN estimation for  $V_{\text{clay}}$  with the additional cubic polynomial fitting procedure; this is a novel approach that allows us to have lateral variations governed by the facies classification. The  $V_p$  is a required parameter for the construction of synthetic seismograms, and because these attain a good match with real seismic traces, this implies that the model is good. We have used the velocity model that was produced in this study as a first guess of an inversion algorithm, obtaining satisfactory results. We expect that the inversion will increase the image resolution in a more realistic velocity distribution and the ANNs represent an attractive tool to generate initial velocity/density models for inversion. Future work will include the information from the two productive wells Tenerife-1, Tenerife-2, and the inlines and crosslines (that are close to these wells) from the 3D seismic acquisition.

## Acknowledgments

We thank Ecopetrol S.A. within the research program 0266-2013 Colciencias Ecopetrol, for providing the well and seismic data used in this research. This work was partially supported by CONACYT México under PROINNOVA projects 251910, 241763 and DGAPA UNAM project number IN100917.

## Data and materials availability

Data associated with this research are confidential and cannot be released.

## Appendix A

### Seismic attributes

Seismic attributes are the result of the properties derived from the complex seismic signal. We take the definitions of seismic attributes given in Taner et al. (1994), Taner (1997, 2001), Chopra and Marfurt (2007), Barnes (2015), and Azevedo-Guerra (2009), and we start with a collection of 19 seismic attributes (including time). These attributes are based on the definition of the complex trace given by

$$C(t) = s(t) + ih(t), \quad (\text{A-1})$$



where  $t$  is the time,  $C(t)$  is the complex trace,  $s(t)$  is the seismic trace, and  $h(t)$  is the Hilbert transform of the seismic trace  $s(t)$ . Rewriting equation A-1 in polar form, we obtain

$$C(t) = E(t) * e^{i\varphi(t)}, \quad (\text{A-2})$$

where

$$E(t) = \sqrt{s(t)^2 + h(t)^2} \quad (\text{A-3})$$

is the amplitude envelope. The instantaneous phase in equation A-2 is given by

$$\varphi(t) = \tan^{-1} \left( \frac{h(t)}{s(t)} \right). \quad (\text{A-4})$$

The variance attribute is given by

$$\sigma_t^2 = \frac{\sum_{j=t-L/2}^{j=t+L/2} w_{j-t} \sum_{i=1}^l (x_{ij} - \bar{x}_j)^2}{\sum_{j=t-L/2}^{j=t+L/2} w_{j-t} \sum_{i=1}^l x_{ij}^2}, \quad (\text{A-5})$$

where  $x_{ij}$  is the sample value at a horizontal position  $i$  and  $j$  is the vertical time sample. The function  $w_{ij}$  is a vertical smoothing factor over a window of length  $L$ . The root-mean-square (rms) amplitude is given by

$$A_{\text{rms}}(t) = \sqrt{\frac{1}{N} \sum_{k=-N/2}^{N/2} [s(t+k)]^2}, \quad (\text{A-6})$$

with  $N$  is the number of time samples in the trace. The instantaneous bandwidth is given by

$$\omega_B(t) = \left| \frac{\frac{dE(t)}{dt}}{E(t)} \right|. \quad (\text{A-7})$$

The sweetness is defined as follows:

$$\text{Sweetness}(t) = \frac{E(t)}{\sqrt{f_a(t)}}, \quad (\text{A-8})$$

where  $f_a(t)$  is the average frequency and the envelope  $E(t)$ . The other attributes considered in this work are attenuation, structural smoothing, structural azimuth dip (seismic attributes are normally extracted vertically along the seismic trace; this traditional approach can introduce artifacts, and this attribute is invariant to dip and azimuth; see Randen et al., 2000), local flatness, iso-frequency, dip illumination, dip deviation, chaos, amplitude, edge detection, edge evidence (from amplitude contrast), amplitude variation with offset, and first derivative =  $ds(t)/dt$ . These seismic attributes are computed from the seismic trace closest to the the borehole, Tenerife-2, using PETREL software.

## References

- Azevedo-Guerra, L., 2009, Seismic attributes in hydrocarbon reservoirs characterization: Master's thesis, Universidade de Aveiro, Department of Geosciences.
- Barnes, A. E., 2015, Handbook of post-stack seismic attributes: SEG.
- Charalambous, C., 1992, Conjugate gradient algorithm for efficient training of artificial neural networks: IEE Proceedings G — Circuits, Devices and Systems, **139**, 301–310, doi: [10.1049/ip-g-2.1992.0050](https://doi.org/10.1049/ip-g-2.1992.0050).
- Chevitarese, D., D. Szwarcman, R. M. D. Silva, and E. V. Brazil, 2018, Seismic facies segmentation using deep learning: AAPG Annual Convention and Exhibition.
- Chopra, S., and K. Marfurt, 2007, Seismic attributes for prospect identification and reservoir characterization: SEG Geophysical Development Series.
- Cooper, M., F. Addison, R. Alvarez, M. Coral, R. Graham, A. Hayward, S. Howe, J. Martinez, J. Naar, R. Peñas, A. Pulham, and A. Taborda, 1995, Basin development and tectonic history of the Llanos basin, Eastern cordillera, and Middle Magdalena Valley, Colombia: AAPG Bulletin, **79**, 1421–1443.
- Cybenko, G., 1989, Approximation by superpositions of a sigmoidal function: Mathematics of Control Signals and Systems, **2**, 303–314, doi: [10.1007/BF02551274](https://doi.org/10.1007/BF02551274).
- Di, H., Z. Wang, and G. Alregib, 2018, Deep convolutional neural networks for seismic salt-body delineation: AAPG Annual Convention and Exhibition.
- Durrant, P. J., 2001, winGamma™: A non-linear data analysis and modelling tool with applications to flood prediction: Ph.D. thesis, University of Cardiff.
- Ecopetrol, S., 1983, Evaluación de yacimientos del campo Tenerife: Technical report, Ecopetrol S.A.
- Evans, D., and A. J. Jones, 2002, A proof of the Gamma test: Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, **458**, 2759–2799.
- Evans, D., and A. J. Jones, 2008, Non-parametric estimation of residual moments and covariance: Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, **464**, 2831–2846, doi: [10.1098/rspa.2007.0195](https://doi.org/10.1098/rspa.2007.0195).
- Evans, D., A. J. Jones, and W. M. Schmidt, 2002, Asymptotic moments of near neighbour distance distributions: Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, **458**, 2839–2849, doi: [10.1098/rspa.2002.1011](https://doi.org/10.1098/rspa.2002.1011).
- Faraji, M. A., A. Kadkhodaie, R. Rezaee, and D. A. Wood, 2017, Integration of core data, well logs and seismic attributes for identification of the low reservoir quality units with unswept gas in the carbonate rocks of the world's largest gas field: Journal of Earth Science, **28**, 857–866, doi: [10.1007/s12583-017-0800-2](https://doi.org/10.1007/s12583-017-0800-2).
- Gao, D., 2012, Implications of Fresnel-zone texture for seismic amplitude interpretation: Geophysics, **77**, no. 4, O35–O44, doi: [10.1190/geo2012-0023.1](https://doi.org/10.1190/geo2012-0023.1).
- Gao, D., and H. Di, 2015, Texture inversion from seismic amplitude: Application to facies interpretation, offshore

- Angola, West Africa: AAPG Annual Convention and Exhibition.
- Gastaldi, C., J. Biguenet, and L. D. Pazzis, 1997, Reservoir characterization from seismic attributes: An example from the Peceiko field (Indonesia): *The Leading Edge*, **16**, 263–266, doi: [10.1190/1.1437614](https://doi.org/10.1190/1.1437614).
- Hampson, D. P., J. S. Schuelke, and J. A. Quierin, 2001, Use of multiattribute transforms to predict log properties from seismic data: *Geophysics*, **66**, 220–236, doi: [10.1190/1.1444899](https://doi.org/10.1190/1.1444899).
- Hornik, K., M. Stinchcombe, and H. White, 1989, Multilayer feedforward networks are universal approximators: *Neural Networks*, **2**, 359–366, doi: [10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- Iturrarán-Viveros, U., 2012, Smooth regression to estimate effective porosity using seismic attributes: *Journal of Applied Geophysics*, **76**, 1–12, doi: [10.1016/j.jappgeo.2011.10.012](https://doi.org/10.1016/j.jappgeo.2011.10.012).
- Iturrarán-Viveros, U., A. M. Muñoz-García, and J. O. Parra, 2018, Petrophysical seismic images obtained with Artificial Neural Networks as prior models for full-waveform inversion: A case study from Colombia: 88th Annual International Meeting, SEG, Expanded Abstracts, 2236–2240, doi: [10.1190/segam2018-2996124.1](https://doi.org/10.1190/segam2018-2996124.1).
- Iturrarán-Viveros, U., and J. O. Parra, 2014, Artificial neural networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data: *Journal of Applied Geophysics*, **107**, 45–54, doi: [10.1016/j.jappgeo.2014.05.010](https://doi.org/10.1016/j.jappgeo.2014.05.010).
- Jones, A. J., 2004, New tools in non-linear modelling and prediction: *Computational Management Science*, **1**, 109–149, doi: [10.1007/s10287-003-0006-1](https://doi.org/10.1007/s10287-003-0006-1).
- Kennett, B., 1983, *Seismic wave propagation in stratified media*: Cambridge University Press.
- Ma, Y. Z., E. Gomez, and B. Luneau, 2017, Integration of seismic and well-log data using statistical and neural network method: *The Leading Edge*, **36**, 64–69.
- Mavko, G., T. Mukerji, and J. Dvorkin, 2009, *The rock physics handbook*, 2nd ed.: Cambridge University Press.
- Parra, J. O., U. Iturrarán-Viveros, J. S. Parra, and P.-C. Xu, 2015, Attenuation and velocity estimation using rock physics and neural network methods for calibrating reflection seismograms: *Interpretation*, **3**, no. 1, SA121–SA133, doi: [10.1190/INT-2014-0175.1](https://doi.org/10.1190/INT-2014-0175.1).
- Priezzhev, I., and E. Stanislav, 2018, Application of machine learning algorithms using seismic data and well logs to predict reservoir properties: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201800920](https://doi.org/10.3997/2214-4609.201800920).
- Randen, T., E. Monsen, C. Signer, A. Abrahamsen, O. J. Hansen, T. Sætter, J. Schlaf, and L. Sonneland, 2000, Three-dimensional texture attributes for seismic data analysis: 70th Annual International Meeting, SEG, Expanded Abstracts, 668–671, doi: [10.1190/1.1816155](https://doi.org/10.1190/1.1816155).
- Restrepo, P., 1995, Late Precambrian to early Mesozoic tectonic evolution of the Colombian Andes, based on new geochronological, geochemical and isotopic data: Ph.D. thesis, University of Arizona.
- Roy, A., A. S. Romero-Pelez, T. J. Kwiatkowski, and K. J. Marfurt, 2014, Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico: *Interpretation*, **2**, no. 1, SA31–SA47, doi: [10.1190/INT-2013-0077.1](https://doi.org/10.1190/INT-2013-0077.1).
- Russell, B., D. Hampson, J. Schuelke, and J. Quirein, 1997, Multiattribute seismic analysis: *The Leading Edge*, **16**, 1439–1444, doi: [10.1190/1.1437486](https://doi.org/10.1190/1.1437486).
- Sandoval, J., 2010, *Modelo estructural y atributos sísmicos campo Tenerife*: Instituto Colombiano del Petróleo.
- Shafiq, M. A., M. Prabhushankar, Z. Long, H. Di, and G. Al-regib, 2018, Attention models based on sparse autoencoders for seismic interpretation: AAPG Annual Convention and Exhibition.
- Stefánsson, A., N. Končar, and A. J. Jones, 1997, A note on the Gamma test: *Neural Computing & Applications*, **5**, 131–133, doi: [10.1007/BF01413858](https://doi.org/10.1007/BF01413858).
- Taner, M. T., 1997, Seismic trace attributes and their projected use in prediction of rocks properties and seismic facies: Technical report, Rock Solid Images.
- Taner, M. T., 2001, Attributes revisited: Technical report, rock solid images: Technical report, CSEG Recorder.
- Taner, M. T., J. S. Schuelke, R. O'Doherty, and E. Baysal, 1994, Seismic attributes: Revisited: 64th Annual International Meeting, SEG, Expanded Abstracts, 1104–1106.
- Tiab, D., and E. C. Donaldson, 2004, *Petrophysics*, 2nd ed.: Elsevier.
- Towsey, M., D. Alpsan, and L. Sztriha, 1995, Training a neural network with conjugate gradient methods: IEEE International Conference on Neural Networks, 373–378.
- Veillard, A., O. Morère, M. Grout, and J. Gruffeille, 2018, Fast 3D seismic interpretation with unsupervised deep learning: Application to a potash network in the North Sea: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201800738](https://doi.org/10.3997/2214-4609.201800738).
- Velasquez-Espejo, A. J., 2012, 3-D multicomponent seismic characterization of a clastic reservoir in the Middle Magdalena Valley Basin, Colombia: Master's thesis, Colorado School of Mines.



**Ursula Iturrarán-Viveros** received a B.S. (1994) in mathematics from the Universidad Nacional Autónoma de México (UNAM) and a Ph.D. (1998) in computer science from the Imperial College of Science Technology and Medicine. From 1999 to 2008, she worked at the Instituto Mexicano del Petróleo. Since 2010, she has been working at the Facultad de Ciencias UNAM. Her main interests include numerical modeling of wave propagation and machine learning for geophysical applications.



Andrés M. Muñoz-García received a B.S. (2008) in physical engineering and a master's degree (2010) in engineering from the Universidad Nacional de Colombia. He is a Ph.D. candidate in the Department of Geosciences at the Universidad Nacional de Colombia. His main interests include petrophysics, rock physics, quantitative seismic analysis using machine learning, and electromagnetic techniques to study sedimentary deposits of mineralization.



**Jorge O. Parra** received an M.S. and Ph.D. in geophysics from the Colorado School of Mines, as well as a mining civil engineering degree from the University of Chile. He is a consultant at JPGeosciences and was a technical advisor at SwRI. He has contributed to advances in electromagnetics, resistivity, borehole geophysics, poroelastic-

ity, and anisotropy, and seismic modeling and interpretation. He has more than 100 publications and four patents. He is a member of SEG, SPWLA, and he is an editorial member of the *Journal of Applied Geophysics*. He jointly received Best Poster Presentation awards at the SEG 2001 and SPWLA 2002 annual meetings.



**Josué Tago** received a B.S. in chemical engineering from the Instituto Tecnológico de Celaya, an M.S. in computer science and industrial mathematics from the Centro de Investigación en Matemáticas, and a Ph.D. in earth sciences from UNAM. He is an associate professor at the Facultad de Ingeniería at UNAM. He is interested in mathematical and computational modeling of geophysical systems.