Journal of Applied Geophysics 107 (2014) 45–54

Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: www.elsevier.com/locate/jappgeo

Artificial Neural Networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data

Ursula Iturrarán-Viveros [a,*], Jorge O. Parra [b]

[a] Facultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Escolar S/N, Coyoacán C.P., 04510 México D.F., Mexico
[b] Southwest Research Institute, Division of Applied Physics, San Antonio Texas, USA

ARTICLE INFO

Article history:
Received 11 January 2014
Accepted 5 May 2014
Available online 20 May 2014

Keywords:
The Gamma test
Seismic attributes
Artificial Neural Networks
Permeability
Porosity
Attenuation

ABSTRACT

Permeability and porosity are two fundamental reservoir properties which relate to the amount of fluid contained in a reservoir and its ability to flow. The intrinsic attenuation is another important parameter since it is related to porosity, permeability, oil and gas saturation and these parameters significantly affect the seismic signature of a reservoir. We apply Artificial Neural Network (ANN) models to predict permeability (k) and porosity (ϕ) for a carbonate aquifer in southeastern Florida and to predict intrinsic attenuation ($1/Q$) for a sand–shale oil reservoir in northeast Texas. In this study, the *Gamma test* (a revolutionary estimator of the noise in a data set) has been used as a mathematically non-parametric nonlinear smooth modeling tool to choose the best input combination of seismic attributes to estimate k and ϕ, and the best combination of well-logs to estimate $1/Q$. This saves time during the construction and training of ANN models and also sets a lower bound for the mean squared error to prevent over-training. The Neural Network method successfully delineates a highly permeable zone that corresponds to a high water production in the aquifer. The Gamma test found nonlinear relations that were not visible to linear regression allowing us to generalize the ANN estimations of k, ϕ and $1/Q$ for their respective sets of patterns that were not used during the learning phase.

1. Introduction

Artificial Neural Networks are mathematical tools design to perform complex pattern recognition tasks and they have been employed to quantify patterns and estimate parameters in many geophysical applications, see for example Poulton (2001), Sandham and Leggett (2003). Reservoir characterization is a process for quantitatively describing various reservoir properties in spatial variability using all the available field data. These properties have a significant impact on petroleum field operations and reservoir management. Committee of Neural Networks has been successfully used to estimate porosity and permeability form well logs, see Bhatt and Helle (2002) and Helle et al. (2001). We choose Artificial Neural Network models (ANNs) to estimate reservoir properties for two applications with two different data sets. In the first application, we integrate multi-attributes from surface seismic data with well-log permeability (k) and porosity (ϕ) to produce permeability and porosity images for a carbonate aquifer in southeastern Florida. The second application only involves well-log data from one well in an oil reservoir in northeast Texas and a reference estimation of the intrinsic attenuation ($1/Q$) extracted from sonic full-waveforms used to train an ANN. Similar to any other statistical and mathematical model, ANN models have also some disadvantages. Having a large number of input variables is one of the most common problems for their development because they are not engineered to eliminate superfluous inputs. It is critical to devise a systematic feature selection scheme that provides guidance on choosing the most representative features for estimation of petrophysical parameters. This paper presents an input feature selection scheme based on the *Gamma test* (GT), that saves time during the training phase of the ANNs and it sets a lower bound for Mean Square Errors (MSEs) to avoid over-fitting. The Gamma test is a mathematically proven smooth non-linear modeling tool with a wide variety of applications that helps modelers to choose the best input combination before calibrating and testing models, therefore it reduces the input selection uncertainty. Given a data set, the GT will be able to tell us how well we can predict k, ϕ and $1/Q$ using any model that is likely to be operating. It can delineate the complex non-linear relationships if there are any. There are other techniques to reduce the number of input variables such as multi-linear regression and principal component analysis (PCA), see Malhi and Gao (2004). However when using PCA we could not estimate a lower bound for the MSE, nor estimate the number of training samples needed to achieve a reasonable output. Studies in other fields show interesting comparisons between plain ANN, the GT-ANN and PCA-ANN. Findings favor GT-NN and PCA-NN over plain ANN, see Gholami and Moradzadeh (2011). In a different approach Díaz-Viera et al. (2006) use copulas to model permeability values associated with the

* Corresponding author. Tel.: +52 55 56 22 54 11.
E-mail addresses: ursula@ciencias.unam.mx (U. Iturrarán-Viveros), jparra@swri.org (J.O. Parra).

http://dx.doi.org/10.1016/j.jappgeo.2014.05.010
0926-9851/

secondary porosity in carbonate formations. In Kisi and Shiri (2011), Shiri and Kisi (2012) and Shiri et al. (2013) genetic programing, adaptive neuro-fuzzy systems, support vector machines and Artificial Neural Network models are used to forecast ground-water depth.

Seismic data are routinely and effectively used to estimate the structure of reservoir bodies, but play no role in the essential task of estimating the spatial distribution of reservoir properties. On the other hand, well-log data gives precise information on reservoir properties at specific field locations with high vertical resolution. The integration of well-log and seismic data and their seismic attributes for reservoir characterization has been well documented in the literature, see for example Gastaldi et al. (1997), Russell et al. (1997) and Hampson et al. (2001). Once the Neural Networks are trained to a satisfactory level they may be used with a suitable set of input attributes corresponding to the 2-D seismic section to obtain inter-well estimations of k and ϕ. The second assignment is to estimate intrinsic attenuation using Artificial Neural Networks. There are various methods to obtain intrinsic attenuation, one of them is given in Parra et al. (2006b) but this algorithm requires a sonic full-waveform log. These full-waveform logs are not always acquired (that is the case of old drilled wells or also they might be financially unviable) which impedes to have an estimation of $1/Q$. We use an estimation of $1/Q$ derived from sonic full-waveform data in a well and train an ANN with half of the available data. The other half of the data set is used to test the ANN, obtaining good results since their fit with respect to the target curve $1/Q$ has good correlation. Hence the trained neural network was able to generalize and we have reliable estimates of $1/Q$. The ultimate goal would be to apply this trained ANN to other nearby wells where there are no sonic full-waveform logs to obtain trustworthy estimations of $1/Q$. This is beyond the scope of this paper but it is the motivation behind this second application of ANN. We believe that this is the first work that applies ANN to estimate intrinsic attenuation.

2. Study areas and data sets

P-wave surface seismic reflection data, acquired from a carbonate aquifer underlying Palm Beach County, from Hillsbore test Site, South Florida, U.S., are used to delineate flow units in a proposed aquifer storage and recovery (ASR) horizon, see Fig. 1. Within the proposed ASR horizon, located in the upper Floridan aquifer, low-permeability zones consist of sandstones and highly permeable zones are carbonates with interconnected vuggy porosity. The image in Fig. 2 provides the vuggy porosity structure at the core scale, it gives an idea as how the vuggy porosity is distributed in the carbonate aquifer, see Parra et al. (2003) and Parra et al. (2006a). A productive water zone of the aquifer outlined by the well-logs was selected for analysis and interpretation. A total of 1060 m (3475 ft) of reflection *P*-wave surface seismic was acquired. Reflections containing frequencies of 300 Hz were recovered. A number of continuous reflectors are observed through the section, especially between 0.1 and 0.5 s (61–430 m or 200–1740 ft). Events at and below 0.5 s (approximately 518 m or 1700 ft) show an anticlinal structure between stations 135 and 200. Below 0.7 s (about 823 m or 2700 ft) the reflections become discontinuous and are harder to follow. Well-logs and cores from a test well were used to create a lithological column between depths 229 m (750 ft) and 381 m (1250 ft). The permeability log is derived from the nuclear magnetic resonance (NMR). *P*-wave and *S*-wave velocity logs are obtained from full waveform sonic data. These logs are given in Fig. 3 together with the density, permeability and porosity. The well PBF10 in Fig. 1(a) that we are using to build the ANNs is located at trace number 44 of the seismic

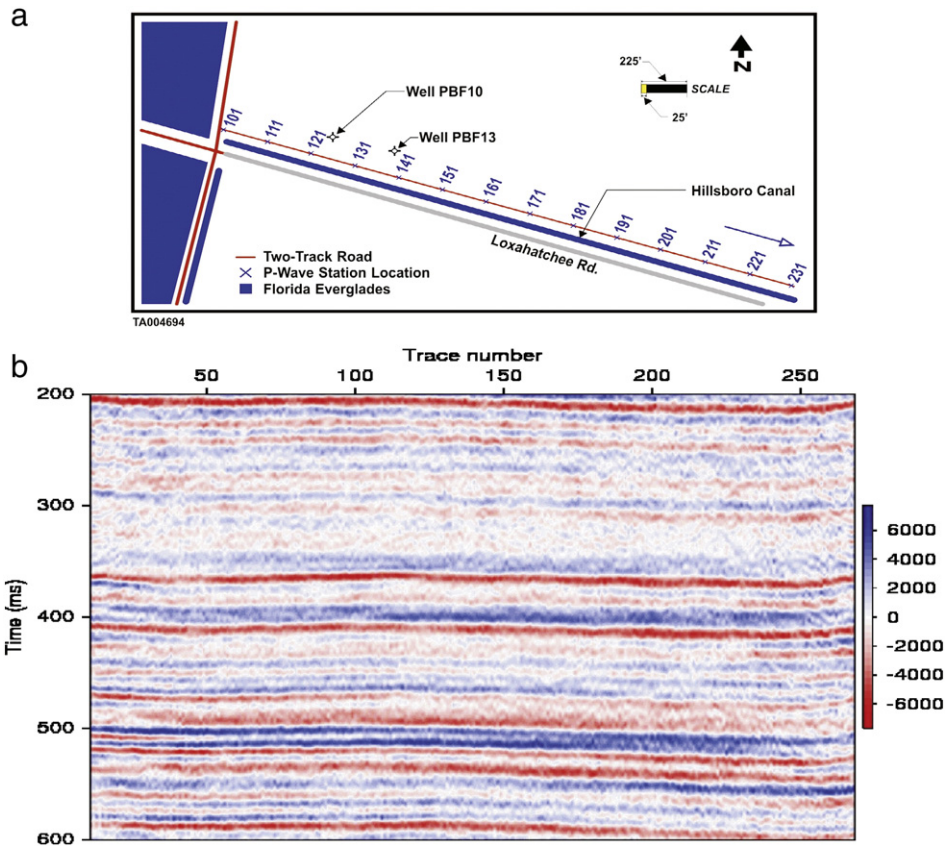

Fig. 1. (a) The location of seismic experiments at the western Hillsboro aquifer storage recovery (ASR) site near Boca Raton, Florida. One *P*-wave profile of 1060 m (3475 ft) was acquired 9.1 m (30 ft) from the line containing test wells PBF10 and PBF13. At the surface, the wellheads are approximately 100 m (330 ft) apart. The distance between stations is 7.6 m (25 ft). We only have access to the data for one well (PBF10). Taken from Parra et al. (2006a). (b) Observed seismic surface reflection illustrated at a stacked common depth point section.

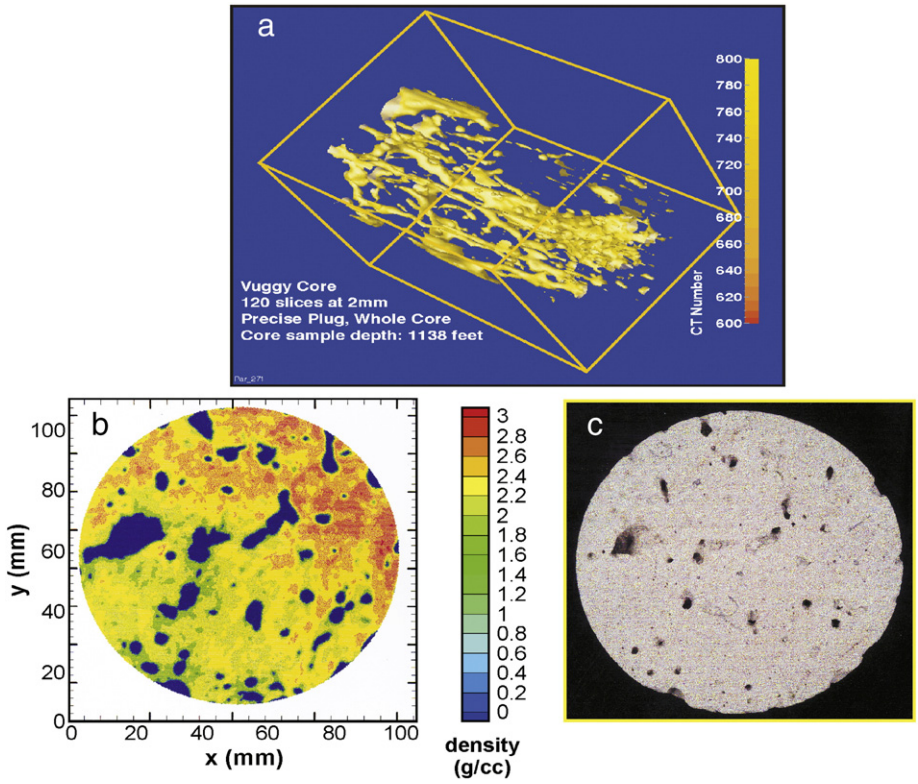

Fig. 2. (a) 3-D density image of a vuggy carbonate core from the Floridian aquifer in South Florida. The image shows connected, or touching vugs in the vertical direction of a full diameter core sample (in the direction of the borehole where the core was obtained at a depth of 1138 ft). (b) 2-D density image obtained from the core and usual to produce 3-D density image. (c) Photograph of the end of the core. Taken from Parra et al. (2006a).

section. We perform depth-to-time conversion of the well-log data and upscale with interpolation of the well-logs. Other methods such as Backus averaging could be use in the future, see Backus (1962) and Lindsay and Koughnet (2001). We sought a relationship between log-derived physical properties and seismic attributes and we follow the same approach as the one given in Iturrarán-Viveros, (2012). A seismic attribute is a derivative of a basic seismic measurement that may be extracted along a horizon or summed over a time window. Several factors go into seismic multi-attribute analyses, such as the weight assigned to each attribute (for traditional linear regression analysis) in the final equation and the type and number of attributes to use. Seismic attributes have been widely applied to the interpretation of geologic structure, stratigraphy and rock/pore fluid properties. We will describe the attributes we use in a later section.

The second application of Artificial Neural Networks is to estimate intrinsic attenuation ($1/Q$). Here we use only well-logs from one well and our target is an estimation of intrinsic attenuation derived from full-wave sonic data. The sonic log was acquired in an oil reservoir in the Waggoner ranch in northeast Texas. The site geology consists of a sand–shale sequence. Based on well-log data, the lithological units identified are sand (quartz), limestone (calcite), shale, and sandy shale (shale with a mixture of sand-to-silt-sized quartz or calcite grains).

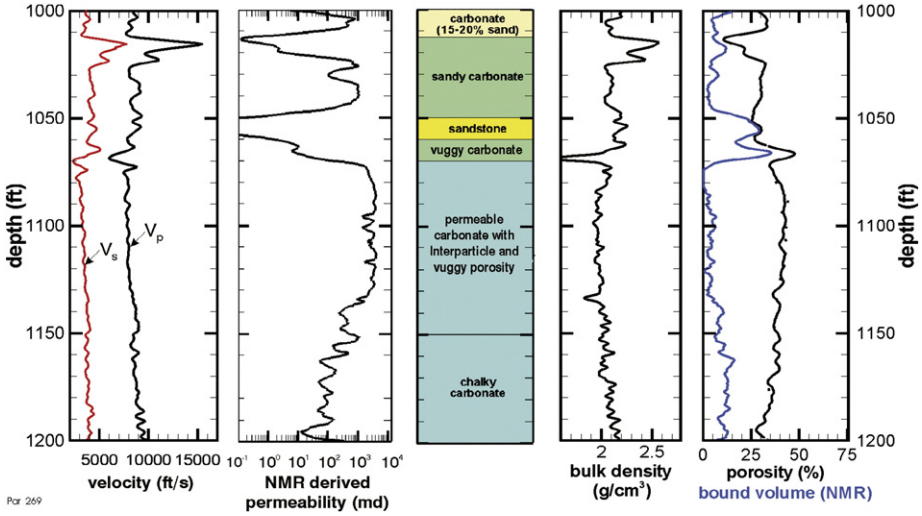

Fig. 3. Well logs from well PBF10, south Florida. Comparison of V_p, V_s, density, and gamma ray log with lithology, permeability, and porosity. Taken from Parra et al. (2006b).

Shale laminations are present in all sandstones, and the limestone is a mudstone with very little porosity. There are two sandstones with moderate permeability and 20% porosity, one at 2409–2419 ft (approximately 734–737 m) and one at 2489–2499 ft (about 758–761 m). Both have a lamination of 10–20% clasts. There is an oil producing zone at depth of 2409 ft (approximately 734 m) where there is an anomaly of interest which does not present the same type of attenuation signature as that near limestone/shale transition zones. Both selection of best input seismic attributes and best input well-logs to estimate images of k and ϕ and to estimate $1/Q$, respectively is done via the Gamma test. Because we are taking a data-driven approach, any relationships will be, by definition, data-dependent.

3. The Gamma test

The Gamma test (GT) is a non-linear modeling and analysis tool which 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 GT, that estimates the Γ statistic) 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 is an entirely non-parametric technique and its results apply *regardless of the particular methods subsequently used to build the model*. The Gamma test brings together two important ideas — it is a *mean squared error estimator* that applies to all *smooth non-linear models*. For non-linear modeling, there are numerous techniques for curve fitting, including Neural Networks, see Bishop (1995), local linear regression and support vector machines. Historically the construction of non-linear models from sampled data has been very much a subjective process. This in part stems from the enormous diversity of possible modeling techniques and the difficulty of assessing the quality of the data. There has been no way to estimate goodness of fit for non-linear models without knowing the model, now there is. If the Gamma statistic Γ is small there is a strong predictive relationship between the input variables and the output. If Γ is large there is no smooth predictive relationship between inputs and output — the inputs are irrelevant to the output. In this way the Γ is similar to the MSE around a regression line. However, linear regression measures goodness of fit to a line. Γ measures goodness of fit to any and all smooth curves.

Suppose we are given a set of input–output data

$$\{x_1(i), x_2(i), \ldots, x_m(i), y_i\} = \{(\mathbf{x}_i, y_i) | 1 \le i \le M\} \tag{1}$$

where m is the dimension of the input space and M is the number of data points. We think of the vector $\mathbf{x} = (x_1, x_2, \ldots x_m) \in \mathbb{R}^m$ as the *input*, confined to some closed bounded set $C \subset \mathbb{R}^n$, and (without loss of generality) the corresponding scalar $y \in \mathbb{R}$, as the *output*, and asked: to what extent the output is determined by the input? Suppose that the underlying relationship between input and output can be decomposed into a smooth component and a noisy component. We write this as:

$$y = f(\mathbf{x}) + r \tag{2}$$

where $\mathbf{x}$ is an input vector and f is a smooth function with bounded first and second derivatives, this is the only required assumption. In particular, we remark that the Gamma test is not directly applicable to problems involving categorical data. Variable r represents that part of the output that cannot be accounted for by any smooth data model. We can assume that the mean of r is equal to zero since any constant bias can be absorbed into the unknown function f. The Gamma test provides an estimate of the variance $\mathrm{Var}(r)$ of r in $O(Mln(M))$ time. Once we know what part of the data variance is determined, and what part is due to noise, fitting a curve to the data is considerably easier and more reliable. A formal proof and a detailed description of this algorithm can be found in Evans and Jones (2002).

4. Seismic attributes to estimate k and ϕ

A seismic attribute can be define as a transform, either linear or non-linear of the seismic trace. Seismic attributes are specific measurements of geometric, kinematic, dynamic or statistical features. The more seismic attributes available, there is more confusion selecting the appropriate ones. A common practice is identifying which seismic attributes to use, solely on the strength of their observed correlations with reservoir properties. As with any parameter, when the sample size is small, the uncertainty about the value of the true correlation can be quite large. In order to avoid these problems we use the GT to guide us on what is the best choice of inputs (seismic attributes) to construct a smooth model to estimate k and ϕ. We use instantaneous attributes derived from seismic data. These attributes conform a class called sample-based attributes. These correspond to transforms of the input trace in such a way as to produce another output trace with the same number of samples as the input. A comprehensive review on seismic attributes can be found at Chopra and Marfurt (2006) and for a classification see Cheng and Sidney (1997). We take definitions of attributes as given in Taner et al. (1979) and Taner et al. (1994). These attributes are based on the definition of the complex trace given by:

$$C(t) = s(t) + ih(t) \tag{3}$$

where t is the time, $C(t)$ is the complex trace, $s(t)$ is the seismic trace and $h(t)$ is the Hilbert transform trace. Rewriting Eq. (3) in polar form we obtain:

$$C(t) = E(t) * e^{i\varphi(t)} \tag{4}$$

where

$$E(t) = \sqrt{s(t)^2 + h(t)^2} \tag{5}$$

is the amplitude envelope (Env). It represents the total instantaneous energy and its magnitude is of the same order as the seismic trace. The envelope relates to acoustic impedance contrasts and the instantaneous phase is given by:

$$\varphi(t) = \tan^{-1}(h(t)/s(t)). \tag{6}$$

This is a good indicator of lateral continuity, it can be use to compute the phase velocity and we also include its first and second derivatives. Eq. (4) can also be written as:

$$C(t) = E(t) * (cos\varphi(t) + isin\varphi(t)). \tag{7}$$

Another instantaneous attribute can be computed differentiating with respect to time the instantaneous phase as follows:

$$\boldsymbol{\omega}(t) = \frac{d\varphi(t)}{dt} = \frac{s(t)h'(t) - s'(t)h(t)}{s^2(t) + h^2(t)}, \tag{8}$$

we call it the *instantaneous frequency* (IFR), it relates the wave propagation and depositional environments. It might be use as a fracture zone indicator and also to estimate seismic attenuation. The *weighted mean frequency* (WMF) is given by:

$$\overline{\boldsymbol{\omega}}(t) = \frac{\sum_T E(t) * \boldsymbol{\omega}(t)}{\sum_T E(t)} \tag{9}$$

Table 1
Two combinations of attributes to obtain best estimations of k and ϕ. We write a 0 if the corresponding input is not considered or 1 otherwise.

Mask	Time	(t)	Hilb	Env	Phase	IFR	WMF	TBI	IBW	Decay	QF	ACC	Semb	Γ
Mask1 (k_1)	1	0	1	1	0	0	0	0	1	0	0	1	0	6.346e−6
Mask2 (k_2)	1	1	1	0	0	1	0	0	1	0	0	0	0	8.691e−6
Mask1 (ϕ_1)	1	1	1	1	0	0	0	1	0	1	0	1	1	6.786e−6
Mask2 (ϕ_2)	1	0	1	1	1	0	1	0	1	1	1	1	1	1.207e−5

where T is the smoothing time window. This attribute indicates longer wavelength variations, provides more robust/smoothed estimate of instantaneous frequency and is less prone to noise. The *thin bed indicator* (TBI) is computed as the difference between the instantaneous and the time-averaged frequencies:

$$t_b(t) = \omega(t) - \overline{\omega}(t) \tag{10}$$

indicates thin beds, when laterally continuous and $\overline{\omega}(t)$ are given in Eq. (8). The *instantaneous acceleration* (ACC) is the time derivative of the instantaneous frequency:

$$a(t) = \frac{d\omega(t)}{dt}. \tag{11}$$

This attribute accentuates bedding differences. The *instantaneous band width* (IBW) is given by:

$$\sigma_f^2(t) = \frac{d\omega(t)}{dt}/2\pi E(t). \tag{12}$$

This equation measures the absolute value of the rate of change of envelope amplitude. The instantaneous bandwidth is one of the high-resolution character correlators and shows overall effects of seismic character changes. Barnes (1992) suggests that the *instantaneous quality factor* (QF) can be defined by the expression:

$$q(t) = -\pi\omega(t)/d(t) \tag{13}$$

where $d(t)$ is the *instantaneous decay rate*, which is defined as follows:

$$d(t) = \frac{E'(t)}{E(t)}. \tag{14}$$

The decay rate can take both positive and negative values. Hence the instantaneous quality factor is the ratio of instantaneous frequency to twice the instantaneous bandwidth. The semblance is a measure of coherent power existing between a number of traces versus the total power of all traces, as given by:

$$Semb(t) = \frac{\sum_{\tau=-N/2}^{N/2}\left[\sum_{m=1}^{M} f_m(t+\tau)\right]^2 - \sum_{\tau=-N/2}^{N/2}\sum_{m=1}^{M} f_m^2(t+\tau)}{\sum_{\tau=-N/2}^{N/2}\sum_{m=1}^{M} f_m^2(t+\tau)} \tag{15}$$

where $f_m(t)$ is the mth trace of the gather and N is the length of computational window. We consider a trace and its two next neighbor traces. i.e. we use $M = 3$.

5. Selection of a suitable set of inputs to train Artificial Neural Networks

To perform a data analysis to estimate k and ϕ (at the seismic scale) we consider the attributes described in Section 4. We exploit the advantages of the Gamma test to select the best combination of attributes to train a Neural Network to predict the target well-log estimations of k and ϕ given by the available well PBF10. This reduces training time for Neural Networks and can substantially improve the resulting model. The number of possible combinations of seismic attributes to be use as inputs is $2^{13}-1$. We define a *Mask* as a string of 0s and 1s with length the total number of inputs that one wants to examine via the GT. For the case of seismic attributes we have 13 different attributes so the Masks are of length 13. For the case of the well-logs to estimate $1/Q$ we have 9 different well-logs so the Masks in this case are of length 9. Each binary digit represents one particular attribute or well-log. A "1" in a Mask indicates to include that specific attribute/well-log for the combination of inputs and "0" indicates not to include it. Therefore the mask: 1111111111111 means that we include all the 13 possible attributes as inputs for our model to estimate either k or ϕ. The concept of a mask helps us to clearly indicate any particular combination of attributes/well-logs. We then consider those combinations of attributes or well-logs (*Masks*) that produce the lowest Γ statistic, see Jones (2004). For the given data in each one of the two problems (to estimate k, ϕ from seismic attributes or $1/Q$ from conventional well-logs) we perform a data analysis using the GT to obtain the best combinations of seismic attributes given in Table 1 to estimate k and ϕ and the best combination of well-logs to estimate the target log $1/Q$ given in Table 2.

We could have listed more possible combinations, but for simplicity we kept only the best two for k and ϕ.

For the case of estimating $1/Q$ we need to choose the best selection of inputs among the available well-logs, which are shown in Fig. 4. In Table 2 we have the four best combinations of well-logs to estimate $1/Q$ (the best is the one with the lowest Γ statistic).

Next, we proceed to analyze how many data we need to use for training and follow with the training of the Neural Networks.

5.1. The M-test

An M-test (see Končar (1997)) computes a sequence of Gamma statistics Γ_M for an increasing number of points M. The value Γ at which the sequence stabilizes serves as the estimate for $\mathrm{Var}(r)$, and if M_0 is the number of points required for Γ_M to stabilize then we will need at least this number of data points to build a model that is capable of predicting the output y, in Eq. (2) within a mean squared error of Γ. Therefore, the M-test is very useful for sparse datasets modeled using Artificial Neural Networks (ANNs). Fortunately, a single Gamma test is normally a relatively fast procedure so that running an M-test with a suitably selected step size is not a prohibitively time intensive procedure. We take the mask

Table 2
Four best combinations of well-logs to obtain best estimations of $1/Q$. We write a 0 if the corresponding input is not considered or 1 otherwise.

Mask	Depth	Gamma ray	V_p	V_s	Density (ρ)	Porosity (ϕ)	Sand	Shale	Calcite	Γ
Mask1	1	1	1	0	1	1	0	0	0	4.694e−5
Mask2	0	1	0	1	0	1	1	1	0	0.000862
Mask3	0	1	1	1	1	1	0	1	0	0.001438
Mask4	0	1	1	1	1	1	1	0	1	0.001652

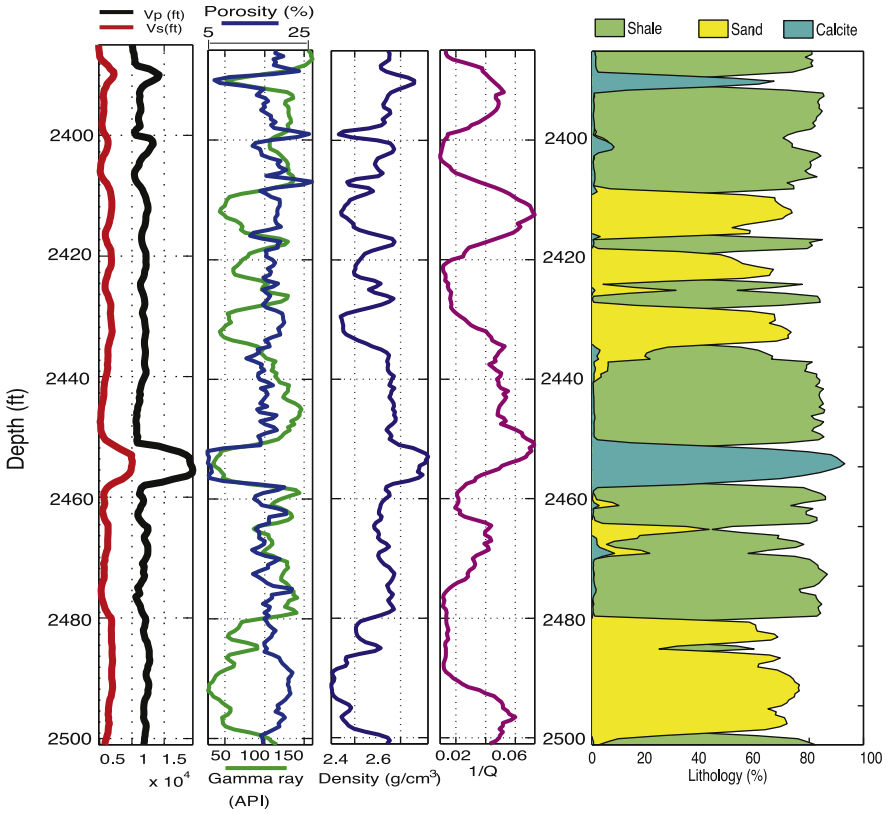

Fig. 4. Well-logs and lithology used to determine via the GT relevant inputs to estimate $1/Q$. The estimated $1/Q$ will be target curve to the Neural Network to train to learn this curve from the relevant well-logs chosen by the GT.

(given in Table 1) that produce the lowest *MSError* for porosity and run the M-test to obtain a curve that becomes stable (does not have big jumps) around 300 data points, see Fig. 5(a). The same experiment is done for permeability Fig. 5(b) and see that for anywhere around 350 data points we can obtain an adequate model. Since this test depends on the way one selects each data point (this is done randomly) then every realization will produce a different result. Hence we carry out each of these tests for 40 times and then take the average.

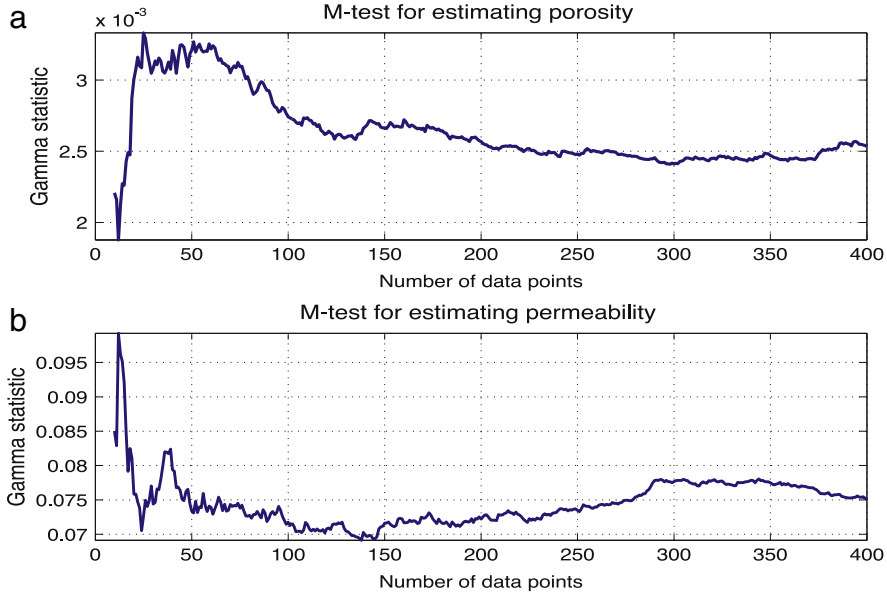

Fig. 5. The M-test is used to show how the *Gamma statistic* estimate varies as more data is used to compute it. Eventually if enough data is used the *Gamma statistic* should asymptote to the true noise variance on the output for which it has been computed. This test can also tell us how much data we likely need to obtain a model of a given quality, in the sense of predicting with a *MSError* around noise level. We take the mask that produce the lowest *MSError* for porosity (a) and run the M-test to obtain a curve that becomes stable (does not have big jumps) around 300 data points. The same experiment is done for permeability (b) and see that for anywhere around 350 data points we can obtain an adequate model. Since this test depends on the way one selects each data point (this is done randomly) then every realization will produce a different result. Hence we carry out each of these tests for 40 times and then take the average.

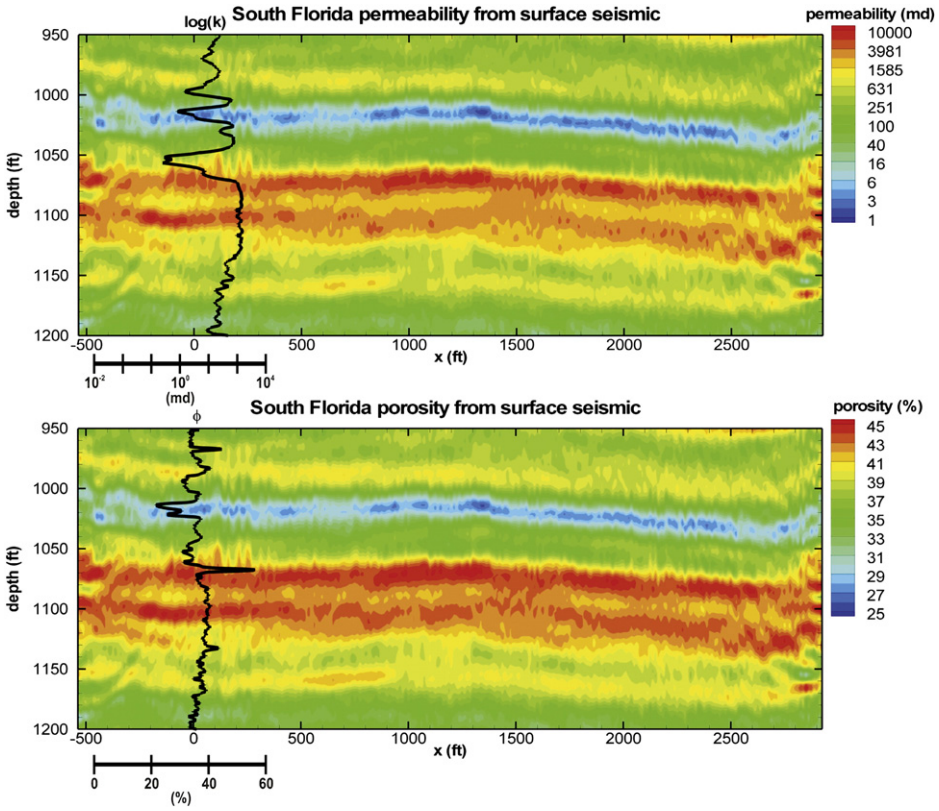

Fig. 6. Permeability (top) and porosity (bottom) estimated with linear regression. Taken from Parra et al. (2006a).

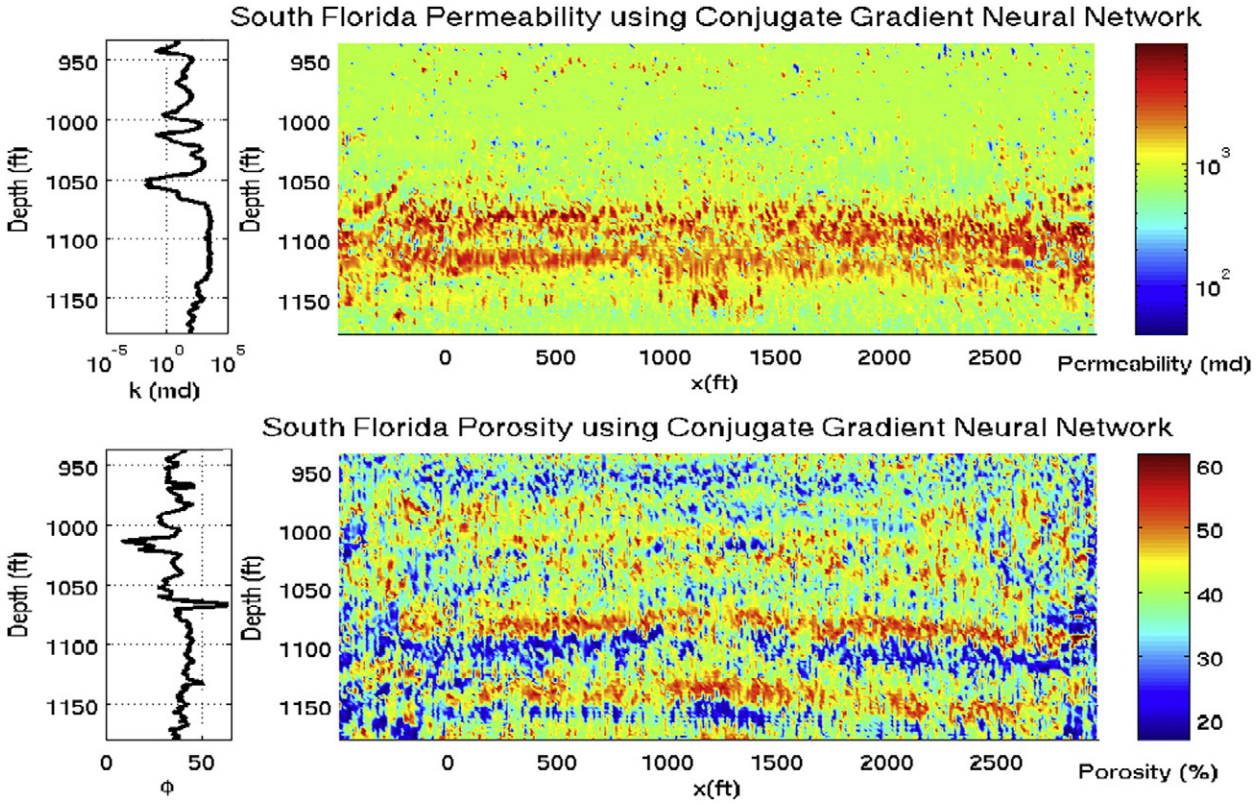

Fig. 7. Permeability (top) and porosity (bottom) estimated with Conjugate Gradient Neural Networks. Two Neural Networks were trained, one using mask k_1 and the other using mask ϕ_1 in Table 1, to estimate k and ϕ respectively.

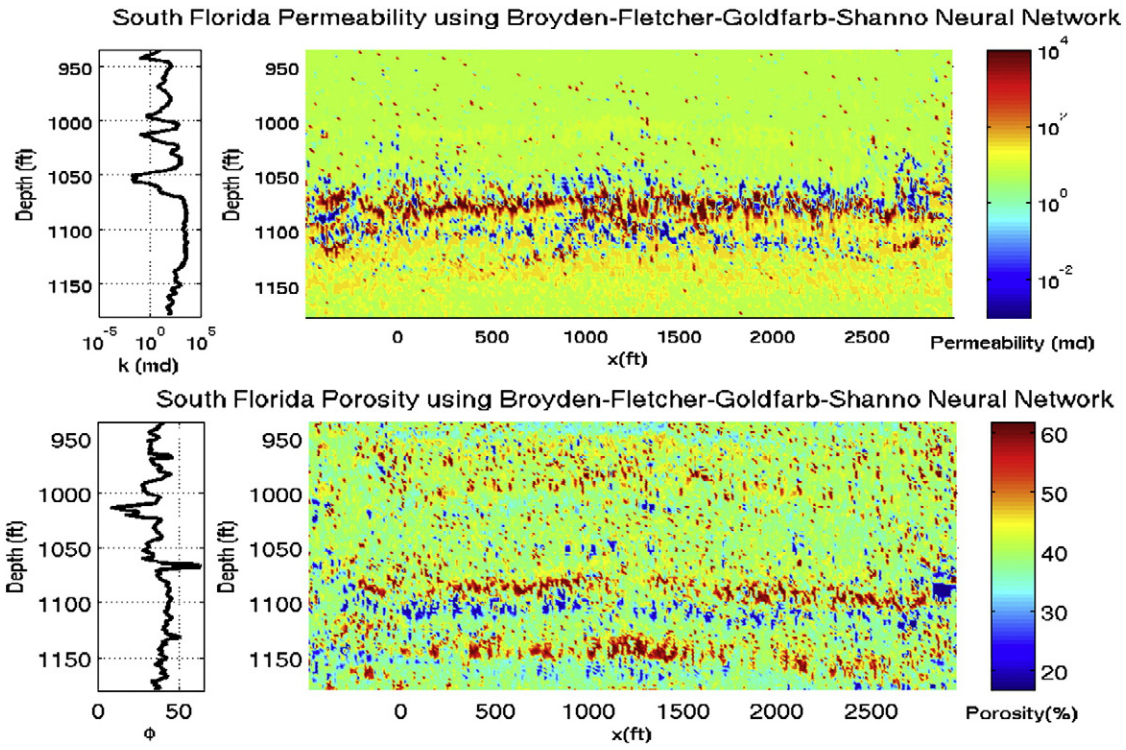

Fig. 8. Permeability (top) and porosity (bottom) estimated with Broyden–Fletcher–Goldfarb–Shanno Neural Networks. Two Neural Networks were trained, one using mask k_1 and the other using mask ϕ_1 in Table 1, to estimate k and ϕ respectively.

6. Training Artificial Neural Networks

The strength of Artificial Neural Networks (ANNs) derives from their capability to infer complex, non-linear, underlying relationships without any a priori knowledge of the model. However, in reality, the fuller the priori knowledge the better the end result is likely to be. We know that feed-forward networks with few as one hidden layer can act as universal approximators for continuous functions over a compact set, see Cybenko (1989) and Hornik et al. (1989). In practice it is usually most effective to use two hidden layers. Plainly, the precise architecture required will depend on the complexity of the surface to be modeled. Assuming there is adequate data, once one has a suitable architecture, training down to the Gamma statistic should be straightforward. If training fails to reach the Gamma statistic, or takes excessive time, this is invariably because the network architecture is unsuited to the function being modeled and the number of units in the hidden layers should be increased accordingly. Multi-layer 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 two options: one is to use Conjugate Gradient networks (CGs) or to use a quasi-Newton approach where one of the most used approximations to the inverse Hessian is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) equation, see Bishop (1995). Both CG and BFGS offer a faster convergence than traditional learning algorithms such as Backpropagation, and they make use of standard numerical optimization techniques. In CG algorithms for fast convergence, a search is performed along conjugate directions. All CG algorithms start out by searching in the steepest descent direction on the first iteration. The BFGS learning algorithm is an algorithm of the type Levenberg–Marquardt designed to approach second-order training speed without having to compute the Hessian matrix. One simple measure if the performance of a model on unseen data for which the measured outputs are known is the mean-squared error over the test data. If $\{y_i : i \in U\}$ is a previous unseen set of measured values of an output and $\{\hat{y}_i : i \in U\}$ is a set of predictions for y_i then the mean-squared error *MSError* of the predictions is given by:

$$MSError = \frac{1}{|U|}\sum_{i \in U}\left(\hat{y}_i - y_i\right)^2. \tag{16}$$

When constructing a non-parametric non-linear model, the natural tendency is to try to minimize the mean-squared error of the model against the training data. However, if there is significant intrinsic noise in the data then training down to *MSError* closes to zero, regardless the noise level will be counterproductive. This is referred to as *over-fitting* the model. The Gamma test provides an excellent guide to prevent form over-training (over-fitting), since the Γ statistic represents the *MSError* where we need to stop training.

7. Numerical results

From the seismic section in Fig. 1(b) we consider only a time window that contains 401 time samples, which corresponds to the ASR horizon (from 200 to 400 ms) and 278 traces. After running the Gamma test and obtained the best combination of attributes (given in Table 1) to train ANNs to estimate k and ϕ on the well. To estimate k and ϕ we use Neural Networks trained either using CG or BFGS learning algorithms. We have a four-layered networks with a variable number of inputs (corresponding to Mask1 named k_1 and ϕ_1), 10 units in the two middle layers and one output, either k or ϕ. To determine the number of layers and neurons in hidden layers in ANNs is a really hard problem. The more internal structure a network has, the better that network will be at representing complex solutions. On the other hand, too much internal structure is slower, may cause training to diverge, or lead to overfitting. There are multiple empirical rules to apply, see Elisseeff and Paugam-Moisy (1997), Lawrence et al. (1996). In our both applications, the number of chosen hidden neurons represents a trade-off between the ability of the Neural Network to generalize and to keep the Mean Squared Error small (with respect to what the Gamma test

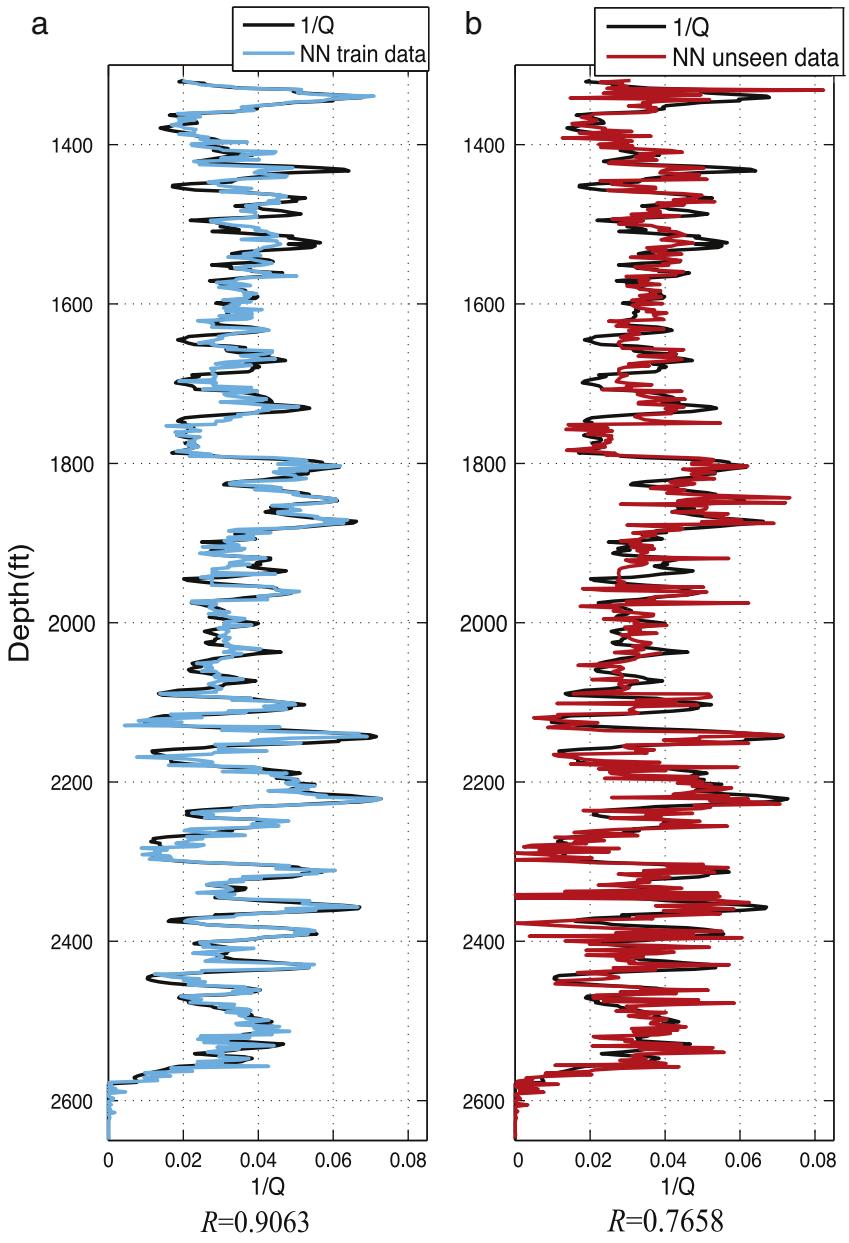

Fig. 9. We have divided the well-log data for the available well into two data sets. The training data set and the testing data set. From depth [1380–2650] ft we have taken every other point as the training set and leaving the rest as the testing data set. This strategy allows us to have a good sampling. We trained the Artificial Neural Network using the Backpropagation algorithm with a momentum $\boldsymbol{\alpha} = 0.15$ and a learning rate $\boldsymbol{\eta} = 0.3$. The Neural Network is a three-layer network with five inputs given in Table 2 for the Mask1 and two middle layers each one of them with eight neurons. The output is the target $1/Q$ given in Fig. 4. In (a) we have the test of the Neural Network on the data set used to train it. In (b) we have the blind test with the data unseen by the Neural Network during the training phase. The correlation between the target curve and the response obtained by the Neural Network is $R = 0.9063$ for the trained set and $R = 0.7658$ for the unseen data. This gives us confidence on the results generated by the Neural Network since the Neural Network shows its capability to generalize.

indicates). We show the porosity and permeability images obtained with these two learning paradigms (BFGS and CG). These images delineate the mayor permeability structure of the vuggy carbonate. However, the image in Fig. 8 provides information on the high permeability (interconnected vugs) and low permeability zone unconnected vugs, or permeability barriers within the carbonate rock. Therefore image depicted in Fig. 8 gives better practical results for identifying lateral heterogeneities in this geological unit. The permeability image in Fig. 7(top) captures the variability of the carbonate structure but it does not capture those zones of interconnected conduits and permeability barriers that are observed in the cores. The regression method resembles the well log data, but it does not give much information on the lateral variability, see Fig. 6. These regression images are taken from Parra et al., (2006a) and they are computed as: $k = 7 \times 10^{10} \exp(-1.1 \times 10^{-3} Z)$, where k is in millidarcies and Z is the impedance in $(g/cm^3)(feet/s)$, the goodness of fit is $R^2 = 0.51$. The porosity image in Fig. 6(bottom) is computed as follows: $\phi = 80 - 0.0024Z$, where ϕ is the percentage of porosity, Z is the impedance in $(g/cm^3)(feet/s)$ and the goodness of fit is $R^2 = 0.55$. On the other hand Neural Network models provide several processing schemes that may not agree with the geology. As a simple rule one can provide to the interpreter is to run various Neural Network models and select the images that resemble the most of the regression image. Thus the regression image will delineate the large structure by following the logs, and the selected Neural Network will not only capture small structures within the scale of the heterogeneity but will also provide information on large scale heterogeneities. The regression estimates might be used as a general trend to guide Neural Network estimations for porosity and permeability.

However, for the case of intrinsic attenuation, we were unable to obtain good estimates of $1/Q$ using linear regression.

With respect to the estimation of $1/Q$ we have used Backpropagation learning algorithm to train an Artificial Neural Network with a momentum $\alpha = 0.15$ and a learning rate $\eta = 0.3$. The Neural Network is a four-layer network with five inputs given in Table 2 for the Mask1 and two middle layers each one of them with eight neurons. This is the learning algorithm that produced the best results, reaching an MSE = 0.1876. The target output is the curve $1/Q$ given in Fig. 4. The training data set was divided into two halves taking with respect to depth every other data point for training and leaving the rest for the blind test. In Fig. 9(a) we have the test of the Neural Network on the data set used during the training phase. In Fig. 9(b) we have the blind test with the data unseen by the Neural Network during the training. The correlation between the target curve and the response obtained by the Neural Network is $R = 0.9063$ for the trained set and $R = 0.7658$ for the unseen data. This gives us confidence on the results generated by the Neural Network since the Neural Network shows its capability to generalize.

8. Conclusions

The main result drawn from these experiments is to use the Gamma test as a powerful tool for non-linear modeling and non-linear model building. The Artificial Neural Networks we obtained provide good and realistic estimates of k, ϕ and $1/Q$. The methodology shown here represents an attractive alternative to build non-linear models, prevents from over-fitting and assesses the quality of data before any attempt to build a model, given an idea about the quality of data and the quality of the resulting model. The resulting images for k and ϕ obtained with the ANNs show strong correlation between high permeability and the vuggy carbonate unit, which is formed by interconnected vuggy and matrix porosity (see core image of Fig. 2). Although we do not know how the conduits are distributed, with the seismic (impedance) we have been able to map the major structures and with Neural Networks we captured the heterogeneities of the conduits formed by the interconnected vuggy porosity net and structures that form barriers of low permeability (shown in blue). Estimations of $1/Q$ give a glimpse of the capability of ANN to generalize and we have successfully applied this trained ANN to estimate $1/Q$ for those nearby wells, in the same northeast Texas oil field, that do not have full waveform sonic logs. These estimations enabled us to use them as an additional formation evaluation tool to quantify oil, gas, and water saturations of sand in a reservoir, see Parra et al. (2014). We should keep in mind that the application of this technique is data dependent so the best combinations of attributes or well-logs obtained for these cases might not be the best choice for other cases. This technique could be applied to estimate other well-log properties such as saturation. If the innate errors in the input data exceed the models' capability, it would be very difficult for the model to perform, no matter how good the model itself is. For this purpose, the Gamma test employed in this research would have substantial potential to help modelers in solving the uncertainty issues in reservoir characterization and in general modeling processes.

9. Acknowledgments

We would like to thank the comments and suggestions from two anonymous reviewers. The South Florida Water Management District provided the well logs and cross-well reflection seismic data set. We thank Todd Thomas and the W.T. Waggoner Estate for providing the data. This work was partially supported by DGAPA-UNAM under project IN116114-2 and CONACYT México under program PROINNOVA project number 212923. UI-V thanks Pilar Ladrón de Guevara, for her crucial help to locate useful references.

References

Backus, G.E., 1962. Long-wave elastic anisotropy produced by horizontal layering. J. Geophys. Res. 67, 4427–4440.
Barnes, A.E., 1992. The calculation of instantaneous frequency and instantaneous bandwidth (short note). Geophysics 57, 1520–1524.
Bhatt, A., Helle, H.B., 2002. Committee neural networks for porosity and permeability prediction from well logs. Geophys. Prospect. 50, 645–660.
Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, U.S.A.
Cheng, Q., Sidney, S., 1997. Seismic attribute technology for reservoir forecasting and monitoring. Lead. Edge 16, 445–446.
Chopra, S., Marfurt, K., 2006. Seismic attributes — a historical perspective. Geophysics 70 (5), 3SO–28SO.
Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst. 2, 303–314.
Díaz-Viera, M., Anguiano-Rojas, P., Mousatov, A., Kazatchenko, E., Markov, M., 2006. Stochastic modeling of permeability in double porosity carbonates applying a Monte-Carlo simulation method with t-copulas. SPWLA 47th Annual Logging Symposium. Veracruz Mexico, pp. 2006–OO, June 4–7.
Elisseeff, A., Paugam-Moisy, H., 1997. Size of multilayer networks for exact learning: analytic approach. Advances in Neural Information Processing Systems 9. The MIT Press, Cambridge MA, pp. 162–168.
Evans, D., Jones, A.J., 2002. A proof of the Gamma test. Proc. Roy. Soc. London A 458, 2759–2799.
Gastaldi, C., Biguenet, J., Pazzis, L.D., 1997. Reservoir characterization from seismic attributes: an example from the Peciko field (Indonesia). Lead. Edge 16, 263–266.
Gholami, R., Moradzadeh, A., 2011. Support vector regression for prediction of gas reservoirs permeability. J. Min. Environ. 2 (1), 41–52.
Hampson, D.P., Schuelke, J.S., Quierin, J.A., 2001. Use of multiattribute transforms to predict log properties from seismic data. Geophysics 66, 220–236.
Helle, H.B., Bhatt, A., Ursin, B., 2001. Porosity and permeability prediction from wireline logs using artificial neural networks: a north sea case study. Geophys. Prospect. 49 (4), 431–444.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 356–366.
Iturrarán-Viveros, U., 2012. Smooth regression to estimate effective porosity using seismic attributes. J. Appl. Geophys. 76, 1–12.
Jones, A.J., 2004. New tools in non-linear modelling and prediction. Comput. Manag. Sci. 1, 109–149.
Kisi, O., Shiri, J., 2011. Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. Comput. Geosci. 37 (10), 1692–1701.
Končar, N., 1997. Optimization Methodologies for Direct Inverse Neurocontrol. (Ph.D. thesis) , Imperial College of Science Technology and Medicine, University of London.
Lawrence, S., Giles, C., Tsoi, A., 1996. What size neural network gives optimal generalization? Convergence properties of backpropagation. Tech. Rep., University of Maryland, College Park, Technical Report UMIACS-TR-96-22 and CS-TR-3617. Institute for Advanced Computer Studies.
Lindsay, R., Koughnet, R.V., 2001. Sequential Backus averaging: upscaling well logs to seismic wavelengths. Lead. Edge 20 (2), 188–191.
Malhi, A., Gao, R.X., 2004. PCA-based feature selection scheme for machine defect classification. IEEE Trans. Instrum. Meas. 53 (6), 1517–1525.
Parra, J.O., Hackert, C.L., Bennett, M., Collier, H.A., 2003. Permeability and porosity images based on NMR, sonic, and seismic reflectivity: application to a carbonate aquifer. Lead. Edge 22 (12), 1102–1108.
Parra, J.O., Hackert, C.L., Bennett, M.W., 2006a. Permeability and porosity images based on *P*-wave surface seismic data: application to a south Florida aquifer. Water Resour. Res. 42, 1–14.
Parra, J.O., Hackert, C.L., Xu, P.-C., Collier, H.A., 2006b. Attenuation analysis of acoustic waveform in a borehole intercepted by a sand-shale sequence reservoir. Lead. Edge 25 (2), 186–193.
Parra, J.O., Iturrarán-Viveros, U., Parra, J.S., 2014. Attenuation and Velocity Estimation using Rock Physics and Neural Network Methods for Calibrating Reflection Seismograms (Submitted to Interpretation).
Poulton, M.M., 2001. Computational Neural Networks for Geophysical Data Processing. Pergamon Press, Amsterdam (335 pages).
Russell, B., Hampson, D., Schuelke, J., Quirein, J., 1997. Multiattribute seismic analysis. Lead. Edge 16, 1439–1443.
Sandham, W., Leggett, M., 2003. Geophysical applications of Artificial Neural Networks and Fuzzy Logic. Springer Science and Business Media, New York.
Shiri, J., Kisi, O., 2012. Wavelet and neuro-fuzzy conjunction model for predicting water table depth fluctuations. Hydrol. Res. 43 (3), 286–300.
Shiri, J., Kisi, O., Yoon, H., Lee, K., Nazemi, A., 2013. Predicting ground-water level fluctuations with meteorological effect implications—a comparative study among soft computing techniques. Comput. Geosci. 56, 32–44.
Stefánsson, A., Končar, N., Jones, A.J., 1997. A note on the Gamma test. Neural Comput. & Applic. 5, 131–133.
Taner, M.T., Koheler, F., Sheriff, R.E., 1979. Complex seismic trace analysis. Geophysics 44, 1041–1106.
Taner, M.T., Schuelke, J., O'Doherty, R., Baysal, 1994. Seismic attributes: revisited. 64rd Ann. Internat. Mtg. Soc. Expl. Geophys, pp. 1104–1106.