Geofísica Internacional (2025) **64-3**: 1657 - 1675

Petrophysical evaluation of clastic formations in boreholes with incomplete well log dataset by using joint inversion technique and machine learning algorithms

F. Santana-Roman[1], A. Aquino-Lopez[1], M. Romero-Salcedo†[1], R. Del Valle-Garcia[2] and O. Campos-Enriquez[3]

Abstract

A succesful petrophysical evaluation of shaly-sand formations requieres: 1) the availability of high quality well log data and, 2) a petrophysical model that successfully represents the geological conditions of the rocks. Unfortunately, it is not always possible to fulfill these conditions, and in many cases the set of well logs is incomplete.

To determine petrophysical parameters (i.e., volumes of laminar, structural and disperse shale) in clastic rocks from incomplete well log data we followed three approaches which are based on a hierarchical model, and on a joint inversion technique: 1) Available well log data (excluding the incomplete well log) are used to train machine learning algorithms to generate a predictive model; 2) the first step of the second approach machine learning algorithms are used to generate the missing data which are subsequently included a joint inversion; 3) in the third approach, machine learning process is used to estimate the missing data which are subsequently included in the prediction of the petrophysical properties. The supervised learning paradigm we used was in a joint based on different regression models (linear, decision trees, and kernel).

A performance analysis of the three approaches is conducted with synthetic data (representing real conditions of clastic formations from an oil field in southern Mexico). We simulated gamma ray, deep resistivity, P-wave travel time, bulk density and neutron porosity logs by means of a hierarchical petrophysical model for clastic rock to accomplish a controlled analysis. The three different approaches were applied without P-wave travel time data to analyze the impact of the missing information.

In general, the results indicate an adequate petrophysical parameter determination in each of the approaches. Metric evaluations indicate that the best performance was obtained by the second approach followed by approaches one and three. The correct estimation of the volumes of shale distribution could not be correctly resolved by any of the three applied methods but the total shale content could accurately be predicted which suggests that there is a non-uniqueness problem.

Key words: clastic formations, well logs, petrophysical joint inversion, machine learning.

Resumen

La incertidumbre en la evaluación petrofísica de las formaciones areno-arcillosas depende de la calidad de los datos de registros geofísicos de pozo disponibles y del modelo petrofísico que represente adecuadamente las condiciones geológicas de la roca, pero desafortunadamente, no siempre se cumplen con estas condiciones.

Para determinar los parámetros petrofísicos (volumenes de arcilla laminar, estructural y dispersa) en rocas clásticas en casos de falta de información de registros de pozo, hemos desarrollado tres enfoques basados en un modelo jerárquico de la roca y en la técnica de inversión conjunta: 1) Registros geofísicos de pozo disponibles (excluyendo los registros de pozo incompletos) son usados para entrenar los algoritmos de aprendizaje automático para generar un modelo predicitivo; 2) un segundo enfoque comprende un primer paso en el que se utilizan algoritmos de aprendizaje automático para generar datos de registros geofísicos faltantes y posteriormente, la técnica de inversión conjunta es aplicada; 3) en el tercer enfoque, el proceso de aprendizaje automático se usa para estimar los datos faltantes, los cuales son posteriormente usados para predecir las propiedades petrofísicas. En este trabajo usamos el paradigma de aprendizaje supervisado en diferentes modelos de regresión: lineal, árboles de decisión y kernel.

Se presenta un caso de estudio con datos sintéticos (que representa condiciones reales en formaciones clásticas en un campo petrolero en el sur de México) para comparar los tres enfoques descritos. Modelamos los registros de rayos gamma, resistividad profunda, tiempo de tránsito de la onda P, densidad y porosidad de neutrones usando un modelo petrofísico jerárquico para rocas clásticas para realizar un análisis controlado. Los tres diferentes enfoques fueron aplicados sin los datos de tiempo de tránsito de onda P para analizar el impacto de la falta de esta información.

En general, los resultados indican una determinación adecuada de los parámetros petrofísicos en cada uno de los enfoques analizados. Las métricas de evaluación indican que el mejor desempeño se obtuvo con el enfoque dos, seguido del uno y el tres. La correcta estimación de los volúmenes de las distribuciones de arcilla no pudo ser determinada por ninguno de los tres métodos aplicados, pero el contenido total de arcilla se pudo predecir con precisión, lo que sugiere que existe el problema de no unicidad.

Palabras clave: formaciones clásticas, registros geofísicos de pozo, inversión petrofísica conjunta, aprendizaje automático.

Received: April 30, 2024; Accepted: April 4, 2025; Published on-line: July 1, 2025.

Editorial responsibility: Dra. Silvia Raquel García-Benítez

* Corresponding author: A. Aquino-López, aaquino@imp.mx
[1] Instituto Mexicano del Petróleo, Posgrado, Ciudad de México (CDMX), México.
[2] Instituto Mexicano del Petróleo, Gerencia de Investigación en Exploración, Ciudad de México (CDMX), México.
[3] Consultor Independiente, Geofísica Aplicada, Ciudad de México, México.
[4] Universidad Nacional Autónoma de México, Instituto de Geofísica, Ciudad de México, México.

F. Santana-Román, A. Aquino-López, M. Romero-Salcedo†, R. Del Valle-García, O. Campos-Enríquez

https://doi.org/10.22201/igeof.2954436xe.2025.64.3.1803

1. Introduction

Clastic rocks are economically important because they comprise around 60% of the world reservoir formations (Bjørlykke and Jahren, 2010). The geological complexity of clastic formations, petrophysical models and quality and completeness of measured borehole data generate uncertainty in the petrophysical evaluation of geologically complex clastic formations, especially in the determination of volume and spatial distribution pattern of shale (laminar, structural, and disperse).

The shale spatial distribution affects rock petrophysical properties such as porosity or permeability (Han *et al.*, 1986; Thomas and Stieber, 1975; Wilson and Pittman, 1977) as well as the physical quantities measured by the well logging tools (Aquino-López *et al.*, 2015; Han, 1986; Minear, 1982; Tosaya and Nur, 1982; Wyllie *et al.*, 1958).

Clastic formation evaluation has been undertaken based on different petrophysical models, each one idealizing the rocks by considering different parameters. The most popular petrophysical model to estimate porosity was proposed by Archie in 1942, which is to be used when porosity logs are not available, and mainly to determine water saturation in clean sand rocks. Archie´s petrophysical model has been additionally used in the evaluation of clastic formations with shale content (Bust *et al.*, 2013; Worthington, 2000, 2001, 2011), but in real conditions it is not appropriate for this purpose.

Over the years, by incorporating the internal rock structure, Archie's model, was modified to be able to evaluate shaly-sand formations. Some modifications are based on the electrical length path through the rock (Winsauer etal., 1952); internal geometry of porous media (Perez-Rosales, 1982) and particle shape (Wyllie and Gregory, 1953).

Other modifications of the Archie's model focused on determining water saturation, takes into account shale content distribution. Most known petrophysical models consider dispersed shale volume (Waxman and Smits, 1968); laminar shale distribution (as the Indonesian model, Poupon *et al.*, 1954); free water within the pore spaces of the reservoir rock and the bound water within the clay matrix (Clavier *et al.*, 1984); homogeneous mixtures of sorted sand and natural shale that works regardless of shale distribution (Shedid and Saad, 2017; Simandoux, 1963). These models consider empirical parameters such as the cementation exponent "m" derived from core measurements which renders petrophysical interpretation difficult.

To adequately determine the water saturation of shaly-sand formations, it is necessary to estimate the shale distribution by using previously developed petrophysical models (Aquino-López *et al.*, 2011; Minear, 1982; Thomas and Stieber, 1975; Zwennes, 2017). New techniques and models that allow the evaluation of shaly-sands formations taking into account the shale distribution have been developed, in particular, to simulate P- and S- wave velocity (Minear, 1982); there are volumetric petrophysical model for laminated shaly-sands combining isotropic and anisotropic models properties of formations (Mezzatesta *et al.*, 2002); one that includes all volumes of shale distribution by using conventional and unconventional well logs (Mezzatesta *et al.*, 2006). Recently, a hierarchical petrophysical model of shaly-sand formations has been elaborated, which comprises three levels of homogenization where the model parameters are the spatial distribution of shale, sand, porosity and fluid saturation (Aquino-López *et al.*, 2011).

The set of inversion techniques comprise a large number of mathematical and statistical tools to obtain ("invert") model parameters from the measured, tools that are based on forward modeling and inverse theory (Meju, 1994; Snieder and Trampert, 1999; Zhdanov, 2002). They have been widely used to obtain rock properties from a chosen petrophysical model; one of their first applications comprised their application, through linear equations, together with conventional well logs to conduct a volumetric analysis (Mitchell and Nelson, 1988). In the last years the inversion technique has been applied to: a) invert single well logs, such as the array-induction tools, to find petrophysical properties such as porosity and permeability (Alpak *et al.*, 2006; Torres-Verdín *et al.*, 2006), b) to apply acoustic and electrical logs in microstructure evaluation of carbonate formation (Kazatchenko *et al.*, 2004), c) in the joint inversion of several types of well logs to evaluate carbonates and organic shales formations (Heidari and Torres-Verdín, 2014), d) to estimate shale-gas formations (Heidari and Torres-Verdín, 2013), e) to determine microstructure in carbonate formations (Kazatchenko *et al.*, 2007) and f) to calculate volumes of shale distribution in clastic formations (Aquino-López *et al.*, 2011; 2015).

The joint inversion technique is suitable for predicting model parameters by using different borehole physical measurements (gamma ray, electrical resistivity, neutron porosity, bulk density, photoelectric factor, P- and S- wave travel time), and as already mentioned, its adequate application depends on the quality and quantity of information available. It is well known that in old oil fields it is necessary to deal with incomplete borehole data, mainly without P-wave travel time.

Petrophysical interpretation based on missing well log information can be solved by using machine learning algorithms, which create systems with the ability to learn without being explicitly programmed, that is, automatically and without human intervention (Koza *et al.*, 1996), to preserve and to generate knowledge through examples and experience (Mitchell, 1996).

To develop a prediction algorithm, it is necessary to analyze the available data to discover patterns and reproduce them in another dataset. The algorithm learns from complete well log information and then can to predict data in wells with incomplete

information. Examples of prediction of reservoir properties based on artificial intelligence methods include: 1) neural networks applied to measurement while drilling (MWD) and wireline logs (Bhatt and Helle, 2002), 2) prediction of density, resistivity and neutron using wells logs and location data in sand formations (Rolon *et al.*, 2009), 3) prediction of electrical conductivity (Nguyen-Sy *et al.*, 2021), 4) creation of synthetic neutron logs by using well log data (Ghavami, 2011), 5) prediction of shear sonic log by using different machine learning algorithms (Bukar *et al.*, 2019).

In this paper we analyze the performance of three different approaches for determining lacking P-wave well log, and the parameters of a hierarchical petrophysical model, based on supervised machine learning, and well log joint inversion as a workflow to deal with missing well log datasets in evaluation of shaly-sand formations.

2. Fundamentals of machine-learning

Machine learning is part of the artificial intelligence scientific field, including interdisciplinary relationships between mathematics, statistics, and computer science. The concept "machine learning" was introduced in 1959 as a part of artificial intelligence. It considers that computers have the ability to learn without being explicitly programmed (Samuel, 1959).

The machine learning algorithms can identify and extract the complex structure of the data. Among the different types of machine learning techniques, the most popular are supervised and unsupervised learning processes (James *et al.*, 2013).

- In unsupervised learning the data set is composed only by a feature vector (x_i) comprising several instances or (N) numbered samples. The data are not labeled, and the goal is to find structures and patterns inside the samples to create a new representation that led to the best data interpretation.

$$\{(x_i)\}_{i=1}^{N}$$

- In supervised learning the data set uses a set of known output data (labeled data, y_i) to make a predictive model using a feature vector (x_i).

$$\{(x_i, y_i)\}_{i=1}^{N}$$

Two different approaches can be used as a part of supervised learning techniques:

- Classification: consists of assigning a label to a data vector, these labels are defined as a finite number of classes, which can be binary (two classes), $y \in \{0,1\}$ or multi-classes, $y \in \{0,1,2,3,\ldots, n\}$;
- Regression: consists in the prediction of a label which is a continuous numerical value based on a feature vector, $y \in R$

The prediction of physical and petrophysical properties results in numerical continuous values obtained by means of a regression model as part of supervised learning technique. According to the "No free lunch" theorem (Wolpert and Macready, 1997), there is not a universal machine learning algorithm applicable to every prediction problems and we will work with three different model prediction groups: linear models, decision tree-based models and kernel models (Figure 1).

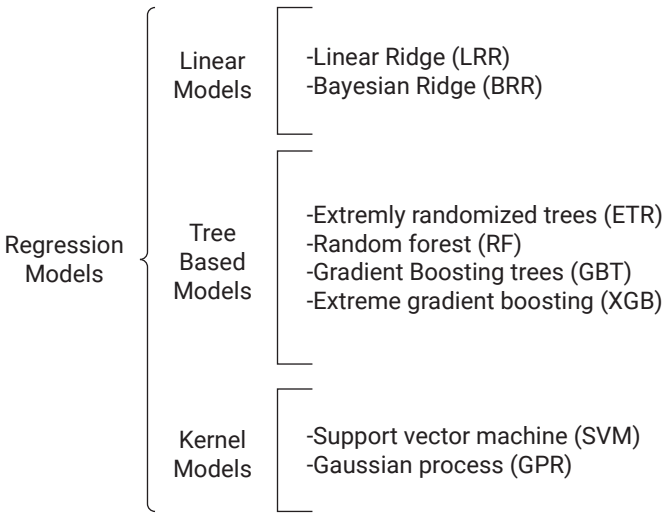

Figure 1. Three different groups of regression models and their algorithms used in this analysis.

2.1 Linear Models

Linear regression is a very straightforward approach to quantitatively predict the labeled data y_i from a predictor variable x_i. This model assumes that there is approximately a linear relationship between x_i and y_i (James *et al.*, 2013). We can write this linear relationship as:

$$y(X,w) = w_0 + w_1X_1 + \cdots + w_kX_k \quad (1)$$

where
$y(X,w)$: is the response model (label).
X_i: are the characteristics of input data (features).
w: weights are constants representing the intercept and slope in the linear model.

In this work we used: 1) linear ridge regression (LRR) which controls the tradeoff between fitting and overfitting during linear regression by adding a penalty term to the original equation (shrinkage penalty term) (Bishop, 2006; Hoerl and Kennard, 1970; Shalev-Shwartz and Ben-David, 2014) and 2) Bayesian ridge regression (BRR) which assigns a probabilistic model to the regression problem which will avoid the over-fitting problem of maximum likelihood (Bishop, 2006).

2.2 Tree Based Models.

Decision tree-based models are defined as an acyclic flowchart used to make a decision (Figure 2). A main node is divided by means of an if/then condition, which will examine the feature vector and their relationships with the label vector. The main node (root node) branches and represents the output of a condition, each one leads to a decision node or a leaf node that represents the end of the test (Friedl and Brodley, 1997).

Four different tree models were used in this work: 1) The Random Forest (RF) is an ensemble method based on decision trees; the model introduces extra randomness when splitting a node; it searches for the best feature among a random subset of features (Géron, 2019; James *et al.*, 2013). 2) Extremely randomized trees (ETR) are based on an ensemble method by using many decision trees, the model makes trees even more random by also using random thresholds and it uses the whole learning sample, and it is much faster to train than regular random forests (Géron, 2019; Geurts *et al.*, 2006). 3) Gradient boosting tree (GBT) and 4) Extreme gradient boosting (XGB) are based on ensemble methods and the main idea of boosting is to train predictors sequentially, where each tree is grown using information from previously grown trees trying to correct its predecessor (Géron, 2019; James *et al.*, 2013).

2.3 Kernel models

The term “kernels" is used to describe inner products in the feature space that allow that linear parametric models can be recast into an equivalent ‘dual representation’ in which the predictions are also based on linear combinations (Bishop, 2006; Boser *et al.*, 1992; Shalev-Shwartz and Ben-David, 2014). Therefore, it represents a linear version of any non-linear data-based algorithm.

For models we used in this paper, which are based on a fixed non-linear mapping of the feature space, $\varphi(x)$, to a higher dimensional space, the kernel can be expressed as $K(x_i,x_j) = \varphi(x_i)^T \varphi(x_j)$ (Bishop, 2006).

Figure 3 on the left depicts two data sets that are not linearly separable, on the right side of the figure, it can be seen that in a

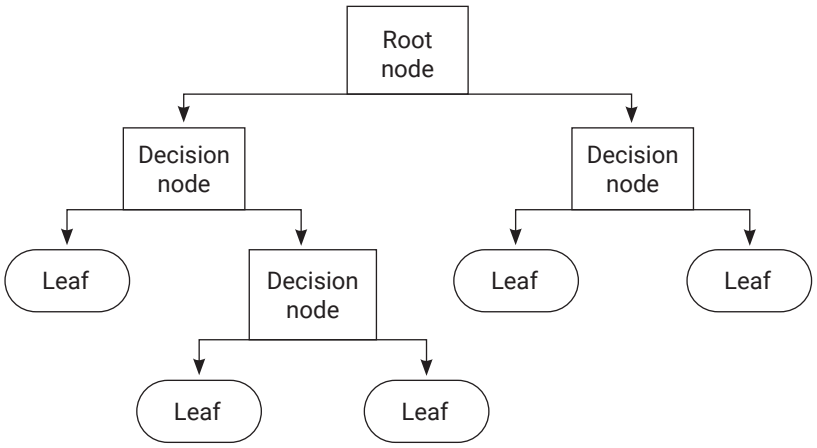

Figure 2. Scheme of application of a decision tree, modified from Friedl and Brodley (1997).

higher dimension space, it is possible to separate the data using a linear relationship.

Two different kernel models were used in this work: 1) Support vector machines (SVM) that allow to map the data into a higher dimensional feature space via nonlinear mapping, after which a regression can be performed in this feature space (Wang and Hu, 2015). 2) Gaussian process (GPR), in these models a prior probability distribution over functions is directly defined and the inference takes place in the space of functions (Bishop, 2006; Rasmussen and Williams, 2006).

The regression models are useful for the prediction of numerical values, unfortunately, there is no preferred model (i.e., better or worse than another) since each one has significant different weakness or robustness's, for example:

- Linear models are useful for making quick predictions because their implementation is easy and the overfitting is low, however, they cannot model complex or non-linear relationships and presents errors when outliers are present in the data.
- Decision tree models are good at learning complex or non-linear relationships and are robust to noise and overfitting, however, trees can be difficult to interpret because they can be too complex.
- Kernel-based models are useful to predict complex or non-linear relationships and are robust to noise and maximize the solution margins, however, finding a suitable kernel function can be difficult in addition to the longer processing time.

3. Petrophysical model

The high heterogeneity in shaly-sand formations poses problems to a correct petrophysical evaluation. Managing to include the rock composition, considering the main components and their distribution, in a model as much as closest to the reality provides an opportunity for a correct evaluation of these formations. In this study, the used petrophysical model considers three hierarchical levels (Figure 4). The first one is related to components inside the pores, the second at sandstone rock and the third level considers intercalations of laminar shales and sandstones (Aquino, 2011).

The petrophysical hierarchical model is composed by:

- First level of homogenization: A pore space composed by water, hydrocarbons (oil and gas) and dispersed shale.
- Second level of homogenization: A sandstone composed by quartz, structural shale, and pores.
- Third level of homogenization: Composed by intercalations of sandstone and laminar shale.

The modelling of physical effective properties on the first and second homogenization levels were achieved by using the micromechanical method of the effective media approximation (EMA). EMA has been applied for modelling acoustic velocities and electrical conductivity in carbonates rocks (Kazatchenko *et al.*, 2004, 2007; Markov *et al.*, 2005) and clastic rocks (Aquino-López *et al.*, 2011, 2015). This method is used to simulate effective properties considering the content and shape of different components of the model at the first and second homogenization levels. On the third homogenization level a transversely isotropic medium composed by layers of shale and sandstones is considered.

The petrophysical parameters considered in the model were obtained through a joint inversion of conventional well-log data. The procedure consists in the minimization of a cost function (equation 2), that comprises a weighted sum of the squares differences of experimental and predicted data (Aquino-López *et al.*, 2015; Aquino-López *et al.*, 2011; Kazatchenko *et al.*, 2007)

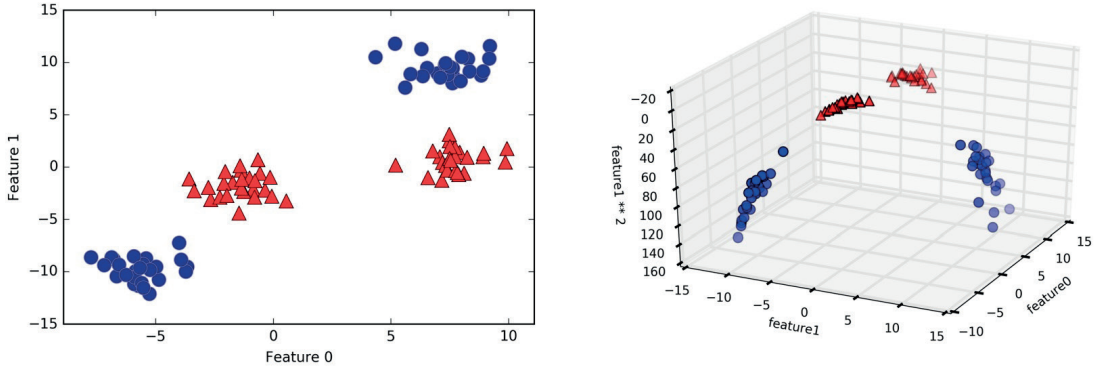

Figure 3. Data linearly separable after a transformation into a higher dimensional space (modified from Müller and Guido, 2016).

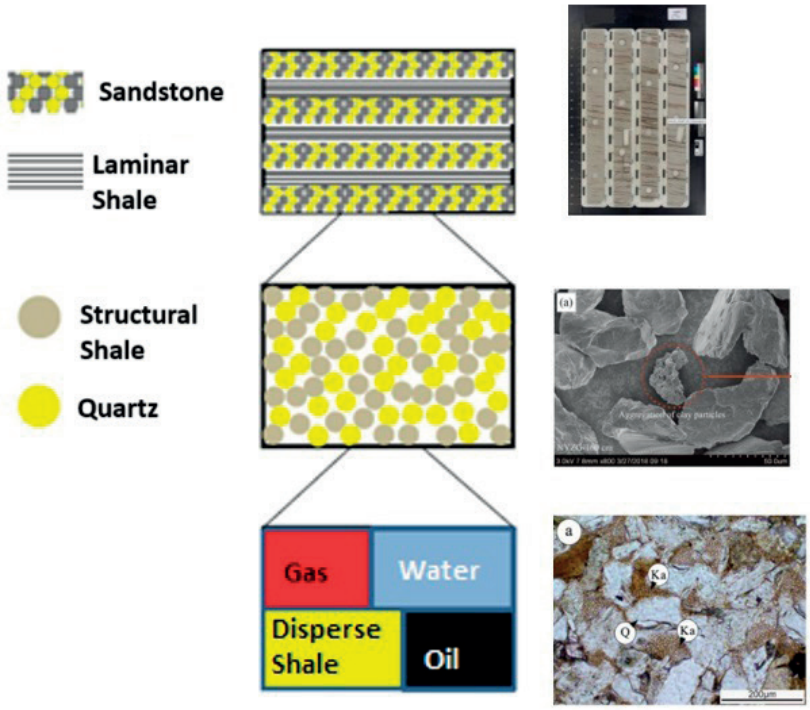

Figure 4. On the left side are represented the three-level hierarchical models for shaly-sand formations based on the size of the components (Aquino, 2011). On the right side the photographs show the components at different levels. First level pore-filling by shale (Kaolinite) (Zhang *et al.*, 2017). Second level quartz and clay-sized particles (Zeng *et al.*, 2020). Third level laminar shale (Hajizadeh *et al.*, 2019).

$$F(Z_i) = W_d \|d(m) - d_{obs}\|^2 + \lambda(W_m \|m - m_o\|^2) \quad (2)$$

Where:

d_{obs} is a vector of measured data that include gamma ray, deep resistivity, bulk density, P-wave travel time, S-wave travel time, neutron porosity.

$d(m)$ is a vector of simulated data.

m is a vector of the model petrophysical parameter that includes laminar, structural, dispersed shales, gas, oil and water volumes.

m_o is the vector of the initial model.

W_d weight coefficients of data.

W_m weight coefficients for model components.

λ regularization parameter.

4. Methodology

To establish the best approach to determine petrophysical parameters, we simulate, by using the hierarchical model, the synthetic well logs data (gamma ray, deep resistivity, neutron porosity, bulk density, P- and S- wave travel time). The advantage of this treatment is that we know the response of physical properties of a controlled synthetic model (Figure 5).

The methodology applied in this work is focused on the determination of the petrophysical properties of the rocks when well log information is incomplete, and it comprises the application of the three described different approaches that lead us to the prediction of petrophysical properties (laminar, structural and dispersed shale and fluids volumes) using machine learning techniques and hierarchical petrophysical model for clastic rocks.

The implementation process of the three mentioned approaches for estimating petrophysical parameters followed the steps as described below:

4.1 Data analysis and preparation.

Many factors as enviromental conditions, tool failure and human errors affect the well log information. These incorrect data are labeled in the well log with the flag -999. These values and low quality data must be deleted because they generate error in machine learning models. The process of correcting inaccuracies, clean anomalous data, classiffy and organization of the available data constitutes a big part of the data preparation and their analysis. Once we have the clean and the available data identified, we create our training and validation templates which will be used in our prediction process.

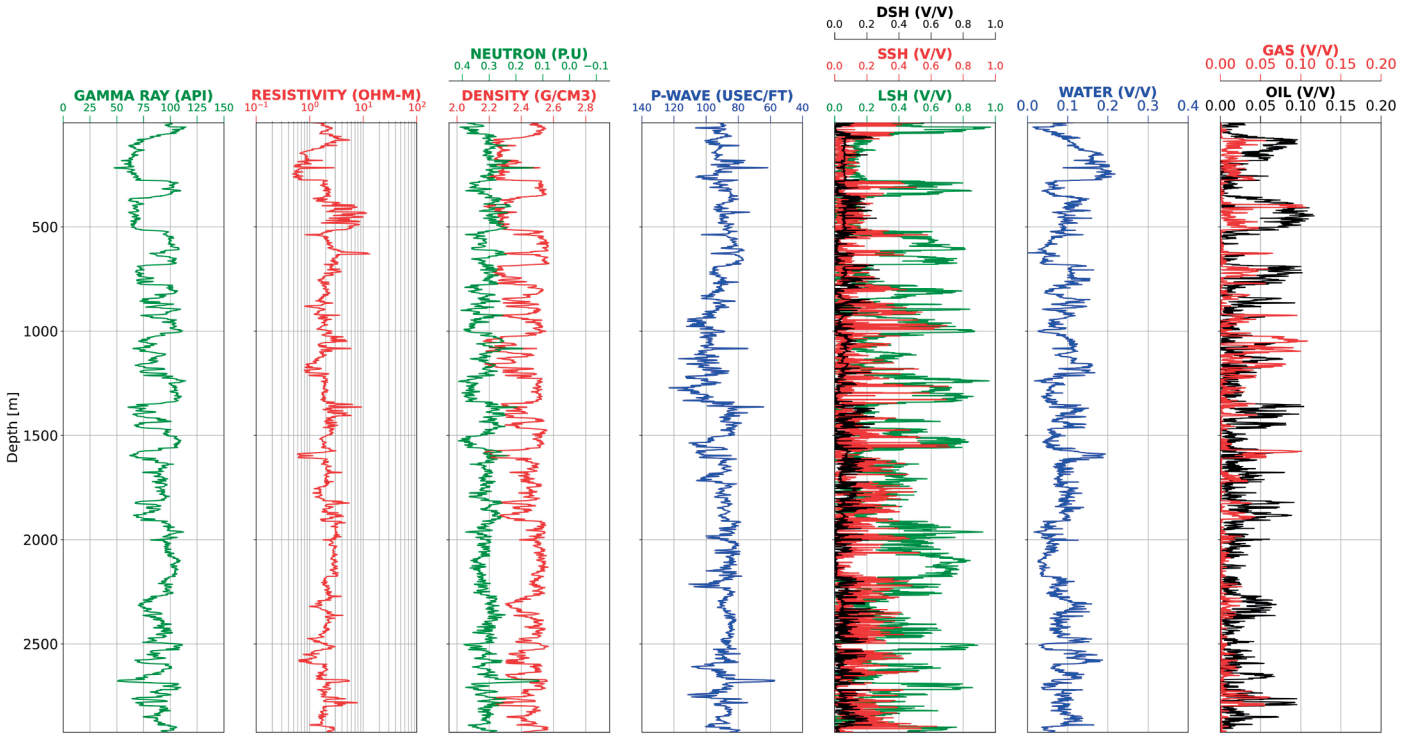

Figure 5. Well log data and related petrophysical properties. Track 1: gamma ray, track 2: deep resistivity, track 3: neutron porosity and bulk density, track 5: shale distribution (laminar, structural and disperse), track 6: volume of water, track 7: volume of oil and gas.

4.2 Data transformation.

Well log data sets contain different types of measurements with varying scale magnitudes, for example, gamma ray has a value range from 0 to 300 API units, resistivity from 0.0001 to 200000 Ohm-m, porosity from 0 to 0.4, density from 2 to 3 gr/cc. These large range magnitud variations need to be scaleted due inputs with larger range values may significantly affect the outputs of models (Anemangely *et al.*, 2019). We use the standarization (z-scores normalization) to scale our data and is useful to compare data from different samples or populations, it is defined as:

$$z = \frac{x - \mu}{\sigma} \tag{3}$$

Where:
z= standardized variable
x= original data
μ=mean of the population
σ= standard deviation of the population.

4.3 Data Partitioning

In this step we randonmly separate our training data in two segments, train and test datasets. Train data is used to train machine learning models and test data is used to test the trained model and it is never used to train the model. A 70 to 30 split ratio is commonly used to partition the data for training and testing purposes respectively. One of the most popular partitioning ratio is 70% training subdata and 30% test subdata and it was used in this paper.

4.4 Metric evaluation

Once the training data has been processed with different machine learning algorithms it is necessary to asses the performance of each model to select the best one. To do that, we choose three different performance metrics (equations 4-6). Root-mean-square error (RMSE) and a normalized root-mean-square error to compare errors between data with different scale magnitudes. RMSE and NRMSE are measured in the units of y and it is not always clear what constitutes good RMSE and NMRSE values. We use the coefficient of determination (R^2) as

an alternative measure of fit and is independent of the scale of y (James *et al.*, 2013).

$$R^2 = 1 - \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2} \quad (4)$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y} - y)^2}{n}} \quad (5)$$

$$NRMSE = \frac{\sqrt{\sum_{i=1}^{n} \frac{(\hat{y} - y)^2}{n}}}{\bar{y}} \quad (6)$$

Where
$\hat{y}$ The predicted value
y The real values
$\bar{y}$ The mean of y values
n Number of samples

4.5 Approaches to apply Machine Learning for predicting petrophysical parameters

This study was focused to analyze the performance of machine learning (ML) methods in the case of incomplete data, namely when a physical parameters was lacking. The physical parameter was the P-wave travel time. Three approaches were considered to predict petrophysical parameters considered in the petrophysical model. For prediction of P-wave travel time and petrophysical parametrs by using machine learning, we used the models mentioned above (LRR, BRR, ETR, RF, GBT, XGB, SVM and GPR). These approaches are the following ones:

4.5.1 Approach 1 (Petrophysical prediction by using ML without P-wave travel time)

This case comprises a data set where the P-wave travel time is absent. In this approach, the P-wave travel time was not generated. To predict the petrophysical parameters we trained machine learning models with available information and evaluate their performance to select the optimal one (Figure 6).

4.5.2 Approach 2 (Prediction of P-wave travel time by using ML and petrophysical parameters determination by joint inversion)

The second approach comprises an application of the technique of machine learning followed by the joint inversion procedure to estimate petrophysical parameters of the clastic rocks. In this case as a first step (Figure 7), we generate the missing log (P-wave travel time) by means of machine learning technique. To do that, we train the machine learning models and evaluate their performance to select the best one. After that, we use the validation data to predict the well log missing.

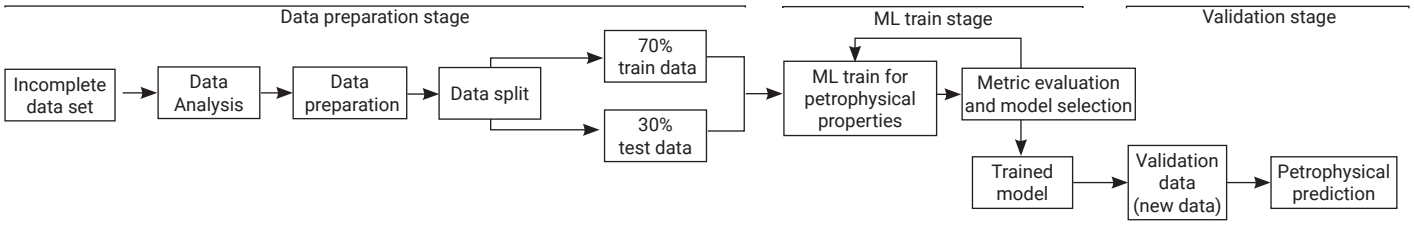

Figure 6. Proposed steps for prediction of petrophysical parameters (approach 1).

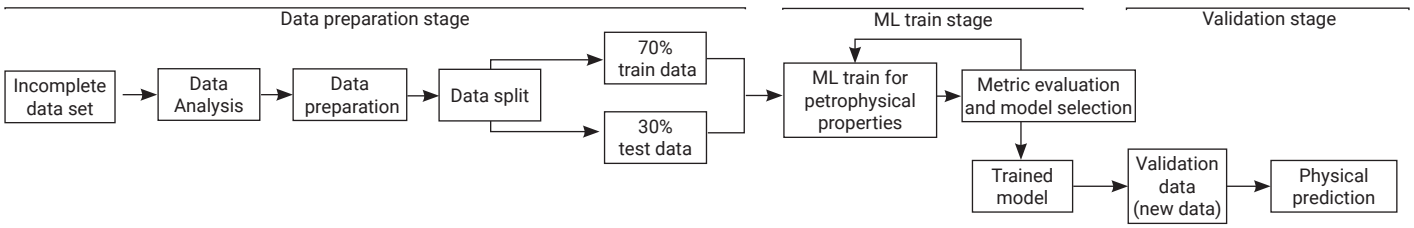

Figure 7. Proposed steps for prediction of missing well log data.

In the step 2, once we have a complete dataset, conformed by original data (gamma ray, deep resistivity, neutron porosity, bulk density and S- wave travel time) and machine learning predicted P-wave synthetic data, we applied the joint inversion procedure to estimate petrophysical parameters of the clastic rocks (Figure 8).

4.5.3 Approach 3 (Petrophysical prediction with a complete data set by using ML)

In this approach, we applied a double machine learning process. The first one is focused to obtain P- wave travel time data similarly to step one of approach 2 (Figure 7).

Once a complete data set is available, composed of original data (all logs except P-wave travel time) and predicted machine learning data (P-wave travel time), we applied a second machine learning process. In this step, we train our machine learning models and select the best one. After that we use the complete data and obtain the petrophysical parameters of the hierarchical model of clastic rocks (Figure 6).

5. Application

The three developed approaches to determine the hierarchical model petrophysical parameters (volumes of laminar, structural and dispersed shales and fluids) were applied to three synthetic well logs data (well-1, well-2 and well-3), each one is composed of 2,924 samples (5,848 for training and 2,924 for testing). The synthetic data was separated into two groups. The first one is composed by the well-1 and well-3 and was considered for the training process. The second one is composed by the well-2, which was used for validation (Figure 9).

Two additional parameters were considered (shale volume and hydrocarbon volume) amounting to a total of five well log data and nine petrophysical parameters. Basic statistical parameters (minimum, maximum, standard deviation and mean) for every physical property resulting of the training and test data are shown for gamma ray, electrical resistivity, neutron porosity, bulk density and P-wave travel time (Figure 10). All properties for training and testing data and all statistics are very close, which means that data are correlated, and it is expected to be good determination of petrophysical parameters.

Figure 11 shows the statistics of the real petrophysical parameters derived from the train and test wells. It is noteworthy that there is a higher variation in the petrophysical parameters than in the well log data.

5.1 Proposed Case of Study

We first tested the three different approaches described above on the synthetic data shown in Figure 5. As already mentioned, we considered that P-wave transit time was not available. Next, the results of applications of the three approaches in the deter-

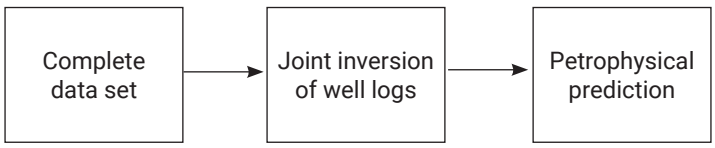

Figure 8. Step two for prediction of petrophysical parameters (approach 2).

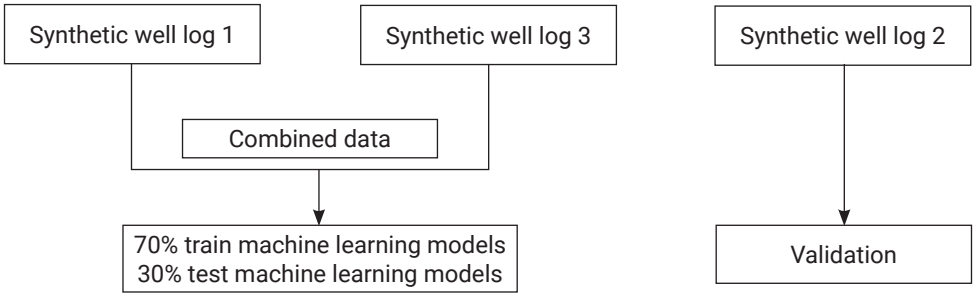

Figure 9. Partitioning of the data used to develop the three different approaches.

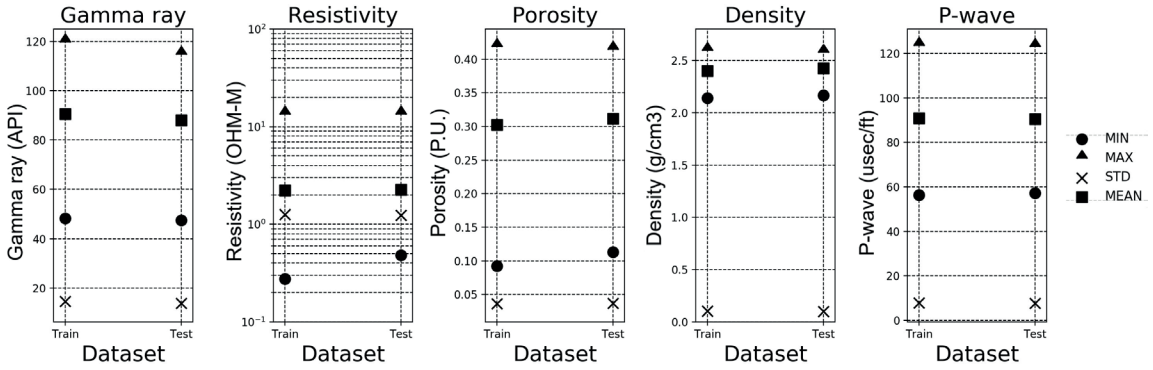

Figure 10. Statistical description of train and test datasets of well log data (gamma ray, resistivity, neutron porosity, density and P-wave travel time).

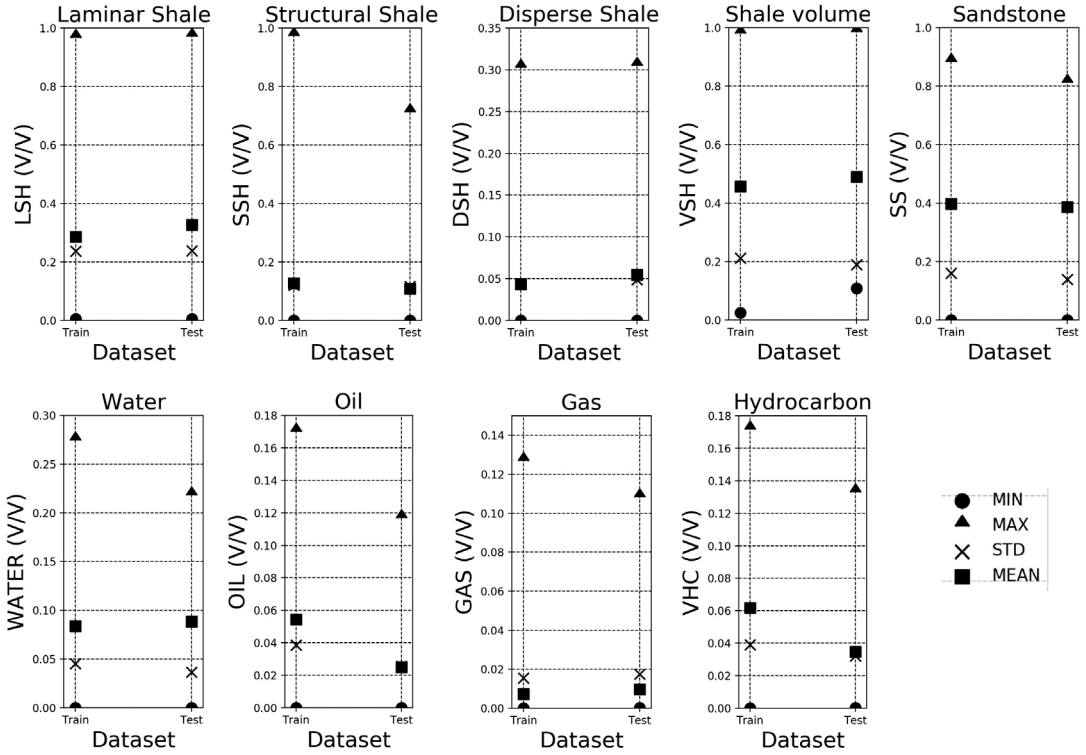

Figure 11. Statistical description of train and test datasets of petrophysical parameters described by the model.

mination of petrophysical properties are presented. Only the coefficient of determination (R^2) is described to select the best prediction because it is more informative than RMSE and NRMSE in regression problems (Chicco *et al.*, 2021). In cases where R^2 is not informative, all the metrics described above will be used.

5.1.1 Results with approach 1 (Petrophysical prediction by using ML without P-wave travel time)

As mentioned before, approach 1 was applied to determine the hierarchical petrophysical parameters without requiring an estimation of the missing well log data. In the training stage, we used as feature data bulk density, electrical resistivity, neutron porosity and gamma ray logs, and as label data the petrophysical parameters describing the hierarchical model (Figure 12). Figure 12 was elaborated to show the workflow followed to predict the petrophysical parameters with approach 1.

Once the models were trained for each of the properties to be predicted, their performance evaluation was assessed, and the best learning machine models were selected to be used in the validation stage. Figure 13 shows the results of training evaluation using R^2 score in three different groups. Group a) shows the

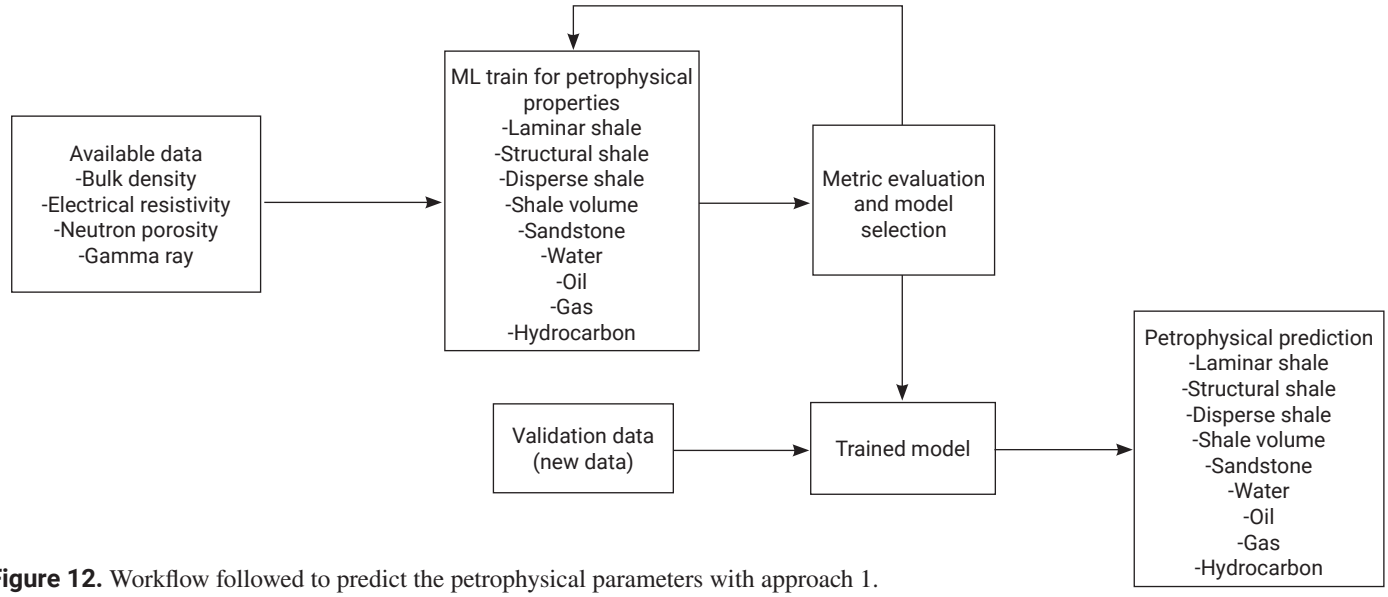

Figure 12. Workflow followed to predict the petrophysical parameters with approach 1.

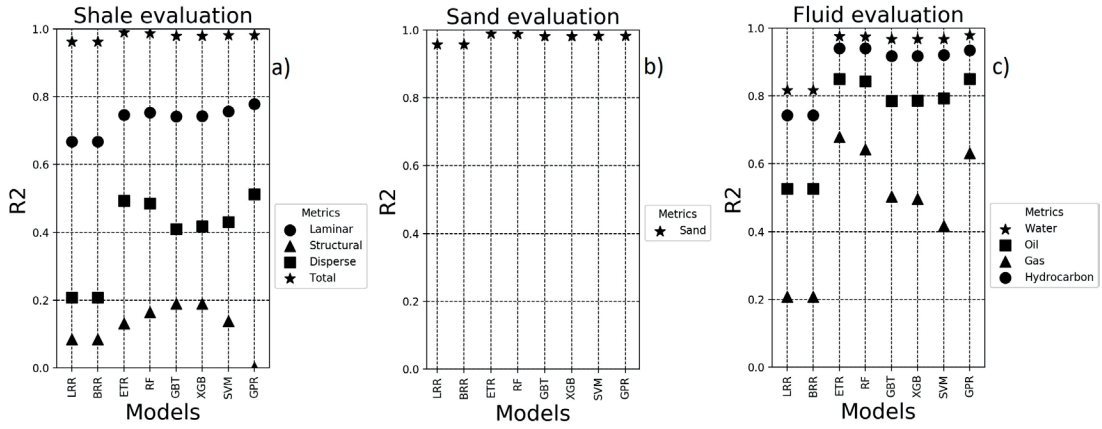

Figure 13. R^2 score in training stages for petrophysical parameter determination by using the different machine learning models.

evaluation of shale and its distribution. The total shale shows the best training (R^2) compared to the shale distribution especially structural and disperse. Group b) shows sandstone evaluation where training is adequate in almost all models. Group c) shows fluid evaluation. Water shows good R^2 training in almost all the models, hydrocarbon volume shows a good R^2 training compared to oil and gas volumes.

After metric evaluation and model selection, the validation data was input to the model to predict the petrophysical properties. In Figure 14 it can be observed the matching degree between the actual and predicted properties.

The predicted data generated by the models during validation presents a high fitting degree in most cases, Table 1 shows the performance of validation.

5.1.2 Results with approach 2 (Prediction of P-wave travel time by using ML and petrophysical parameters determination by joint inversion)

As mentioned before, in approach 2, a requirement is to generate the missing well log data (step 1) to proceed to the petrophysical parameters determination of the hierarchical model. In this training step we used as feature data the bulk density, electrical resistivity, neutron porosity and gamma ray logs, and as label data the P-wave travel time of wells one and three (Figure 15). Figure 15 was elaborated to show the workflow followed to predict de P-wave travel time.

Eight different machine learning models (LRR, BRR, ETR, RF, GBT, XGB, SVM, and GPR) were applied to obtain the

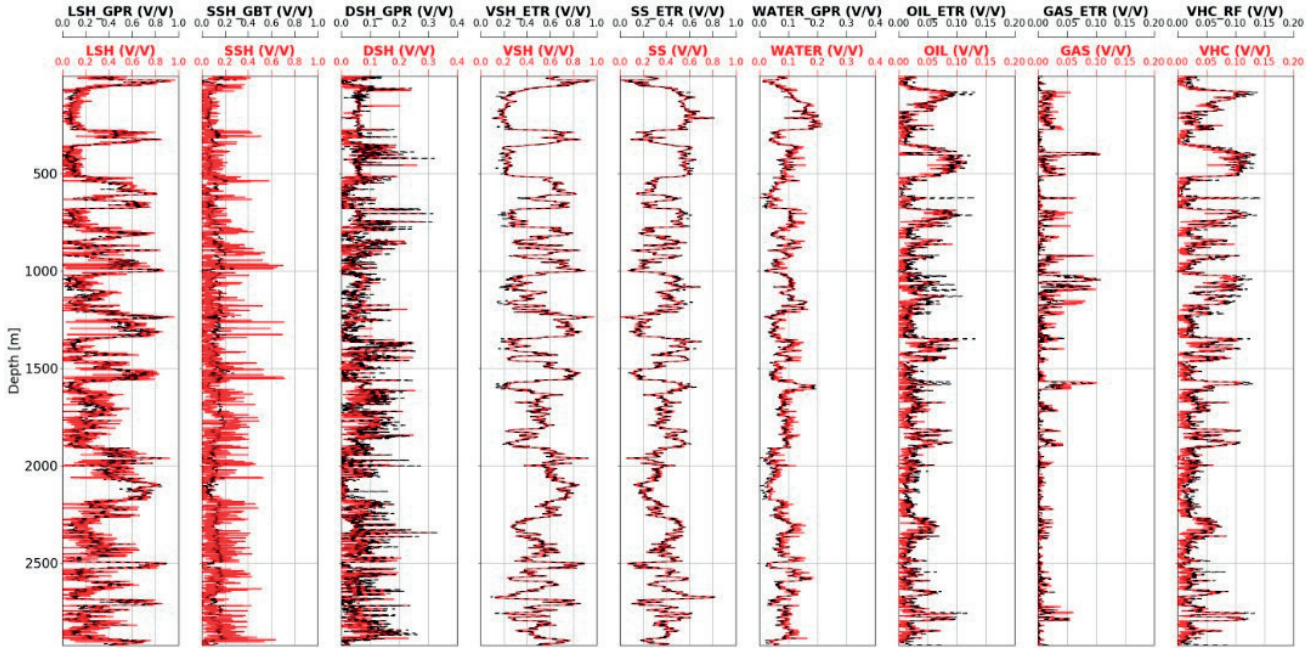

Figure 14. Comparison of results of the validation process following approach 1. Predicted values match quite well to the well log measured values.

Table 1. Performance of petrophysical evaluation following approach 1.

Model	Parameter	R^2	RMSE	NRMSE
GPR	Laminar shale (LSH)	0.8143	0.1025	0.3138
GBT	Structural shale (SSH)	0.1905	0.1042	0.9612
GPR	Disperse shale (DSH)	0.4398	0.0362	0.6638
ETR	Sandstone (SS)	0.9887	0.0147	0.2706
GPR	Water	0.9244	0.0100	0.1129
ETR	Oil	0.6713	0.0145	0.5842
ETR	Gas	0.8205	0.0074	0.7579
ETR	VSH = LSH + SSH + DSH	0.9895	0.0193	0.0394
RF	VHC = Oil + Gas	0.8619	0.0119	0.3430

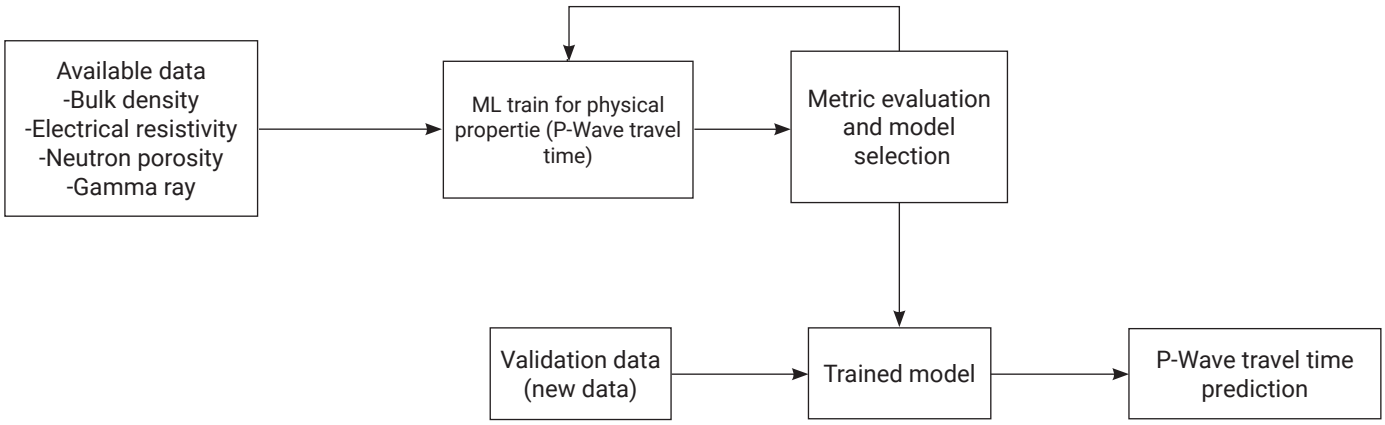

Figure 15. P-Wave travel time prediction by using available physical data.

missing well log data. The selection of the best model was based on those having the lowest RMSE and NRMSW, and the highest R^2. The performance of these machine learning methods is depicted in Figure 16.

Once the best model was selected (ETR), this trained learning model was applied to the validation data (well 2) to find the well log of the transit time. Table 2 shows the performance of the P-wave transit time prediction. The left side of Figure 17 shows the comparison between the real and predicted data, and in its right side it is shown a cross plot showing the dispersion of the predicted data.

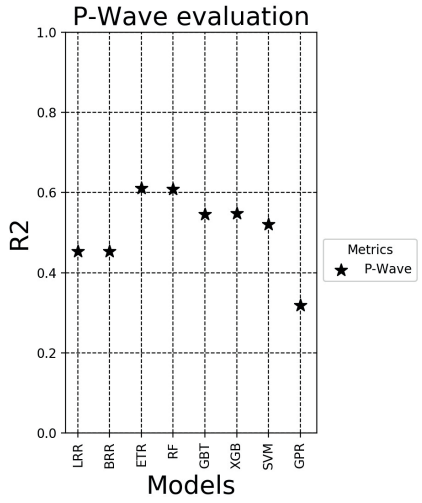

Figure 16. Comparison of results of the validation process following approach 2. ETR, and RF methods yielded the best approximations. GPR resulted in a poor fit.

Once the set of well log data was complete, the numerical joint inversion process described at the end of section 4 (i.e., equation 2) was applied to obtain the petrophysical properties of the hierarchical model. Figure 18 shows the difference between the actual and predicted data.

Prediction of petrophysical data using approach 2 provide better results than those obtained from approach 1, except for the spatial distribution of shale. Table 3 shows the performance after the inversion process.

5.1.3 Results with approach 3 (Petrophysical prediction with a complete data set by using ML)

In approach 3, the prediction of petrophysical properties includes in its first phase an estimate of the missing well log (P-wave travel time in this case). For the training stage, we used the bulk density, electrical resistivity, neutron porosity, gamma ray logs, while P-wave constitutes the feature data, and the petrophysical properties describing the petrophysical hierarchical model as the label data (Figure 19). Figure 19 was elaborated to show the workflow followed to predict petrophysical properties by using an incomplete data set.

Table 2. Performance of transit time prediction.

Well log	Model	R^2	RMSE	NRMSE
IDT	ETR	0.758315	3.7622	0.041521

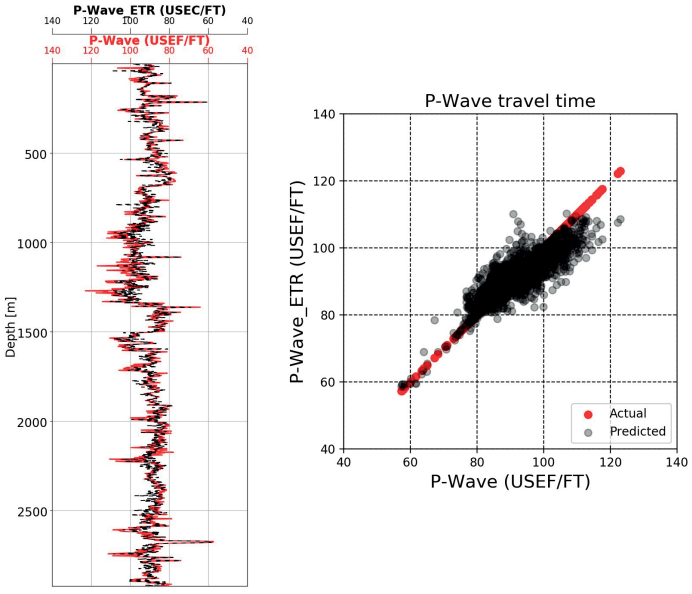

Figure 17. Right side: Comparison between real and predicted P-wave travel time following approach 2. Left side: Cross-plot of predicted and measured travel times.

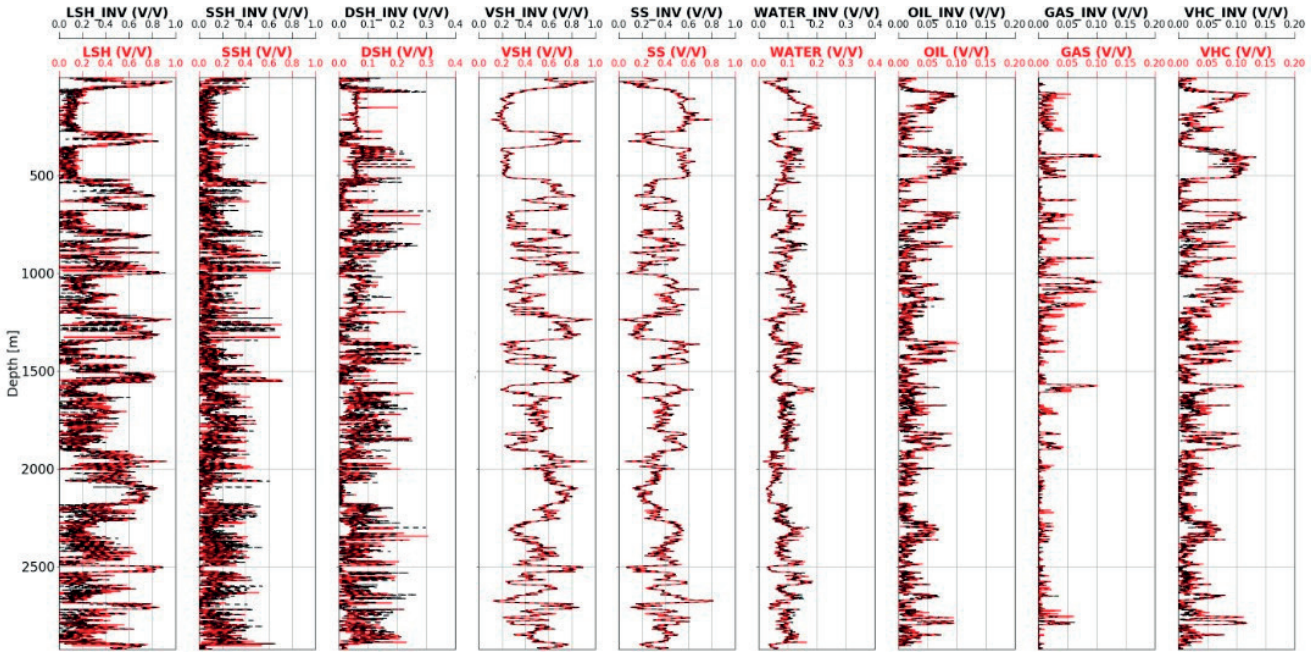

Figure 18. Comparison of results of the validation process following approach 2. Predicted values match quite well to the well log measured values.

Table 3. Performance measurements of petrophysical estimations following approach 2.

Model	Parameter	R^2	RMSE	NRMSE
Inversion	LSH	0.68845	0.13289	0.40652
Inversion	SSH	-0.1776	0.12579	1.15941
Inversion	DSH	0.14972	0.04467	0.81784
Inversion	SS	0.9927	0.0119	0.03078
Inversion	Water	0.92202	0.01017	0.11482
Inversion	Oil	0.8641	0.00937	0.37567
Inversion	Gas	0.87497	0.00621	0.6328
Inversion	VSH	0.99744	0.00956	0.01952
Inversion	VHC	0.90637	0.00982	0.28248

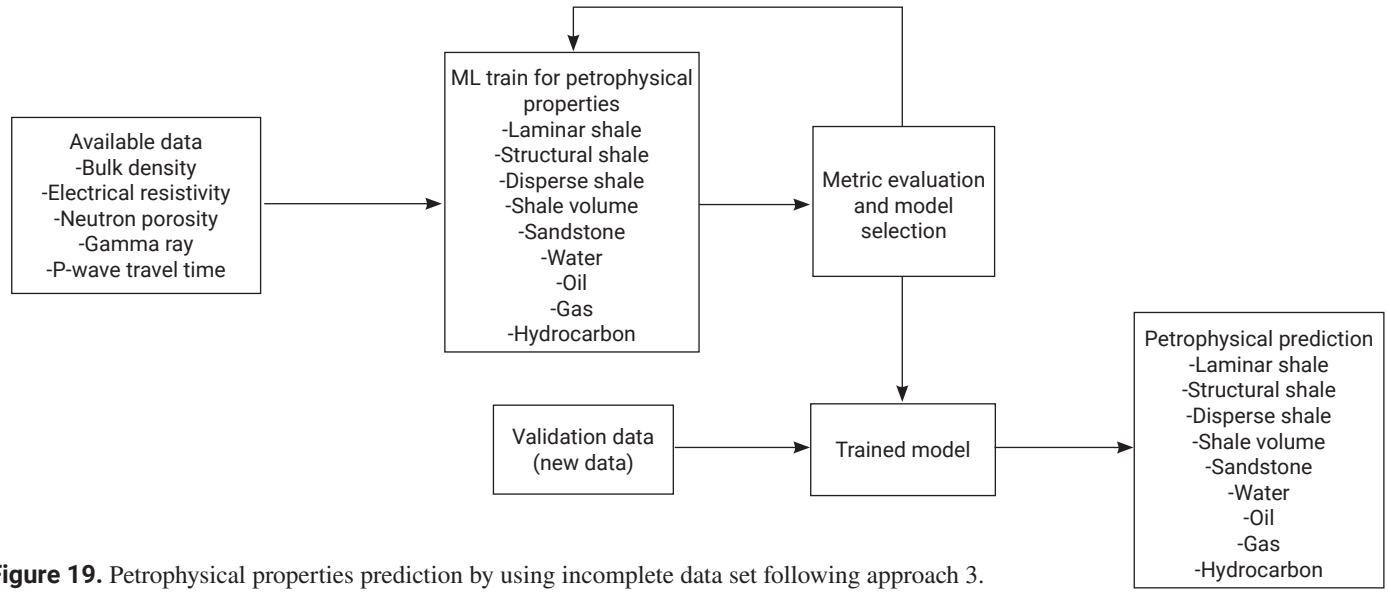

Figure 19. Petrophysical properties prediction by using incomplete data set following approach 3.

Figure 20 shows the R^2 score of the results of training evaluation of three different groups. Group a) shows the evaluation of shale and its distribution. Group b) shows sandstone. Group c) shows fluid evaluation. Model training using a complete data set showed superior performance (compared to approach one) on all petrophysical properties. The best models were used in the validation stage for every petrophysical parameter (Table 4).

Once the best models were determined according the results shown in Table 4, they were applied to the validation data to predict the petrophysical properties. Figure 21 enables a comparison between the actual and predicted data and Table 4 shows their respective performances.

5.1.4 Discussion

The results shown above for the three different approaches indicate that they have different performances. Table 5 shows that an approach may be better than another, for example, approach 3 showed a lower performance compared to approaches 1 and 2.

In the training stage approach 3, the use of an extra piece of information (P-Wave travel time) lead to a better performance metrics compared to approach 1 where the missing well log were not generated, however in the validation stage, approach 1 showed a better performance metrics compared to approach three.

Our results indicate that it is not possible to determine without high uncertainty levels, the correct values for shale distribution, nevertheless, for the well log analysis, the results are promising. A very interesting fact is that the calculation of the total volume of shale, that showed an accurate prediction, suggests that there is an equivalence problem which should be analyzed in more detail in future research.

The application of different machine learning models indicates that there is not a unique model to optimally predict all the petrophysical properties in agreement with Wolpeet and Macready (1977) that stated that some methods performed very good in predicting specific petrophysical properties, but no good agreement was obtained in all the properties.

In this study, the best performances were obtained in the following order: first by approach two, then for approach one and finally for approach three. The results also showed that the performance of the three different approaches was not enough

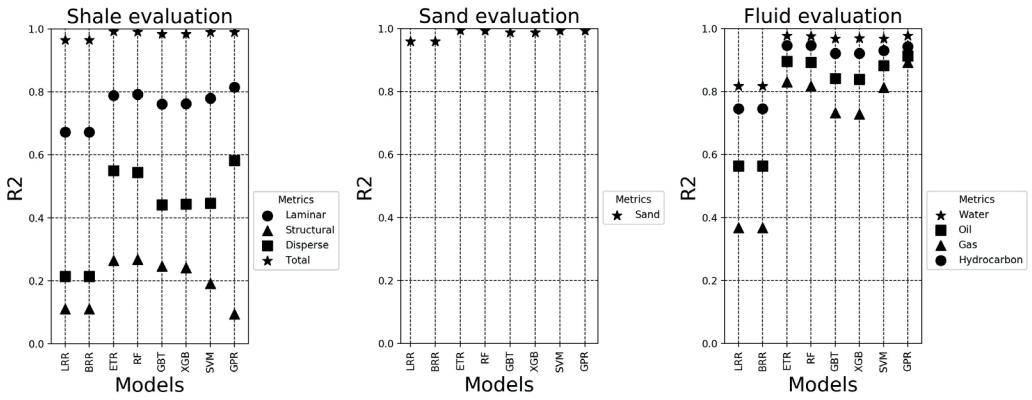

Figure 20. Training results of machine learning models by approach 3.

Table 4. Performance measurements of petrophysical estimations following approach 3.

Model	Parameter	R^2	RMSE	NRMSE
GPR	LSH	0.7739	0.1131	0.3462
RF	SSH	0.2189	0.1024	0.9442
GPR	DSH	0.3986	0.0375	0.6877
ETR	SS	0.9867	0.0160	0.0413
ETR	Water	0.8984	0.0116	0.1310
ETR	Oil	0.6091	0.0158	0.6370
GPR	Gas	0.7632	0.0085	0.8707
ETR	VSH	0.9875	0.0210	0.0430
ETR	VHC	0.8713	0.0115	0.3311

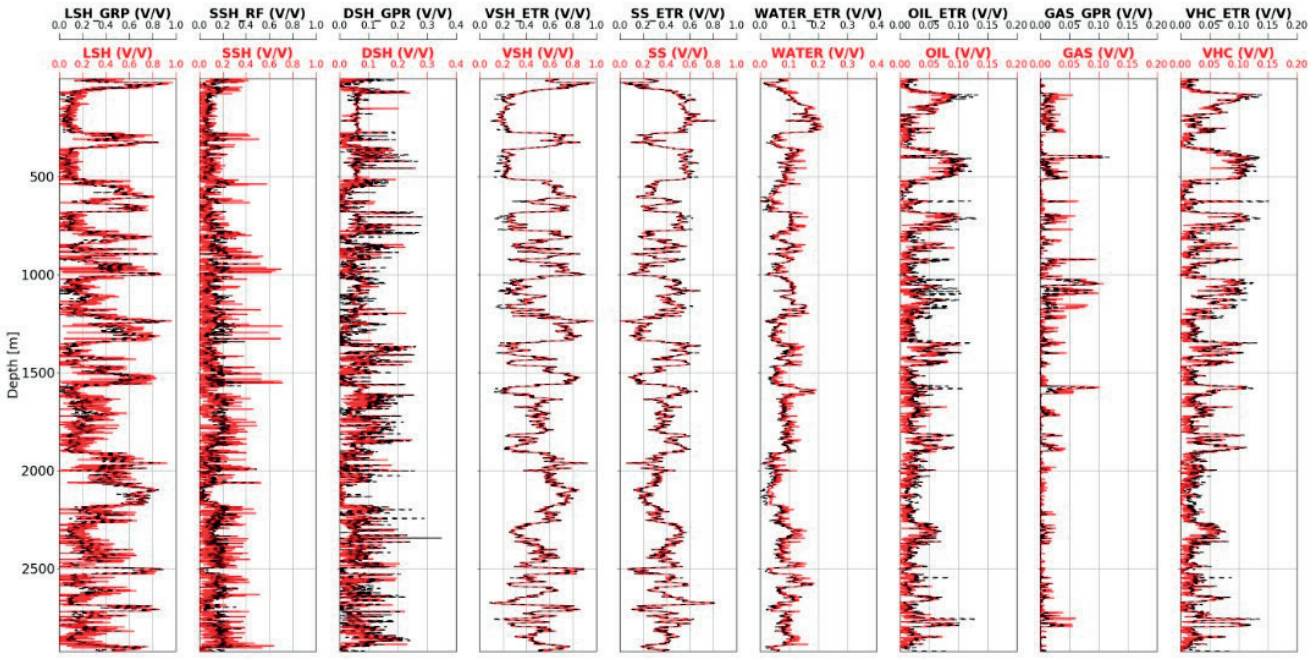

Figure 21. Comparison of results of the validation process following approach 3. Predicted values match quite well to the well log measured data.

Table 5. Comparison of the different approaches used, shaded the best ones.

	Approach 1			Approach 2			Approach 3		
	R^2	RMSE	NRMSE	R^2	RMSE	NRMSE	R^2	RMSE	NRMSE
LSH	0.8143	0.1025	0.3138	0.68845	0.1328	0.40652	0.7739	0.1131	0.3462
SSH	0.1905	0.1042	0.9612	-0.1776	0.1257	1.15941	0.2189	0.1024	0.9442
DSH	0.4398	0.0362	0.6638	0.14972	0.0446	0.81784	0.3986	0.0375	0.6877
SS	0.9887	0.0147	0.2706	0.9927	0.0119	0.03078	0.9867	0.016	0.0413
Water	0.9244	0.01	0.1129	0.9220	0.0101	0.11482	0.8984	0.0116	0.131
Oil	0.6713	0.0145	0.5842	0.8641	0.0093	0.37567	0.6091	0.0158	0.637
Gas	0.8205	0.0074	0.7579	0.8749	0.0062	0.6328	0.7632	0.0085	0.8707
VSH	0.9895	0.0193	0.0394	0.9974	0.0095	0.01952	0.9875	0.021	0.043
VHC	0.8619	0.0119	0.343	0.9063	0.0098	0.28248	0.8713	0.0115	0.3311

for the prediction of the spatial distribution of shale, especially in structural and dispersed shale (R^2<0.44), but for the total shale content estimation they presented a high level of accuracy in the three different approaches (R^2>0.98).

6. Conclusions

This performance analysis of the results obtained with application of machine learning methods to wells with an incomplete dataset indicates that petrophysical prediction with a reasonable degree of accuracy is possible by using machine learning algorithms together with a joint inversion technique following the described approaches. Furthermore, it was demonstrated that the best performances were obtained in the following order: first by approach two, then for approach one and finally for approach three and they can be used when any well log data is not available. The results also showed that the performance of the three different approaches was not enough for the prediction of the spatial distribution of shale especially in structural and dispersed shale (R^2<0.44), and that for the total shale content estimation it presented a high level of accuracy in the three different approaches (R^2>0.98).

The prediction of the P-wave travel time log can be calculated with a high level of certainty by machine learning models. The decision tree-based machine learning models provided the best performance, and the synthetic P-wave travel time log could be used for predicting petrophysical properties (approaches 2 and 3), showing favorable results.

Using a complete dataset (approach 2) by previously generated P-wave travel time log with a robust petrophysical model together with a well log joint inversion technique, provided better results than approaches 1 and 3 for predicting petrophysical properties.

7. Acknowledgments

We thank the Mexican Petroleum Institute, in whose 'Academic Posgraduate Program' frame this study was fulfilled. Special thanks to all anonymous reviewers for their important comments and technical corrections to improve this paper.

8. References

Alpak, F. O., Torres-Verdín, C., and Habashy, T. M. (2006). Petrophysical inversion of borehole array-induction logs: Part I—Numerical examples. *Geophysics*, 71(4), F101-F119. doi: **https://doi.org/10.1190/1.2213358**

Anemangely, M., Ramezanzadeh, A., Amiri, H., & Hoseinpour, S.-A. (2019). Machine learning technique for the prediction of shear wave velocity using petrophysical logs. *Journal of Petroleum Science and Engineering*, 174, 306-327. doi: **https://doi.org/10.1016/j.petrol.2018.11.032**

Aquino, A. (2011). *Inversión conjunta de registros de pozos para la evaluación petrofísica en formaciones areno-arcillosas anisótropas.* [Tesis de doctorado]. Instituto Mexicano del Petróleo.

Aquino-López, A., Mousatov, A., and Markov, M. (2011). Model of sand formations for joint simulation of elastic moduli and electrical conductivity. *Journal of Geophysics and Engineering*, 8(4), 568-578. doi: **https://doi.org/10.1088/1742-2132/8/4/009**

Aquino-López, A., Mousatov, A., Markov, M., and Kazatchenko, E. (2015). Modeling and inversion of elastic wave velocities and electrical conductivity in clastic formations with structural and dispersed shales. *Journal of Applied Geophysics*, 116, 28-42. doi: **https://doi.org/10.1016/j.jappgeo.2015.02.013**

Archie, G. E. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Transactions of the AIME*, 146(01), 54-62. doi: **https://doi.org/10.2118/942054-G**

Bhatt, A., and Helle, H. B. (2002). Determination of facies from well logs using modular neural networks. *Petroleum Geoscience*, 8(3), 217-228. doi: **https://doi.org/10.1144/petgeo.8.3.217**

Bishop, C. M., 2006. *Pattern Recognition and Machine Learning*. Springer, New York, NY.

Bjørlykke, K. and Jahren, J. (2010). Sandstones and sandstone reservoirs. En A. Per Avseth, Jan Inge Faleide, Roy H. Gabrielsen, Nils-Martin Hanken, Kaare Høeg, Jens Jahren, Martin Landrø, Nazmul Haque Mondol, Jenø Nagy and Jesper Kresten Nielsen (Eds.), Petroleum Geoscience: From Sedimentary Environments to Rock Physics. (pp. 113-140). Springer, Berlin, Heidelberg. doi: **https://doi.org/10.1007/978-3-642-02332-3_4**

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers*. [Conference Paper]. Proceedings of the fifth annual workshop on Computational learning theory - COLT '92, 144-152. doi: **https://doi.org/10.1145/130385.130401**

Bukar, I., Adamu, M. B., & Hassan, U. (2019). *A machine learning approach to shear sonic log prediction*. [Conference Paper]. SPE Nigeria Annual International Conference and Exhibition. doi: **https://doi.org/10.2118/198764-MS**

Bust, V. K., Majid, A. A., Oletu, J. U., and Worthington, P. F. (2013). The petrophysics of shale gas reservoirs: Technical challenges and pragmatic solutions. *Petroleum Geoscience*, 19(2), 91-103. doi: **https://doi.org/10.1144/petgeo2012-031**

Chicco, D., Warrens, M. J., and Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. doi: **https://doi.org/10.7717/peerj-cs.623**

Clavier, C., Coates, G., & Dumanoir, J. (1984). Theoretical and experimental bases for the dual-water model for interpretation of shaly sands. *Society of Petroleum Engineers Journal*, 24(02), 153-168. doi: **https://doi.org/10.2118/6859-PA**

Friedl, M. A., and Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3), 399-409. doi: **https://doi.org/10.1016/S0034-4257(97)00049-7**

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow*. (2a ed.). O'Reilly Media, Inc.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42. doi: **https://doi.org/10.1007/s10994-006-6226-1**

Ghavami, F. (2011). Developing synthetic logs using artificial neural network: Application to Knox County in Kentucky.

Hajizadeh, A., AL Mudaliar, K., and Tewari, R. D. (2019). Designing a pragmatic solution for complex numerical modeling problem in thinly laminated reservoirs. *Journal of Petroleum Exploration and Production Technology*, 9(4), 2831-2844. doi: **https://doi.org/10.1007/s13202-019-0662-5**

Han, D. H. (1986). *Effects of porosity and clay content on acoustic properties of sandstones and unconsolidated sediments*. [Ph.D. Thesis]., Stanford University, Department of Geophysics, Stanford, USA.

Han, D. H., Nur, A., and Morgan, D. (1986). Effects of porosity and clay content on wave velocities in sandstones. *Geophysics*, 51(11), 2093-2107. doi: **https://doi.org/10.1190/1.1442062**

Heidari, Z., & Torres-Verdín, C. (2013). Inversion-based method for estimating total organic carbon and porosity and for diagnosing mineral constituents from multiple well logs in shale-gas formations. *Interpretation*, 1(1), T113-T123. doi: https://doi.org/10.1190/INT-2013-0014.1

Heidari, Z., and Torres-Verdín, C. (2014). Inversion-based detection of bed boundaries for petrophysical evaluation with well logs: Applications

to carbonate and organic-shale formations. *Interpretation*, 2(3), T129-T142. doi: **https://doi.org/10.1190/INT-2013-0172.1**

Hoerl, A. E., and Kennard, R. W. (1970). Ridge Regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67. doi: **https://doi.org/10.1080/00401706.1970.10488634**

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to statistical learning*. Springer New York. doi: **https://doi.org/10.1007/978-1-4614-7138-7**

Kazatchenko, E., Markov, M., and Mousatov, A. (2004). Joint inversion of acoustic and resistivity data for carbonate microstructure evaluation. *Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description*, 45(02).

Kazatchenko, E., Markov, M., Mousatov, A., & Pervago, E. (2007). Joint inversion of conventional well logs for evaluation of double-porosity carbonate formations. *Journal of Petroleum Science and Engineering*, 56(4), 252-266. doi: **https://doi.org/10.1016/j.petrol.2006.09.008**

Koza, J. R., Bennett, F. H., Andre, D., and Keane, M. A. (1996). *Automated design of both the topology and sizing of analog electrical circuits using genetic programming*. In: Gero, J.S., Sudweeks, F. (Eds.). Artificial Intelligence in Design '96 (pp. 151-170). Springer Netherlands. doi: **https://doi.org/10.1007/978-94-009-0279-4_9**

Markov, M., Levine, V., Mousatov, A., and Kazatchenko, E. (2005). Elastic properties of double-porosity rocks using the differential effective medium model. *Geophysical Prospecting*, 53(5), 733-754. doi: **https://doi.org/10.1111/j.1365-2478.2005.00498.x**

Meju, M. A. (1994). *Geophysical data analysis: understanding inverse problem theory and practice.* Society of Exploration Geophysicists. doi: **https://doi.org/10.1190/1.9781560802570**

Mezzatesta, A. G., and Méndez, E. R. (2006) *A Novel Approach to Numerical Integration of Conventional, Multi-Component Induction, and Magnetic Resonance Data in Thinly Bedded Sand-Shale Systems*. [Sesión de conferencia]. SPWLA Annual Logging Symposium 1, Old CSD Building, KDMIPE Campus, Kaulagarh Road, Dehradun, Uttarakhand, India.

Mezzatesta, A. G., Mollison, R. A., and Frost, E. (June, 2002). *Laminated Shaly Sand Reservoirs-An Interpretation Model Incorporating New Measurements*. [Presentación de paper]. In SPWLA Annual Logging Symposium, Oiso, Japan.

Minear, J. W. (September, 1982). *Clay models and acoustic velocities*. [Presentación de paper]. SPE Annual Technical Conference and Exhibition (SPE-11031). SPE. doi: **https://doi.org/10.2118/11031-MS**

Mitchell, T. (1996). Machine Learning (First). McGraw-Hill Education.

Mitchell, W. K., and Nelson, R. J. (June,1988). A practical approach to statistical log analysis. [Presentación de paper]. In SPWLA Annual Logging Symposium.

Müller, A. C., and Guido, S. (2016). Introduction to machine learning with Python (First edit). O'Reilly Media, Inc.

Nguyen-Sy, T., To, Q.-D., Vu, M.-N., Nguyen, T.-D., and Nguyen, T.-T. (2021). Predicting the electrical conductivity of brine-saturated rocks using machine learning methods. *Journal of Applied Geophysics*, 184, 104238. doi: **https://doi.org/10.1016/j.jappgeo.2020.104238**

Pérez-Rosales, C. (1982). On the relationship between formation resistivity factor and porosity. *Society of Petroleum Engineers Journal*, 22(04), 531-536. doi: **https://doi.org/10.2118/10546-PA**

Poupon, A., Loy, M. E., and Tixier, M. P. (1954). A contribution to electrical log interpretation in shaly sands. *Journal of petroleum Technology*, 6(06), 27-34. doi: **https://doi.org/10.2118/311-G**

Rasmussen, C. E. and Williams, C. K., (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT press.

Rolon, L., Mohaghegh, S. D., Ameri, S., Gaskari, R., and McDaniel, B. (2009). Using artificial neural networks to generate synthetic well logs. Journal of Natural Gas Science and Engineering, 1(4-5), 118-133. doi: **https://doi.org/10.1016/j.jngse.2009.08.003**

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229. doi: **https://doi.org/10.1147/rd.33.0210**

Shalev-Shwartz, S., and Ben-David, S. (2014). *Understanding machine learning*. Cambridge University Press. doi: **https://doi.org/10.1017/CBO9781107298019**

Shedid, S. A., and Saad, M. A. (2017). Comparison and sensitivity analysis of water saturation models in shaly sandstone reservoirs using well logging data. *Journal of Petroleum Science and Engineering*, 156, 536-545. doi: **https://doi.org/10.1016/j.petrol.2017.06.005**

Simandoux, P. (1963). Measures dielectrique en milieu poreux, application a mesure de saturation en eau, etude des massifs argileaux. *Revue de l'Inst. Francais du petrole*, 193-215.

Snieder, R., and Trampert, J. (1999). *Inverse problems in geophysics*. In: Wirgin, A. (Eds.). Wavefield Inversion. International Centre for Mechanical Sciences (pp. 119-190). Springer Vienna. doi: **https://doi.org/10.1007/978-3-7091-2486-4_3**

Thomas, E.C. and Stieber, S. J. (June, 1975). *The distribution of shale in sandstones and its effect on porosity*. [Sesión de conferencia]. Trans SPWLA 16th Annual Logging Symp.

Torres-Verdín, C., Alpak, F. O., and Habashy, T. M. (2006). Petrophysical inversion of borehole array-induction logs: Part II—Field data examples. *Geophysics*, 71(5), G261-G268. doi: **https://doi.org/10.1190/1.2335633**

Tosaya, C., and Nur, A. (1982). Effects of diagenesis and clays on compressional velocities in rocks. *Geophysical Research Letters*, 9(1), 5-8. doi: **https://doi.org/10.1029/GL009i001p00005**

Wang, J., and Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, 93, 41-56. doi: **https://doi.org/10.1016/j.energy.2015.08.045**

Waxman, M. H., and Smits, L. J. M. (1968). Electrical conductivities in oil-bearing shaly-sands. *Society of Petroleum Engineers Journal*,

8(02), 107-122. doi: **https://doi.org/10.2118/1863-A**

Wilson, M. D., and Pittman, E. D. (1977). Authigenic clays in sandstones: recognition and influence on reservoir properties and paleoenvironmental analysis. *Journal of Sedimentary Research*, 47(1), 3-31. doi: **https://doi.org/10.1306/212F70E5-2B24-11D7-8648000102C1865D**

Winsauer, W. O., Shearin Jr, H. M., Masson, P. Y., and Williams, M. (1952). Resistivity of brine-saturated sands in relation to pore geometry. *AAPG Bulletin*, 36(2), 253-277. doi: **https://doi.org/10.1306/3D9343F4-16B1-11D7-8645000102C1865D**

Wolpert, D. H., and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67-82. doi: **https://doi.org/10.1109/4235.585893**

Worthington, P. F. (2000). Recognition and evaluation of low-resistivity pay. *Petroleum geoscience*, 6(1), 77-92. doi: **https://doi.org/10.1144/petgeo.6.1.77**

Worthington, P. F. (2001). The influence of formation anisotropy upon resistivity-porosity relationships. *Petrophysics*, 42(02).

Worthington, P. F. (2011). The petrophysics of problematic reservoirs. *Journal of Petroleum Technology*, 63(12), 88-97. doi: **https://doi.org/10.2118/144688-JPT**

Wyllie, M. R. J. and Gregory, A. R. (1953). Formation factors of unconsolidated porous media: Influence of particle shape and effect of cementation. *Journal of petroleum technology*, 5(04), 103-110. doi: **https://doi.org/10.2118/223-G**

Wyllie, M. R. J., Gregory, A. R., and Gardner, G. H. F. (1958). An experimental investigation of factors affecting elastic wave velocities in porous media, *Geophysics*, 23(3), 459-493. doi: **https://doi.org/10.1190/1.1438493**

Zeng, L., Yi, S., Zhang, W., Feng, H., Lv, A., Zhao, W., Luo, Y., Wang, Q., and Lu, H. (2020). Provenance of loess deposits and stepwise expansion of the desert environment in NE China since ~1.2 Ma: Evidence from Nd-Sr isotopic composition and grain-size record. *Global and Planetary Change*, 185, 103087. doi: **https://doi.org/10.1016/j.gloplacha.2019.103087**

Zhang, Y. L., Bao, Z. D., Zhao, Y., Jiang, L., Zhou, Y. Q., and Gong, F. H. (2017). Origins of authigenic minerals and their impacts on reservoir quality of tight sandstones: Upper Triassic Chang-7 Member, Yanchang Formation, Ordos Basin, China. Australian Journal of Earth Sciences, 64(4), 519-536. doi: **https://doi.org/10.1080/08120099.2017.1318168**

Zhdanov, M. S. (2002). Geophysical inverse theory and regularization problems (V. 36). Elsevier.

Zwennes, J.W., (2017), *Shale distribution quantification in a sandstone reservoir using density porosity and neutron porosity log data* [M.S. Thesis]., University of Louisiana at Lafayette.