FULL LENGTH ARTICLE

Prediction of permeability and porosity from well log data using the nonparametric regression with multivariate analysis and neural network, Hassi R'Mel Field, Algeria

Baouche Rafik [a],*, Baddari Kamel [b]

[a] University of Boumerdes, Algeria
[b] Department of Geophysics, University of Bouira, Algeria

Abstract Most commonly, to estimate permeability, we can use values of porosity, pore size distribution, and water saturation from logging data and established correlations. One benefit of using wireline log data to estimate permeability is that it can provide a continuous permeability profile throughout a particular interval.

This study will focus on the evaluation of formation permeability for a sandstone reservoir in the reservoir formations of Hassi R'Mel Field Southern from well log data using the multivariate methods. In order to improve the permeability estimation in these reservoirs, several statistical regression techniques have already been tested in previous work to correlate permeability with different well logs. It has been shown that statistical regression for data correlation is quite promising. We propose a two-step approach to permeability prediction that utilizes non-parametric regression in conjunction with multivariate statistical analysis. First we classify the well log data into electrofacies types. A combination of principal component analysis, model-based cluster analysis and discriminant analysis is used to characterize and identify electrofacies types. Second, we apply non-parametric regression techniques to predict permeability using well logs within each electrofacies. Three non-parametric approaches are examined via alternating conditional expectations (ACE), generalized additive model (GAM) and neural networks (NNET) and the relative advantages and disadvantages are explored. The results are compared with three other approaches to permeability predictions that utilize data partitioning based on reservoir layering, lithofacies information and hydraulic flow units. An examination of the error rates associated with discriminant analysis for uncored wells indicates that data classification based on electrofacies characterization is more robust compared to other approaches.

* Corresponding author.
E-mail address: R_baouche@yahoo.fr (B. Rafik).

These methods are tested and compared at the heterogeneous reservoirs in Triassic formations of Hassi R'Mel. The results show that permeability prediction is improved by applying variable selection to non-parametric regression ACE while tree regression is unable to predict permeability.

In comparing the relative predictive performance of the three regression methods, the alternating conditional expectations with ACE method appears to outperform the other two methods.

1. Introduction

The knowledge of permeability is critical to developing an effective reservoir description. Permeability data can be obtained from well tests, cores or logs. Normally, using well log data to derive estimates of permeability is the lowest cost method. In the Hassi R'Mel area (Fig. 1), the principal reservoir is located in the deposits of Triassic age. This field of

Southern directed anticline structure is located at the Western end of the Triassic province of Algeria (Fig. 1). The principal studies carried out in this formation are mainly focalized in electrofacies characterization and non-parametric regression ACE techniques [7] and are proven to be successful in predicting permeability in highly heterogeneous sandy reservoirs.

The prediction of absolute permeability is a key element in reservoir descriptions and has a direct impact on, among others, [14], effective completion designs, successful water injection programs and more efficient reservoir management. Traditionally simple linear regressions of core porosity – permeability data or empirical formulae using various log responses (usually porosity, clay volume and water saturation) have been used to predict permeability. These empirical methods only apply locally and ignore the fact that there is no theoretical basis for a relationship between porosity and permeability. In addition the scatter of the data about the regression line is explicitly ignored and implicitly attributed to measurement error or lower order variability in reservoir characteristics. Amaefule et al. [1] state that a practical and theoretically correct methodology is proposed for identification of hydraulic flow units based on a modified Kozeny–Carman equation [6,5,2,12,1]. In doing so they make extensive use of core data to describe the complex variations in pore geometry and the identification of "flow zone indicators" (FZI) from hydraulic units, which are, in turn, implicitly linked too much improved permeability estimations [24].

This article investigates the usefulness of the mathematical concepts and the predictive capabilities of multivariate analysis [23] with a view to better estimation of facies using the tree regression methods and to improving permeability estimates for the Hassi R'Mel Field in Triassic Formations.

In this context, we utilize the stepwise algorithm method to further improve permeability estimates. Stepwise regression is a kind of stepwise algorithm used in linear regression. We first demonstrate the availability of the stepwise algorithm by applying stepwise regression [10,11] to Hassi R'Mel Southern (Triassic). A method combining the stepwise algorithm and ACE is proposed and applied to the Hassi R'Mel Field (Triassic). Results are compared with those from regression without variable selection.

Figure 1   Field location of Hassi R'Mel. Sands (yellow), limestone (blue), shale (green), dolomite (Magenta), Silts (light brown), Sebkha (dark brown).

2. Methodology

The applications of the results obtained during the facies analysis, starting from the available log data for the Triassic formations of Hassi R'Mel Southern made it possible, first, to define ten electrofacies (sandstone, shale, dolomite, evaporate, andesite and clay), in the wells of Triassic formations. Manual facies analysis of Triassic well log data was carried out in ten
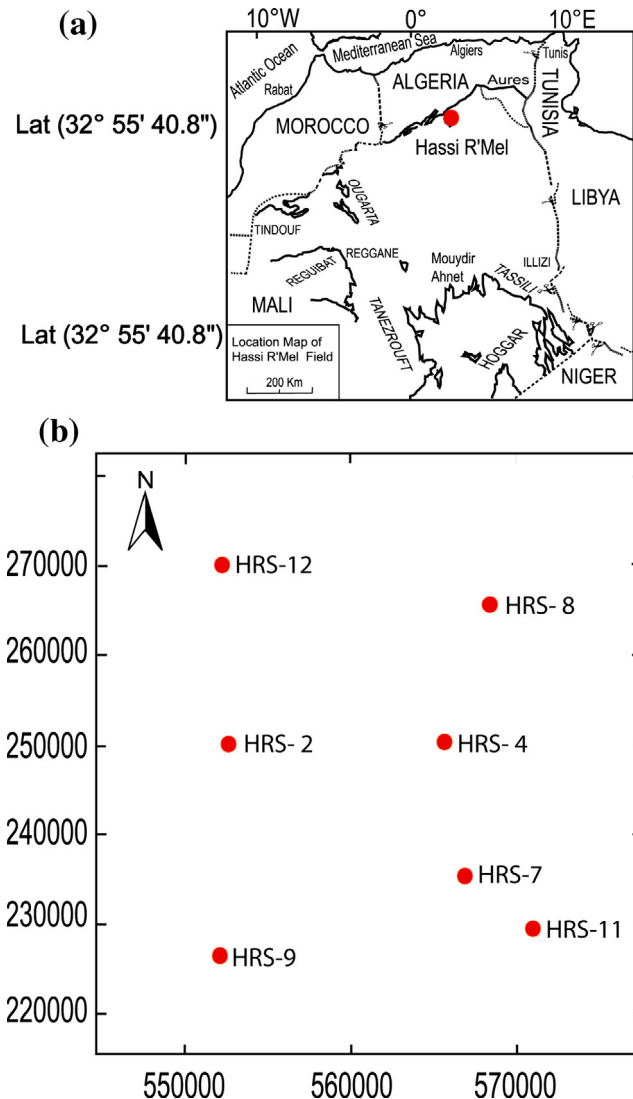
surveys [3]. This allowed us, in a first stage, to define ten lithologies. Next, after developing a model on a well using the "Petrolog" software, and checking it on other wells, a semi-automatic data processing was carried out on seven other wells. The facies analysis in the Hassi R'Mel Field is shown in Fig. 2.

This approach was developed in two stages for the prediction of permeability using the non-parametric regression in conjunction with multivariate statistical analysis [17]. In a first step, a classification of log data in many electrofacies types is made in accordance with the unique characteristics of well log data and measures reflecting minerals and lithofacies in
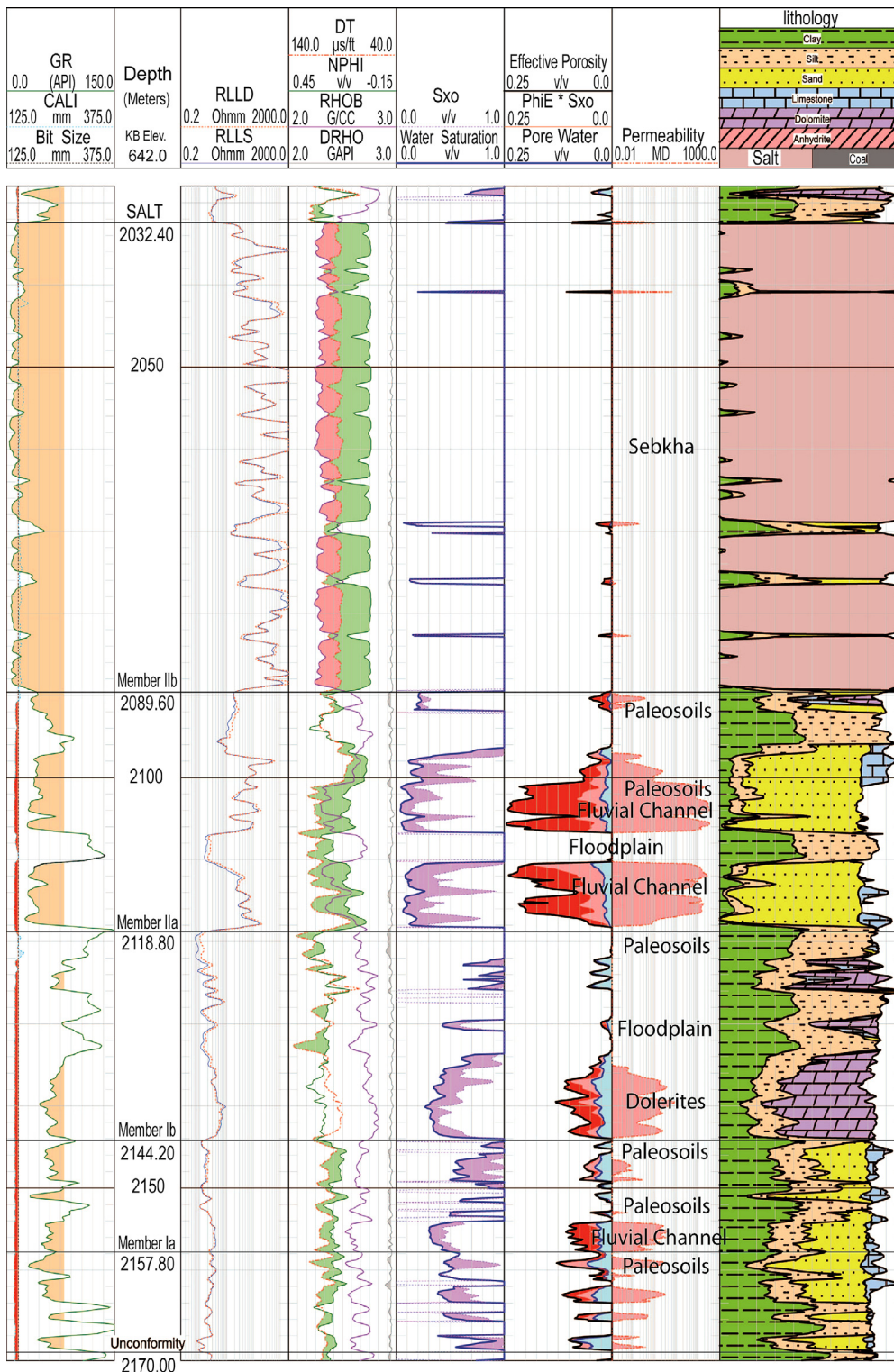

Figure 2   Facies analysis in Triassic of Hassi R'Mel [3].

the interval studied. A combination of principal component analysis, cluster analysis based on a model and discriminant analysis [16] is used to characterize and identify the types of electrofacies. Second, we apply nonparametric regression techniques to predict permeability using well logs in each electrofacies. Three non-parametric approaches are examined:ACE, GAM, and NNET and the relative advantages and disadvantages are explored.

3. Data available

The data presented in this analysis have been gathered from 5 wells. We left out two cored wells (HRS-7 and HRS-8) to verify our correlations using blind tests. The data presented in this analysis have been gathered from 7 wells. The consideration of the data quality and field-wide availability, a suite of well logs is selected for the analysis. In this field we have 7 well logs: caliper, gamma ray (GR), and three different resistivity (LLD, LLS, and MSFL), acoustic transit-time (DT), neutron (NPHI), and density (RHOB), and photoelectric logs (PEF). Among the 7 well logs, only 6 logs (GR, RLLD, DT, NPHI, RHOB, and SW) are chosen for characterizing the electrofacies groups.

After partitioning of well log responses into electrofacies groups, statistical regression techniques are applied to model the correlation between permeability and well log responses within the partitioned groups. In this study, three non-parametric techniques are examined using ACE, GAM, and NNET and their relative predictive performances are assessed. In neural network modeling, the 927 sample data set from 7 wells was divided into two subsets for training and supervising. The supervising data set is used for testing whether the neural net can generalize from the training data set. For each electrofacies group, we modeled the networks containing the optimal number of nodes in the hidden layer which produce the least mean square error for the supervising data set.

Table 1a and b compares regression errors for the three models used for developing the correlations. These errors are summarized in terms of mean squared error (MSE) and mean absolute error (MAE) during regression. The GAM appears to fit the data better compared to the other models.

4. Analyses

4.1. Electrofacies analysis

Recent methods used in electrofacies classification are mainly based on techniques for identifying groups from responses logging data with the same characteristics. (Fig. 3). For this purpose, the process results of principal component analysis can be shown in Table 2. First principal component (PC1) appears to indicate porosity (NPHI) of the formation while second principal component (PC2) shows a stronger correlation with density (RHOB) readings. The eigenvectors of the covariance matrix ($\sum$) provides coefficients of the principal component transformation.

4.1.1. Principal component analysis

Principal component analysis (PCA) is probably the oldest and best known of the techniques of multivariate analysis. It was first introduced by Pearson [19], and developed independently by [13]. Like many multivariate methods, it was not widely used until the advent of electronic computers, but it is now well entrenched in virtually every statistical computer package.

Principal components constitute an alternative form of displaying the data [15], thereby allowing better knowledge of its structure without changing the information. In addition, because the total variance in a data set can be defined as the sum of the variances associated with each principal component, the first few principal components that explain most of the variation in the original variables are often useful to reveal

Table 1a   Shows the Mean Square Errors (MSE) and Mean Absolute Errors (MAE) for porosity.

	GR (API)	ΔT (μs/ft)	Rhob (cc)	RLLD (Ω m)	NPhi (%)	SW (%)
MSE	1420.377	4485.003	177.4659	349071.5	230.4	224.171
MAE	33.329	66.021	11.2626	109.9	13.4	13.136
MRSE	0.496	0.696	34.1510	2.5	630535.9	4185.298
MRAE	0.681	0.829	4.8987	0.9	212.8	45.430
R^2	0.275	0.343	0.4992	−0.1	0.5	0.494

Erreurs quadratiques moyennes (MSE) et erreurs absolues moyennes (MAE) pour la porosité. MSE: Erreur quadratique moyenne (Mean square error); MAE: Erreur absolue moyenne (Mean absolute error).
Erreurs quadratiques moyennes (MSE) et erreurs absolues moyennes (MAE) pour la perméabilité. Mêmes notations qu'au Table 7a.

Table 1b   Shows the Mean Square Errors (MSE) and Mean Absolute Errors (MAE) for permeability. Same notations as Table 7a.

	GR (API)	ΔT (μs/ft)	Rhob (cc)	RLLD (Ω m)	NPhi (%)	SW (%)
MSE	193258.4	171866.0	223655.5	514328.3	225,150	225,010
MAE	317.8	289.0	355.7	415.0	358	358
MRSE	110.7	24.7	45558.6	5268.2	474,475,956	4,928,654
MRAE	7.5	3.5	156.8	32.1	4785	1295
R^2	0.1	0.4	−0.1	−0.1	−0	−0

Erreurs quadratiques moyennes (MSE) et erreurs absolues moyennes (MAE) pour la porosité. MSE: Erreur quadratique moyenne (Mean square error); MAE: Erreur absolue moyenne (Mean absolute error).
Erreurs quadratiques moyennes (MSE) et erreurs absolues moyennes (MAE) pour la perméabilité. Mêmes notations qu'au Table 7a.
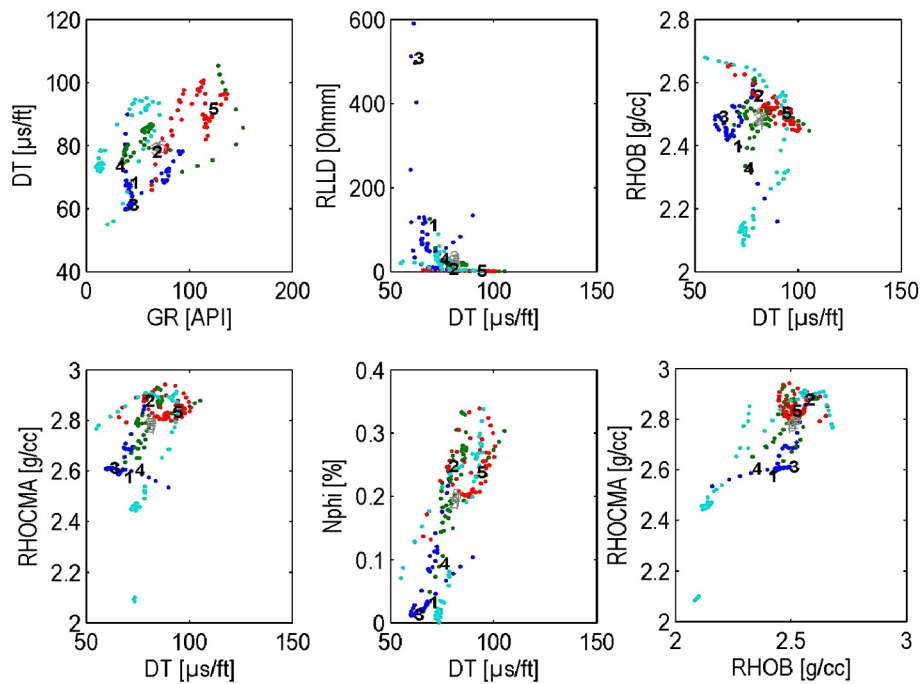
Figure 3 Electrofacies – Well Hassi R'Mel – Triassic. (Red) paleosols, (Blue) Sebkha, (Green) Floodplain, (Magenta) Fluvial, (Black) Dolerites.

Table 2 First principal component analysis.

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7
GR	0.810222	−0.221781	0.112247	−0.326747	0.254331	0.332007	0.008781
DT	0.928652	−0.085377	−0.060521	0.064263	0.149098	−0.298393	0.106090
RHOB	0.802908	0.415180	−0.075180	0.373197	0.141656	−0.053781	−0.122791
RLLD	−0.771005	0.191981	−0.209594	0.454785	0.309703	0.135997	0.059369
NPHI	−0.017782	−0.613081	0.664005	0.427526	0.008350	−0.002590	−0.007632
SW	0.431749	0.779061	0.310636	0.177623	−0.216682	0.163703	0.069734
CPOR	0.525146	−0.521029	−0.491840	0.356037	−0.239513	0.162650	0.016158
Eigen value	3.220466	3.220466	3.220466	3.220466	3.220466	3.220466	3.220466
Contribution (%)	46.0067	46.0067	46.0067	46.0067	46.0067	46.0067	46.0067
Total variance (%)	46.00666	46.00666	46.00666	46.00666	46.00666	46.00666	46.00666

the structure in the data. This can reduce the dimensionality of the problem and complexity in the cluster and discriminant analysis.

4.1.2. Cluster analysis

This method can provide better performance compared to traditional methods such as the nearest neighbor and k-mean clustering methods, which are often not capable of identifying the groups (sebkha, river channel, paleosols, floodplains and dolerite) that are either overlapping or of varying sizes and shapes.

Another advantage of the model-based approach is that there is an associated Bayesian criterion for assessing the model [22]. Bayesian criterion is associated with this method to evaluate the model [22] where the number of clusters without the subjective judgments necessary in other conventional cluster analysis techniques can be estimated.

4.1.3. Discriminant analysis

It is a statistical analysis to predict a categorical dependent variable (called a grouping variable) by one or more continuous or binary independent variables (called predictor variables). Discriminant analysis is used when groups are known a priori (unlike in cluster analysis). Each case must have a score on one or more quantitative predictor measures, and a score on a group measure. In simple terms, discriminant function analysis is classification the act of distributing things into groups, classes or categories of the same type.

This technique is based on the assumption that an individual sample arises [8] from one of g populations or groups $(\Pi_1, \ldots \Pi_g, \; g > 2)$. If each group is characterized by a group-specific probability density function $fc(\mathbf{x})$ and the prior probability of the group/c is known, then according to the Bayes theorem, the posterior distribution of the classes given the observation x is as follows (Eq. (1)):

$$p(c/x) = \frac{\pi_c p(x/c)}{p(x)} = \frac{\pi_c f_c(x)}{p(x)} \alpha \pi_c f_c(x) \tag{1}$$

4.1.4. Permeability correlation

The non-parametric regression techniques which do not require a priori assumptions about the functional forms to model the data in this case may be applied if electrofacies are known. In this case correlations between permeability and log responses can be established for each kind of electrofacies. There are generally three different nonparametric approaches that address this issue: GAM, ACE, and Nnet.

4.2. The Generalized additive models (GAM)

Regression model has the general form (Eq. (2)):

$$E\big(y/x_1, x_2, \ldots, x_p\big) = \alpha + \sum_{i=1}^{p} \varphi_1(x_l) + \varepsilon \tag{2}$$

where χ_L is the predictor and φ_L is function of predictors. Thus additive models replace the problem of estimating a function of a p-dimensional variable χ by one of estimating p separate one-dimensional functions, φ_L. Such models are attractive if they fit the data since they are far easier to interpret than a p-dimensional multivariate surface.

Algorithm: The response transformation models generalize the additive model by allowing transformation of the response variable y. The models have the following general form as (Eq. (3)):

$$\theta(y) = \alpha + \sum_{l=1}^{p} \phi_l(x) + \varepsilon \tag{3}$$

The ACE algorithm, originally proposed by Breiman and Friedman [4], provides a method for estimating optimal transformations for multiple regressions that result in a maximum correlation between a dependent (response) random variable and multiple independent (predictor) random variables. Such optimal transformations can be derived by minimizing the variance of a linear relationship between the transformed response variable and the sum of transformed predictor variables.

4.2.1. Feed-forward neural networks (NNET)

A feed forward neural network is an artificial neural network where connections between the units do *not* form a directed cycle. This is different from recurrent neural networks. The feed forward neural network was the first and simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes.

Usually, the feed forward neural net is arranged in multiple layers as shown in Fig. 7: Six input layers, two output layers and more hidden layers. Each layer contains a number of nodes (also called neuron's processing units) which are connected to each node in the preceding layer by simple weighted links. Except for nodes in the input layer, each node multiplies its specific input value by the corresponding weight and then sums up all the weighted inputs. Sometimes a constant (the 'bias' term) can be involved in the summation. The final output from the node is calculated by applying an activation function (transfer function) to the sum of the weighted inputs as shown in Eq. (4):

$$y_j = f_o\left(\alpha_j + \sum_{h=1}^{nh} w_{hj} f_h\left(\alpha_h + \sum_{i=1}^{ni} w_{ih} x_i\right)\right) \tag{4}$$

where **y** is the vector of output variables, **x** is the vector of input variables, 6 is the bias term, w_{ij} are the connection weights on the link from i to j node, nh is the number of hidden nodes, and ni is the number of input variables.

The original activation function proposed by McCulloch and Pitts [18] is a step function, the so called threshold function (with $f(x) = I$ $(x > 0)$). However, the artificial neuron model has been expanded to include other functions such as the sigmoid, piecewise linear, and Gaussian. In this study, a logistic function is used as the activation function, which gives values that range from 0 to 1 as (Eq. (5)):

$$f(x) = \frac{\exp(x)}{1 + \exp(x)} \tag{5}$$

A critical aspect in a neural network is the learning process of forcing a network to yield a particular output (response) for a specific input (signal). For multilayer feed forward neural networks, a more powerful supervised learning algorithm, called back propagation, can be employed to recursively adjust the connection weights so that the difference between the predicted and the observed outputs is as small as possible.

4.2.2. Permeability in uncored wells

Many of the Petrophysical models used for simulation studies are based on the classical approach of cross plotting the logarithms of permeability versus porosity (Fig. 8) and then, by fitting a regression line on this plot, predicting the permeability through the reservoir rock with correlation coefficient ($R^2 = 0.029$). This approach is critical when used to model permeable rocks, as it implies two misleading concepts. First, it considers the relationship between the logarithms of core permeability versus core porosity as linear. Secondly, using log porosities on this plot to predict the permeabilities would imply a scaling agreement between the macroscopic level (core plug) and the megascopic level (log data). Discretizing the reservoir into units such as layers and blocks, and assigning values of all pertinent physical properties to these blocks will give a better agreement with the reservoir heterogeneity.

4.3. Classification based on stratigraphical zones

The Saharan Triassic consisted of varied continental environments, namely, fluvial, flood plain, lake, Sebkha or Evaporate, and Eolian [3]. This Triassic is composed of three Formations: I, II and III. At the top of Formation I, the lower series of strata of the Triassic, there are intercalations of volcanic rocks represented by dolerites. Sedimentation interspersed with periods of no deposition is thought to have occurred during the Triassic continental; resulting in a deposition of ground-level strata and the development of a more or less intense pedogenesis. At the top of Formation II, the eruptive on the level of the lower member II_a at the base and constituted primarily of dolerites but the roof is primarily by a channel in a fine sandy filling. The member II_b is characterized by fluvial facies of channels prevailing and evolving to the top in a complex of playa or evaporates.

Formation III is characterized primarily by an evaporate facies of sebkha type halite.

4.4. Classification based on Lithofacies: Error rate from all classifications

The facies studies started with a core analysis, and was carried out on the Triassic formations of the Southern Hassi R' Mel. This required a core analysis in the laboratory of Sonatrach (Algiers), as well as the combination of the different well log data: Gamma Ray, neutron and sonic log, their fittings in sequences and their vertical evolution. The study of the sedimentary figures, representing the hydrodynamic and physical conditions of the sedimentation environment, was also taken into account.

Well log responses can be also partitioned using lithofacies that are usually determined by core descriptions and thin section analysis for all the wells (2, 4, 7, 8, 9, 11 and 12). The Field of Hassi R'Mel can be described by 10 different lithofacies types as shown in (Table 6).

The crossval function (Matlab 12a, 2012) can estimate the misclassification error for both LDA (Linear Discriminant Analysis) and QDA (Quadratic Discriminant Analysis) using the given data partition for the two wells HRS-7 and HRS-8, where the QDA has a slightly larger cross-validation error than LDA.

5. Results

5.1. Principal component analysis

In the scatter plot (Fig. 4), the relationship between reservoir properties and the 3 major principal components generated from the 5 well logs can be explored. First principal component (PC1) appears to indicate porosity of the formation while second principal component (PC2) shows a stronger correlation with gamma ray readings.

Fig. 5a shows the projection of the well log variables on the factors plane (1x2), where the contribution of the factor 1 is of 46.01% and 21.71% for the second factor.

For the PC1 and PC2 (Table 3a), with porosity, the components are given by,

$$PC1 = 0.3994GR - 0.2456LLD + 0.4688NPHI$$
$$- 0.4329RHOB - 0.5153DT - 0.1417SW$$
$$+ 0.3328CPOR$$

$$PC2 = 0.3426GR - 0.4122LLD + 0.1544NPHI$$
$$+ 0.3402RHOB + 0.0746DT + 0.6803SW$$
$$- 0.3156CPOR$$

Fig. 5b shows the projection of the well log variables on the factors plane (1×2), where the contribution of the factor 1 is of 48.22% and 35.61% for the second factor.

For the PC1 and PC2 (Table 3b), with permeability KH, the components are given by,

$$PC1 = 0.4818GR - 0.2642LLD + 0.3283NPHI$$
$$+ 0.4406RHOB - 0.1455DT + 0.5331SW + 0.3009KH$$

$$PC2 = 0.0930GR - 0.4067LLD + 0.4906NPHI$$
$$- 0.3862RHOB + 0.6197DT + 0.0280SW - 0.2259KH$$

Figure 4 Scatter plot of GR, DT, RLLD, RHOB, RHOCMA and NPHI, SW.

Table 3b Eigenvector spreadsheets of Hassi R'Mel Upper Triassic for core Permeability (**KH**)

Eigenvector spreadsheet Number of components is 2			
	Variable number	Component 1	Component 2
DT (μs/ft)	1	−0.145534	0.619764
GR (API)	2	0.481838	0.093089
LLD (Ω mm)	3	−0.264245	−0.406760
NPHI (cc)	4	0.328384	0.490641
RHOB (g/cc)	5	0.440670	−0.386256
SW (cc)	6	0.533144	0.028047
CPERM (mD)	7	−0.300927	0.225976

basis of a measure of similarity or dissimilarity between the groups (Fig. 6a, b). In this study, the grouping based on a model, a hierarchical clustering technique by agglomeration is used.

5.3. Discriminant analysis

The observations should be allocated into the group with the maximal posterior probability $p(c/\mathbf{x})$. The discriminant analysis used in this study requires prior classification (Table 4) of the data into relatively homogenous subgroups whose characteristics can be described by the statistical distributions [9] of the grouping variables associated with each subgroup.

5.4. Permeability correlation

The optimal transformations are derived solely based on the data sets and can be shown to result in a maximum correlation in the transformed space. The transformations do not require a priori assumptions of any functional form for the response or predictor variables and thus, provide a powerful tool for exploratory data analysis and correlation. The correlations of vectors in matrix representing the predictor values (well logs) and the predicted value (core permeability KH or CPerm and core porosity CPor) are summarized in Table 5a and b.

5.5. Feed forward neural network

In this study, we used a typical feed forward model that consists of one input layer, one hidden layer and one output layer. In addition, the optimum number of hidden nodes is determined by trial and error.

This technique is applied to highly shaly sand reservoirs in the Triassic formations of Hassi R'Mel, Sahara Algeria. Conventional multiple regressions did not provide satisfactory permeability and porosity predictions in this field. The superior predictive ability of our proposed approach is verified using blind tests. We also, compared our results with the neural network method for permeability and porosity prediction based lithofacies characterization and the results are better (Fig. 7).

5.6. Permeability in uncored wells

Some flow units in Well HR-7 and the permeability porosity relationship constructed for the unique HFUs are illustrated in Figs. 8 and 9. The samples in the scatter plot are grouped
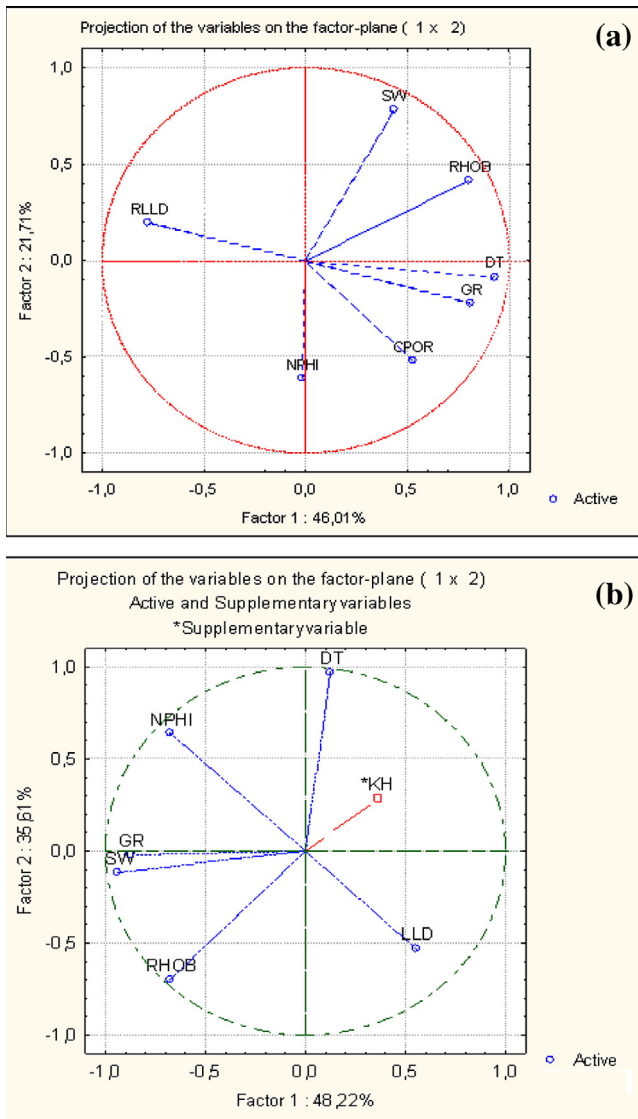

Figure 5 (a) Projection of the variables on the factor-plane (1 × 2) with porosity CPor. (b) Projection of the variables on the factor-plane (1 × 2) with permeability KH.

5.2. Cluster analysis

The goal of the cluster analysis is to establish classes of data groups, homogeneous and isolated from the outside on the

Table 3a Eigenvector spreadsheets of Hassi R'Mel Upper Triassic for core Porosity (**CPor**).

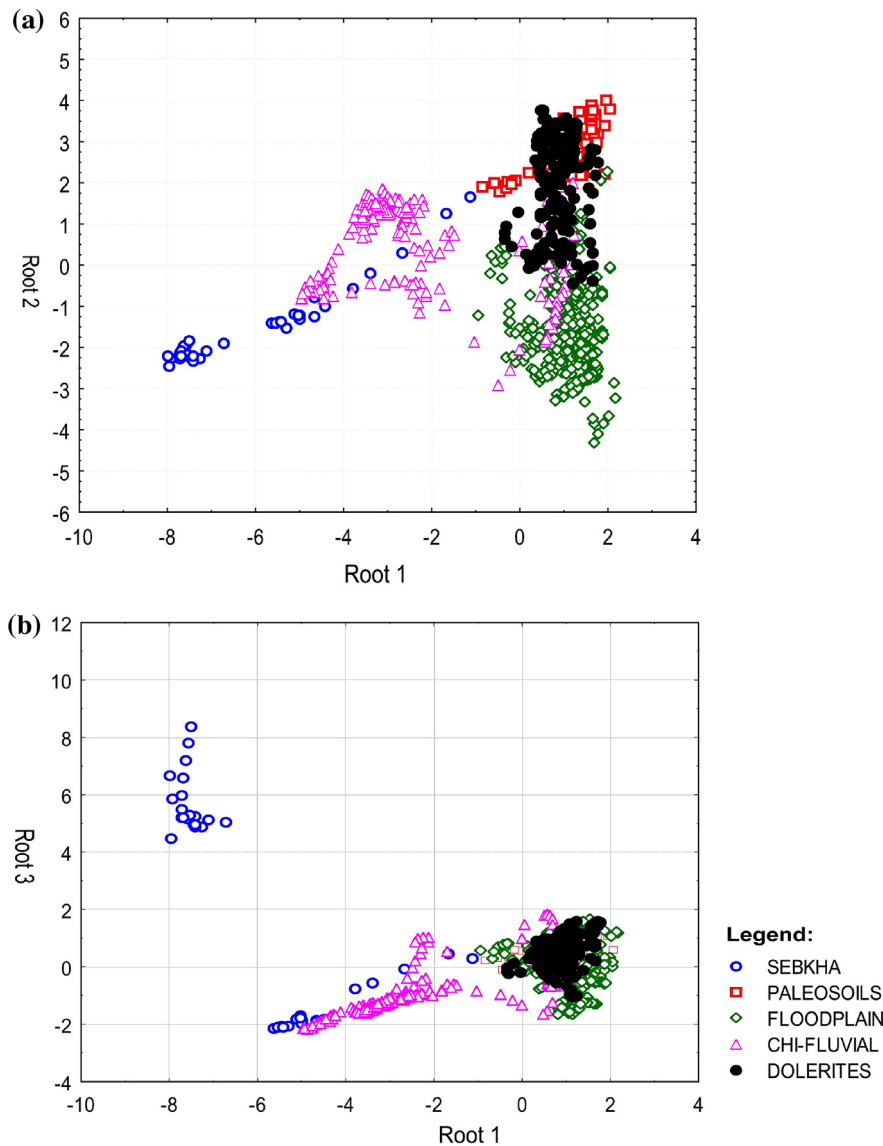| Eigenvector spreadsheet Number of components is 2			
	Variable number	Component 1	Component 2
GR (API)	1	0.399453	0.342616
DT (μs/ft)	2	0.491891	0.074640
RHOB (g/cc)	3	−0.432947	0.340221
RLLD (Ω mm)	4	−0.245687	−0.418269
NPHI (cc)	5	0.468838	0.154481
SW (cc)	6	−0.141704	0.680331
CPOR (cc)	7	0.332846	−0.315642

Figure 6 (a) Cluster analysis of facies environments (Root 1 vs Root 2). (b) Cluster analysis of facies environments (Root 1 vs Root 3).

by HFUs after calculating FZI in uncored well using Eqs. (6)–(11) from well log data, permeability can be determined for each HFU (mean FZI value) using Eq. (12).

A result of calculated permeability versus core permeability with depth in uncored wells is shown in Fig. 10. These predicted permeability profiles were obtained by assuming that this well had been logged only and did not have core data, while in reality core data were available at well. This was done only to check how accurately the HFU method would predict permeability in these well if they had not been cored. As shown, the permeability profiles of the log-derived HFU agree with core data.

Table 4 Classification matrix rows: observed classifications columns: predicted facies classifications. Well: (HRS-7).

		ldaCVerr	qdaCVerr	Perf	Tr_Perf	Val_Perf	Test_Perf
HRS-7	HFUs	0.4052	0.4125	0.7425	0.5266	1.0002	0.9255
	LITHO	0.4444	0.3919	0.7535	0.6496	1.0285	0.9622
	Facies	0.4450	0.3871	0.7337	0.6894	0.9048	0.7693
HRS-8	HFUs	0.4129	0.4135	0.7140	0.4828	0.7797	0.8358
	LITHO	0.4444	0.3919	0.7535	0.6496	1.0285	0.9622
	Facies	0.4189	0.3221	0.1954	0.2181	0.1415	0.1439

5.7. Classification based stratigraphy

The estimation of the permeability by application of the correlation equation established in each area can be applied directly to the log responses in corresponding predefined area without the use of discriminant analysis. All the Permeability and porosity predictions based on the zonal classification for the well HRS-7 as: KH (core permeability) can be shown in Fig. 10: KH-NN (Neural permeability) and PERMC (FHU permeability). The same results were obtained for the well HRS-8 (Figs. 10 and 11).

5.8. Classification based Lithofacies

The Means with error plot of core porosity and core permeability, associated with the lithofacies can be shown in Fig. 12a and b. An examination of this plot will show that variables evaporate, shale, shale dolomite, shale sand and low sand shale are the least influential in determining the first principal component while sands plays the most significant important role. The predictive performance of different classifiers is defined as an error rate which represents the fraction of training or a test set that have been misclassified. The error rates

Table 5a Correlations of vectors in matrix representing the predictor values (well logs) and the predicted value (core permeability KH (CPERM).

	DT (µs/ft)	GR (API)	RLLD (Ω mm)	NPHI (cc)	RHOB(g/cc)	SW (cc)	KH (mD)
DT	1.000000	−0.089607	−0.339964	0.571517	−0.758730	−0.224298	0.309288
GR	−0.089607	1.000000	−0.330645	0.580730	0.563389	0.731145	−0.256745
RLLD	−0.339964	−0.330645	1.000000	−0.530421	−0.011256	−0.457333	0.047765
NPHI	0.571517	0.580730	−0.530421	1.000000	0.046107	0.554921	−0.090472
RHOB	−0.758730	0.563389	−0.011256	0.046107	1.000000	0.708240	−0.495770
SW	−0.224298	0.731145	−0.457333	0.554921	0.708240	1.000000	−0.380541
KH	0.309288	−0.256745	0.047765	−0.090472	−0.495770	−0.380541	1.000000

Table 5b Correlations of vectors in matrix representing the predictor values (well logs) and the predicted value core porosity (**CPor**).

|	GR (API)	DT (µs/ft)	RHOB (g/cc)	(Ω mm)	NPHI (cc)	SW (cc)	CPOR (%)
GR	1.000000	0.619490	0.138185	−0.313039	0.153794	0.150801	0.275141
DT	0.619490	1.000000	0.075433	−0.327866	0.102853	0.084975	0.342660
RHOB	0.138185	0.075433	1.000000	0.002034	0.998923	0.999061	0.499198
RLLD	−0.313039	−0.327866	0.002034	1.000000	−0.006388	−0.014186	−0.120998
NPHI	0.153794	0.102853	0.998923	−0.006388	1.000000	0.998692	0.515081
SW	0.150801	0.084975	0.999061	−0.014186	0.998692	1.000000	0.493719
CPOR	0.275141	0.342660	0.499198	−0.120998	0.515081	0.493719	1.000000
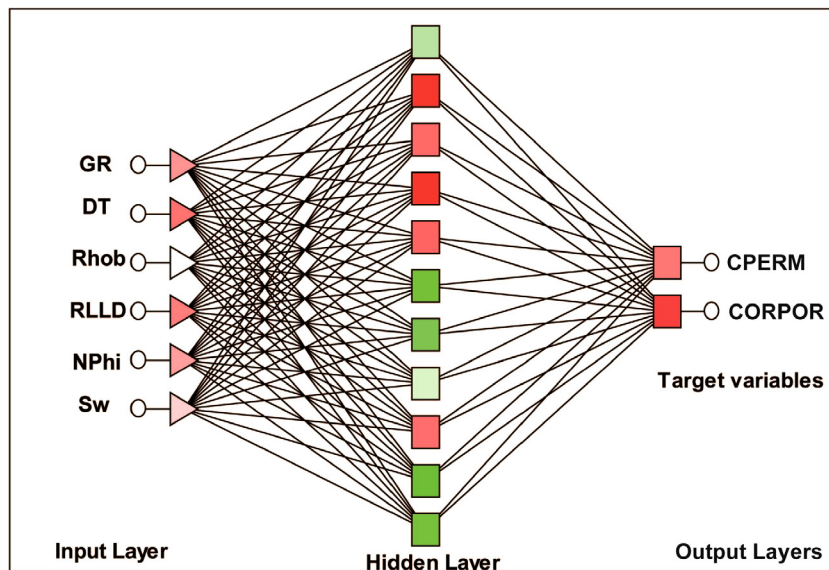
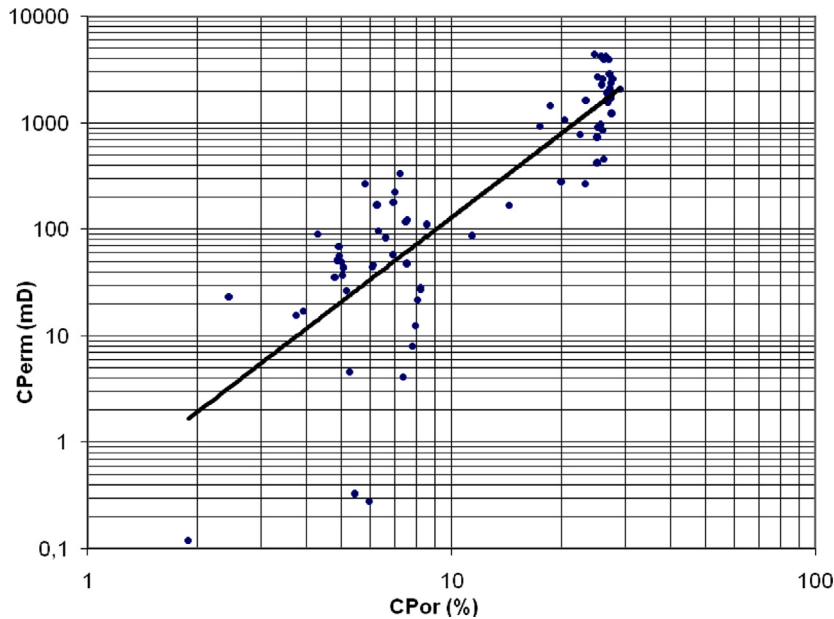**Figure 7** Feed forward neural net in multiple layers.

Figure 8 Core porosity/core permeability vs. Depth, HRS-7, in triassic formations of HRS. ($R^2 = 0.717$ and $Y = 0.314X^{2.602}$).

valuated for this context are summarized in (Tables 7 and 6) for the wells HRS-7 and HRS-8. It shows that a simpler model may get comparable or better performance than a more complicated model of other classifications (Table 7). The error rate associated with the classification of HFUs is the highest among all techniques considered in this study.

The quadratic discriminant analysis (QDA) for the data used shows the computed "ldaCVerr" and the "qdaCVerr" errors for linear and quadratic discriminant analysis and repre-

sents the expected prediction error on an independent set. To this end, the discriminant analysis based on lithofacies data performs worse than classifiers based on the electrofacies information.

Among the three non-parametric methods, the ACE model appears to be the best models in terms of their predictive of porosity and permeability. The Neural Network model tends to overestimate or underestimate in some intervals of the formation. Especially the poor predictive performance in the blind well HRS-8 indicates that a trained neural network will not necessarily provide accurate predictions for the data that were not used during training. Therefore, the GAM model produces good results for most intervals in the blind wells.

6. Comparison for other techniques for permeability prediction

The Hydraulic Unit concept [1] was selected for subdividing the reservoir into distinct petrophysical types. Each distinct reservoir type has a unique Flow Zone Indicator (FZI) value.

According to [20], a hydraulic flow unit is a continuous body over a specific reservoir volume that practically possesses consistent petrophysical and fluid properties, which uniquely characterize its static and dynamic communication with the wellbore.

This technique is based on a modified Kozeny–Carman [5] (cited in Amaefule et al. [1]) and the concept of mean hydraulic radius (Eq. (6)):

$$k = \left\{ \frac{1}{2\tau^2 S_g^2} \right\} \left(\frac{\varphi_e^3}{(1 - \varphi_e)^2} \right) \qquad (6)$$

S_{gv}: may also be defining as the surface area of grains exposed to fluid per unit volume of solid material.

Flow zone indicator depends on geological characteristics of the material and various pore geometry of a rock mass; hence, it is a good parameter for determining hydraulic flow
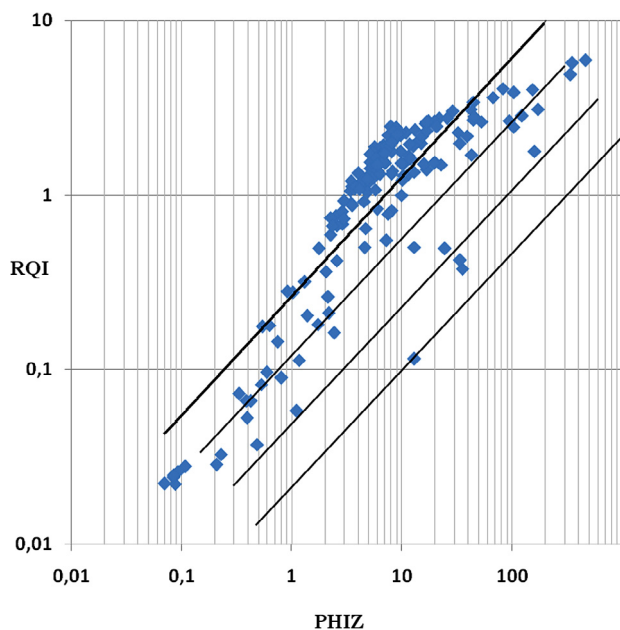

Figure 9 Scatter plot of log RQI versus log PHIZ when one HFU is used to fit the whole data. Well: HRS-7 ($R^2 = 0.738$ and $Y = 0.262X^{0.684}$).

units (HFUs). Flow zone indicator is a function of reservoir quality index and void ratio.

Amaefule et al. [1] addressed the variability of Kozeny's constant by dividing Eq. (1) by the effective porosity, φ_e and taking the logarithm:

Defining the flow zone indictor FZI (µm) as (Eq. (7)):

$$FZI = \frac{1}{S_g \tau \sqrt{F_s}} \qquad (7)$$

Reservoir quality index RQI (µm) as (Eq. (8)):

$$RQI = 0.0314 \sqrt{\frac{K}{\varphi_e}} \qquad (8)$$

Normalized porosity φ_z (fraction) as (Eq. (9)):

$$\varphi_Z = \left(\frac{\varphi_e}{1 - \varphi_e} \right) \qquad (9)$$

Therefore (Eq. (10)) becomes:

$$RQI = FZI \times \varphi_Z \qquad (10)$$

Taking the logarithm of both sides of (Eq. (11)) yields:

$$LogRQI = LogFZI + Log\varphi_Z \qquad (11)$$

On a log–log plot of RQI versus φ_z, all samples with similar FZl values will lie on a straight line with unit slope, in Fig. 9, the value of the FZI constant can be determined from the intercept of the unit slope straight line at $\varphi_z = 1$. Samples that lie on the same straight line have similar pore throat attributes and, thereby, constitute hydraulic unit. The permeability of a sample point is then calculated from a pertinent HFU using the mean FZI value and the corresponding sample porosity using the following (Eq. (12)):

$$K = 1014 \times FZI^2 \frac{\varphi_e}{(1 - \varphi_e)^2} \qquad (12)$$

However, it is worth-mentioning that given the true porosity and true HFUs (based on core data), the predicted permeability shows almost perfect agreement with the true permeability. So, the principal difficulty appears to be the identification of hydraulic flow units in uncored wells. In Table 7,
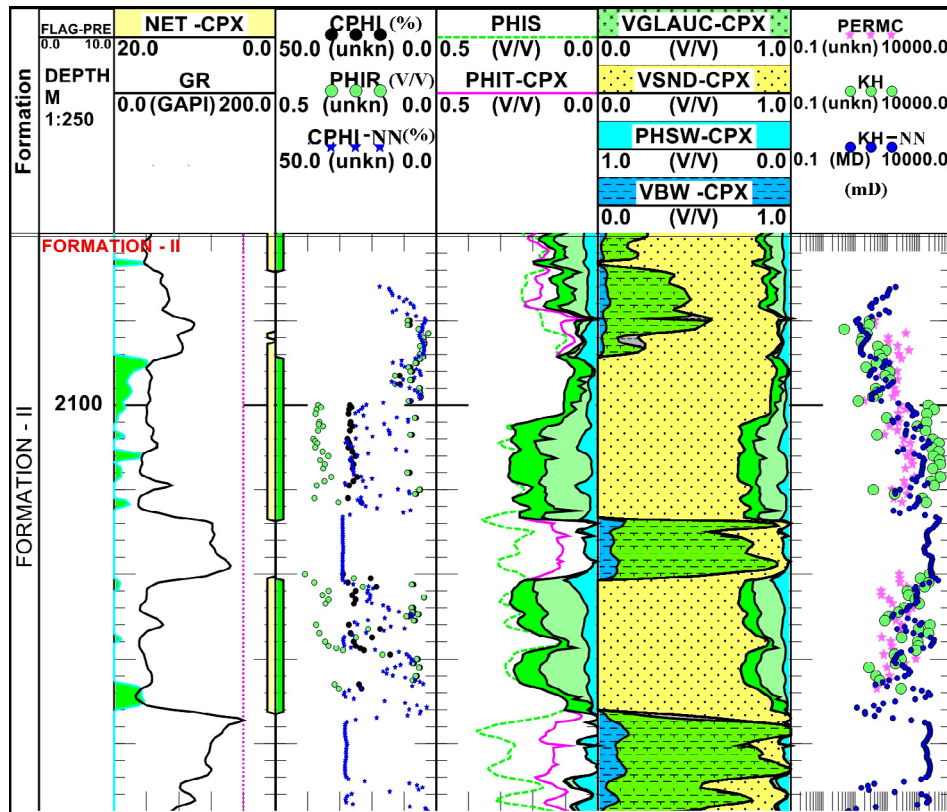

Figure 10 Permeability and porosity predictions based on the zonal classification for the well HRS-7 as: KH (core permeability). Legend: Composite log of well log predictions for HR-167. U1, U2, M1, M2, L1, L2 and L3: in-situ formations; GR [API]: gamma ray; Rhob [g/cm³]: density; Nphi [%]: neutron porosity; ΔT [ms/ft]: sonic log; RT [Ω m]: true resistivity, CPerm [mD]: permeability; subscript nn relates to predicted values from neural network. The horizontal arrows represent the levels of samples used in the training. Fuzzy logic Bin number distributions within the Hassi R'Mel Triassic reservoir for the associated (a) core porosity (CorPor or CPor) and (b) core permeability (CPerm); gamma ray (GR), sonic (ΔT), neutron porosity (Nphi), density (Rhob), water saturation (Sw) and true resistivity (RT); Bin number: predicted values. For notations see Appendix A and Nomenclature. VGLAUC-CPX: Volume of Glauconite in complex lithology, VSND–CPX: Volume of Sandstones in complex lithology, PHSW-CPX: Water volume in complex lithology, VBW-CPX: Volume of Bulk Water in complex lithology.
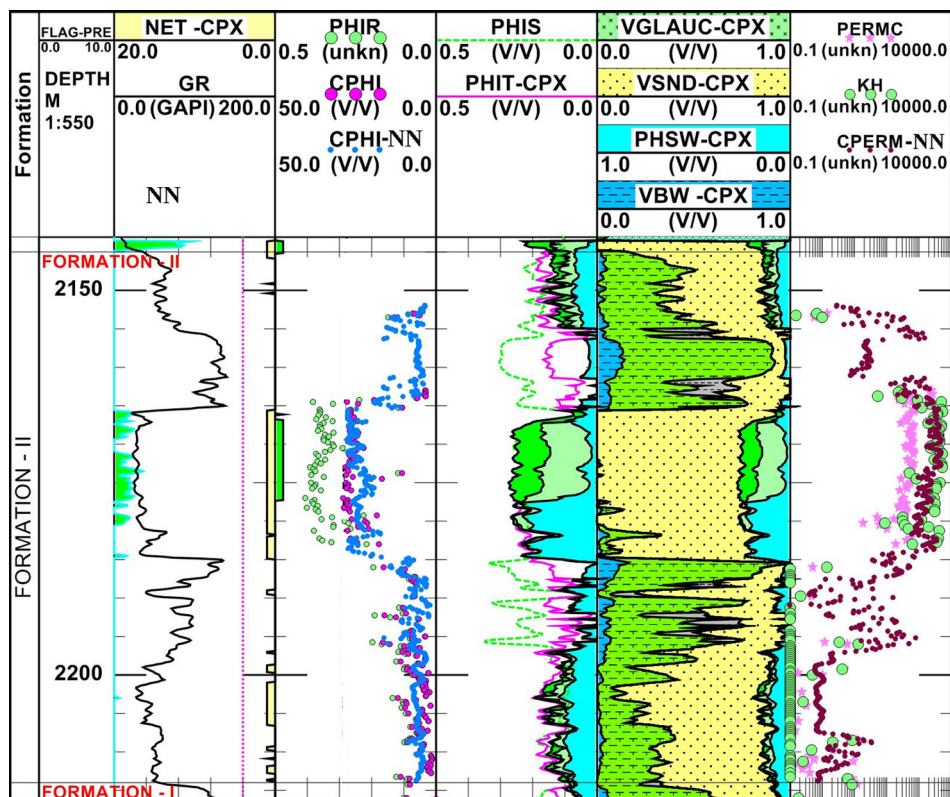
Figure 11 Permeability and porosity predictions based on the zonal classification for the well HRS-8 as: KH (core permeability). Legend: Composite log of well log predictions for HR-167. U1, U2, M1, M2, L1, L2 and L3: in-situ formations; GR [API]: gamma ray; Rhob [g/cm³]: density; Nphi [%]: neutron porosity; ΔT [ms/ft]: sonic log; RT [Ω m]: true resistivity, CPerm [mD]: permeability; subscript nn relates to predicted values from neural network. The horizontal arrows represent the levels of samples used in the training. Fuzzy logic Bin number distributions within the Hassi R'Mel Triassic reservoir for the associated (a) core porosity (CorPor or CPor) and (b) core permeability (CPerm); gamma ray (GR), sonic (ΔT), neutron porosity (Nphi), density (Rhob), water saturation (Sw) and true resistivity (RT); Bin number: predicted values. For notations see Appendix A and Nomenclature. VGLAUC-CPX: Volume of Glauconite in complex lithology, VSND–CPX: Volume of Sandstones in complex lithology, PHSW-CPX: Water volume in complex lithology, VBW-CPX: Volume of Bulk Water in complex lithology.

the predictive performances of the four approaches are compared using linear and quadratic errors for both the two wells HRS-7 and HRS-8.

All the results of Permeability and porosity prediction are plotted in the Fig. 12a and b where the three methods used in this work shows the well log data correlation with porosity and permeability calculated from the HFU (PERMC) and others porosity/permeability as (KH-NN) from FFNET with core data (KH).

7. Discussion and conclusion

The use of nonparametric regression methods in the Electrofacies Characterization study allows more significant results in the model clay sandstone oil reservoir. Therefore,

1. The predicted permeability profiles obtained from the data recorded in the depth, by assuming that this well had been logged only and did not have core data, while in reality core data were available at well. An examination of the error rates associated with Discriminant Analysis for core porosity and core Permeability showing the summary Goodness of Fit for uncored wells indicates that data classification based on electrofacies characterization look more robust compared to other methods.
2. The simplest approach uses flow zones or reservoir layering. Other approaches have used lithofacies information identified from cores, electrofacies derived from well logs, and the concept of HFUs.
3. Non-parametric regression offer better technical means to better approximate the permeability in the shaly sand reservoirs, even in intervals where there is no relationship between permeability and logs.
4. The best performance of permeability prediction result of the discrimination ability of the pattern recognition and reliability correlation models and this has, in advance; explained the methods for predicting porosity and permeability calculated from partitioning and correlation of data. The identification of errors from HFU and lithofacies can thus lead to incorrect results in the predictions of porosity and permeability.

To improve further the permeability and porosity predictions,

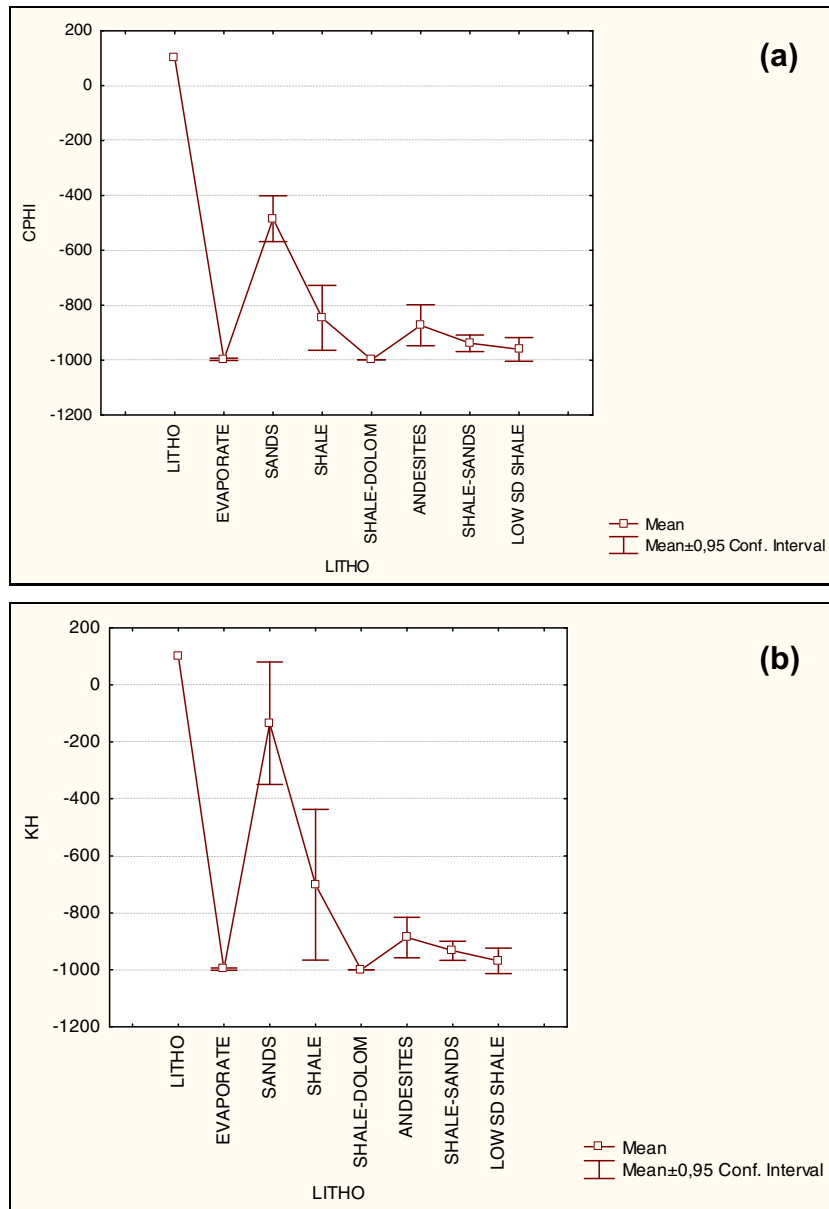- The results of treatment should be checked carefully.

Figure 12   (a) Means with error plot of porosity versus lithology. (b) Means with error plot of permeability versus lithology.

Table 6   Facies description represented by 10 different litho-facies types in Hassi R'Mel Triassic.

Facies number (Facino)	Facies description
1.	Halite
2.	Shale halite
3.	Sand
4.	Low shale sand
5.	Shale sand
6.	Dolomitic shale
7.	Dolomitic shale sand
8.	Low dolomitic shale sand
9.	Sandy shale
10.	Dolerites

- The number of the Hydraulic flow units must be optimum in the current study.
- Calculations FZI and HFU settings must first be made.
- We need to ensure that a better quality of permeabilities is obtained practically and good effect compaction corrections have been introduced into the petrophysical analysis stage.

Acknowledgements

Table 7 Predictive performances of different classifiers are compared using linear and quadratic errors for the two wells HRS-7 and HRS-8.

		ldaCVerr	qdaCVerr	Perf	Tr_Perf	Val_Perf	Test_Perf
HRS-7	HFU	0.4052	0.4125	0.7425	0.5266	1.0002	0.9255
	LITHO	0.4444	0.3919	0.7535	0.6496	1.0285	0.9622
	Facies	0.4450	0.3871	0.7337	0.6894	0.9048	0.7693
HRS-8	HFU	0.4129	0.4135	0.7140	0.4828	0.7797	0.8358
	LITHO	0.4444	0.3919	0.7535	0.6496	1.0285	0.9622
	Facies	0.4189	0.3221	0.1954	0.2181	0.1415	0.1439

Appendix A

K: is the permeability (μm^2),

φ_e: is the effective porosity (fraction),

FZI: is the Flow Zone Indicator (μm),

ACE: Alternating Conditional Expectation,

GAM: General Additive Models,

DT: Sonic travel time (Delta Time, $\mu sec/ft$),

NNET: Neural Network,

PHIT: Total Effective Porosity

RHOCMA: Matrix Density

SUM: Summation

LdaCVerr: Linear discriminant CV error

QdaCVerr: Quadratic discriminant CV error

Perf: Performance

Tr_Perf: Train Performance

Val_Perf: Validation Performance

Test_Perf: Test Performance

SS: Sandstones

LS: Limestones

Dol: Dolomite

Dt_f = Fluid transit time ($\mu s/ft$)

D_t = Transit time ($\mu s/ft$)

ρ_b = Bulk density (g/cc)

ρ_f = Fluid density (g/cc)

Φ_{Nf} = Neutron fluid (l.p.u.%))

Φ_N = Neutron (l.p.u.%))

Root1 = Principal Composant Axe 1 (PC1)

Root2 = Principal Composant Axe 2 (PC2)

Root3 = Principal Composant Axe 3 (PC3)

Dt_{MAA} = Apparent matrix transit time – including shale ($\mu s/ft$)

ρ_{MAA} = Apparent matrix density – including shale (g/cc)

Φ_{TAA} = Apparent total porosity – including shale (%)

CorPor (Cpor): Core Porosity (%)

CPerm: Core Permeability (mD)

KH: Core Permeability (mD)

PHIR: Porosity obtained from FHU (%)

NET-CPX: Net Volume in complex lithology

CPHI-NN: Core porosity neural network (%)

PHIS: Porosity sonic (%)

PHIT-CPX: Total porosity in complex lithology (%)

PERMC: Permeability obtained from FHU (mD)

KH-NN: Core permeability neural network (mD)

VGLAUC-CPX: Volume of Glauconite in complex lithology

VSND–CPX: Volume of Sandstones in complex lithology

PHSW-CPX: Water volume in complex lithology

VBW-CPX: Volume of Bulk Water in complex lithology

X, Y: Coordinates (UTM)

HRS: Hassi R'Mel Southern

K: Permeability (mD)

K_v: Vertical permeability (mD)

K_h: Horizontal permeability (mD)

Φ: Porosity (%)

V_{sh}: Shale fraction (%)

S_{wi}: Initial water saturation (%)

μ: Viscosity (cp)

V_{ap}: Apparent velocity (m/s)

V_{ac}: Actual velocity (m/s)

P: Pressure (psi)

L: Length

DLL: Dual Laterolog (Ω m)

LDT: Litho density tool (Pe)

Rhob: Bulk density (g/cc)

CNL: Compensated neutron log (%)

GR, SGR: Gamma ray (API)

SGR: Spectrometry Gamma ray (API)

MSFL: Micro Spherically Focused Log (Ω m)

BHC: Borehole compensated ($\mu s/ft$)

CAL: Caliper (in.)

nnt: Neural network

mlr: Multi linear regression

MSE: Mean square error

MAE: Mean absolute error

MRSE: Mean root Square error

MRAE: Mean root absolute error

R^2: correlation coefficient

References

[1] J.O. Amaefule, M. Altunbay, D. Taib, D.G. Kersey, D.K. Keelan, Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells, Paper SPE 26436 Prepared for Presentation at the 68th Annual Technical Conference and Exhibiton of SPE Held in Houston, Texas, 3–6 October, 1993.

[2] J.W. Amyx, D.M. Bass Jr., R.L. Whiting, Petroleum Reservoir Engineering, McGraw-Hill Book Co, New York City, 1960.

[3] R. Baouche, A. Nedjari, Africa Geosci. Rev. 17 (2) (2010) 151–219.

[4] L. Breiman, J. Friedman, D. Pregibon, Y. Vardi, A. Buja, R. Kass, E. Fowlkes, J. Kettenring, J. Am. Stat. Assoc. 80 (391) (1985) 580–619.

[5] P.C. Carman, Flow of Gases Through Porous Media, Academic Press Inc, New York City, 1956.

[6] P.C. Carmen, Trans. AIChE. 15 (1937) 150–166.

[7] A. Datta-Gupta, G. Xue, S.H. Lee, Nonparametric Transformations for Data Correlation and Integration: From Theory to Practice, in: Reservoir Characterization: Recent Advances, AAPG Datapages, 1999.

[8] J. Friedman, A. Barron, X. Xiao, G. Golubev, R. Hasminskii, L. Breiman, A. Bujaz, D. Duffy, T. Hastie, R. Tischirani, Ann. Stat. 19 (1) (1991) 1–141.

[9] J. Friedman, C. Roosen, An introduction to multivariate adaptive regression splines, Stat. Methods Med. Res. 4 (3) (1995) 197–217.

[10] Guoping Xue, A. Datta-Gupta, Peter Valko, T. Balsingame, Optimal transformations for multiple regression: application to permeability estimation from well logs, SPE 35412 Presented at the Improved Oil Recovery Symposium, Tulsa, Ok, 21 April, 1996.

[11] T. Hastie, R. Tibshirani, Generalized Additive Models, Chapman & Hall/CRC, New York, 1990.

[12] J.R. Hearst, P.H. Nelson, F.L. Paillet, Well Logging for Physical Properties, John Wiley & Sons, New York City, 2000.

[13] H. Hotelling, J. Educ. Psychol. 24 (1933) 417–441, 498–520.

[14] J.L. Jensen, L.W. Lake, Optimization of regression-based porosity-permeability predictions, in: CWLS 10th Symposium, Calgary, Alberta, Canada, 1985.

[15] Jiang Xie, Improved Permeability Prediction Using Multivariate Analysis Methods, Thesis Master of Science, Submitted to the Office of Graduate Studies of Texas A&M University, 2008.

[16] Jude O. Amaefule, Mehmet Altunbay, Djebbar Tiab, David G. Kersey, Dare K. Keelan, Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/ wells, Paper SPE 26436 Presented at the 68th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers held in Houston, Texas, 3–6 October, 1993.

[17] S.H. Lee, K. Arun, A. Datta-Gupta, SPE Reservoir Eval. Eng. 5 (3) (2002) 237–248.

[18] W.S. McCulloch, W. Pitts, Bull. Mathe. Biophys. 5 (1943) 115.

[19] K. Pearson, Phil. Mag. 2 (6) (1901) 559–572.

[20] D. Tiab, Advances in Petrophysics, Vol. 1-Flow Units, Lecture Notes Manual, University of Oklahoma, 2000.

[21] A. Timur, An Investigation of Permeability, Porosity, & Residual Water Saturation Relationships For Sandstone Reservoirs, The Log Analyst IX (4), SPWLA-1968-VIXn4a2, 1968.

[22] D. Wang, M. Murphy, J. Data Sci. 2 (4) (2004) 329–346.

[23] W.A. Wendt, S. Sakurai, P.H. Nelson, Permeability prediction from well logs using multiple regression, in: L.W. Lake, H.B.J. Caroll (Eds.), Reservoir Characterization, Academic Press, New York, 1985.

[24] G. Xue, A. Datta-Gupta, P. Valko, T. Blasingame, SPE Reservoir Eval. Eng. 12 (2) (1997) 85–94.