Petroleum and Coal

Article	Open Access

Litho-Fluid Facies Modeling by Using Logistic Regression

Mohammed F. Gouda, Ahmed M. A. Salim

Department of Geosciences, Universiti Teknologi Petronas, Malaysia

Abstract

Rock-physics monitoring plays a crucial role in forecasting the facies distribution within the subsurface. However, litho-fluid estimation is challenging due to the randomness of the relationship between reservoir properties and facies distribution. As a result, some facies show similar responses to rock properties and attributes. Data-quality, as well as the elasticity difference between different facies, are the main factors affecting the efficiency of facies-discrimination. This study aims at predicting the distribution of gas-sand, wet-sand, and shale from elastic properties by using logistic regression. The effect of sand and gas distributions on fourteen elastic attributes has been tested to reduce the facies model's variables and determine the best lithology and fluid predictors. The results show that the near and far elastic impedances are the best lithology predictors, while the best fluid predictors are the Mu-Rho, Lambda-Rho/Mu-Rho, and Poisson's ratio. Accordingly, the coefficients of both models have been estimated by Markov-Chain Monte-Carlo simulation to calculate the sand and gas probabilities. Eventually, the two models have been merged by using appropriate cut-offs to turn the probabilities into facies estimates. The correct-classification-rate values of the gas-sand, wet-sand, and shale are 0.92, 0.81, and 0.91, respectively.

Keywords: *Facies; Litho-fluid; Classification; Logistic regression; MCMC.*

1. Introduction

A direct hydrocarbon indicator (DHI) is one of the conventional approaches that show some amplitude-related anomalies due to the effect of litho-fluid content on seismic velocities [1-4]. However, such indicators are, sometimes, misleading because the change of amplitude can be related to various reasons rather than the facies change [5-6]. Another facies indicator is called Amplitude-Versus-Offset analysis (AVO) [7], which is a technique to analyze the prestack seismic data by monitoring the amplitude-change with offset. More pre-stack attributes proved efficient detection of facies such as the Elastic-Impedance (EI) [8-9] and Extended-Elastic Impedance (EEI) [10]. However, those attributes are affected by data-quality issues such as the tuning effect and seismic anisotropy.

Unlike amplitude-attributes, elastic properties are reliable facies detectors that can discriminate various litho-fluid classes [11-13]. For instance, the LMR theory shows how to use Lambda-Rho and Mu-rho to identify different lithologies and fluids [14]. Besides, Young Modulus and Poisson's ratio have been used to derive three facies in the Niobrara formation, Powder River basins [15]. The integration of the elastic and acoustic impedances has been proved to be an efficient way to discriminate brine-sand, hydrocarbon-sand, and shale in the North Sea, Gulf of Mexico, and Australia [16]. However, even some facies show the same elastic response, such as sandstone and pure-marl, which show the same P-wave velocity as well as the mixed marl-chalk and basal-chalk that have, mostly, similar ranges of P- and S-wave velocities [17]. Some studies detect facies by using petrophysical properties, such as porosity [18]. However, the extraction of the petrophysical properties from seismic data is challenging and risky.

Statistical modeling and machine learning can be applied to various rock properties to forecast facies distribution accurately. One of the statistical methods to estimate facies distribution is the Bayesian-classification theory [19- 21]. Also, various unsupervised-learning algorithms

can predict the facies distribution without any dependency on the availability of a known response variable. Examples of unsupervised-algorithms are the self-organizing map (SOM) [22-24] and the principal component analysis (PCA) [25]. Besides, supervised-learning algorithms play an essential role in facies classification. For instance, the Multi-Layer Feed-Forward Neural Networks (MLFN) has been used to forecast the depositional facies based on the gamma-ray, density, neutron, and resistivity logs [26]. The bagged-tree (BT) or bootstrap aggregating algorithm, can generate more accurate facies estimates than the MLFN [24].

The limitations of the statistical methods appear in heterogeneous media, where there is a significant degree of randomness in the relationships between reservoir properties and facies distribution. Also, many studies apply statistical models to a large number of variables resulting in over-parameterization. Other studies may postulate over-fitted models due to the false selection of facies predictors. Moreover, the combination of various lithologies and fluids may enhance the degree of randomness between variables. Accordingly, an efficient variable-selection procedure is needed to reduce the ambiguity in facies detection and avoid over-parameterization.

This study uses logistic-regression to study the sensitivity of fourteen elastic attributes to facies distribution in a three-class medium consisting of gas-sand, wet-sand, and shale. The idea of the model is to estimate the sand and gas probabilities from the scaled elastic-attributes by estimating the posterior means of the model's coefficients. The probability density curve of each variable's coefficient determines whether the variable is a strong facies predictor. The selected lithology and fluid predictors are used to postulate two models, which should be combined to estimate the facies distribution based on the cut-off value of each model.

2. Methodology

The objective of this study is to determine which elastic properties can discriminate litho-fluid classes. One way to do so is to fit several models with various combinations of the elastic properties and then compare the models by determining the most effective combination, which leads to the maximum separation between the litho-fluid classes. Another way to select the appropriate variables is to fit a linear regression model between the response variable (facies) and the elastic properties such that the priors of the model's coefficients favor values near zero. Therefore, a variable is considered a weak facies-predictor if its coefficient has a near-zero mean. On the other side, if a coefficient's mean is far from zero, the variable of the coefficient is assumed as a strong predictor. The coefficients are obliged to have a near-zero value by centering and scaling each variable by using the "scale" function in R. Because the response variable is discrete, the logistic-regression approach has been chosen for modeling.

2.1. Logistic-regression modeling

Logistic regression is one of the linear regression families that can model discrete variables by providing model estimates that represent the probability of occurrence of each possible event [27]. The likelihood of the logistic-regression model follows a Bernoulli distribution, as shown below:

$Y_k | \varphi_k \sim Bernoulli(\varphi_k)$ 　　　　　　　　　　　　　　　　(1)

where Y_k is a discrete event at an observation (k); while φ_k is the probability of occurrence of this event. Accordingly, a model's variables can be related to the value of φ_k by using the following link-function:

$log(\varphi k / 1 - \varphi k) = \beta_0 + \beta_1 (X_1)_k + .. \beta_n (X_n)_k$ 　　　　　　(2)

where β_0 is the intercept; β_1 to β_n are the model's coefficients; and $(X1)k$ to $(Xn)k$ are the model's variables at an observation (k). Therefore, the φk can be obtained, as shown below:

$\varphi k = (1/1 + e - (\beta_0 + \beta_1 (X_1)_k + .. \beta_n (X_n)_k))$ 　　　　　(3)

The coefficients of this model can be obtained, according to the Bayes' theory [28], from the joint distribution of the likelihood and priors, as shown below:

$P(b|y) = P(y|b)P(b)/P(y)$ 　　　　　　　　　　　　　　　(4)

where P(b|y) is the posterior probability of the coefficient (b) given the likelihood (y), which represents the logging data observations; P(y|b)P(b) is the joint distribution of the likelihood

and the priors of the coefficients; and P(y) is the marginal distribution of the likelihood (y). Therefore, the posterior probability is equivalent to the joint distribution of the likelihood and priors, as shown below:

$$P(b|y) \propto \int P(y|b)P(b)db \qquad (5)$$

In some cases, the joint probability of the model is too complex to be integrated. That's why the Markov-Chain Monte-Carlo Simulation can be used to obtain the posterior distribution of such models [29].

2.2. Markov-Chain Monte-Carlo Simulation(MCMC)

A Markov-chain (MC) is a chain of numbers such that each number depends on the previous one in the sequence. If $\{X_1, X_2, X_3, .., X_t\}$ is a Markov-chain, where $\{1, 2, 3, ..t\}$ are successive points in time, the probability of the variables can be expressed, according to the chain role, as shown below:

$$P(X_1, X_2, X_3, .., X_t) = P(X_1).P(X_2|X_1).P(X_3|X_2, X_1), .., P(X_i|X_{t-1}, X_{t-2}, .., X_2, X_1) \qquad (6)$$

Assuming that the transition probabilities are not time-dependent, the probability at the time (t) depends only on the value of $\{X_{t-1}\}$:

$$P(X_1, X_2, X_3, .., X_t) = P(X_1).P(X_2|X_1). P(X_3|X_2), .., P(X_t|X_{t-1},) \qquad (7)$$

The stationary distribution of the MC is the initial distribution at which the transition probability doesn't change in any given state. The idea of the MCMC is to draw multiple random values from a proposed distribution such that the sequence of all simulations is a Markov-chain, and the stationary distribution of that chain is the posterior distribution. According to the law of large numbers [30], the MC should converge to the true mean of the posterior distribution which can be assigned to the coefficient. Therefore, each variable's coefficient in the litho-fluid facies models will be obtained by calculating the posterior mean of all realizations drawn from the proposed distribution.

2.3. Solving for sand and gas probabilities

After getting the litho-fluid facies models' coefficients, the sand and gas probabilities can be calculated by the equation (3.9). The lithology variable is a two-class categorical variable that has the value 1, for sand, and 0 for shale. Similarly, the fluid variable has the value 1, for gas-sand, and 0 for wet-sand. Figure 1 shows how the predicted sand and gas probabilities relate to the lithology and fluid variables, respectively. The dashed lines should be the cut-offs that best discriminate the litho-fluid facies classes.
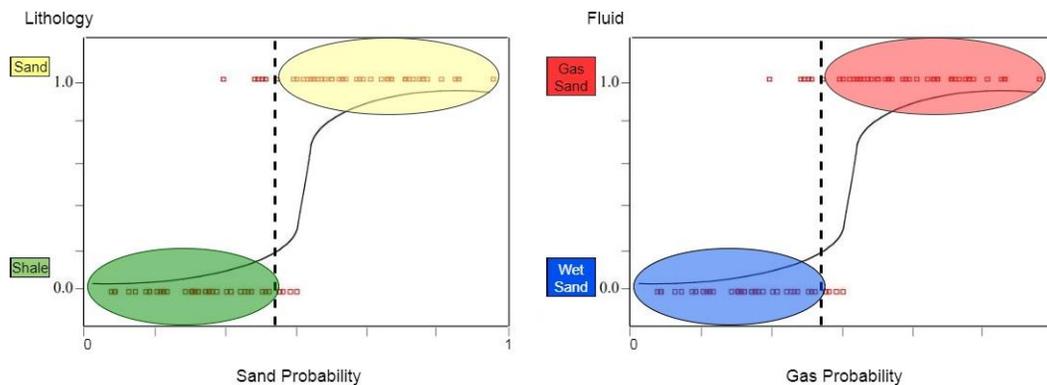

Figure 1. Plots of the (a) Sand probability Vs. the lithology variable, (b) Gas probability Vs. the fluid variable

3. Results and discussion

A sensitivity analysis has been applied to fourteen variables, which can be classified into three groups, as shown in Table 1. The separate modeling of each group helped to prevent over-parameterization and reduce the number of linearly-related covariates that affect the model's accuracy. The purpose of the sensitivity analysis is to reduce the variables and classify

them into lithology predictors to estimate the sand probability, and fluid predictors to forecast the gas probability.
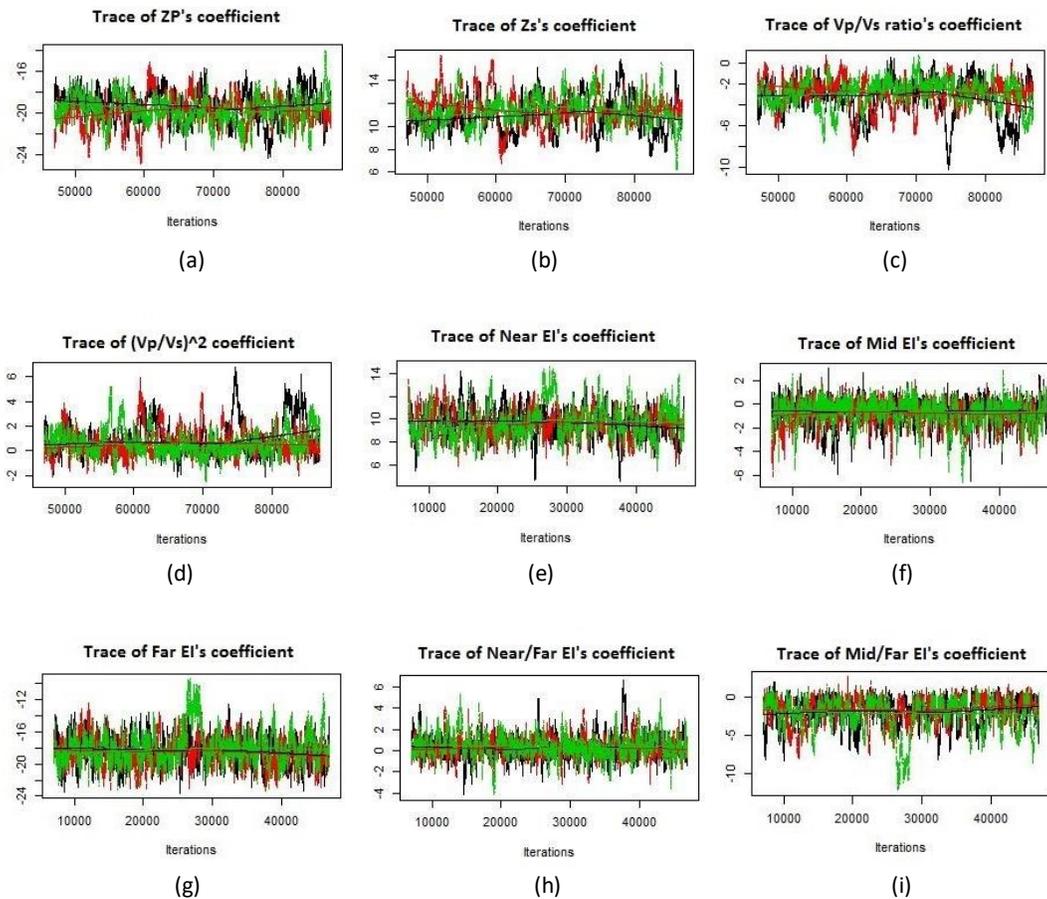
Table 1. The preliminary variables of the litho-fluid facies model

Group	Variables
1	P-Impedance (Zp), S-Impedance (Zs) , Vp /Vs , and (Vp /Vs)2
2	Near-EI, Mid-EI, and Far-EI, Near-EI/Far-EI, and Mid-EI/Far-EI
3	Lambda-Rho (LR), Mu-Rho (MR), LR/MR ratio (LMR), Poisson's Impedance (PI), and Poisson's Ratio (PR)

The variable selection has been achieved by R-Studio, firstly, by centering and scaling the explanatory variables, to quantify the effect of each variable on each of the sand and gas probabilities, and, secondly, by postulating various models with different sets of variables and selecting the best model according to the correct classification rate of each model.

3.1. Variable selection for the lithology model

After Beginning with the lithology model, the Markov-chains have been diagnosed for autocorrelation, as shown in Figure 2. After a gradual increase of iterations from 5,000 to 40,000, there is still high autocorrelation for some variables, such as the Mu-Rho and Poisson ratio. However, the chains' convergence is acceptable for the variable-selection procedure, as the simulation's objective is to determine whether the coefficients' means are close to zero or far from it.
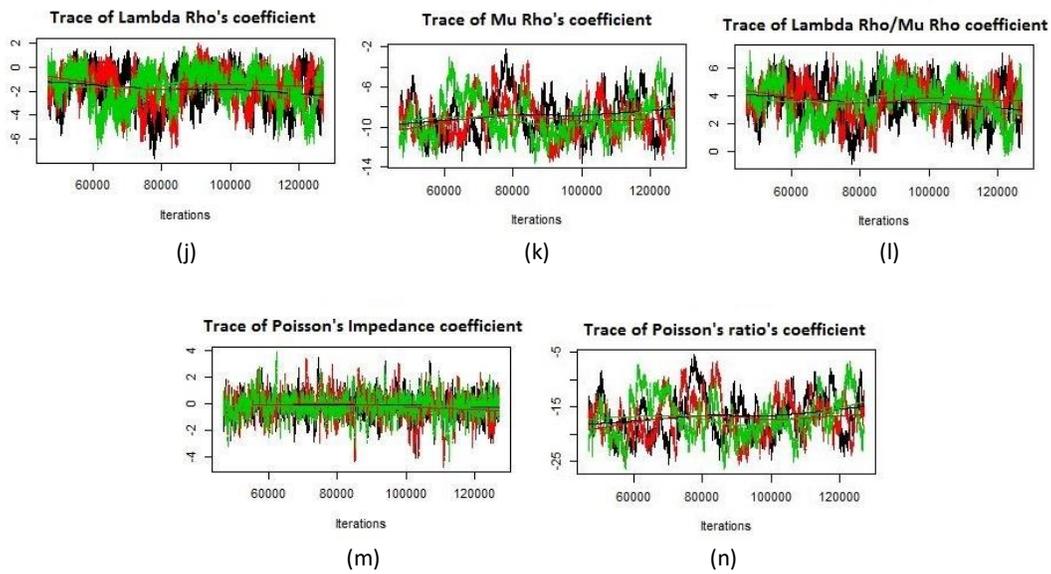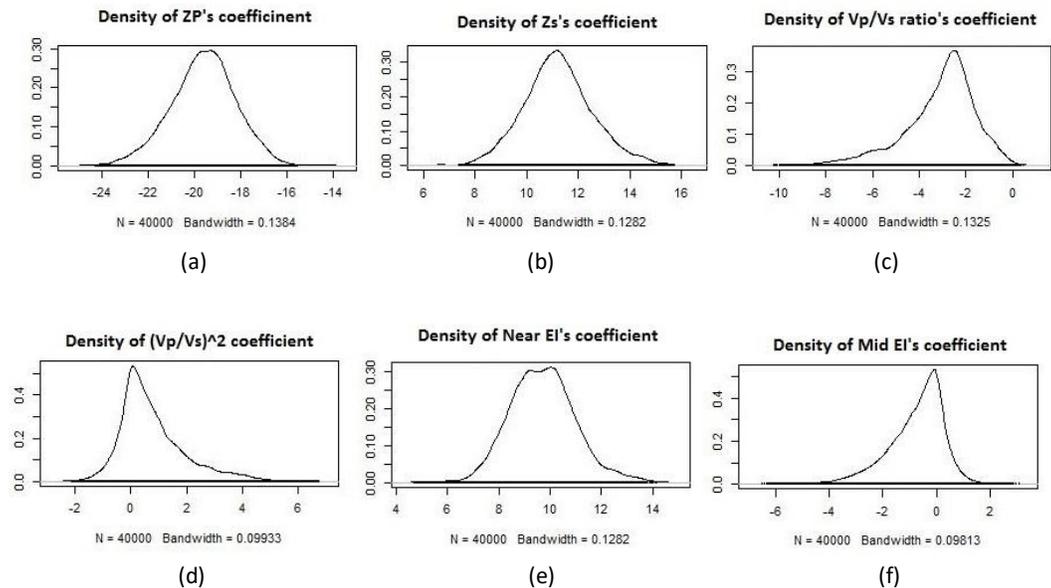
Figure 2. The trace-plots of the Markov-chains that simulates the coefficients of the lithology model's variables

The probability density curves of the variables' coefficients are shown in Figure 3. The posterior means of the coefficients of the squared Vp/Vs, mid-EI, near/Far-EI, mid/Far-EI, LR, and PI are close to zero, which means that those variables are not strong lithology-predictors. That's why these variables have been excluded from the lithology modeling. On the other hand, the posterior means of the coefficients of Zp, Zs, Vp/Vs, Near-EI, Far-EI, Mu-Rho, Lambda-Rho/ Mu-Rho, and Poisson ratio are significantly away from zero which indicates a significant relationship with the lithology distribution. Consequently, these variables are considered for lithology modeling.
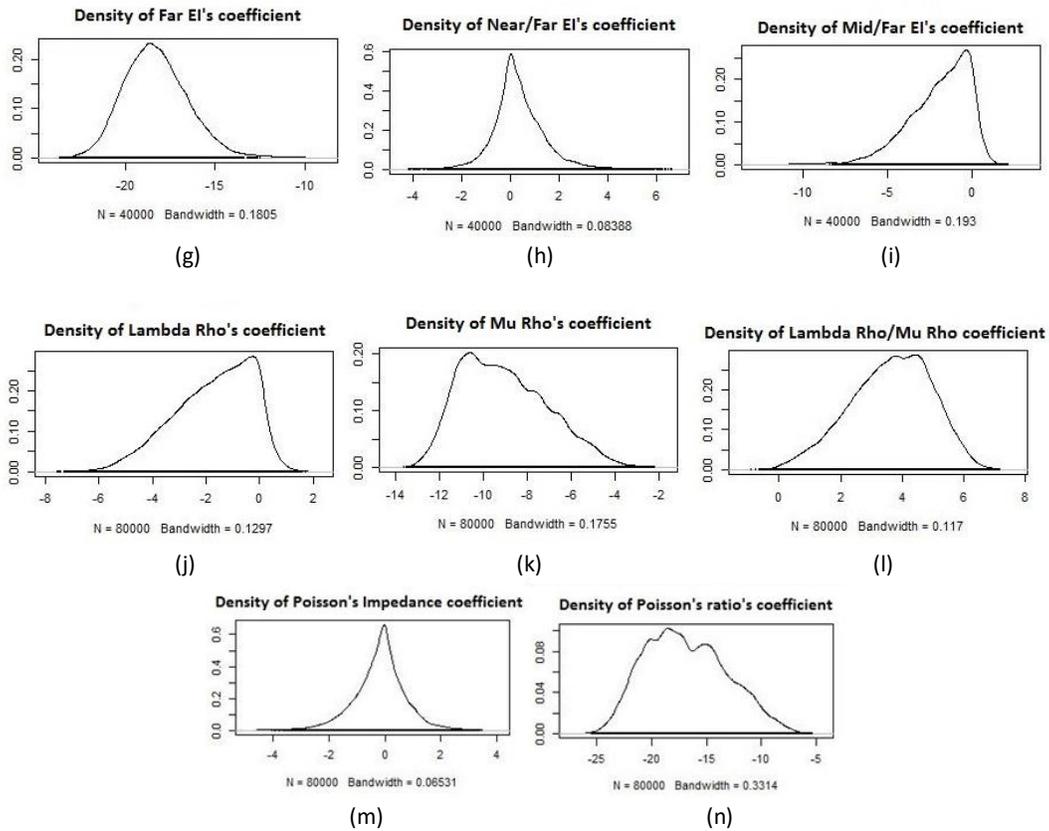
Figure 3. The probability density curves of lithology model's coefficients showing the effect of each rock property on lithology distribution

The variables Zp, Vp/Vs, Far-EI, Mu-Rho, and Poisson ratio have negative coefficients. This means that an increase in these properties indicates a low sand probability and high shale probability. On the other hand, the Zs, Near-EI, and Lambda-Rho/ Mu-Rho ratio have positive coefficients. In other words, an increase in these properties raises the sand probability and reduces the shale probability.

3.2. Variable selection for the fluid model

The Markov-chains of each coefficient are plotted in Figure 4, where the number of iterations is on the X-axis, while the value of each coefficient is on the Y-axis. All coefficients show relatively low autocorrelation, which is obvious through the trace-plots, where the convergence of the Markov-chains happens at 20,000 to 25,000 iterations. However, the coefficient of the Poisson's ratio shows very good convergence of the chains at 5000 iterations.

The probability density curves of variables' coefficients are shown in Figure 5. The Zp, Vp/Vs, near-EI, mid/Far-EI, and LR seem to have a weak relationship with the fluid distribution because their coefficients have posterior means centered on zero. However, the variables Zs, squared Vp/Vs, Mid-EI, Far-EI, near/Far-EI ratio, MR, PR, LMR show higher coefficients' values away from zero which means a strong relationship with the fluid variable.

The variables Zs, squared Vp/Vs, near/Far-EI ratio, MR, LMR, PI, and PR have negative coefficients. This means that an increase in these attributes indicates a decrease in the probability of gas and an increase in the probability of water. On the other hand, the Mid-EI, Far-EI, Poisson's impedance, and Poisson's ratio have positive coefficients. In other words, an increase in these properties enhances the gas-sand probability and reduces the wet-sand probability.
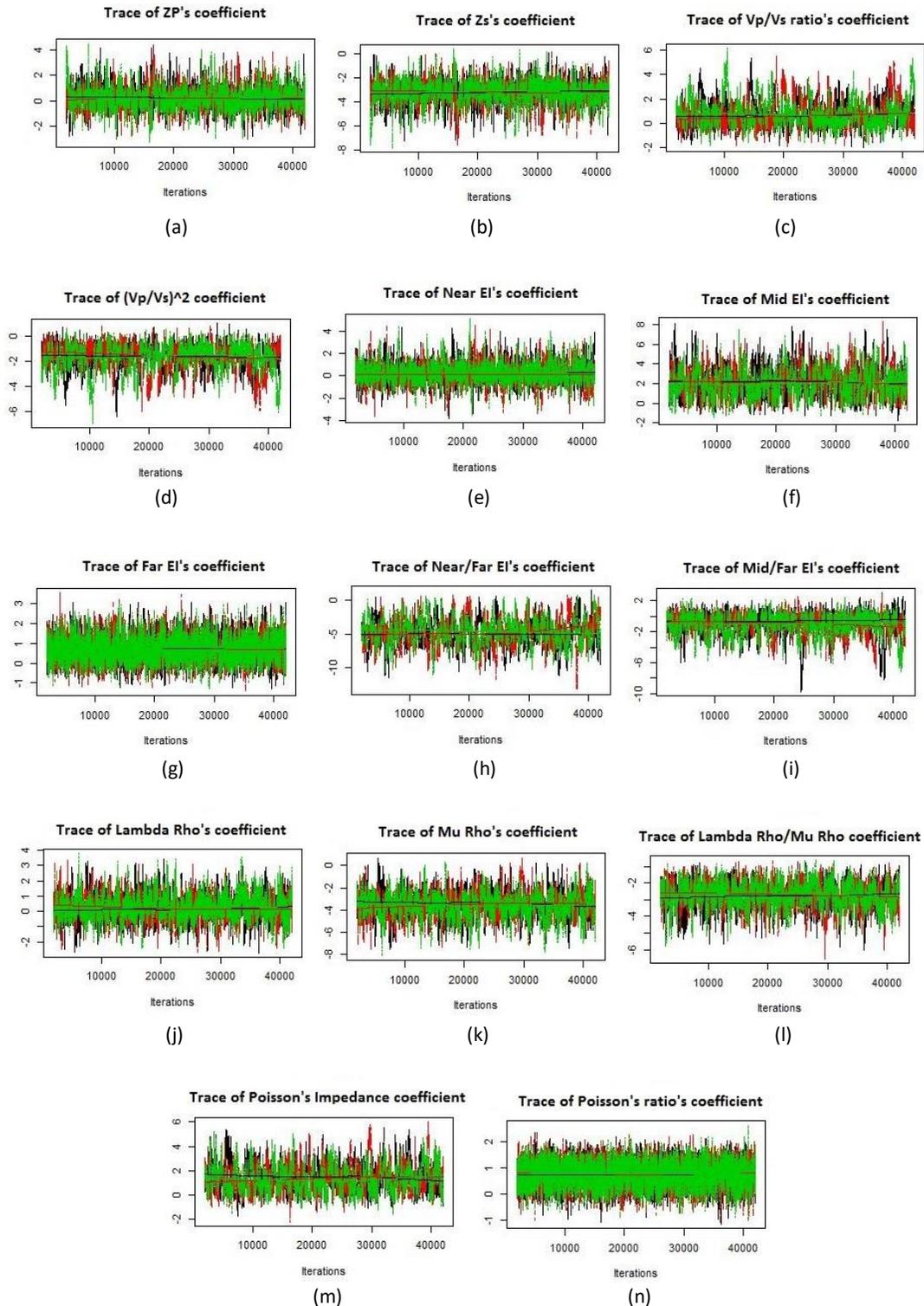
Figure 4. The trace-plots of the Markov-chains that simulates the coefficients of the fluid model's variables

It can be concluded that the acoustic domain is not enough for litho-fluid facies modeling. For instance, the P-impedance and Vp/Vs ratio can discriminate sand and shale rocks but can't identify the fluid type. Unlike the acoustic domain, the elastic attributes can solve this ambiguity such that both the litho-fluid facies classes will be separated efficiently. In addition to that, a combination of some properties may lead to a better separation than a single property.

As a result, different combinations of variables have been used to create various models to forecast the distribution of the litho-fluid facies.
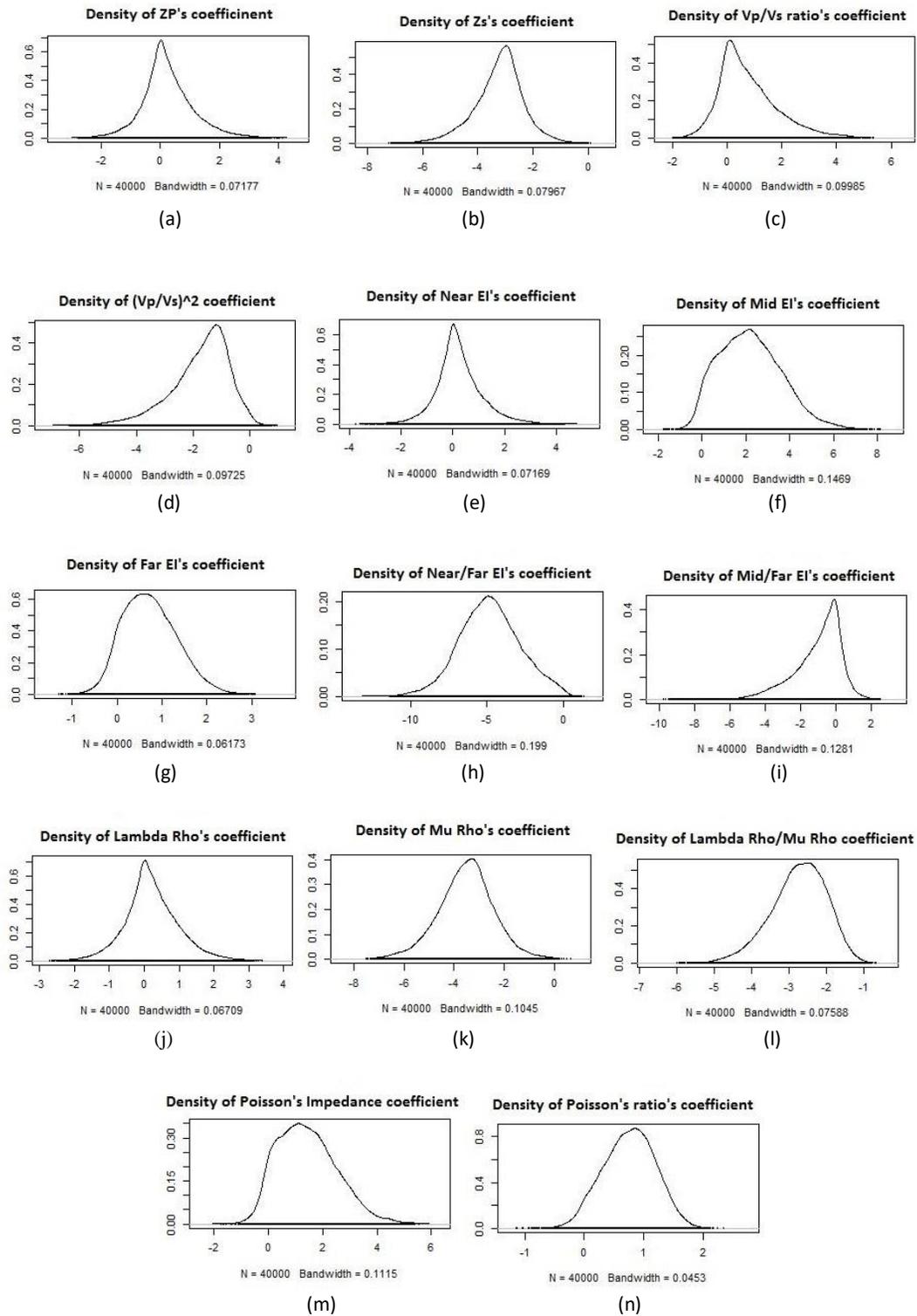

Figure 5. The probability density curves of fluid model's coefficients showing the effect of each rock property on fluid distribution

3.3. Lithology modeling

The selected lithology predictors have been arranged to postulate different logistic-regression models from which the best model should be selected. Six models have been created to forecast the sand probability, which equals (1-shale probability). After filtering the data points from thin layers and coal beds, the total number of data points collected from well (8) is 3559 points, which are classified into 2526 shale samples, having the value (0), and 1033 sandstone samples having the value (1). The data has been randomly divided into 2659 points for training and 900 for validation. The training data includes 1929 shale samples, and 730 sandstone samples, while the validation data consists of 597 shale samples and 303 sandstone samples.

The six models are nominated from FM-1 to FM-6, as shown in Table 2, where each model consists of different sets of explanatory variables. The objective of lithology modeling is to estimate the coefficients' means to calculate the probability of sand at each observation. A convenient cut-off value is selected to differentiate between the sand and shale observations for each model. Finally, the correct classification rate is calculated to determine the most accurate model.

Table 2. The explanatory variables of the six lithology models; FM-1 to FM-6

Model Name	Explanatory Variables
FM-1	Zp, Zs, and Vp/Vs
FM-2	Zp and Vp/Vs
FM-3	Near-EI and Far-EI
FM-4	Vp/Vs, Near-EI, and Far-EI
FM-5	Mu-Rho, Lambda-Rho/Mu-Rho, and Poisson's ratio
FM-6	Mu-Rho and Poisson's ratio

The next step is to estimate the coefficients of each model by MCMC simulation, such that three chains are fully converged to the posterior mean of each coefficient at a specific number of iterations. The posterior means of the coefficients are shown in Table 3.

Table 3 The posterior means of the coefficients of the lithology models

Model Name	Intercept	($\beta 1$)	($\beta 2$)	($\beta 3$)
FM-1	0.6988	-19.893	11.5432	-2.1789
FM-2	0.7849	-9.0109	-6.5863	N/A
FM-3	0.9626	8.5906	-20.0666	N/A
FM-4	1.029	-0.2036	3.345	-14.676
FM-5	2.05	-11.255	5.053	-21.239
FM-6	1.488	-10.116	-13.837	N/A

The sand probabilities have been calculated for each model. The predicted sand probabilities of the training data are plotted against the lithology variable, as shown in Figure 6. As expected, the sand probability seems generally high, where the lithology variable equals 1, indicating a sandstone observation, and it's relatively low, where the lithology variable equals 0, meaning a shale observation. To specify the exact lithology for each observation, a cut-off is set where the highest correct-classification rates (CCR) of the sandstone and shale are reached as shown in Table 4.

Table 4. The cut-off values and correct classification rates of the lithology models applied to the training data

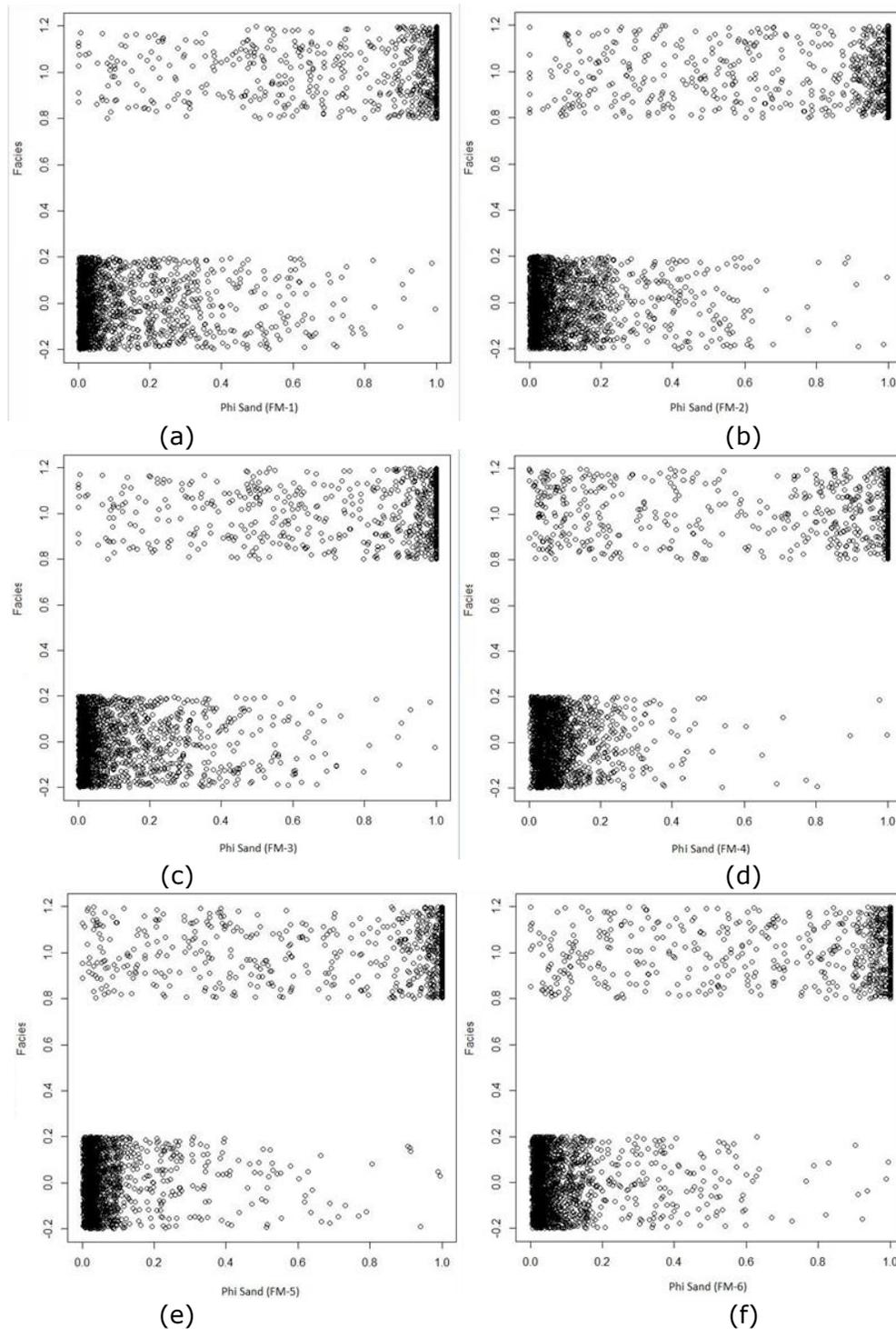| Model Name	Cut-off	CCR of sandstone	CCR of shale	Total CCR
FM-1	0.3	0.91	0.91	0.91
FM-2	0.3	0.88	0.93	0.905
FM-3	0.3	0.91	0.92	0.915
FM-4	0.2	0.81	0.93	0.87
FM-5	0.2	0.90	0.93	0.915
FM-6	0.2	0.90	0.92	0.91

Figure 6. The plots of the predicted sand probabilities of the training data, on the X-axis, and the lithology variable, on the Y-axis, for the lithology models from FM-1 to FM-6 corresponding to figures from a to f, respectively

The best cutoff points lie between the sand probabilities from 0.2 to 0.3, such that the sample is most likely to be sandstone if the sand probability is higher than the cutoff. The models FM-1, FM-3, FM-5, FM-6 show good correlation for both sandstone and shale with total CCR values 0.91, 0.915, 0.915, and 0.91, respectively. The models FM-2 and FM-4 have resulted in slightly low CCR values for sandstone, which are 88% and 81%, respectively.

The models have been applied to the validation data to determine the best model. Figure 7 shows the plots of the predicted sand probabilities of the validation data against the lithology variable. Table 5 shows the best cutoff as well as the CCR value of the sandstone and shale for each model.
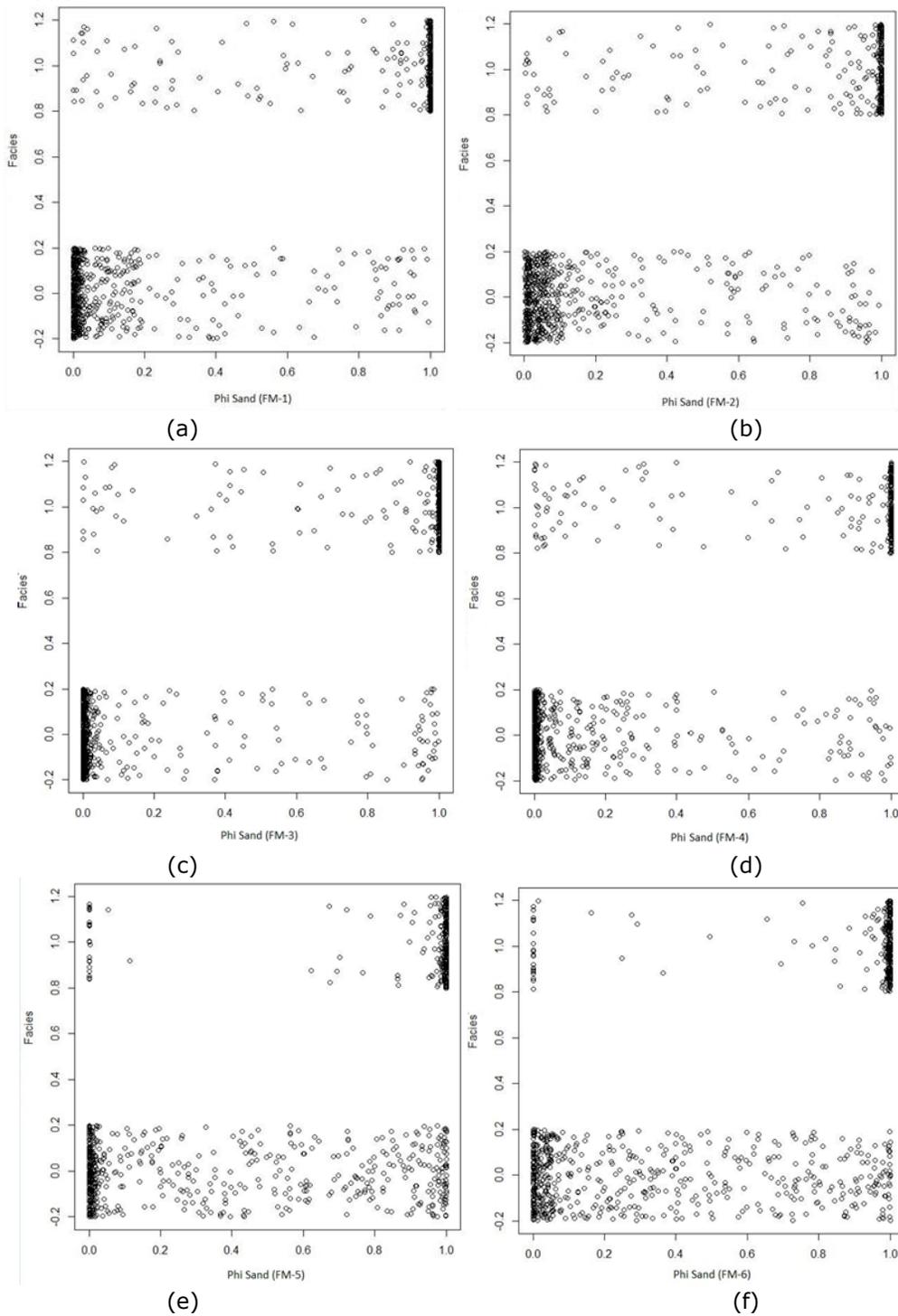

Figure 7. The plots of the predicted sand probabilities of the validation data, on the X-axis, and the lithology variable, on the Y-axis, for the lithology models from FM-1 to FM-6 corresponding to figures from a to f, respectively

Table 5. The cut-off values and correct classification rates of the lithology models applied to the validation data

Model Name	Cut-off	CCR of sandstone	CCR of shale	Total CCR
FM-1	0.3	0.88	0.86	0.87
FM-2	0.4	0.88	0.85	0.865
FM-3	0.45	0.89	0.89	0.89
FM-4	0.6	0.82	0.92	0.87
FM-5	0.8	0.90	0.79	0.845
FM-6	0.8	0.90	0.85	0.875

Unlike the models FM-4, FM-5, and FM-6, the training and validation data of the models FM-2 and FM-3 have, mostly, similar cutoffs. The total CCR value of the model FM-3, for the validation data, is 0.89, which is the highest among those of the models; FM-1, FM-2, FM-4, FM-5, and FM-6, which are 0.87, 0.865, 0.87, 0.845, 0.875, respectively.

Table 6 shows the total CCR values of all data points, including the training and validation data. Accordingly, the model FM-1 has been selected to be the best lithology model due to the high CCR value (0.9), compared to the other models, and the similarity of the cutoff values of both training and validation data, which indicates the stability of the model.

Table 6. The total CCR values of the lithology models applied to the training and validation data

Model name	FM-1	FM-2	FM-3	FM-4	FM-5	FM-6
Total CCR	0.89	0.885	0.9	0.87	0.88	0.89

3.4. Fluid modeling

The same steps of lithology-modeling apply to the fluid model. Four logistic regression models have been created to predict the probability of gas-sand and wet-sand. The total number of the collected sandstone samples is 1035 points, which are classified into 778 water samples, having the value 0, and 257 gas samples having the value 1. The data have been randomly divided into 713 points for training and 322 for validation. The training data includes 533 water samples, and 180 gas samples, while the validation data consists of 245 water samples and 77 gas samples. The explanatory variables of the four fluid models, from FLM-1 to FLM-4, are shown in Table 7.

Table 7. The explanatory variables of the four fluid models; FLM-1 to FLM-4

Model Name	Explanatory Variables
FLM-1	Zs and squared Vp/Vs
FLM-2	Mid-EI, Far-EI, Near/Far-EI
FLM-3	Mu-Rho, Lambda-Rho/Mu-Rho, and Poisson's ratio
FLM-4	Lambda-Rho/Mu-Rho, Poisson's Impedance, and Poisson's ratio

Table 8 shows the predicted posterior means of the coefficients. Accordingly, the gas probability has been calculated for the training data and plotted against the fluid variable, as shown in Figure 8. The gas probability is, generally, high, where the fluid variable equals 1, indicating gas, and, generally, low, where the fluid variable equals 0, meaning water. The cut-off value and CCR of the training data have been obtained, for each model, as shown in Table 9.

Table 8. The posterior means of the coefficients of the fluid models

Model Name	Intercept	(β_1)	(β_2)	(β_3)
FLM-1	-2.0861	-2.9419	-0.9383	N/A
FLM-2	-2.1909	2.1222	0.5859	-5.4184
FLM-3	-2.703	-3.641	-1.683	1.011
FLM-4	-3.266	-9.583	10.703	-0.896

Table 9 The cut-off values and correct-classification rates of the fluid models applied to the training data

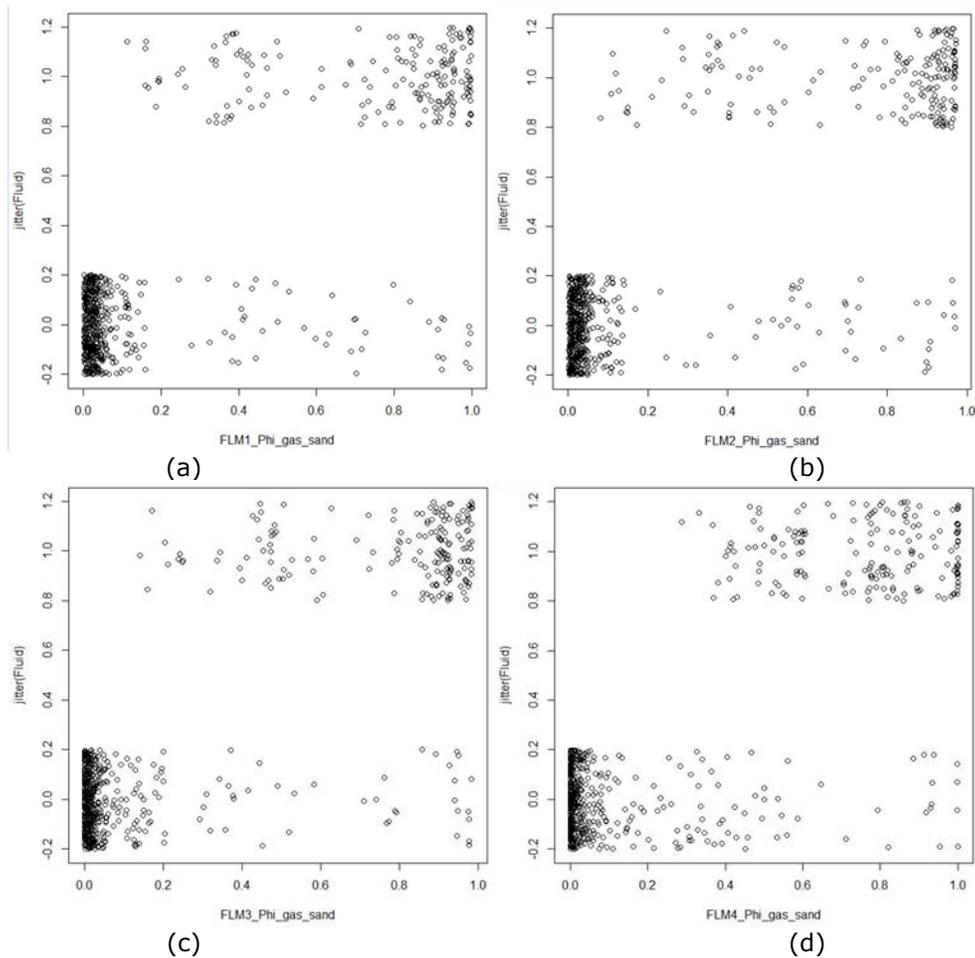| Model Name	Cut-off	CCR of gas	CCR of water	Total CCR
FLM-1	0.3	0.93	0.92	0.91
FLM-2	0.3	0.92	0.92	0.92
FLM-3	0.3	0.95	0.93	0.94
FLM-4	0.4	0.97	0.94	0.955

(a) (b) (c) (d)

Figure 8. The plots of the predicted gas probabilities of the training data, on the X-axis, and the fluid variable, on the Y-axis, for the fluid models from FLM-1 to FLM-4 corresponding to figures from a to d, respectively

The best cut-off value is 0.3 for all models except for that of the model FLM-4, which is 0.4. All models show excellent CCR values when applied to the training data. The model FLM-4 results in the best total CCR value (0.955) compared to those of the models FLM-1, FLM-2, and FLM-3, which are 0.91, 0.92, 0.94, respectively. However, when applying the models to the validation data, as shown in Figure 9, the models FLM-2 and FLM-4 show shallow CCR values for gas (0.38 and 0.35 respectively) and high CCR for water (0.7 and 1 respectively), as shown in Table 10. On the other hand, the models FLM-1 and FLM-3 led to a good correlation for both gas and water, where the total CCR values are 0.855 and 0.85, respectively.

Table 10. The cut-off values and correct-classification rates of the fluid models applied to the validation data

Model Name	Cut-off	CCR of gas	CCR of water	Total CCR
FLM-1	0.3	0.84	0.87	0.855
FLM-2	0.3	0.38	0.7	0.54
FLM-3	0.3	0.86	0.84	0.85
FLM-4	0.5	0.35	1	0.675

(a)　　　　　　　　　　　　　(b)
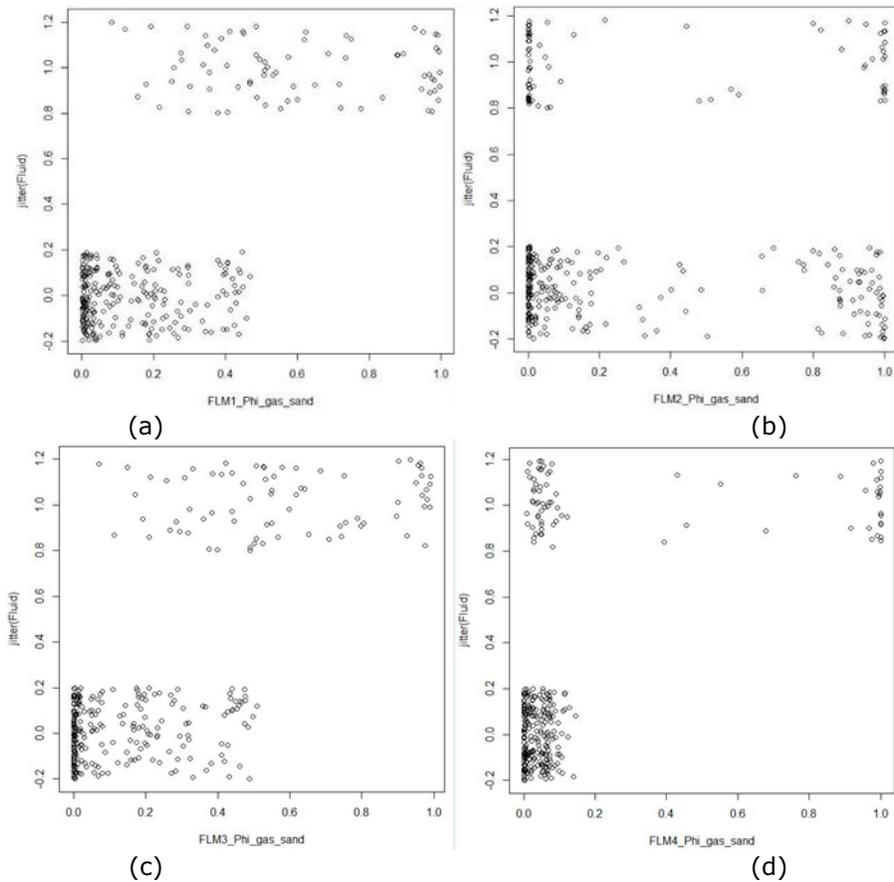
(c)　　　　　　　　　　　　　(d)

Figure 9. The plots between the predicted gas probabilities of the validation data, on the X-axis, and the fluid variable, on the Y-axis, for the fluid models from FLM-1 to FLM-4 corresponding to figures from a to d respectively

All the fluid models, except for the model FLM-4, have the same cut-off value (0.3) when applied to the training and validation data. Table 11 shows the total CCR value for each model applied to all data, including the training and validation data. Accordingly, the model FLM-3 has been selected to be the best fluid model due to the high CCR value (0.9), compared to the other models.

Table 11. The total CCR values of the fluid models applied to all data

Model Name	FLM-1	FLM-2	FLM-3	FLM-4
Total CCR	0.88	0.73	0.9	0.815

3.5. The integrated Litho-Fluid facies model

The next step is to merge between the two models; FM-3 and FLM-3, to create an integrated litho-fluid facies model by using a straightforward function created in Matlab. The outputs of the function consist of the predicted facies and the correct classification rate of each litho-fluid class. The inputs of the function are shown below:
- The lithology predictors (Near-EI & Far-EI) and fluid predictors (MR, LMR, & PR).
- The facies log, which is a categorical variable having the values; 1, 2, and 3 for the gas-sand, wet-sand, and shale, respectively.
- The cut-off values of the litho-fluid facies models, such that the first cutoff value belongs to the sand probability and discriminates the shale and sandstone in the lithology model, while the other cutoff value belongs to the gas-sand probability and discriminates the gas-sand and wet-sand in the fluid model.

The combined facies model has resulted in the correct classification rates; 0.92, 0.81, and 0.91 for the gas-sand, wet-sand, and shale, respectively. These results seem better than those of a previous study [16], that uses statistical classification to differentiate between the gas-sand, wet-sand, and shale by using the near-offset acoustic and far-offset elastic impedances. Figure 10.a shows a plot of the acoustic and elastic impedances, where facies classification seems ambiguous. Figure 10.b shows that the success rates of gas-sand, wet-sand, and shale are about 80%, 75%, and 52%, respectively, for the acoustic impedance (AI) only, while the success rates are 93%, 75%, and 65%, respectively, in case of using both the elastic and acoustic impedances as facies predictors.
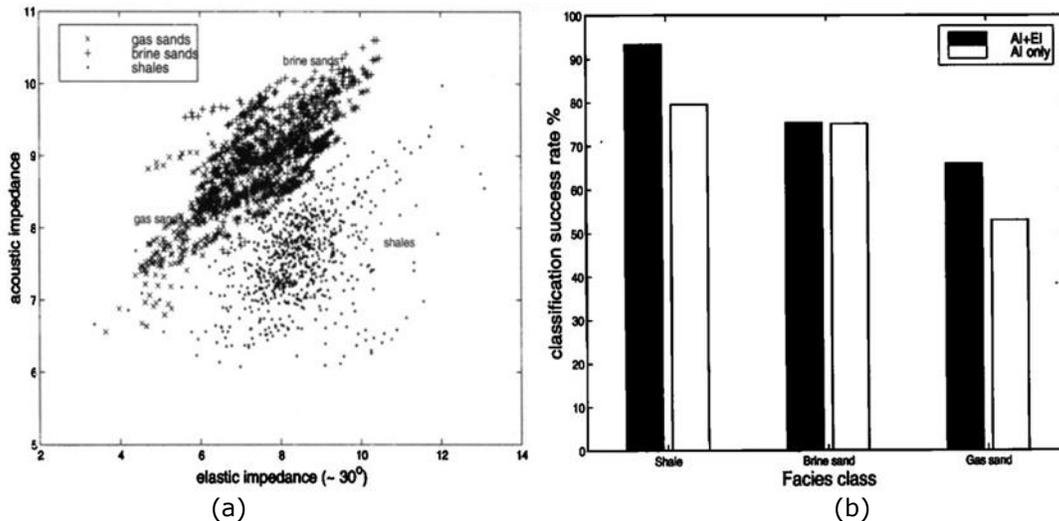

(a)   (b)

Figure 10. Facies classification by using the elastic and acoustic impedances: (a) The plot of the acoustic impedance Vs. elastic impedance, (b) The success-rate histogram of facies classification [16]

The results seem less-accurate than those of the current study, which confirms the advantage of logistic regression and MCMC simulation over the traditional statistical classification. However, both studies proved that the acoustic impedance is not enough for facies detection and that the elastic impedance adds more information that enhances the distinguishability of the facies classes.

4. Conclusions

The variable selection applies to fourteen elastic properties from which the best lithology predictors have been selected to be the near and far elastic impedances, while the best fluid predictors are the Mu-Rho, Lambda-Rho/Mu-Rho, and Poisson's ratio. The lithology and fluid models have been separately postulated by logistic regression. Both models have been merged to predict the distribution of the litho-fluid classes.

Assuming a three-class medium, the correct classification rates of the gas-sand, wet-sand, and shale are 0.92, 0.81, and 0.91, respectively. These results are based on the cut-off values of the sand and gas probabilities, which are 0.45 and 0.3, respectively. However, these values may change from well to another according to the degree of heterogeneity of the field.

The separate modeling of lithologies and fluids is vital to reduce the ambiguity of facies detection. Moreover, selecting the model's predictors is crucial to guarantee the stability of the model. Also, the acoustic domain is not enough to forecast the facies distribution. On the other hand, the combination of acoustic and elastic attributes is a great way to discriminate the different facies classes accurately.

References

[1]   Ensley RA. Direct hydrocarbon detection with P-and SH-wave seismic data. Inter. SEG Mtg., Abs. with Biographies, 349351, 1983.

[2]   Ensley RA. Evaluation of direct hydrocarbon indicators through comparison of compressional- and shear-wave seismic data: a case study of the Myrnam gas field, Alberta. Geophysics, 50(1):37–48, 1985.

[3]   Swan HW, Castagna JP, and Backus MM. Properties of direct AVO hydrocarbon indicators. Offset Dependent Reflectivity-Theory and Practice of AVO anomalies. Investigations in Geophysics, 8:78–92, 1993.

[4]   Sheriff RE. Seismic stratigraphy. Springer Science & Business Media, 2012.

[5]   Shamsuddin SZ, Hermana M, Ghosh DP, and Salim AMA. Reducing uncertainties in hydrocarbon prediction through application of elastic domain. In IOP Conference Series: Earth and Environmental Science, volume 88, page 012004. IOP Publishing, 2017.

[6]   Yang J, Chang X, and Yang Z. The solution of non-gas bright-spot and non-bright-spot gas identification: Elastic prediction. In SEG Technical Program Expanded Abstracts 2015, pages 595–599. Society of Exploration Geophysicists, 2015.

[7]   Ostrander WJT. Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence. Geophysics, 49(10):1637–1648, 1984.

[8]   Connolly P. Elastic impedance. The leading edge, 18(4):438–452, 1999.

[9]   Veeken PCH and Rauch-Davies M. AVO attribute analysis and seismic reservoir characterization. First Break, 24(2):41–52, 2006.

[10]   Whitcombe DN, Connolly PA, Reagan RL, and Redshaw TC. Extended elastic impedance for fluid and lithology prediction. Geophysics, 67(1):63–67, 2002.

[11]   Di Luca M, Salinas T, Arminio JF, Alvarez G, Alvarez P, Bolivar F, and Marín W. Seismic inversion and AVO analysis applied to predictive-modeling gas-condensate sands for exploration and early production in the lower Magdalena Basin, Colombia. The Leading Edge, 33(7):746–756, 2014.

[12]   Karbalaali H, Shadizadeh SR, and Riahi MA. Delineating hydrocarbon bearing zones using elastic impedance inversion: a Persian Gulf example. Iranian Journal of Oil & Gas Science and Technology, 2(2):8–19, 2013.

[13]   Yoong AA, Lubis LA, and Ghosh DP. Application of simultaneous inversion method to predict the lithology and fluid distribution in "X" Field, Malay Basin. In IOP Conference Series: Earth and Environmental Science, volume 38, page 012007. IOP Publishing, 2016.

[14]   Goodway B, Chen T, and Downton J. Improved AVO fluid detection and lithology discrimination using lamé petrophysical parameters; "$\lambda\rho$", "$\mu\rho$", & "λ/μ fluid stack", from P and S inversions. In SEG Technical Program Expanded Abstracts 1997, pages 183–186. Society of Exploration Geophysicists, 1997.

[15]   Kiche Y, Lyes O, Balogh D, and Ouenes A. Lithology constrained elastic inversion: Application to Niobrara brittleness estimation. In SEG Technical Program Expanded Abstracts 2016, pages 4916–4920. Society of Exploration Geophysicists, 2016.

[16]   Mukerji T, Jørstad A, Mavko G, and Granli JR. Near and far offset impedances: Seismic attributes for identifying lithofacies and pore fluids. Geophysical research letters, 25(24):4557–4560, 1998.

[17]   Harryandi S. Facies modeling using 3D pre-stack simultaneous seismic inversion and multi-attribute probability neural network transform in the Wattenberg field, Colorado. PhD thesis, Colorado School of Mines. Arthur Lakes Library, 2017.

[18]   Tabatabaei H and Poursamad R. Evaluation of Asmari reservoir properties via petrophysical logs in Mansour-Abad oil field, Sw of Iran. Petroleum & Coal, 61(1), 2019.

[19]   Doyen P. Seismic reservoir characterization: An earth modelling perspective, volume 2. EAGE publications Houten, 2007.

[20]   Hampson DP, Russell BH, and Bankhead B. Simultaneous inversion of pre-stack seismic data. In SEG Technical Program Expanded Abstracts 2005, pages 1633–1637. Society of Exploration Geophysicists, 2005.

[21]   Babasafari AA, Ghosh DP, Salim AMA, Ratnam T, Sambo Ch, and Rezaee S. Petro-elastic modeling for enhancement of hydrocarbon prediction: Case study in Southeast Asia. In SEG Technical Program Expanded Abstracts 2018, pages 3141–3145. Society of Exploration Geophysicists, 2018.

[22]   Liu R, Zhang B, and Wang X. Patterns classification in assisting seismic-facies analysis. In SEG Technical Program Expanded Abstracts 2017, pages 2127–2131. Society of Exploration Geophysicists, 2017.

[23]   Zhao T, Li F, and Marfurt KJ. Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification. Interpretation, 5(2): T163–T171, 2017.

[24]   Keynejad S, Sbar ML, and Johnson RA. Assessment of machine-learning techniques in predicting lithofluid facies logs in hydrocarbon wells. Interpretation, 7(3): SF1–SF13, 2019.

[25]   Roden R, Smith T, and Sacrey D. Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. Interpretation, 3(4):SAE59–SAE83, 2015.

[26]   Kapur L, Lake LW, Sepehrnoori K, Herrick DC, Kalkomey CT. Facies prediction from core and log data using artificial neural network technology. In SPWLA 39th annual logging symposium. Society of Petrophysicists and Well-Log Analysts, 1998.

[27]   Kleinbaum DG and Klein M. Analysis of matched data using logistic regression. Logistic regression: A self-learning text, pages 227–265, 2002.

[28]   Stuart A. Kendall's advanced theory of statistics. Distribution theory, 1, 1994.

[29]   Banerjee S, Carlin BP, and Gelfand AE. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, 2014.

[30]   Gao R, and Sheng Y. Law of large numbers for uncertain random variables with different chance distributions. Journal of Intelligent & Fuzzy Systems, 31(3):1227–1234, 2016.

To whom correspondence should be addressed: Ahmed M. A. Salim, Department of Geosciences, Universiti Teknologi Petronas, Seri Iskander, Perak, 32610, Malaysia, E-mail: mohammed_17002692@utp.edu.my