Multiple response nonlinear regression applied to simultaneous density, vp and vs rock physics calibration and related statistical analysis

Erick C. S. Talarico (Petrobras S.A.), Hélio Lopes, Simone D. J. Barbosa, Sinésio Pesco (PUC-Rio)

Abstract

The present paper focuses on rock physics analysis of well data, using a four-step workflow: model selection, calibration, analysis, and prediction. It is shown how different rock physics recipes from literature, as well as custom made, can be interactively created to predict elastic properties (v_P, v_S, and density) from rock textural properties (total porosity, shale volume, etc.), ambient conditions (effective pressure, etc.) and fluid content. In the presented workflow, rock physics calibration is performed by a coupled non-linear least-squares scheme, so that all the elastic properties are simultaneously taken into account. Further, statistical analysis of the calibrated model is presented as a tool to assist model selection (such as, granular models and effective medium models) and to support comparison between rock layers (for example, comparing pore aspect ratio or cementation of two different reservoirs). Finally, the workflow includes a Markov Chain Monte Carlo algorithm to simultaneously estimate both the model's parameters and the prediction uncertainties. Proper rock physics analysis, as presented in this paper, is a fundamental input for any geophysical interpretation and uncertainty assessment, from exploration prospects to reservoir characterization.

Introduction

Many published papers show the importance of using rock physics models for consistently predicting rock properties and rock types, [1-10]. For example, Grana [5] uses statistical rock physics to model the relationship between petrophysical and elastic properties, and applies it to drive seismic inversion to petrophysical properties and facies.

The present work focuses on the rock physics modelling problem. Thus, it poses an important tool to assist subsequent steps in quantitative seismic interpretation.

In this paper, we make four contributions:

1 – Introduction of an interactive and flexible framework and respective data-structure for the creation of rock physics recipes, from basic rock physics formulas, such as modified Hertz-Mindlin [11] and Hashin-Shtrikman [12].

2 – Application of multiple response non-linear least-squares to calibrate theoretical rock physics models. The methodology employed here is fully adequate to seismic quantitative interpretation problem, since it calibrates density, v_P and v_S simultaneously.

3 – Application of statistical tools for analyzing rock physics models and their fit.

4 – Application of Markov Chain Monte Carlo (MCMC) technique to estimate simultaneously the uncertainty in the model's parameters and predicted properties.

Method

The workflow presented in this paper, comprises four stages: model selection, calibration, analysis, and prediction. The geoscientist might feel the need to run the first three steps in this workflow multiple times until a satisfactory model is selected. The four steps are described next.

A: Rock Physics model construction

The rock physics equations available in the literature, will usually yield no reasonable result, unless coupled with other equations. For example, soft sand model consists, basically, of chaining Hertz-Mindlin, Hashin-Shtrikman and Gassmann equations [12].

A tree is the most appropriate data structure to create more complex rock physics recipes from basic rock physics equations (for example, when modeling shaly sandstones, as in Figure 1). In the tree, equations are nodes, and they may receive multiple arrows from parent nodes. The tree leaves are the inputs and parameters of the recipe. The inputs are the petrophysical well logs, and the parameters are additional scalars, such as Mineral Bulk and Shear Modulus, coordination number, etc. The parameters and inputs may come into the model through any equation node.

The user must indicate the initial guess for the parameters and their bounds β^{min}, β^{max}.

The equation nodes may have multiple outputs, because some rock physics equations output more than one property simultaneously, for example, bulk and shear modulus. At the root node of the calculus tree, the recipe must have as outcomes, the elastic properties of interest, for instance, the triads: density, v_P and v_S; or density, v_P and Poisson's ratio; etc. The examples in this paper used density, v_P and v_P/v_S.

B: Model Calibration

The mathematical tool used for the calibration step is non-linear multiple response (or multivariate) regression. Bates and Watts [13] give a broad introduction to non-linear regression, and, more specifically, to multiple response non-linear regression. In this kind of regression, one is

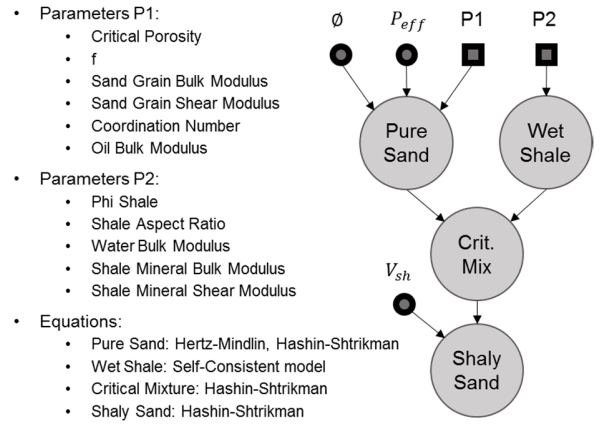

Figure 1: Workflow for predicting Bulk and Shear Modulus of shaly sandstones, according to the soft-sand model. *Phi* is the total porosity, P_{eff} is the effective pressure, and V_{sh} is the Shale Volume, well logs. The equations are the underlying functions used in each calculus node.

interested in minimizing the sum of squared errors coming from different variables (in the present paper, density, v_P and v_P/v_S). The error of each variable is normalized by its corresponding standard deviation, and then the sum of the normalized errors through all the samples and all the variables is performed. Equations (1) and (2) illustrate the above multiple response least-squares formulation, where the index i runs through all the samples in the dataset.

$$\begin{bmatrix} \rho_i \\ v_{P_i} \\ v_{P_i}/v_{S_i} \end{bmatrix} = F\left(\emptyset_i, V_{sh_i}, P_{eff_i} | \beta_1, \beta_2, \dots, \beta_k\right) + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix} \quad (1)$$

$$U(\beta) = \frac{1}{2{\sigma_1}^2}\sum_i [\rho_i - F_1(x_i|\beta)]^2 + \frac{1}{2{\sigma_2}^2}\sum_i \left[v_{P_i} - F_2(x_i|\beta)\right]^2 + \frac{1}{2{\sigma_3}^2}\sum_i \left[{}^{v_{P_i}}/_{v_{S_i}} - F_3(x_i|\beta)\right]^2 \quad (2)$$

Since these standard errors are not known a priori, they must be estimated in an iterative manner. At each iteration, the regression performs the following two steps:

- Given estimated residual standard deviations for each dependent variable, use some non-linear optimization approach to optimize the model parameters. In this paper, it is used the *least_squares* function from Python Scipy package [14].
- Given estimated model parameters, the standard deviation for the errors are recalculated.

In order to consider a more complex and realistic error structure, further development is required. For instance, to take into account the finite and yet different resolutions of different well-logs, depth-correlated errors must be modelled. In addition, since, both petrophysical, and elastic logs are prompt to measurement errors, error-in-variable models must be applied, as [15] exemplifies.

C: Statistical analysis

Usual statistical tools must be used to analyze the results of every rock physics calibration. Two types of analyses are reinforced herein: goodness-of-fit, and significance tests. Rawlings et al. [16] give a broad introduction to least-squares problem and statistical tools.

Goodness-of-fit statistics are the usual residual standard deviations. Using this analysis, the geoscientist can compare how different rock physics models perform on a given dataset. These statistics might be used, for example, to decide which parameters are excluded (by keeping them as internal constants) or included in the original model.

ANOVA is a useful tool to compare how much the residuals decrease (or increase), with the new (or excluded) parameters. It uses an F-test on the residuals' sum of squared errors to evaluate the statistical significance of their variation.

It is suggested that model selection aims at the model which presents the smaller residuals, and also which does not have many parameters (Occam's Razor). Some indices, like AIC (Akaike Information Criterion) help support the model selection procedure, but the best choice relies on the geoscientist experience, and particular problem, as will be exemplified.

Significance tests are applied to distinguish between the parameters of the same rock physics recipe, calibrated to two different datasets, or between one rock physics model parameters and some hypothesized values. These analyses are performed by t and F tests, for single and multiple hypothesis testing, depending on the number of parameters being tested, as explained in Rawlings [16], and Paternoster [17]. These tests depend on an estimate of the parameters uncertainties, which will be dealt with in the next section.

Beyond significance tests, which compare the expected values of the parameter estimates alone, the user may be interested, in practice, in how big the difference between the predictions of two models is. The effect size is the difference of the model predictions divided by the pooled residual standard deviations, and it is a measure of the strength in the difference of the models' predictions relative to their accuracies.

D: Prediction

The most important use of a rock physics model is to predict elastic properties given a new geologic setup. Using rock physics prediction, the geoscientist can, for example, simulate seismic response for a reservoir with smaller porosity, or under greater effective stress (greater overburden), or even saturated with a different fluid.

On the other hand, to yield a robust prediction, it is of utmost importance to estimate the prediction uncertainty, or confidence band. The confidence band is obtained either approximately by linearization of the model, as demonstrated in Rawlings [16] and Bates [13], or by Markov Chain Monte Carlo (MCMC), as discussed in [18], [19], [20], [21], and [22]. Grana [23] applies the concept to rock physics models, taking into account the input

uncertainties, while the present paper focuses on the parameter uncertainties.

The MCMC method takes the nonlinearities into account and is more advisable if the rock physics model is highly non-linear with respect to the parameters, or if the optimum parameters are too close to their bounds. In the non-linear regression problem, the conditional probabilities are given by Equations (3) to (5).

$$p\left(\beta\middle|\{(y_i,x_i)\}_{i=1}^{N},\sigma_1,\sigma_2,\sigma_3\right) \propto exp\{-U(\beta)\}\prod_{j=1}^{m} Unif\left(\beta_j^{min},\beta_j^{max}\right) \quad (3)$$

$$p\left(\sigma_1,\sigma_2,\sigma_3\middle|\{(y_i,x_i)\}_{i=1}^{N},\beta\right) = \prod_{k=1}^{3}\mathcal{IG}\left(\sigma_k\middle|N+1,\frac{1}{2}SSE_k\right) \quad (4)$$

$$p\left(y\middle|x,\{(y_i,x_i)\}_{i=1}^{N},\beta,\sigma_1,\sigma_2,\sigma_3\right) = \mathcal{N}(y|F(x_i|\beta);\sigma^2) \quad (5)$$

Thus, the sampling method uses a Gibbs strategy, where each iteration consists of three steps (analogous to the strategy used by Sinay [21]):

- Sampling the parameter values. Since Equation (3) is non-linear in the parameters, we use an approximate transition proposal given by a Gaussian centered on the previous sampled vector of parameters, and with covariance equal to a scaled version of the linearized parameter covariance. This sampling step follows the Metropolis-Hastings criteria for acceptance, as explained in Gilks et a. [18].
- Sampling of the residuals standard deviations, given the inverse gamma distribution in Equation (4). Where $N+1$ is the shape parameter, and $\frac{1}{2}SSE_k$ is the scale parameter.
- Sampling the predicted elastic properties, given the Gaussian in Equation (3).

Results and Discussion

To apply aforementioned concepts, a dataset was downloaded from [24]. From this dataset we used digital logs plus static pressure from an unknown well.

Shale volume, and total porosity logs were interpreted from GR, and NPHI, respectively.

The fluid densities were estimated from the static measurements, and the fluid bulk moduli were estimated from both the static pressures and the temperature measurements, by using Batzle and Wang's formulas [25].

The rock physics model was based on the Thomas-Steiber model for dispersed clay sand-shale mixture, described in Marion [26] and Mavko et. al. [12]. This model considers that some clay fills the pore space of the sandstone (shaly sand regime); and, as the clay volume increases beyond the critical porosity, the rock becomes matrix supported (sandy shale regime).

For the density prediction, simple weighted mean of the component materials was used. For the elastic properties, the calculation workflow depends upon the shale volume. If the shale volume is smaller than 40%, then the shaly sand workflow is considered, otherwise sandy shale is used.

Both workflows depend on the wet shale estimates for bulk and shear moduli. They are obtained using self-consistent model, explained in Wu [27], and Mavko et. al. [12].

The sandy shale bulk and shear moduli were calculated using Hashin-Shtrikman [12] lower bound average between sand grain properties and wet shale properties as in [4].

The shaly sand bulk and shear modulus were calculated using soft-sand model [4] (given in Figure 1): modified Hertz-Mindlin to find the properties of clean sand at the critical porosity, followed by modified Hashin-Shtrikman lower bound interpolation to find the clean sand properties at the formation porosity, and Gassmann fluid inclusion, for the clean sand saturated with the formation fluid.

Finally, to calculate the shaly sand properties, a Hashin-Shtrikman lower bound average is used to interpolate between the saturated clean sandstone and the critical mixture properties. Where the critical mixture, is a sandstone, with all its pore volume filled with clay, and its properties are calculated with a Hashin-Shtrikman lower bound between clean sandstone and wet shale properties.

The model was calibrated for the interval covering two reservoirs. The parameters of the model are listed below, in Table 1. The black parameters are fixed, and the red bold ones were optimized. As an exercise, we start by adjusting seven parameters for both reservoirs, and in the next paragraphs, we will compare the calibrations to each reservoir, and then study the effect of reducing the number of free parameters.

Figure 2 shows the fit using the optimized parameters, and the confidence bands at 80% confidence level.

Table 1: Parameters of the rock physics model. Those subject to optimization are in red. The last columns are the confidence bands, at 80% level, for the parameters, which were calculated using MCMC technique.

Parameters	Values	P10	P90
f	1.0	0.98	1.00
Sand Critical Porosity (%)	64	61.9	66.8
Clay Density (g/cm³)	2.58	2.57	2.59
Clay Bulk Modulus (GPa)	23.4	22.9	24.3
Clay Shear Modulus (GPa)	9.9	9.7	10.0
Sand Grain Bulk Modulus (GPa)	38.8	35.6	40.4
Sand Grain Shear Modulus (GPa)	26.5	25.9	28.0
Quartz Density (g/cm³)	2.75	-	-
Coordination Number	6	-	-
Pure wet Shale Porosity (%)	20	-	-
Shale pore aspect ratio	0.05	-	-
Water Bulk Modulus (GPa)	2.8	-	-
Water Density (g/cm³)	1	-	-
Oil Bulk Modulus (GPa)	0.7	-	-
Oil Density (g/cm³)	0.7	-	-
Gas Bulk Modulus (GPa)	0.3	-	-
Gas Density (g/cm³)	0.35	-	-

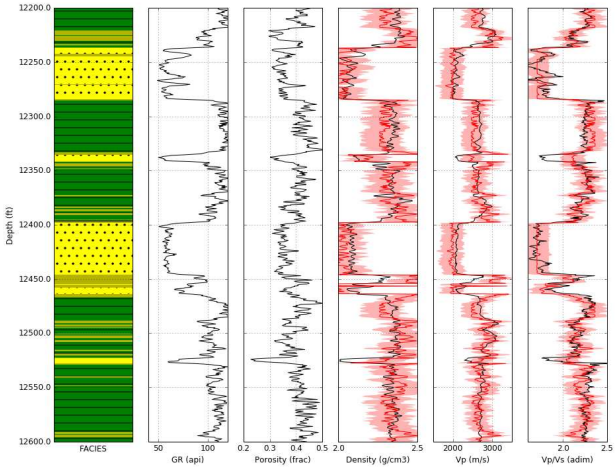

Figure 2: Well logs from the case study. From left to right: rock types (green is shale, brown is intermediate V_{sh} facies, and yellow is sandstone). Then the Gamma Ray curve, the Neutron Porosity, Density, v_P, and v_S. The red curves are the predicted elastic curves, and the red region is the 80% confidence band.

As a first exercise, two new calibrations were performed using two different datasets. One is the top oil reservoir (around 12250 ft depth), while the other is the lower oil reservoir (around 12400 ft depth), both illustrated in Figure 2. The aim is to compare both reservoirs' microstructures, which are represented by their model parameters. Avseth et al. [1] apply the same reasoning, point wisely, by jointly inverting the v_S and porosity logs to cementation, and sorting, which they use in their rock physics model. In this paper, we treat the parameters as constant values representative of the entire dataset, thus we can use them to compare two different datasets.

Based on the statistical tests, the two reservoirs have different microstructures. For example, their calibrated clay densities are 2.61 g/cm^3 (0.01 g/cm^3 standard deviation with 394 degrees of freedom) and 2.56 g/cm^3 (0.01 g/cm^3 standard deviation with 334 degrees of freedom) respectively, and the null hypothesis of equality between these two parameters must be rejected, according to the t-test. The global null hypothesis that all the parameters are equal is rejected, according to the F-test. Table 2 summarizes the comparison.

Table 2: Comparison of the two reservoirs. The last column gives the significance t-tests at a 95% confidence level for each model parameter. The bottom row, shows the result of the F-test for the whole set of parameters.

Parameters	R1	R2	t-tests
f	1	1	-
Sand Critical Porosity (%)	64.5	62.7	-
Clay Density (g/cm^3)	2.61	2.56	Reject
Clay Bulk Modulus (GPa)	26.3	20.8	Reject
Clay Shear Modulus (GPa)	10.6	9.7	Reject
Sand Grain Bulk Modulus (GPa)	31.7	50.0	Reject
Sand Grain Shear Modulus (GPa)	26.0	24.9	-
F-test: Reject			

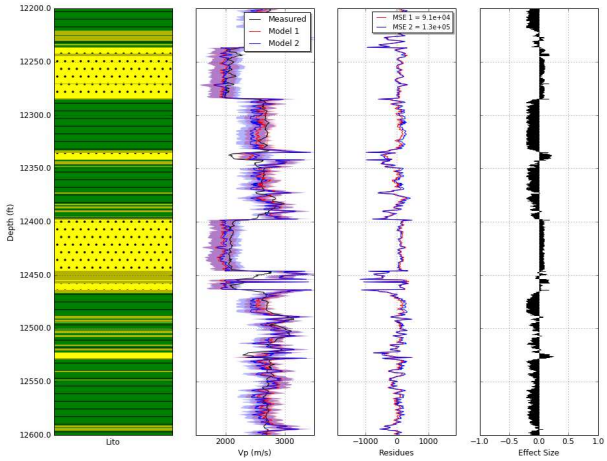

Figure 3: Comparison between the predictions of the rock physics models calibrated for the two reservoirs (Model1 refers to the upper reservoir and Model 2 to the lower reservoir). The effect sizes is insignificant in the reservoir interval.

On the other hand, the effect size is much smaller than a unit at the reservoirs, for all the predicted variables, as illustrated in Figure 3 for v_P. Thus, it can be concluded that the differences in the upper and lower reservoirs' parameters yield insignificant impact on their elastic behavior. From a practical point of view, both calibrations could be used for both reservoirs.

Another exercise is to select between models. It is compared the difference between using the aforementioned dispersed clay model, and a laminar shale model, which considers that the rock is composed of a interbedding of clean sand and shale layers (as used by Avseth et al. [1].

Figure 4 shows the effect size of the difference between the dispersed and laminar models, for v_P. The effect sizes are significant in the entire well interval, and additionally, it yields an improved fit for the lower reservoir. The residuals of the second model have less outliers, as it describes better the intermediate facies' elastic behavior. This analysis gives evidence to use the laminar model over the original dispersed clay model.

The present workflow is intended to help the geophysicist understand the underlying rock physics behavior of the rock. It does not, however, substitute more direct analysis. Thus, further laboratory rock analysis is reinforced to help increase confidence in the model choice, and its parameter values.

As a third exercise, the confidence bands, calculated by the MCMC method, are compared with the confidence bands estimated by linear approximation. Figure 5 shows the marginal 95% confidence ellipse for the parameters clay bulk and shear moduli. The Rock Physics model considered is the dispersed clay model, calibrated for the whole interval in Figure 2.

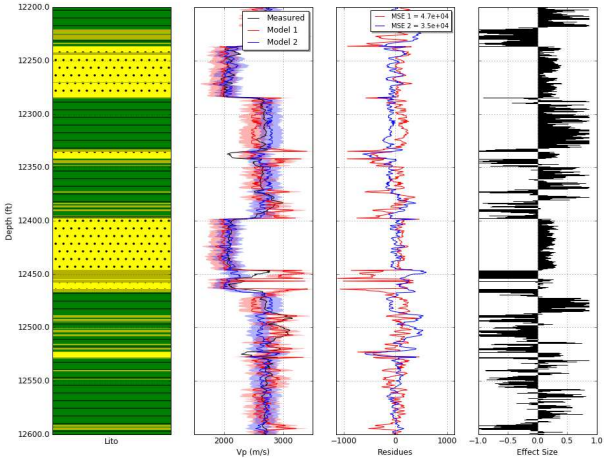

Figure 4: Comparison between the predictions of two rock physics models calibrated for the same dataset (Model1 is the dispersed clay model, and 2 is the laminar shale model). The effects sizes are significant throughout the complete well interval, and Model 2 yield an improved fit in the lower reservoir.

The black ellipse is the confidence calculated by linear approximation. The red points in Figure 5 are the parameter values sampled with MCMC technique, and the green ellipse is calculated from fitting a Gaussian to the sampled points. The black dot is the optimum parameter vector. In this example, both ellipses are similar, showing that rock physics model is approximately linear in these parameters.

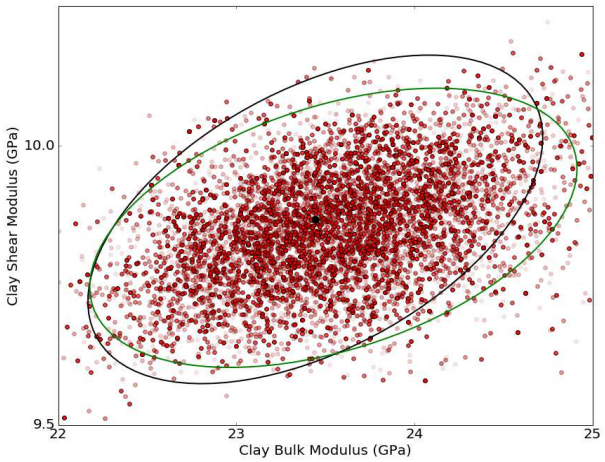

Figure 5: Comparison of the MCMC (green) and the linear approximation (black) 95% marginal confidence region for clay bulk and shear moduli. The red dots are 50000 samples drawn from the MCMC. The black dot is the optimum parameter vector. It can be seen that both methods yield similar confidence regions.

Conclusions

This paper showed how to quantitatively calibrate and analyze rock physics models. Using appropriate statistical tools and data structure the authors believe that it is possible to extract more reliable information from the rock measurements, than based on simple crossplot analysis or linear empirical relations.

Using the proposed data structure the user can easily implement famous rock physics recipes, or change them to create custom made models.

The paper demonstrated how to use nonlinear least-squares to simultaneously calibrate all the elastic properties.

The proposed framework enables the user to model simultaneously rock physics parameter and prediction uncertainties, through MCMC.

The framework makes it possible to compare the microstructure of two different reservoirs, by calibrating the same rock physics model to both reservoirs and comparing their parameters, and predictions. It is also showed how to use statistical tools to compare the performance of different rock physics models for the same reservoir.

Effect size is a statistical tool, which was demonstrated to yield more interpretable results than usual significance t and F tests.

The rock physics calibration described in this paper can be used to support quantitative seismic interpretation, as shown in [5-10].

Acknowledgments

The authors are grateful to Petrobras for the time available for the development of the present work, through the Master's program of Erick Talarico. The authors thanks all the colleagues in Petrobras, in special, Adriano Barreto, Daniel Bosco, Elionardo Pintas, Henrique Fraquelli, José Lira, and Matheus Cáfaro for the valuable discussions. Finally, the authors are grateful to Dr. Ekwere J. Peters for making available his dataset.

References

[1] P. Avseth, A. Jorstad, A.-j. V. Wijngaarden and G. Mavko, "Rock physics estimation of cement volume, sorting, and net-to-gross in North Sea sandstones," *The Leading Edge,* pp. 98-108, 2009.

[2] P. Avseth and N. Skjei, "Rock Physics modelling of static and dynamic reservoir properties - a heuristic approach for cemented sandstone reservoirs," *The Leading Edge,* pp. 90-96, January 2011.

[3] S. Xu and R. White, "A new velocity model for clay-sand mixtures," *Geophysical Prospecting,* vol. 43, pp. 91-118, 1995.

[4] P. Avseth, T. Mukerji, G. Mavko and J. Dvorkin, "Rock-physics diagnostics of depositional texture, diagenetic alterations, and reservoir heterogeneity in high-porosity siliciclastic sediments and rocks — A review of selected models and suggested work flows," *Geophysics,* vol. 75, pp. A31-A47, September 2010.

[5] D. Grana, "Bayesian inversion methods for seismic reservoir characterization and time-lapse studies," 2013.

[6] T. Coléou, F. Allo, R. Bornard, J. Hamman and D. Caldwell, "Petrophysical Seismic Inversion," in *SEG/Houston 2005 Annual Meeting*, 2005.

[7] H. Hammer, O. Kolbjornsen, H. Tjelmeland and A. Buland, "Lithology and fluid prediction from prestack

seismic data using a Bayesian model with Markov process prior," *Geophysical Prospecting,* 2011.

[8] T. Mukerji, A. Jørstadz, P. Avseth, G. Mavko and J. R. Granli, "Mapping lithofacies and pore-fluid probabilities in a North Sea reservoir: Seismic inversions and statistical rock physics," *Geophysics,* vol. 66, pp. 988-1001, July 2001.

[9] R. Bachrach, "Joint estimation of porosity and saturation using stochastic rock-physics modeling," *Geophysics,* vol. 71, pp. O53-O63, September 2006.

[10] P. Avseth, T. Mukerji and G. Mavko, Quantitative Seismic Interpretation, Paperback ed., Cambridge University Press, 2010.

[11] R. Bachrach and P. Avseth, "Rock physics modeling of unconsolidated sands: Accounting for nonuniform contacts and heterogeneous stress fields in the effective media approximation with applications to hydrocarbon exploration," *Geophysics,* vol. 73, pp. E197-E209, November 2008.

[12] G. Mavko, T. Mukerji and J. Dvorkin, The Rock Physics Handbook, Second ed., Cambridge University Press, 2009.

[13] D. Bates and D. Watts, Nonlinear Regression Analysis and Its Applications, John Wiley & Sons, 1988.

[14] "SciPy.org," [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html. [Accessed 13 12 2016].

[15] A. Simkooei and S. Jazaeri, "Weighted total least squares formulated by standard least squares theory," *Journal of Geodetic Science,* pp. 113-124, April 2012.

[16] J. O. Rawlings, S. G. Pantula and D. A. Dickey, Applied Regression Analysis: A Research Tool, second edition ed., Springer, 1932.

[17] R. Paternoster, R. Brame, P. Mazerolle and A. Piquero, "Using the correct statistical test for the equality of regression coefficients," *Criminology,* vol. 36, pp. 859-866, 1998.

[18] W. R. Gilks, S. Richardson and D. J. Spiegelhalter, Markov Chain Monte Carlo in Practice, Springer-Scinece +Business Media B. V., 1996.

[19] S. Ross, Simulation, 4th ed., Elsevier, 2006.

[20] S. Mazumdar, "Monte Carlo Methods for Confidence Bands in Nonlinear Regression," 1995.

[21] M. S. Sinay and J. S. J. Hsu, "Bayesian Inference of a Multivariate Regression Model," *Journal of Probability and Statistics,* vol. 2014, November 2014.

[22] W. Hu, J. Xie, H. W. Chau and B. C. Si, "Evaluation of parameter uncertainties in nonlinear regression using Microsoft Excel Spreadsheet," *Environmental Systems Research,* 2015.

[23] D. Grana, "Probabilistic approach to rock physics modeling," *Geophysics,* vol. 79, no. 2, pp. D123-D143, 2014.

[24] E. J. Peters, Advanced Petrophysics, Volume 2: Dispersion, Interfacial Phenomena/Wettability, Capillarity/Capillary Pressure, Relative Permeability, Austin, Texas: Live Oak Book Company, 2012. Table B2: Digital Log Data, Available at http://www.advancedpetrophysics.com/data-files.php.

[25] M. Batzle and Z. Wang, "Seismic properties of pore fluids," *Geophysics,* vol. 57, pp. 1396-1408, November 1992.

[26] D. Marion, A. Nurt, H. Yint and D. Han, "Compressional velocity and porosity in sand-clay mixtures," *Geophysics,* vol. 57, pp. 554-563, April 1992.

[27] T. T. Wu, "The Effect of Inclusion Shape on the Elastic Moduli of a Two-phase Material," *International Journal Solids Structures,* vol. 2, pp. 1-8, 1966.

[28] R. D. Cook, C. L. Tsai and B. C. Wei, "Bias in Nonlinear Regression," *Biometrika,* vol. 76, pp. 615-623, 1986.