Gaussian Process Modeling of Well Logs

Andy Rawlinson
IBM Research – Australia
Melbourne, Australia
Email: arawlins@au1.ibm.com

Shrihari Vasudevan
IBM Research – India
Bangalore, India
Email: shrivasu@in.ibm.com

***Abstract*—Well log data typically consist of measurements of different physical features, together with a myriad of characteristics, ranging from varying levels of uncertainty, complexity, completeness and correlation. In the oil and gas industry, further exploration, development and production are contingent upon the information derived from the data. The availability of such data in electronic form allows one to perform more advanced analyses. This paper attempts to show the versatility of Gaussian processes to model well log data and to make predictions and quantify uncertainty at locations for which no data are available.**

Keywords–Gaussian processes, machine learning, well logs

I. Introduction

Well logs consist of measurements of physical characteristics of an oil/gas well as a function of depth and location. Measurements may include density, resistivity, neutron porosity, gamma ray porosity, spontaneous potential etc.

Geological core and well log analysis and interpretation is normally carried out manually by an expert and is a labour intensive process [1]–[3]. These data help geologists gain knowledge about the region of interest and also potential locations of natural resource deposits. This may also include inferring features at locations for which no data are available.

Conventional well log analysis may include cross plots, where scatter plots or pairs or triplets (if points are coloured) of well log data sets, such as Pickett plot (e.g. resistivity-porosity or resistivity-porosity-gamma ray) or Hingle plot (e.g. bulk density-conductivity) [4], [5]. These plots are used to determine fluid saturation characteristics of a formation, where the fluid may contain a mixture of water and hydrocarbons. The relationship between water saturation and various well log data can be expressed through Archie's equation.

Within a single well, different log data sets can have varying degrees of uncertainty, complexity (e.g. fluctuating) and may be incomplete (e.g. gaps). Data from single and multiple wells may also be correlated. Overall, well logs are regarded as a reflection of the underlying resource and are correlated with production data.

Given the high cost of analysing and drilling new wells, and the fact that in some cases data is available in electronic form (historically, much of the well log data is in paper format), a natural question arises as to whether the process of analysing log data sets can be facilitated by more advanced statistical and machine learning techniques. A wide range of such methods have been applied – a small set of examples being where data are assumed to be multifractional Brownian motions and Holder or Hurst exponents computed [6], [7]; Bayesian deconvolution [8]; and continuous wavelet transforms [9].

Current analysis techniques may be inadequate in detecting particular features across multiple data sets. In this paper it is proposed that Gaussian processes (GPs) [10] can be used to address and overcome many of these issues. GPs can be constructed to model uncertain, incomplete, complex, noisy and correlated data from single or multiple sources. They are used to predict particular features and also assign a degree of confidence in those predictions. A related investigation has been performed by [11], using functional principal component analysis (fPCA). GPs can also be used to extract geochemical information from well logs [12].

II. Well Log Data

Well log data contains information about the lithology (i.e. physical characteristics of rocks) of the geological formation at the well location (e.g. see [13]). Some typical logs are listed in Table I, together with plots in Fig. 1. The analysis presented here uses a portion of the data set examined by [11], where they consider well logs from 2020 vertical wells and production data from 702 horizontal wells in a single field. Throughout this paper, values on axes in some plots have been removed to de-identify the data.

III. Gaussian Processes

Gaussian processes are stochastic processes where any finite subset of random variables is jointly Gaussian distributed. They can be viewed as a Gaussian probability distribution in function space and is described by a mean function $\mu(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$, which specifies a distribution over functions. The covariance function (also called the kernel function) attempts to provide a stochastic model of the spatial correlations between the data points. GPs are nonparametric learning techniques that can be applied to a wide variety of problems [10], [14]. They are ideally suited to well logs as they can provide a multi-resolution model of the data.

Well log data type	Description
Caliper	Diameter of bore/well, may be wider in regions of softer rocks due to material collapsing into the well, e.g. Fig 1(A)
Neutron porosity	Indicates amount of hydrogen (e.g. from hydrocarbons and/or water) in the formation, Fig 1(B)
Gamma ray	Records natural radioactivity. Shales have high readings, while sandstones and carbonates have low readings, Fig 1(C)
Deep induction resistivity	Measure of the electrical characteristics of the formation. Also called resistivity in this paper. High/low resistivity is related to presence of hydrocarbons/fluids in the formation, Fig 1(D)
Spontaneous potential	Measure of potential difference between the surface and bore hole at a given depth, Fig 1(E)
Bulk density	Measures porosity and can differentiate between presence of gases and liquids in the formation, Fig 1(F)
Sonic	Measures time taken for sound to travel through the formation.

TABLE I: Examples and descriptions of some well log data types.

Known or input data are used to 'train' the GPs to determine an optimal set of parameters that best describes the data, by minimising the negative marginal likelihood estimate of the problem. This step effectively relates the data set to a small set of parameters (about 5 parameters for the examples in this paper). These parameters are then used in the same GPs to make predictions at other locations.

Advantages of GPs include the ability to make predictions and quantifying uncertainty at any resolution. They can be used in multi-dimensional problems – such as 1-dimension (e.g. data as a function of depth only) or 3-dimensional (e.g. data as a function of latitude, longitude and depth). They are also ideally suited to a wide range of problems where data may be incomplete, noisy and interpolations are needed.

Given a spatial data set, the coordinates of the $i-$th datum are defined by :–

$$\mathbf{x}_i = \begin{cases} z_i & \text{for 1 dimension} \\ (x_i, y_i, z_i) & \text{for 3 dimensions} \end{cases} \tag{1}$$

and the value of a physical quantity recorded at this location is v_i. The length scale matrix is defined by :–

$$\Sigma = \begin{cases} l_z^{-2} & \text{for 1 dimension} \\ \begin{bmatrix} l_x & 0 & 0 \\ 0 & l_y & 0 \\ 0 & 0 & l_z \end{bmatrix}^{-2} & \text{for 3 dimensions,} \end{cases} \tag{2}$$

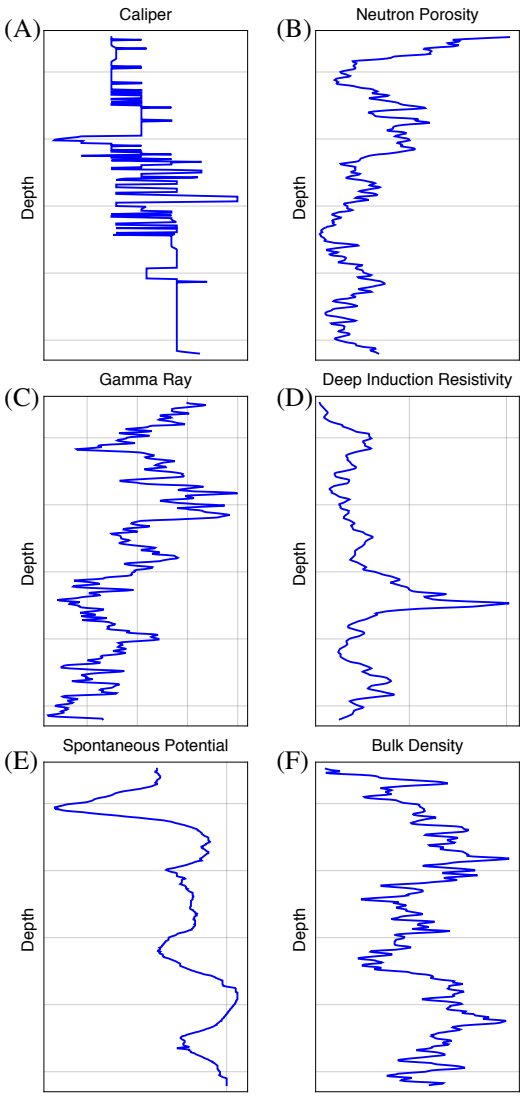

Figure 1: Plots of particular physical features from a single well log. Different features show different levels of noise/uncertainty. Note the apparent anti-correlation between the gamma-ray and deep induction resistivity plots.

with length scale parameters :–

$$\ell = \begin{cases} l_z & \text{for 1 dimension} \\ (l_x, l_y, l_z) & \text{for 3 dimensions,} \end{cases} \tag{3}$$

which are a measure of how the modelled function varies with the coordinates x_i. Let

$$r = (\mathbf{x} - \mathbf{x}')^T \Sigma (\mathbf{x} - \mathbf{x}'), \tag{4}$$

where T denotes the transpose of the matrix. The signal and noise variance of the data are defined by σ_f^2 and σ_n^2 respectively.

Various kernels, with associated kernel parameters θ, can be used for a Gaussian process (GP) :–

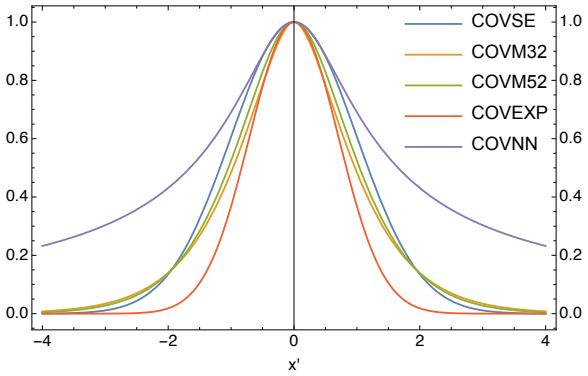

Figure 2: A plot showing the generic differences of 1-dimensional GP kernels. ($COVSE$, $COVM32$, $COVM52$, $COVEXP$, $COVNN$) are the squared-exponential, Matern 3/2, Matern 5/2, exponential and neural network kernels respectively. In all cases, $\mathbf{x} = 0$, $\sigma_f = 0$, $\ell_x = 1$ and for the neural network case, $\beta = 1$.

- *squared-exponential* kernel and $\theta = (\ell, \sigma_f, \sigma_n)$

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left[-\frac{r}{2}\right] \tag{5}$$

- *Matern 3/2* kernel and $\theta = (\ell, \sigma_f, \sigma_n)$

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \sqrt{3r}\right) \exp\left[-\sqrt{3r}\right] \tag{6}$$

- *Matern 5/2* kernel and $\theta = (\ell, \sigma_f, \sigma_n)$

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \sqrt{5r} + \frac{5r}{3}\right) \exp\left[-\sqrt{5r}\right] \tag{7}$$

- *exponential* kernel and $\theta = (\ell, \sigma_f, \sigma_n)$

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left[-\sqrt{r}\right] \tag{8}$$

- *neural network* kernel and $\theta = (\ell, \sigma_f, \beta, \sigma_n)$

$$k(\mathbf{x}, \mathbf{x}') = \frac{6\sigma_f^2}{\pi} \arcsin \left[\frac{\beta + 2\mathbf{x}^T \Sigma \mathbf{x}'}{\sqrt{(1 + \beta + 2\mathbf{x}^T \Sigma \mathbf{x})(1 + \beta + 2\mathbf{x}'^T \Sigma \mathbf{x}')}}\right] \tag{9}$$

  where β is the bias factor.

Figure 2 shows the generic differences between the various kernels.

Let $X = \{\mathbf{x}_i\}$ be the set of training inputs (e.g. depths or locations) and the set of training outputs $\mathbf{v} = \{f(\mathbf{x}_i)\} = \{v_i\}$. The log marginal likelihood of the training outputs $\mathbf{v}$, given the locations X and parameters θ is defined by :–

$$\log p(\mathbf{v}|X, \theta) = -\frac{1}{2}\mathbf{v}^T K^{-1}\mathbf{v} - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi, \tag{10}$$

where $K = K_f + \sigma_n^2$ is the covariance matrix of the noisy targets $\mathbf{v}$, K_f is the covariance matrix for the noise free targets, whose (i, j)-th component is given by :–

$$K_f(i, j) = k(\mathbf{x}_i, \mathbf{x}_j). \tag{11}$$

$|K|$ is the determinant of the matrix K and n is the number of data points used in the training. The kernel $k(\mathbf{x}_i, \mathbf{x}_j)$ can be any of those listed in (5) $\rightarrow$ (9). The three terms of (10) are the data fit, penalty and normalisation terms respectively. Given an input data set X, the optimal set of parameters can be determined by minimising the negative log marginal likelihood (10). A combination of simulated annealing and Broyden-Fletcher-Goldfarb-Shanno (BFGS) Hessian update methods are used to produce the results in this paper.

Using the optimised GP parameters together with the n training inputs X and n outputs $\mathbf{v}$, the expected function values $\bar{f}_*$ and their covariances at n_* test locations $X_* = \{\mathbf{x}_{*i}\}$ are calculated :–

$$\begin{bmatrix} \mathbf{v} \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{12}$$

$$\bar{f}_* = K(X_*, X)\left[K(X,X) + \sigma_n^2 I\right]^{-1}\mathbf{v} \tag{13}$$

$$cov(f_*) = K(X_*,X_*) - K(X_*,X) \times \left[K(X,X) + \sigma_n^2 I\right]^{-1} K(X,X_*). \tag{14}$$

$K(X_*, X)$ is an $n_* \times n$ matrix of covariances evaluated at all test and training points; $K(X_*, X_*)$, $K(X, X_*)$ and $K(X, X)$ are similarly defined. $N(0, [\ldots])$ denotes the normal distribution with zero mean.

Regressing with GPs takes advantage of the fact that training data and test data of a GP are jointly Gaussian distributed. Equations (12)-(14) are the standard GP regression equations for the problem.

GPs have been applied to a large variety of data types with great success. In this paper we demonstrate how they can be applied to well logs data.

IV. GPs on a single well log

These methods are now applied to a single well log 'deep induction resistivity' data set, also referred to as 'resistivity' in this paper. As resistivity is a positive definite quantity, the data is scaled/transformed (in this case, simply take the natural logarithm). The GP learning and prediction algorithms are applied, then the inverse scale/transform process is performed and the results analysed and plotted.

Figure 3(A) shows an example of a resistivity log data set and the set of open circles (about 40 points) used to train, i.e. 'learn', the parameters $(\ell, \sigma_f, \sigma_n, \beta)$ of the different kernels of the GPs. Using this learned set of parameters in the kernel, the resistivity (Fig. 3(B)) and uncertainty (Fig. 4) at any depth are predicted by using (13) and (14) respectively.

A measure of the agreement between the predicted and actual results can be obtained by calculating the relative error. Figure 5 shows how the relative error varies as a function of the amount of data used to train the GPs. It is seen that the neural network (COVNN) and Matern 3/2 (COVM32) kernels generally perform better than other kernels.

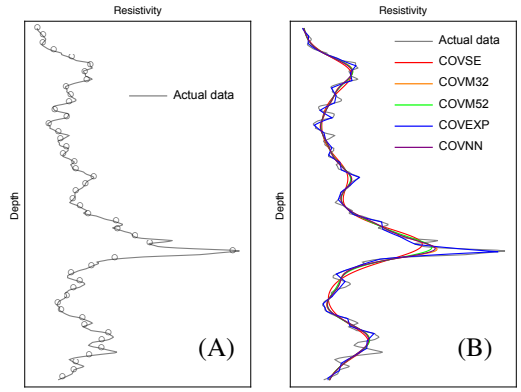

Figure 3: Plot (A) shows the actual deep induction resistivity data, with the open circles forming the set of points (about 10% of the data) used to train the GPs. The solid line is formed by joining the points of the complete data set. Plot (B) shows the values predicted by different kernels.

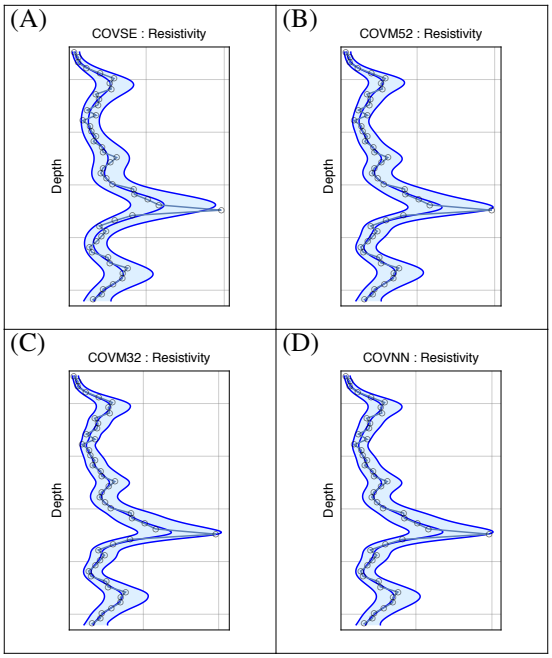

Figure 4: Comparison of different kernels on the same resistivity data set. Kernels considered are (A) squared-exponential (COVSE), (B) Matern-5/2 (COVM52), (C) Matern-3/2 (COVM32) and (D) neural-network (COVNN). In each plot, the central blue line is the GP prediction, while the shaded blue region is the 2σ uncertainty. About 10% of the input data was used to train the GPs.

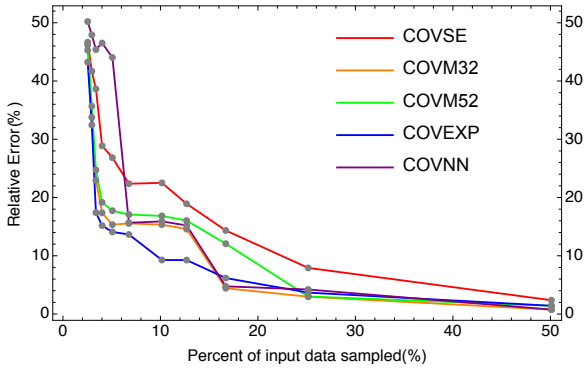

Figure 5: Comparison of relative errors on different kernels as a function of the fraction of input data used for training the GPs. Neural network (COVNN) and Matern 3/2 kernels (COVM32) generally perform better than other kernels.

The question of how accurate are the predictions of GPs can be addressed by considering cross validation of the results. To achieve this, a random sample of 10% of the input data is taken and used to train the GPs. The parameters are then used in the GP to make predictions at the locations of the remaining 90% of the data and then compared with the actual values at those locations. This process is repeated 10 times, each time with a different initial random sample. The mean squared error (MSE) and standard deviation of the squared error (StdDevSE) are then calculated.

Table II shows the results of the cross validation process for different kernels. Results obtained by linear interpolation and inverse distance weighted fitting (IDW) are also included. For linear fitting, a straight lines are connected to the input data points. The inverse distance weighted value v_*^{IDW} at test location $\mathbf{x}_*$ is given by :–

$$v_*^{IDW} = \frac{\sum_i w_i v_i}{\sum_i w_i} \tag{15}$$

with v_i the training outputs at $\mathbf{x}_i$ and weights w_i :–

$$w_i = \frac{1}{d_i^2} \tag{16}$$

where

$$d_i = \sqrt{(\mathbf{x}_i - \mathbf{x}_*).(\mathbf{x}_i - \mathbf{x}_*)}. \tag{17}$$

It can be seen that the neural network (COVNN), Matern 5/2 (COVM52) and Matern 3/2 (COVM32) kernels yield better results than the squared-exponential (COVSE) and exponential (COVEXP) kernels, as well as the linear fitting and inverse distance weighted fitting methods.

V. Kriging from single well to other locations

Given a single well log (e.g. located at $(x, y)_{well} = (0, 0)$), GPs can be used to infer the values of resistivity at locations away from the well by considering 3-dimensional GPs. If the input data to learn the GPs only consists of depth (z) series

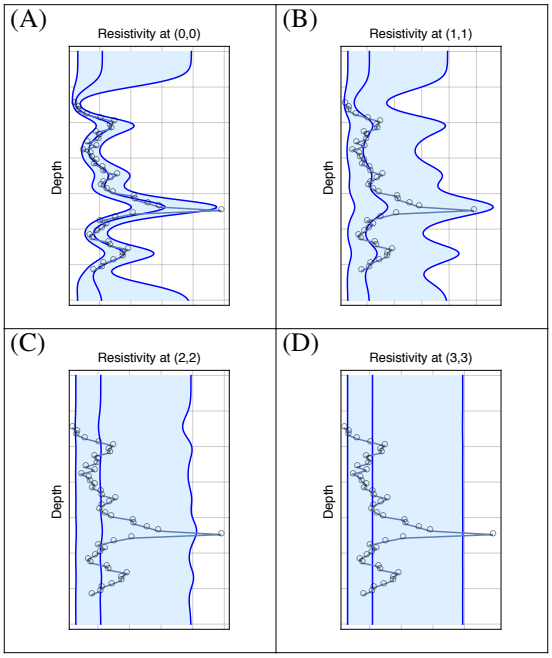

Figure 6: Application of 3D GPs for prediction (central blue line) and uncertainty (2σ, shaded blue region) of resistivity as you move away from well. The well is located at $(x, y)_{well} = (0, 0)$ and predictions are given for locations $(x, y) = (0, 0), (1, 1), (2, 2)$ and $(3, 3)$, measured in some units. A spherically symmetric squared-exponential kernel is used. The open circles and gray line are the resistivity input data at $(x, y, z) = (0, 0, depth)$ (plot (A)) and are shown in plots (B), (C) and (D) for comparison purposes only.

Kernel/method	MSE	StdDevSE
Neural network (COVNN)	0.0617	0.2386
Matern 5/2 (COVM52)	0.0622	0.2484
Matern 3/2 (COVM32)	0.0635	0.2649
Linear fit	0.2342	1.2466
Exponential (COVEXP)	0.2352	1.2444
Squared exponential (COVSE)	0.3769	1.6404
Inverse distance weighted	2.8740	14.662

TABLE II: Kernel/method dependence of cross validation results are ordered according to increasing mean squared error (MSE). Neural network and Matern type GP kernels performed better than other GP kernels, linear interpolation and inverse distance weighting techniques. Note the apparent correlation between the MSE and the standard deviation of the squared error StdDevSE.

data, a spherically (i.e. radially) symmetric 3-dimensional GP kernel can be used to model the data in (x,y,z). This can be considered as a form of kriging, an example of which is given in Fig. 6, where the central blue line denotes the prediction, i.e. mean, of the GPs, while the shaded region is the 2σ uncertainty. In this case we have a cylindrically symmetric result about the about the well.

Note in Fig. 6(A) that the prediction and uncertainty of the GPs at the location of the well is very similar to Fig. 4(A). At locations further and further away from the well the predictions (the central blue line) look more like a straight line whose value approaches the mean of the input data. The non-central location of the prediction (central blue line) relative to the upper and lower limits of the shaded 2σ uncertainty region is due to the fact that the resistivity is a positive definite quantity.

VI. Conclusion

In this paper it was demonstrated how GPs can be adapted to model well log data and show how predictions and uncertainty, or confidence levels, can be calculated. While the focus of the analysis was on deep induction resistivity, the techniques can be applied to any well log data type. Different kernels were explored and an assessment was carried out of the quality of the predictions as a function of the amount of data used to train the GPs. It was found that generally the neural network (COVNN) and Matern type kernels (COVM32 and COVM52) performed better than the squared-exponential kernel (COVSE).

A cross validation analysis was performed to compare GP based predictions with those obtained by linear interpolation fitting and inverse distance weighting. Again, in general, GPs performed better than other methods. Further, the neural network and Matern type GP kernels performed better than other GP kernels.

GP predictions and uncertainties at locations away from the well (i.e. kriging) using a single well log as input data were also explored. In this scenario, a radially symmetric kernel was used, and the range of the effects in the x and y directions is essentially determined by the length scale of the effects in the z-direction.

While emphasis was placed on a single well log in this paper, a future publication will present results on the application of GPs to a range of data integration and fusion problems [15], [16] based on multiple well logs. This will further consolidate the multifaceted capabilities of GPs. Comparison of GPs with other methods, e.g. functional principal component analysis (fPCA) will also be investigated.

References

[1] *Log Interpretation Principles/Applications*. Schlumberger, 1989.
[2] *Log Interpretation Charts*. Schlumberger, 2009.
[3] Q. R. Passey, S. Creaney, J. B. Kulla, F. J. Moretti, and J. D. Stroud, "A practical model for organ richness from porosity and resistivity logs," *The American Association of Petroleum Geologists Bulletin*, vol. 74, no. 12, 1990.
[4] G. Asquith and D. Krygowski, *Basic Well Log Analysis, 2nd Edition*. AAPG, 2006, AAPG Methods in Exploration Series, 16.
[5] D. V. Ellis and J. M. Singer, *Well Logging for Earth Scientists, Second Edition*. Springer, 2008.

[6] S. Gaci and N. Zaourar, "A new approach for the investigation of the local regularity of borehole wire-line logs," *Journal of Hydrocarbons Mines and Environmental Research*, 2010.

[7] S. Gaci, N. Zaourar, M. Hamoudi, and M. Holschneider, "Local regularity analysis of strata heterogeneities from sonic logs," *Nonlin. Processes Geophys.*, vol. 17, pp. 455–466, 2010.

[8] J. A. Christen, B. Sansó, M. Santana, and J. X. Velasco-Hernández, "Bayesian deconvolution of oil well test data using gaussian processes," 2012, https://www.soe.ucsc.edu/research/technical-reports/UCSC-SOE-12-05/download.

[9] L. Briqueu, N. Zaourar, M. Hamoudi, and D. Gibert, "Modelling the stochastic component of geophysical downhole measurements using scaling process. extraction of new attributes by continuous wavelet transform," *Proceedings of the International Conference on Modeling and Simulation (MS'07 Algeria)*, 2007.

[10] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[11] M. Kormaksson, M. R. Vieira, and B. Zadrozny, "A data driven method for sweet spot identification in shale plays using well log data (SPE-173455-MS)," *SPE Digital Energy Conference and Exhibition*, 2015.

[12] D. Cornford, P. Farrimond, and D. S. and Chris Cornford, "Extracting geochemical information from well logs," *26^{th} International Meeting on Organic Geochemistry (IMOG2013)*, 2013.

[13] S. Salehi and B. Honarvar, "Automatic Identification of Formation Lithology from Well Log Data: A Machine Learning Approach," *Journal of Petroleum Science Research*, vol. 3, no. 2, pp. 73–82, 2014.

[14] S. Vasudevan, F. Ramos, E. Nettleton, and H. Durrant-Whyte, "Gaussian Process Modeling of Large Scale Terrain," *Journal of Field Robotics*, vol. 26, no. 10, pp. 812–840, 2009.

[15] S. Vasudevan, "Data fusion using gaussian processes," *Elsevier Journal of Robotics and Autonomous Systems*, 2012.

[16] S. Vasudevan, A. Melkumyan, and S. Scheding, "Efficacy of data fusion using convolved multi-output gaussian processes," *Journal of Data Science*, 2014, based on arXiv report 1210.1928.