Prediction of hydrocarbon depth for seabed logging (SBL) application using Gaussian process

IOP Conf. Series: Journal of Physics: Conf. Series 1132 (2018) 012075 doi:10.1088/1742-6596/1132/1/012075

Prediction of hydrocarbon depth for seabed logging (SBL) application using Gaussian process

Muhammad Naeim Mohd Aris[1,a], Hanita Daud[1] and Sarat Chandra Dass[1]

[1]Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak, Malaysia

E-mail: [a]muhammad_naeim@yahoo.com

Abstract. Seabed Logging (SBL) is an application of marine Controlled-Source Electromagnetic (CSEM) technique to characterize hydrocarbon-filled layers underneath the seabed remotely in deep water regions. This technique maps structure of subsurface electrical resistivity in the offshore environment. Basically, exploration of offshore hydrocarbon is based on the contrast of electrical resistivity between hydrocarbon reservoir and its surrounding sea sediments. Modelling offshore hydrocarbon is a core analysis and time consuming task. Current numerical modelling techniques used in SBL application involve meshes and complicated mathematical equations. Thus, a simple supervised learning method which is Gaussian Process (GP) is proposed to process synthetic SBL data which are generated through Computer Simulation Technology (CST) software to predict the depth of hydrocarbon. This statistical model is able to provide additional hydrocarbon information by utilizing the prior information. 1-dimensional (1-D) forward GP models have successfully been developed to predict the presence of hydrocarbon and as the continuation work, 2-dimensional (2-D) forward GP model is then developed to be used as the standard profile in order to predict the depth of hydrocarbon. This shall give indication that GP can be used as the methodology to predict the depth of hydrocarbon reservoir underneath the seabed.

1. Introduction

Nowadays, oil and gas demand is increasing in world market and this situation has forced petroleum industry to explore hydrocarbon reservoirs in deep water environment. Seabed Logging (SBL) application has been introduced to oil and gas industry to characterize hydrocarbon-filled reservoirs underneath the seabed in deep water regions. This application can make reliable measurements of subsurface resistivity underneath the seabed using marine Controlled-Source Electromagnetic (CSEM) technique. Marine CSEM applications are widely used for offshore hydrocarbon exploration. CSEM has been considered as a vital tool for interpretation of data to reduce ambiguities and exploratory risk since last decade.

This paper fitted Gaussian Process (GP) to synthetic SBL data generated through CST software in order to predict the depth of hydrocarbon underneath the seabed. GP is a non-parametric approach where it places a prior on the space of function, f , directly without parameterizing the function. GP follows Bayesian setting where it begins with a prior distribution and updates the observed data, and then producing the posterior distribution over functions. It is able to handle case of multiple different forms of data as long as covariance function is defined for each data type. This gives good implication that GP is able to be used as a processing tool in SBL application.

2. Literature Reviews

This section describes the background of Seabed Logging (SBL) and Gaussian Process (GP).

2.1. Seabed Logging (SBL)

Seabed Logging (SBL) is a new complementary application of frequency-domain marine CSEM technique for seismic method. According to [1], the use of marine CSEM technique in hydrocarbon exploration is based on assumption that hydrocarbon-saturated formations possess a significant higher electrical resistivity compared to its surroundings. Hydrocarbon is said to have higher electrical resistivity (30-500Ωm) compared to seawater (0.5-2Ωm) and its surrounding sedimentary rocks (2-3Ωm). This technique has been well embraced by some national and huge oil companies such as BP, Statoil, PETRONAS, Shell, Reliance, etc. [2]. It also has been successfully used in different geological structures of offshore environments around the world such as offshore of Brazil, Norway, West Africa and the Gulf of Mexico [3]. Figure 1 shows the corresponding layout of SBL application in offshore environment.

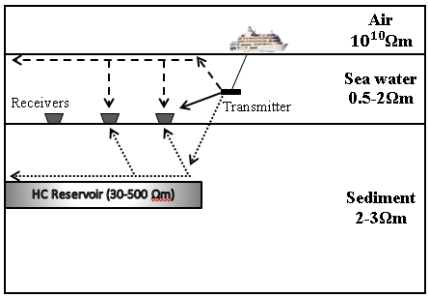

Figure 1. Schematic representation of SBL application.

The basis for this technique is the use of a mobile Horizontal Electrical Dipole (HED) transmitter and an array of receivers, which are magnetic and electric sensors [4] that are positioned on the seabed relative to the source. EM energy is continuously generated by the transmitter during SBL survey. HED transmitter emits low frequency EM signal (typically 0.01-10Hz) into the water column and downward into the seabed. It is typically elevated by a ship at 30m-40m above the seabed [5]. Practically, signals that have been reflected back by various structures are recorded and measured by the seabed receivers.

In SBL application, low frequency EM signal is used to obtain higher wavelength and also to decrease attenuation of the signal [6]. High frequency of EM wave influences the attenuation rate and this affects the outcome as it is the function of distance [7]. The EM wave must be able to well propagate through the different structures in order to provide good information about the subsurface underneath the seabed. SBL application is more suitable in deep seawater due to air-wave phenomenon [8]. This air-wave phenomenon does not contain any information about the seabed resistivity. Researchers in [9] stated that the contribution of air-wave is insignificant when marine CSEM conducted in deeper offshore environment. The details of air-wave can be referred in [10].

2.2. Gaussian Process (GP)

Gaussian Process (GP) is known as a machine of probabilistic kernel that is not only provide the prediction of mean value but also provides variance around the predicted mean as a description of associated uncertainty in the predictions. GP shows error bars for each prediction and provides a full predictive distribution [11]. GP regression has demonstrated outstanding performance in a number of applications [12] and it has been adopted in many scientific and engineering applications [13]. GP regression is a famous statistical modeling approach in computer experiments. It has been widely used as the surrogate model for complicated mathematical models that typically requires longer processing

time [13]. To date, there is no study related to SBL data processing using GP in computer experiment in order to predict the depth of hydrocarbon for SBL application.

Conceptually, GP is a random variables collection, such that any these variables collection will have a joint Gaussian distribution [14]. GP is fully specified by a mean function, $m(x)$, and a covariance function, $k(x,x')$. The collection of random variables, $\{y(x_i); x_i \in D\}$, is said to follow a GP with a mean function, $m(x)$, and a covariance function, $k(x,x')$, if for any finite collection $x_1, x_2, ..., x_n \in D$, where $D = \{(x_i, y_i) \mid i = 1, ..., n\}$. The function is normally distributed and written as $Y = (y(x_1), ..., y(x_n))^T \sim N_n(\mu_n, \Sigma_n)$, where $\mu_n = (m(x_1), ..., m(x_n))$ and $\Sigma_n = (k(x_i, x_j))_{nxn}$. Based on the mathematical definition of GP, it follows that GP is defined over functions.

3. Methodology

This section describes the methodologies involved while conducting this study.

3.1. Synthetic data acquisition using CST software

The SBL models are replicated in rectangular cuboidal shape. There are three basic layers drawn in both models which are air, seawater and sediment. Figures 2(a) and 2(b) are the proposed non-hydrocarbon and hydrocarbon models, respectively, as referred in [15].

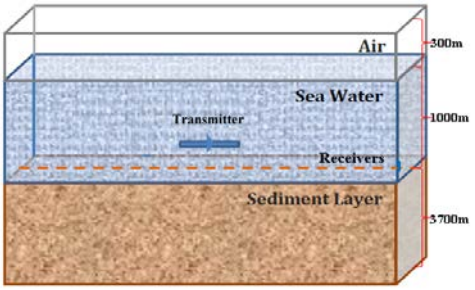

Figure 2(a). Non-hydrocarbon model.

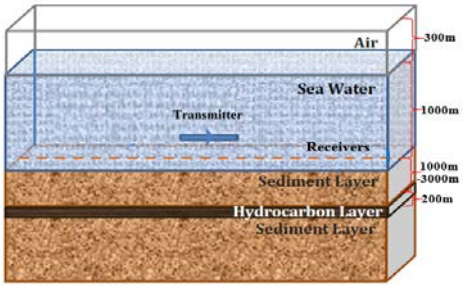

Figure 2(b). Hydrocarbon model.

The SBL models are proposed to have volume of 20×10×5km^3. The length of the models (offset) is assigned as 20km for the purpose of generating the best synthetic SBL data. The thicknesses of air, seawater and sediment for both models are fixed at 300m, 1000m and 3700m, respectively. For hydrocarbon models, the depth of the 200m thick hydrocarbon is varied from 1000m to 3000m with an increment of 250m. HED transmitter, which is 270m long [7] is used in this study. The current and frequency used in the simulations are 1250A and 0.125Hz, respectively. The position of the HED transmitter is fixed at center of 30m above the seabed. An array of receivers is positioned along the seabed (in-line) of the simulation models. The offshore environment is assumed to be free from any disturbances as well as other aspects which may exist in offshore environment and real SBL survey. Specific physical properties for each layer are assigned during the simulation as referred in [7].

Table 1. Physical properties for each layer.

Medium	Relative Permittivity (ε_r)	Electrical Conductivity (S/m)	Thermal Conductivity (W/mK)
Air	1	1.0 x 10^{-11}	0.024
Sea Water	80	1.630	0.593
Sediment	30	1.000	2.000
Hydrocarbon	4	0.002	0.492

3.2. Calibration of GP models

GP is a flexible model that provides effective predictions [16] for the unknown quantity. Assuming that there is a relationship between training set input, x and target outputs, Y, where $x=(s,h)$. s and h represent offset and hydrocarbon depth, respectively. Thus, Gaussian distribution is written as $Y=(y(x_1),...,y(x_n))^T \sim N_n(\mu_n,\Sigma_n)$ where μ_n is the mean and Σ_n is the covariance between outputs, Y, corresponding to training set, x for finite n locations.

GP is fully described by a mean function, $m(x)$, and the covariance function, $k(x,x')$, where $m(x)=\mathrm{E}(y(x_i))$ and $k(x_i,x_j)=\mathrm{E}[(y(x_i)-m(x_i)).(y(x_j)-m(x_j))]$ [17]. GP on target outputs, Y is written as $Y \sim GP[m(x),k(x_i,x_j)]$. Generally, zero mean function is used in the basic prior GP model [18]. Zero mean function prior does not imply zero mean predictive distributions. Thus, Gaussian prior distribution with zero mean is defined as $Y=(y(x_1),...,y(x_n))^T \sim N_n(0,\Sigma_n)$.

According to [16], the correlation between target output and input data is expressed by the value of covariance function. The well-liked choice of covariance function within the kernel machine fields is the squared exponential (SE) covariance function [14]. According to [14], the predicted functional estimates obtained in SE covariance function are smooth and infinitely differentiable. Noise-free GP is considered in this study and the SE covariance function is defined as in [14].

$$k(x,x')=\sigma_f{}^2 \exp\left(-\frac{(x_i-x_j)^2}{2\ell^2}\right) \tag{1}$$

Hyper-parameters are represented by $\theta=\{\sigma_f,\ell\}$ where σ_f and ℓ denote signal variance and characteristic of length scale, respectively. The values for both hyper-parameters must be always positive and if these hyper-parameters are not chosen sensibly, the results will be unacceptable [14]. Typically, the hyper-parameters need to be properly estimated in order to reproduce the best correlation between training set data. This corresponds to the maximization of the log-likelihood function. Given an independently and identically distributed observed data set D is sampled from Gaussian distribution, the negative log-likelihood of those observations [13, 21] is given by

$$\log p(Y \mid x,\theta)=-\frac{1}{2}\left(\log|K|+Y^T K^{-1}Y+n\log(2\pi)\right) \tag{2}$$

where K represents the covariance for inputs training set. The estimate of the hyper-parameters is called as maximum likelihood estimate (MLE).

After obtaining the hyper-parameters, prediction of GP model at test input, x_*, is procured by computing the predictive distribution over functions. Let Y as the known function values of input data, while Y_* denotes a set of function values corresponding to the test input. A covariance matrix of $n \times n$ is defined based on input data, x. The collection of random variables $[y_1,...,y_n,Y_*]$ also known as prior joint distribution [16] is written as $\begin{bmatrix} Y \\ Y_* \end{bmatrix} \sim N_n(0,K_n)$ where $K_n=\begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}$. K_* and K_{**} represent training-test and test set covariance, respectively. Hence, the conditional distribution of which is derived from Multivariate Gaussian Theorem is as follows.

$$Y_* \mid Y \sim N_n\left(K_* K^{-1}Y, K_{**}-K_* K^{-1}\Sigma_*{}^T\right) \tag{3}$$

Once this posterior distribution is acquired, it can be exploited to estimate the predictive values for the test data [19] of offset and hydrocarbon depth. The posterior distribution is $Y_* \sim N_n(m_D,k_D)$ where $m_D=K_* K^{-1}Y$ and $k_D=K_{**}-K_* K^{-1}K_*^T$. m_D and k_D represent the estimated mean Y_* and the variance in the hydrocarbon prediction, respectively. Both equations are the central equation for GP prediction.

4. Results and Discussions

This section elaborates and discusses the results obtained in this study.

4.1. Synthetic SBL data

The synthetic SBL data were generated by CST software. This study only considered electric field (E-field) readings and also noise-free observations over the half of the offset range due to symmetrical simulation models. The offset values were altered for easy interpretation purpose. Figure 3 shows the relationship between offset and magnitude of E-field for each hydrocarbon depth.

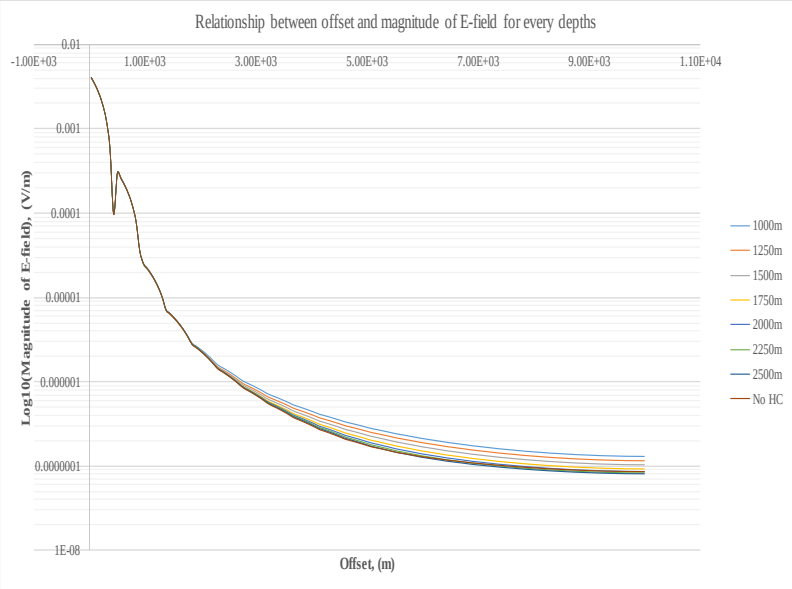

Figure 3. Offset against log10 of magnitude of E-field for each hydrocarbon (HC) depth.

4.2. 1-D forward GP models

This study used GP machine learning (GPML) for MATLAB algorithm written by [21]. Based on figure 4, 46 input data were selected due to the validity of data. GP models were fitted in all hydrocarbon and non-hydrocarbon models and each hydrocarbon model was compared to non-hydrocarbon model to predict the presence of hydrocarbon. Figures 4 and 5 show the comparison plot between the fitted 1-D GP models for hydrocarbon and non-hydrocarbon models.

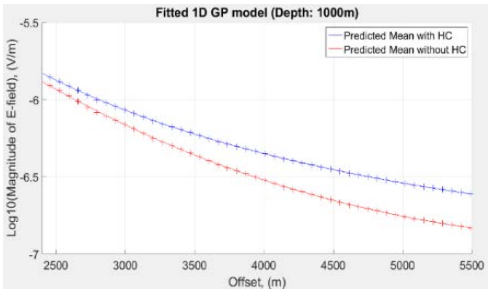

Figure 4. Fitted 1-D GP models for depth of 1000m.

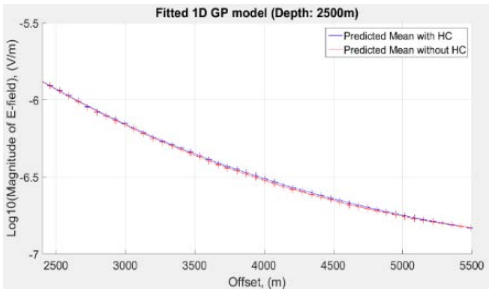

Figure 5. Fitted 1-D GP models for depth of 2500m.

Based on figure 4, at hydrocarbon depth of 1000m, the difference between the predicted mean for both models can be distinguished clearly. Meanwhile, as the hydrocarbon depth increases, the

predicted mean for both models can hardly be distinguished as depicted in figure 5. At this moment, 1-D GP models are very useful since it provides ± two standard deviations around its predicted mean. 1-D GP models were developed to overcome the problem of undistinguished data especially for hydrocarbon depth of 2500m and above. The predictive values for mean and variance of the models were measured at 95% confidence interval. Figures below show the fitted 1-D GP models with $\pm 2\sigma$.

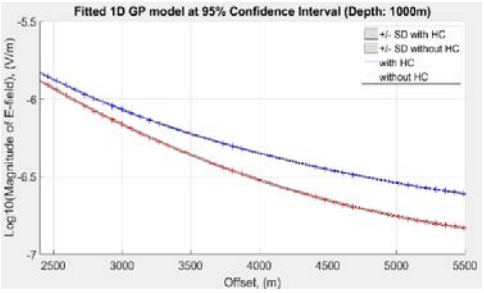

Figure 6. Fitted 1-D GP model at 95% CI for depth of 1000m.

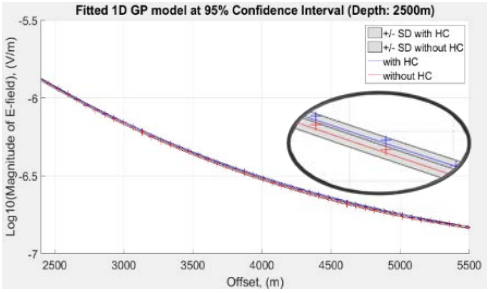

Figure 7. Fitted 1-D GP model at 95% CI for depth of 2500m.

The shaded region in both figures designates $\pm 2\sigma$ at each data. The standard deviation (95% confidence interval) represents the certainty of the presence of hydrocarbon reservoir in the models. As the distance between the variance of hydrocarbon and non-hydrocarbon models become closer, the certainty of existence of hydrocarbon become lower. Based on figure 6, hydrocarbon reservoir can be claimed as certainly present at hydrocarbon depth of 1000m. However in figure 7, both variances start to overlap each other and in this case, the possibility of the hydrocarbon to be absent is high.

4.3. 2-D forward GP model

2-D GP model was developed in order to be the standard profile of the hydrocarbon models for the synthetic SBL data. By creating standards, hydrocarbon depth for any observations from any SBL survey that behave according to it can be predicted. The same 46 input data for each hydrocarbon depth used in 1-D GP models were exercised in the 2-D GP model. The predictive values for $m_D(s,h)$ and $k_D((s,h),(s',h'))$ of data points were calculated. Figure 8 shows the 3-dimensional (3-D) surface plot for the 2-D forward GP model.

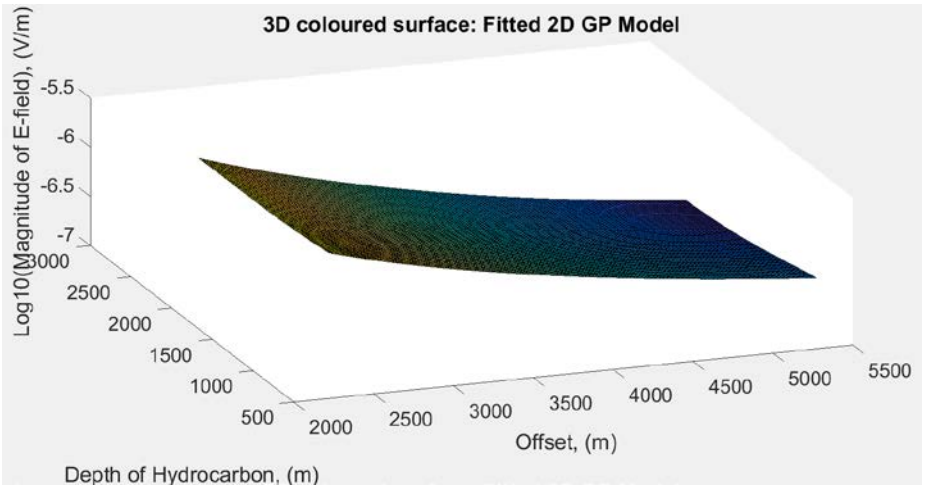

Figure 8. 3-D surface plot of the fitted 2-D GP model.

There are three axes involved in figure 8 which are offset, hydrocarbon depth and log10 of magnitude of E-field. The purpose of using logarithmic scale is to ensure each SBL data can be easily distinguished. Based on the figure, the offset can be observed from 2450m to 5500m while the hydrocarbon depth can be observed from 950m to 2250m. Chronologically, 1-D forward GP models will be fitted first to SBL data in order to confirm the presence of hydrocarbon reservoir. When the hydrocarbon is positively detected at particular offsets, the magnitude of E-field data that behave according to 2-D GP model will be observed to predict the depth of hydrocarbon.

In order to determine the reliability of the developed 2-D GP model, this study provides data analysis with percentage error for five random dataset. Table 2 shows the analysis for hydrocarbon depth of 1000m and table 3 shows the average of percentage error for different hydrocarbon depths.

Table 2. Data analysis for hydrocarbon depth of 1000m

Offset (m)	Log10 of magnitude of E-field generated from CST software (V/m)	Log10 of magnitude of E-field predicted from the 2-D GP model (V/m)	Percentage error (%)
2504.2495	-5.63909852	-5.63827775	0.0146
2834.2163	-5.81858413	-5.81783379	0.0129
3104.1895	-5.95866915	-5.95757926	0.0183
3944.1055	-6.40034439	-6.39903679	0.0204
4124.0869	-6.50812719	-6.50772315	0.0062

Table 3. Average of percentage error for different hydrocarbon depths

Hydrocarbon depth (m)	Average of percentage error (%)
1000	0.0145
1500	0.0118
2000	0.0094

Table 2 is the readings for the log10 of magnitude of E-field generated from CST software and the readings predicted from the 2-D GP model for hydrocarbon depth of 1000m. The percentage errors in table 2 are approaching to zero. It means that the predicted data are fit. For model to be acceptable, researchers in [22] set 5% limit of percentage error average. Based on table 3, the average of percentage error for hydrocarbon depths of 1000m, 1500m and 2000m are less than 5%. This implies that the 2-D GP model is reliable to be used for predicting the hydrocarbon depth for any depths.

5. Conclusions

1-D GP models have successfully been developed to predict the presence of hydrocarbon as in [7]. Based on tables 2 and 3, it was found that the developed 2-D GP model is reliable and it is also able to predict the hydrocarbon depth for SBL application. This study will be improved later by taking considerations of real offshore environment that definitely contains many obstructions and noisy data also will be considered. This study is believed will be useful to the future of oil and gas industry.

Acknowledgement

The authors intend to express their sincere gratitude and appreciation to those who have assisted and hugely contributed towards completing this paper especially to Universiti Teknologi PETRONAS for the opportunities and facilities provided.

References

[1] Wiik T, Hokstad K, Ursin B and Mütschard L 2013 Joint contrast source inversion of marine magnetotelluric and controlled-source electromagnetic data *Geophysics* **78** E315–E327

[2] Feather K 2017 The rapid adoption of seabed logging *Scandinavian Oil and Gas Magazine* **5/6** 37-38
[3] Chiadikobi K C, Chiaghanam O I, Omoboriowo A O, Etukudoh M V and Okafor N A 2012 Detection of hydrocarbon reservoirs using the controlled-source electromagnetic (CSEM) method in the 'Beta' field, deep water offshore Niger Delta Nigeria *Int. J. Sci. Emerg. Technol.* **3** 7-18
[4] Eidesmo T, Ellingsrud S, MacGregor L M, Constable S, Sinha M C, Johansen S and Westerdahl H 2002 Sea bed logging (SBL) – a new method for remote and direct identification of hydrocarbon filled layers in deepwater areas *First Break* **20(3)** 144-152
[5] Daud H, Yahya N and Asirvadam V 2011 Development of EM simulator for sea bed logging applications using MATLAB *Indian J. Mar. Sci.* **40** 267-274
[6] Daud H, Yahya N, Sagayan V and Talib A M 2010 *Magnitude vs low frequency EM waves for sea bed logging applications* IEEE Asia-Pacific Conference on Applied Eletromagnetics, APACE 2010
[7] Mukhtar S M, Daud H and Dass S C 2015 Prediction of hydrocarbon using Gaussian process for seabed logging application *Procedia Computer Science* **72** 225-232
[8] Andréis D and MacGregor L 2007 Controlled-source electromagnetic sounding in shallow water: principles and applications *Geophysics* **73** F21-F32
[9] Kumar A and Lie J E 2008 *Sea bed logging-direct hydrocarbon detection technique in offhsore* exploration 7th International Conference & Exposition on Petroleum Geophysics
[10] Daud H, Yahya N, Asirvadam V and Talib K I 2010 Air waves effect on sea bed logging for shallow water application *Industrial Electronics & Applications (ISIEA) 2010 IEEE Symposium* 306-310
[11] Chan L L T, Liu Y and Chen J 2013 Nonlinear system identification with selective recursive Gaussian process models *Ind. Eng. Chem. Res.* **52** 18276-18286
[12] Schwaighofer A and Tresp V 2002 Transductive and inductive methods for approximate Gaussian process regression *NIPS'02 Proceedings of the 15th International Conference on Neural Information Processing Systems* 977-984
[13] Harari O and Steinberg D M 2014 Oprimal desings for Gaussian process models | via spectral decompostion *J. Stat. Plan. Inference* **154** 87-101
[14] Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* (Cambridge The MIT Press)
[15] Ansari A, Shafie A and Said A 2012 Relationship of resistivity contrast and thickness depth of hdyrocarbon for seabed logging application *IJCSI Int. Comput. Sci. Issues* **9** 214-221
[16] Petelin D, Grancharova A and Kocijan J 2013 Evolving Gaussian process models for prediction of ozone concentration in the air *Simul. Model. Pract. Theory* **33** 68-80
[17] Williams C K I and Rasmussen C E 1996 *Gaussian Process for Regression* (Advances in Neural Information Processing Systems Touretzky D S, Mozer M C and Hasselmo M E Cambridge The MIT) Press **8** 514-520
[18] Liu D, Pang J, Zhou J, peng Y and Pecht M 2013 Prognostic for state of helath estimation of lithium-ion batteries based on combination Gaussian process functional regression *Microelectron. Reliab.* **53** 832-839
[19] Saha S, Saha B, Saxena A and Goebel K 2010 Distributed prognostic health and management with Gaussian process regression *IEEE Aerospace Conference Proceedings* 1-8
[20] Gibson N P, Aigrain S, Roberts S, Evans T M, Osborne M and Pont F 2011 A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy *Mon. Not. R. Astron. Soc.* **419** 2683-2694
[21] Rasmussen C E and Nickisch H 2010 Gaussian processes for machine learning (GPML) toolbox *J. Mach. Learn. Res.* **11** 3011-3015
[22] Daud H, Razali R and Asirvadam V 2015 Sea bed logging applications: predicting hydrocarbon depth using mathematical equations *Applied Mathematical Sciences* **789** 560-565