Gaussian Process Methodology for Multi-Frequency Marine Controlled-Source Electromagnetic Profile Estimation in Isotropic Medium

Muhammad Naeim Mohd Aris [1],[*] , Hanita Daud [1], Sarat Chandra Dass [2] and Khairul Arifin Mohd Noh [3]

[1] Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia; hanita_daud@utp.edu.my
[2] School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, 62200 Putrajaya, Malaysia; s.dass@hw.ac.uk
[3] Department of Geosciences, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia; khairula.nmoh@utp.edu.my
[*] Correspondence: muhammad_naeim@yahoo.com; Tel.: +60-17204-4042

Abstract: The marine controlled-source electromagnetic (CSEM) technique is an application of electromagnetic (EM) waves to image the electrical resistivity of the subsurface underneath the seabed. The modeling of marine CSEM is a crucial and time-consuming task due to the complexity of its mathematical equations. Hence, high computational cost is incurred to solve the linear systems, especially for high-dimensional models. Addressing these problems, we propose Gaussian process (GP) calibrated with computer experiment outputs to estimate multi-frequency marine CSEM profiles at various hydrocarbon depths. This methodology utilizes prior information to provide beneficial EM profiles with uncertainty quantification in terms of variance (95% confidence interval). In this paper, prior marine CSEM information was generated through Computer Simulation Technology (CST) software at various observed hydrocarbon depths (250–2750 m with an increment of 250 m each) and different transmission frequencies (0.125, 0.25, and 0.5 Hz). A two-dimensional (2D) forward GP model was developed for every frequency by utilizing the marine CSEM information. From the results, the uncertainty measurement showed that the estimates were close to the mean. For model validation, the calculated root mean square error (RMSE) and coefficient of variation (CV) proved in good agreement between the computer output and the estimated EM profile at unobserved hydrocarbon depths.

Keywords: multiple frequency marine controlled-source electromagnetic technique; Gaussian process; uncertainty quantification; computer experiment, electromagnetic profile estimation

1. Introduction

Nowadays, the controlled-source electromagnetic (CSEM) technique is a significant application to detect and discover hydrocarbon-filled reservoirs based on the principles of electromagnetic (EM) propagation. For decades, CSEM application has been widely exercised in onshore geophysical exploration (e.g., [1,2]). The efficiency of this application to characterize offshore hydrocarbon reservoirs has also been proven by many oil and gas companies around the world. Li and Key [3] stated that at early stage, marine CSEM application was employed to study the electrical conductivity of the upper mantle and oceanic crust (e.g., [4–8]). Studies related to the commercial application of the marine CSEM technique in offshore hydrocarbon exploration can be found in [9–17]. Previously, the seismic sounding survey, which employs acoustic waves, was solely utilized to map geological structures that

have different acoustic properties [18]. This survey was very important to hydrocarbon exploration due to its capability of providing information of the subsurface. According to [19], seismic data interpretation provides good resolution of the subsurface structures underneath the seabed; however, it has deficiencies. It is said that seismic surveys are unable to distinguish the fluid content inside the reservoirs, whether brine (conductive seawater) or hydrocarbon. Zaid et al. [18] mentioned that seismic sounding is not compatible to the direct detection of the pore fluid reservoirs. Note that EM and seismic techniques are sensitive to two different properties of subsurface; thus, the marine CSEM technique was developed as a complementary interpretational tool to specifically characterize the target reservoirs.

The marine CSEM technique also is referred to as a seabed logging (SBL) application. This is thoroughly described by [10]. This application is particularly able to reduce ambiguities in data interpretation in hydrocarbon exploration. Andreis and MacGregor [16] stated that by studying the reflected EM signal, resistive mediums such as hydrocarbon, gas, and hydrate can be discovered to depths of several kilometers from the seabed. In addition, the resistivity of the subsurface in offshore environments is commonly identified by robust anisotropy because of the sedimentation factor [20]. Note that for a medium that is horizontally stratified, the subsurface is generally less resistive in the horizontal (parallel) direction than in the vertical (perpendicular) direction [20]. Offshore hydrocarbon reservoirs are normally embedded in a high conductive medium unlike the common case of onshore hydrocarbon reservoirs [21]. Hydrocarbon-filled reservoirs are known to have very high electrical resistivity compared to its surroundings, such as of saline water and sedimentary rocks. These structures are very conductive. From [22], hydrocarbon is known to have electrical resistivity between 30 and 500 Ohm meter, whereas the resistivity values of seawater and sediment are 0.5–2 Ohm meter and 1–2 Ohm meter, respectively. If a target reservoir is brine saturated, it is normally a few orders of magnitude less electrically resistive than a hydrocarbon-filled reservoir. From these characteristics, the resistivity of the subsurface can be resolved via data of electric (E-) and magnetic (H-) fields obtained from the marine CSEM survey. The measurement of amplitude and phase of E- and H- fields can be utilized to determine the geological subsurface. Li and Key [3] mentioned that the amplitude and phase of an EM field will vary depending on the resistivity of the structures beneath the seabed, the depth of seawater, and the source–receiver offset.

In the marine CSEM technique, data collected can be interpreted in two different groups depending on the domain—either in time-domain or frequency-domain. Analyzing data in time- or frequency-domains would theoretically give the same output/information [23]. Reyes et al. [24] mentioned that frequency-domain marine CSEM is normally used for the case of oil prospecting. In the frequency-domain application, an antenna/transmitter of towed EM dipole is generally used to generate a low-frequency EM field, and returned/reflected signals recorded by receivers placed on the seabed are utilized for resistivity distribution analysis. The choice of frequency is very crucial in marine CSEM application. In a standard configuration, marine CSEM surveys use a deep-towed horizontal electric dipole (HED) transmitter to emit a low-frequency EM wave which is usually between 0.1 Hz and 10 Hz to an array of seabed receivers, and normally, the transmitter is towed at 30–50 m above the seabed [25]. Practically, the low-frequency of EM waves is used in a deep water environment as the signal transmission due to the fact that low-frequency is able to yield farther penetration through seawater columns into sedimentary rocks. Next, for the receiver, there are two types of receiver configurations—inline and broadside. Inline configuration is when the separation distance is parallel to the direction of antenna, whereas broadside is when the source–receiver offset is normal to the antenna's direction [26]. Due to the characteristics of subsurface conductivity, EM signals spread with a higher rate through the seafloor than through the seawater. The EM energy, which is transmitted from the towed-source, spreads in all directions and is quickly attenuated in conductive medium such as sediment. The occurrence of all possible signal contributions is depicted in Figure 1.
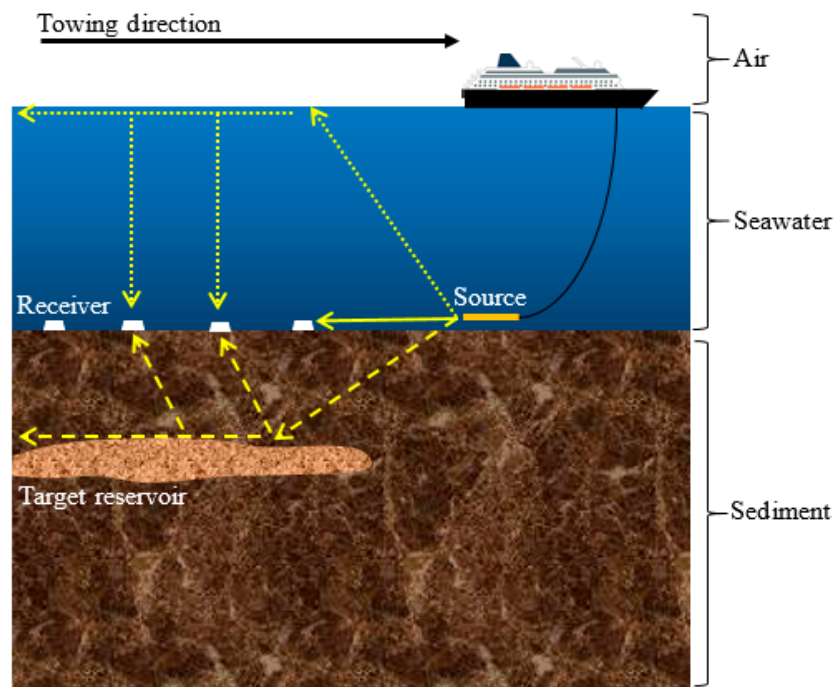
Figure 1. Basic layout of marine CSEM application in hydrocarbon exploration. The source is towed nearly to the seabed receivers. A target reservoir is expected to be embedded in the conductive sedimentary rocks. The EM signal spreads in all direction through the seawater, air–seawater interface, and sediment before being recorded by the EM receivers.

Based on Figure 1, the straight-line arrow denotes the direct-wave travelling from the source to the seabed receiver without any interaction with the geological subsurface beneath the seabed. Researchers in [27,28] mentioned that direct-wave dominates the data collection at short source–receiver separation offsets. Next, the dotted-line arrow represents the reflected and refracted wave from the source upwards to the air–seawater interface and vertically going back through the seawater to the receiver (i.e., air-wave). This wave travels with high rate of velocity (propagates with no attenuation) to the water surface since air is an infinitely electrical resistive medium. Seawater depth can influence the measured EM signal. Air-wave contribution increases as the depth of seawater decreases. Weiss [29] asserted that this contribution becomes significant in a seawater depth of roughly less than 300 m. Both these signals, direct- and air- waves, do not contain any information about hydrocarbon-filled reservoirs. Last signal contribution is denoted as dashed-line arrow. It is known as the reflected and refracted wave (i.e., guided-wave). This wave diffuses outward from the source through seawater column and then through the high resistive formations with less attenuation. The transmitted wave has to enter the formations at certain angle which is between 0° and 11° in order to set the guided mode [28]. This reflected and refracted wave strongly dominates the recordings at intermediate source–receiver offsets (~3 to ~8 km). The detection of the guided-wave is the basis of the marine CSEM survey.

In the context of geophysics forward modeling, electrical resistivity has an important role in oil and gas exploration. Numerical modeling is a crucial component that provides information of the electrical resistivity of the subsurface. There are various computational techniques exercised in EM applications such as the Finite Element (FE), the Finite Difference (FD) and the Method of Moment (MOM) [30], while researchers in [20] have said that FE, FD and integral equation (IE) methods are among the most famous numerical techniques for modeling EM data. According to [20], the FE method is more reliable for EM forward modeling in a complex geological structure when compared to FD and IE methods. Based on the literature, FE is a usual numerical technique exercised in CSEM modeling for hydrocarbon exploration. This computational method uses unstructured grids that can be easily conformed to irregular boundaries compared to the FD method. The MOM is less preferred as well in

marine CSEM data interpretation since this method produces more complex derivations of governing equations than the FE method. The traditional FD method is easier to implement and maintain than the FE method, but the method is based on structured grids. This means that grid refinement is not possible, and hence it affects the overall computational processes [31]. Unstructured grids have long been exercised in various fields such as engineering and applied mathematics; however, this feature has only recently been used in the EM geophysical field, as exemplified by the use of FE method code in marine EM surveys (e.g., [3,32–34]). This feature can realistically replicate the complexities of geological structures [19].

Even though this numerical technique is very powerful, the ad hoc design of meshes in FE is time-consuming. Li and Key [3] mentioned that the most time-consuming task in their code (FE algorithms) is the solutions of the linear equation systems. The study could take a very lengthy computational time if they used all wavenumbers in the mesh refinement for a full solution. Bakr et al. [35] also stated that the most time-consuming tasks in FE are evaluating the integrals and solving the linear equations. For typical simulations, a few million elements are involved in the linear equation systems [31]. According to [24], the execution of real-field simulations in EM problems needs the use of high-performance computing (HPC). This is because typical actual executions require more than hundred thousand realizations which involve millions of degrees of freedom for each process. The computational and memory requirements to solve such solutions may become a serious challenge. It can be more complicated for higher-dimensional EM forward modeling and inversion. Besides forward modeling, inversion is also a powerful way to recover the electrical conductivity profile beneath the seabed given measurements of EM fields acquired from real-field surveys. Not to mention, nowadays, inverse modeling comes with robust inversion schemes and incorporation of more procedures and measurements. This makes it possible to compute the EM fields at the seabed receivers precisely and provide accurate geometry resolution. However, it is said that inversion algorithms tend to be computationally expensive due to the forward modeling schemes. Indeed, an inversion process needs multiple EM forward solutions [35]. Furthermore, in terms of application, the marine CSEM technique generates huge amounts of data (captured by seabed EM receivers with a moving HED source); therefore, processing those data has become a challenging task to many geophysicists [9,36]. Modeling the marine CSEM data is related to the need of accurate representation of very complex geo-electrical models, and the algorithms used should be powerful and fast enough to be applied to repeated use of hundreds of iterations and multiple source–receiver positions. In addition, understanding the noisy CSEM data to quantify the uncertainties involved in EM modeling also is very crucial. Constable and Srnka [14] stated that the economic challenges (e.g., related to drilling) increase as the hydrocarbon exploration moves to deeper offshore environments. Thus, any additional data that can be obtained or collected will be advantageous to the exploration if there is a potential to de-risk a given expectation [19].

In order to seek the most favorable balance between the computational cost involved in the interpretation of EM geophysical data and the accuracy of the modeling, our interest is focused on processing the one-dimensional (1D) frequency-domain marine CSEM data using Gaussian process (GP) algorithms. We propose GP as a methodology for two-dimensional (2D) forward modeling of the marine CSEM technique to provide information on EM profiles when hydrocarbon is present at various depths in isotropic mediums. This forward modeling provides the uncertainty measurement of the estimation in terms of variance. Although the existing CSEM models (the existing numerical modeling techniques) provide robust representation of real-field models, this work has significant contribution for hydrocarbon detection as well. This attempt is very useful and helpful when collected sets of data in CSEM surveys are insufficient for the interpretation of higher-dimensional modeling and inversion. This analysis also could reduce time in the CSEM workflow since forward GP modeling is able to provide uncertainty quantifications without integrating or combining any other numerical quantifiers. On top of that, there is its simplicity, as GP only involves simple equations which means faster computation which only needs basic memory space. Note that this analysis utilizes the simulation

datasets generated through a commercial software, namely, Computer Simulation Technology (CST). Information of the CST software can be found in [37]. The details and literature of GP application are thoroughly elaborated in next section.

2. Statistical Background: Gaussian Process in Computer Experiments

Gaussian Process (GP) is random function which has a property that any finite number of evaluations of the (random) function has a multivariate Gaussian distribution. GP is fully specified by a mean function, $m(x)$, and a covariance function, $k(x, x')$. The Gaussian distribution has mean and covariance values in the forms of vector and matrix evaluations, respectively [38]. Here, x represents all potential independent variables that influences the outputs/responses. GP is a non-parametric and probabilistic method for fitting functional forms based on domain observations. It differs from most of other black-box identification approaches where it does not approximate the modelled system by fitting the parameters of basis function, but rather searches for relationships among the measured data. This non-parametric regression method does not need a fixed discretization. This technique is able to provide predictive mean values and uncertainty of the estimation measured in terms of variance. This variance reflects the quality of the output/information. It is an important numerical measure when it comes to distinguishing GP from the other computational intelligence methods. According to [38], GP is suitable for modeling uncertain processes or data which are unreliable, noisy or contain missing values. GP has been used in many different applications. Studies related to application of GP in various fields can be found in [39–45]. In general, prior belief of spatial smoothness is specified through a covariance defined by similarity characteristics. Training observations are then considered as the realizations from the updated multivariate Gaussian (i.e., posterior). Thus, the conditional realizations from the posterior are simply the testing output at all untried or unobserved points. The mathematics behind this concept are thoroughly explained in the methodology section.

Computer experiments are well-known and not new in science, technology, and engineering. This medium is getting very popular for solving scientific and engineering problems. Nowadays, scientists prefer to use computer simulators rather than doing case studies or conducting any related physical experiments. It can be implemented in any circumstances including experiments that are impossible to do physically, with shorter time taken than in the real situation. Computer experiments are run by means of a complex code and highly developed theories of physics, mathematics, and engineering fields. Sacks et al. [46] described that experimenters usually aim to estimate/predict the output at unobserved input points, optimize the function of the output points, and calibrate the computer code to physical data. To this end, [46] and the subsequent works modelled the output of a computer model, $Y(x)$, based on input x as a sum of regression terms, β_j, and stochastic component, $Z(x)$. $Y(x)$ is defined in Equation (1).

$$Y(x) \ = \ \sum_{j\,=\,1}^{k} \beta_j f_j(x) + Z(x), \qquad (1)$$

where $f_j(x)$ is a known function with $j \ = \ 1, 2, \ldots, k$, and $Z(x)$ is a random process with a zero-mean and a covariance. The most famous choice of the stochastic component, $Z(x)$, is a GP where the distribution of the GP is assumed to be a normal (Gaussian) distribution with a mean and a covariance function. According to [47], GP is used as the surrogate model for any complex mathematical models which consume a lot of time to solve. GP is flexible in representing the computer output, $Y(x)$, and it is feasible to obtain analytical formulas of the predictive distribution and to design the equations.

From the reported literature, there are a few studies calibrating CST computer output of marine CSEM applications with GP (e.g., [48,49]). However, these studies present 1D GP modeling of SBL applications where they only considered univariate independent variables. Besides, the work only focused on predicting the presence of the hydrocarbon layer at a known depth, while we consider various depths of hydrocarbon at observed and unobserved depth levels. This is because the location of hydrocarbon reservoirs can be anywhere and is uncertain in real-field environments. Aris et al. [50] also described forward GP modeling of SBL applications, but the paper only focused on

one transmission frequency and no error was considered in the presented GP modeling. Since CST computer output is assumed to generate very clean data, considering the error in modeling is very important to marine CSEM data processing. Thus, this attempt is novel in two ways; first, it proposes GP methodology to process marine CSEM data calibrated with CST computer output at multiple transmission frequencies at which hydrocarbon is present at all possible depth levels (250–2750 m); second, this is a data-dependent analysis where it utilizes an uncertainty quantification provided by the GP in marine CSEM data processing with error considerations before in-depth analysis. This may enhance the EM data interpretation where the EM profile is estimated with the measurement of variance at various possible depths of the hydrocarbon layer, which helps decision-making for hydrocarbon detection in marine CSEM applications.

3. Methodology

The methodological flow is based on a three-step procedure; (i) synthetic seabed logging (SBL) modeling using Computer Simulation Technology (CST) software, (ii) developing two-dimensional (2D) forward Gaussian Process (GP) models for multiple EM transmission frequencies, and (iii) model validation using the root mean square error (RMSE) and the coefficient of variation (CV).

3.1. Synthetic SBL Modeling Using CST Software

We designed synthetic models of typical marine CSEM application for hydrocarbon exploration which have various depths of hydrocarbon at multiple frequencies by using CST software. The transmissions were tested at frequencies of excitation current of 0.125, 0.25, and 0.5 Hz. Note that in the CST software, Maxwell's equations are discretized using the Finite Integration Method (FIM). FIM solves the Maxwell's equations in a finite calculation domain in grid cells to probe the resistivity contrast. For this study, the SBL model is a three-dimensional (3D) canonical structure, which consists of background layers such as air, seawater, and sediment. The model is designed with an air–seawater interface at $z = 300$ and the seawater thickness is fixed at 1000 m (deep offshore environment). A 200 m-thick horizontal resistive layer (hydrocarbon) is embedded in the sediment layer with various depths from the seabed. The thickness of sediment above the hydrocarbon layer, known as overburden thickness, is varied from 250 m to 2750 m with an increment of 250 m each. The thickness of the overburden layer indicates the depth of hydrocarbon reservoir. Thicker overburden layers mean a deeper location of the hydrocarbon. Both the background and hydrocarbon layers are considered as isotropic. Figure 2 shows the stratified illustration of the horizontal layers used in this study.
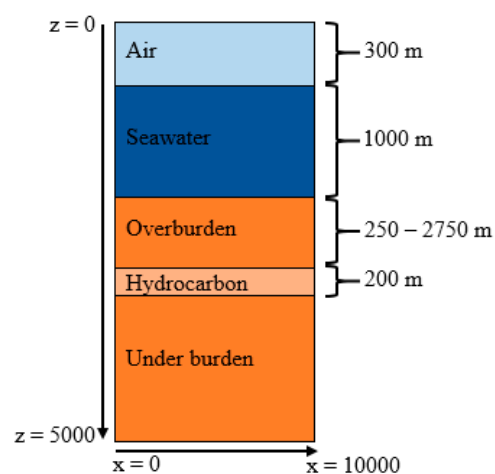

Figure 2. Illustration of SBL models used in this study. The thicknesses of air, seawater, and hydrocarbon are 300, 1000, and 200 m, respectively. The depth of the hydrocarbon layer (thickness of overburden layer) is varied from 250 m to 2750 m with an increment of 250 m. The total height and length of the models are 5000 and 10,000 m, respectively.

The replication of the 3D structures of the SBL model (length, x: 10,000 m; width; y: 10,000 m; height, z: 5000 m) can be referred to in [50]. The electrical conductivities of air, seawater, sediment, and hydrocarbon are tabulated in Table 1. Their properties are taken from [48].

Table 1. Electrical conductivity of every layer considered in the SBL models. Conductivity is the reciprocal of resistivity. The electrical conductivity of the hydrocarbon layer is lower than its surroundings, which are seawater and sediment. It means that the hydrocarbon layer is parameterized with reliable electrical resistivity.

Material	Electrical Conductivity (Sm^{-1})
Air	1.0×10^{-11}
Seawater	1.63
Sediment	1.00
Hydrocarbon	2.0×10^{-3}

The EM signal is transmitted by a HED source located with the orientation of *x*-direction in the seawater with coordinates (5000, 5000, 1270). This means that the inline transmitter pointing along x-axis is positioned at *x*, *y* = 5000, and a height of 30 m above the seabed. The HED source is held stationary at the center of the model. The values of the EM field are measured along an inline profile through the SBL model. In this study, the current strength of the HED source is fixed at 1250 A. The source–receiver separation distances (offsets) are varied along the replication model. An array of 1000 seabed receivers is placed along the seabed at *x* ranges of 0–10,000 m and y = 5000 m. This means, receivers are positioned along the seabed for every 10 m from 0–10,000 m of x-orientation. As a demonstration, Figure 3 is the mesh view of the replication of the marine CSEM model in isotropic medium, replicated by the CST software at a hydrocarbon depth of 250 m.
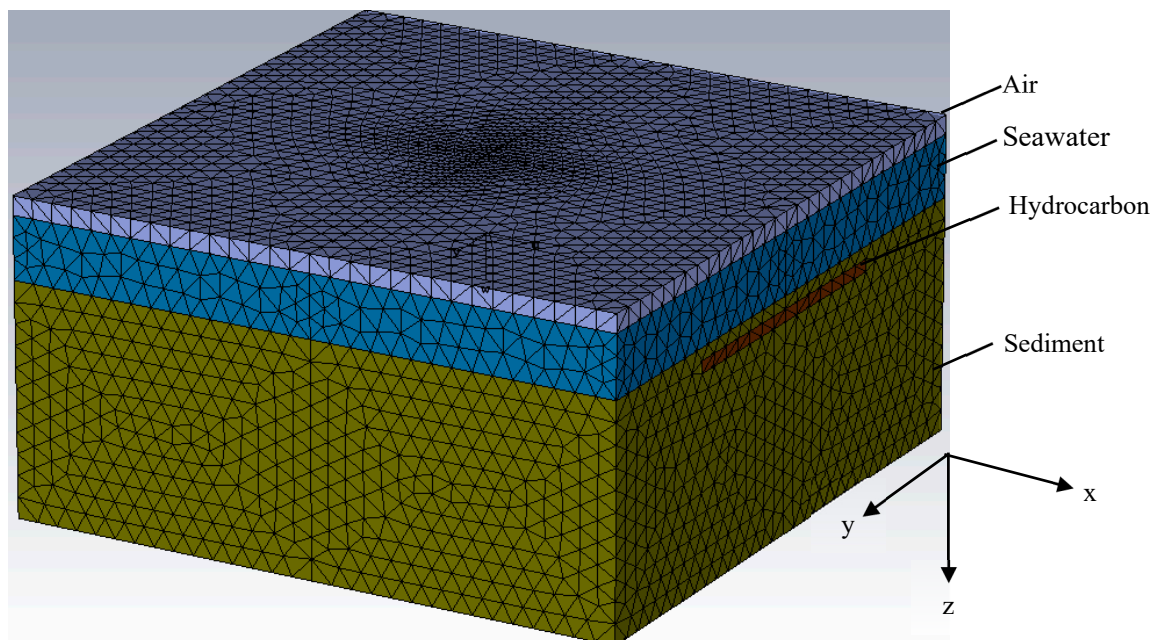

Figure 3. Mesh view of the 3D SBL model at hydrocarbon depth of 250 m replicated by CST software. Both background and hydrocarbon layers are set as isotropic. The hydrocarbon layer is designed with length (x) of 10,000 m, width (y) of 5000 m and height (z) of 200 m.

3.2. Developing 2D Forward GP Models at Multiple EM Transmission Frequencies

Let $Y(x_i)$ be the CST computer output at *k* different input specifications for every frequency used (0.125, 0.25, and 0.5 Hz), where $i = 1, 2, \ldots, k$. The input variable, *x*, can be univariate

or multivariate, but in this study, we exercise a bivariate independent variable. Source–receiver separation distance (offset), s, and depth of hydrocarbon layer from the seabed, h, are considered as input variables where $x = (s, h)$. In this paper, we focus on processing non-normalized CST output, $Y(x)$, which is the magnitude of the E-field (amplitude) obtained from static source–receiver combinations where the transmitter is fixed at the center of the SBL model. For every frequency, we have source–receiver separation distances, $s_i = \{i = 1, 2, \ldots, 210\}$, and hydrocarbon depths, $h_i = \{i = 1, 2, \ldots, 11\}$, which are from 250 to 2750 m with and increment of 250 m each. Thus, in this study, we have $k = 210 \times 11 \times 3 = 6930$ different input specifications of CST computer output that are to be processed.

As mentioned earlier, GP is completely defined by a mean function, $m(x)$, and a covariance function, $k(x, x')$. The GP model on function f with a zero-mean function, $m(x) = 0$, and a covariance function, $k(x, x')$, can be written as

$$f(x) \sim G(0, k(x, x')). \tag{2}$$

An appropriate correlation function for our GP is selected. The choice of the correlation function is very crucial in computer experiments. It governs the smoothness of the sample path realizations of the GP and is dictated by CST computer output. We choose a popular covariance function which is the squared exponential (SE) function. This covariance function has been widely used in many applications of GP regression and it produces smooth functional estimates. The SE is defined in Equation (3).

$$k(x, x') = \sigma_f^2 . e^{\left(\frac{-|x-x'|^2}{2\ell^2}\right)}, \tag{3}$$

where σ_f and ℓ are signal variance and characteristic-lengths scale, respectively. These hyper-parameters need to be properly estimated, and this is usually by optimizing the marginal likelihood. By referring to Bayes' theorem, we assume that very little prior knowledge about these hyper-parameters are known, and this prior knowledge corresponds to the maximization of marginal log-likelihood. We have three different datasets (three different frequencies). For every dataset, there are 210 data points consisting of offset and hydrocarbon depths as the independent variables corresponding to the magnitude of the E-field as the dependent variable. According to [39], two-thirds of the total data should be considered as training data points and the remainder will be the testing points. Thus, in this paper, for every three data points, the first two data are set as the training data points. This procedure was implemented for every frequency used.

The GP regression model is assumed to generally have a relationship of the form $y_i = f(x_i) + \varepsilon$ where the prior joint distribution for the collection of random variables consist of training and testing points are defined as below

$$\begin{bmatrix} m \\ m_* \end{bmatrix} \sim G\left(0, \begin{bmatrix} K_\varepsilon & K_* \\ K_*^T & K_{**} \end{bmatrix}\right). \tag{4}$$

The vector $m \in \mathcal{R}^{n_{train}}$ is observed at spatial locations $x \in \mathcal{R}^{n_{train} \times n_d}$ where $x = (s, h)$. n_{train} denotes the number of training data points generated from CST software, while n_d is the spatial dimension exercised in this paper. Note that for every hydrocarbon depth, $n_{train} = 140$ and $n_d = 2$. Next, $m_* \in \mathcal{R}^{n_{test}}$ is a vector that specifies the predicted values at particular spatial locations $x_* \in \mathcal{R}^{n_{test} \times n_d}$ where $x_* = (s_*, h_*)$. n_{test} is the number of all desired observations (i.e., testing data points) where $n_{test} = 70$ for each depth. With 140 data per depth, a matrix $K \in \mathcal{R}^{n_{train} \times n_{train}}$ is defined using Equation (3) for all pairs involved in the training points. Then, $K_\varepsilon \in \mathcal{R}^{n_{train} \times n_{train}}$ is determined such that

$$K_\varepsilon = K + \sigma^2, \tag{5}$$

where σ^2 represents a diagonal covariance matrix of the specified additive noise at 5%. $K_* \in \mathcal{R}^{n_{train} \times n_{test}}$, and $K_{**} \in \mathcal{R}^{n_{test} \times n_{test}}$ are calculated using Equation (3) where these matrices define the correlation of

training–testing data points and testing–testing data points, respectively. Hence, m_* is predicted at 70 locations of x_* per depth. Here, matrix K_{**} only has input from testing data points and it is derived from prior information. Thus, the posterior conditional GP as in Equation (1), given the information of x_*, x and m, is written as

$$p(m_*|x_*, x, \ m) \ = \ G(m_*|\mu_*, \Sigma_*) \tag{6}$$

Based on the theorem, for every frequency, the Gaussian probability for the random variable m_* with mean, μ_* (i.e., the estimated EM profile at observed and unobserved depths of hydrocarbon), and variance, Σ_* (uncertainty measurement in terms of ± two standard deviations), are defined in Equations (7) and (8), respectively. These equations are the main equations of GP regression.

$$\mu_* \ = \ K_*^T K_\varepsilon^{-1} m \tag{7}$$

$$\Sigma_* \ = \ K_{**} - K_*^T K_\varepsilon^{-1} K_* \tag{8}$$

3.3. Model Validation Using RMSE and CV

To validate our forward GP model, we calculated the root mean square error (RMSE) and coefficient of variation (CV) of the difference between data predicted by GP (estimates) and the data acquired from CST software. RMSE is able to calculate the difference between an estimate and the true value (observation) corresponding to the expected value of root squared loss. CV is calculated as well to evaluate the relative closeness between true values and estimates in percentage. Two random unobserved depths of hydrocarbon which are 900 m and 2200 m were selected for demonstration purposes. The SBL models with these depths of hydrocarbon were simulated separately. The CST computer output was considered as the true values, and the estimate values are the data predicted by GP at the same depths of hydrocarbon. The RMSE and CV between the true values, y_i, and the estimate values, y_i^*, are defined as below

$$RMSE \ = \ \sqrt{\frac{\Sigma\left(y_i - y_i^*\right)^2}{a}}, \tag{9}$$

$$CV \ = \ \frac{RMSE}{\left|\mu_{y_i^*}\right|} \times 100\%, \tag{10}$$

where a is the total number of testing data points, and $\mu_{y_i^*}$ is the absolute average of y_i^*. The estimates from the 2D forward GP model should match the simulation data acquired from the CST software at all observed and unobserved depths very well up until larger offset distances.

4. Results and Discussion

We implemented this analysis by using GP algorithms in a MATLAB code (built-in function) referred from [51]. We considered a typical synthetic frequency-domain marine CSEM study with three different transmission frequencies (0.125, 0.25, and 0.5 Hz), a HED source, and an array of receivers. Note that for every depth of hydrocarbon, the simulation was simultaneously run for the three transmission frequencies (three datasets per simulation). Every simulation process took ~15 min to generate the three different datasets. Hence, total computational time for the CST software to compute the EM fields for 11 hydrocarbon depths was approximately 165 min (~2 h and 45 min). In addition, in order to make the data interpretable, a logarithmic scale with base 10 was applied to the magnitude of the E-field for every frequency since it involves very small values. Figures 4–6 are the CST computer output at all input specifications for every frequency.
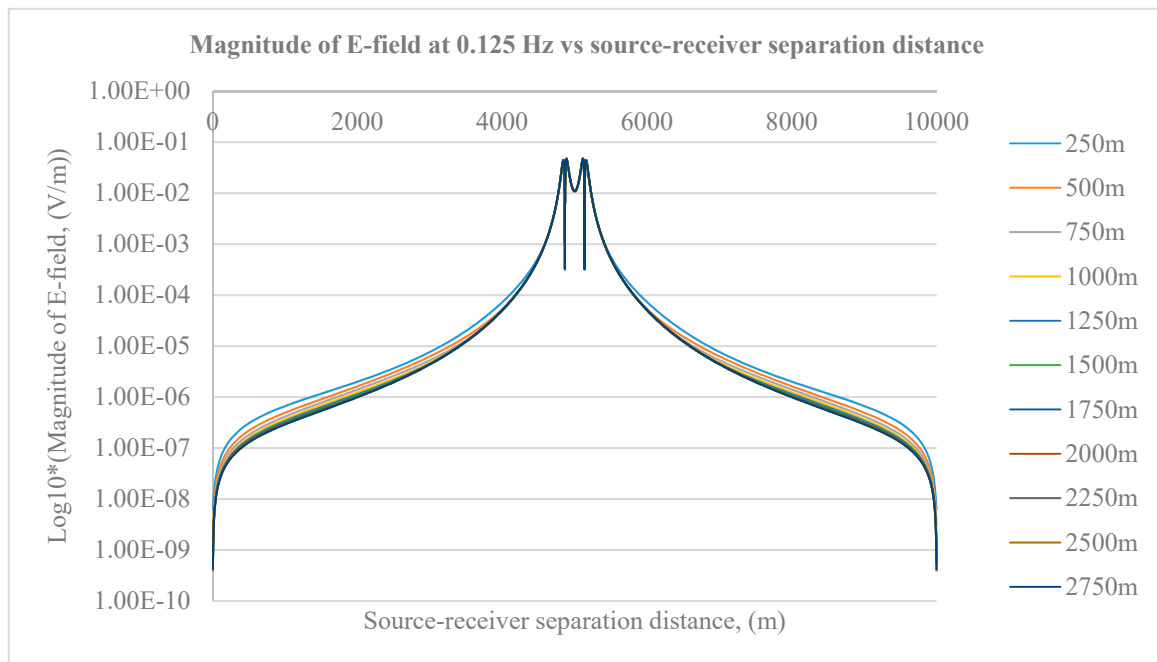
Figure 4. Log10 of magnitude of electric field at 0.125 Hz versus source–receiver separation distance (offset). Different hydrocarbon depths yield different EM responses. The offset is from 0 m (left of the SBL model) to 10,000 m (right of the SBL model).
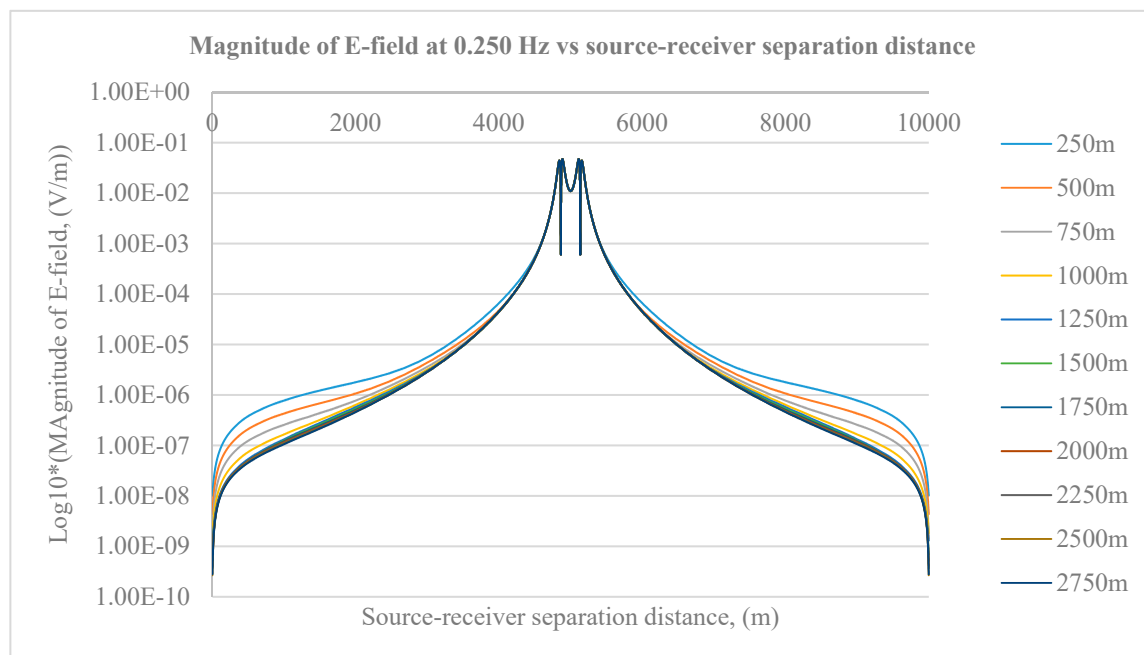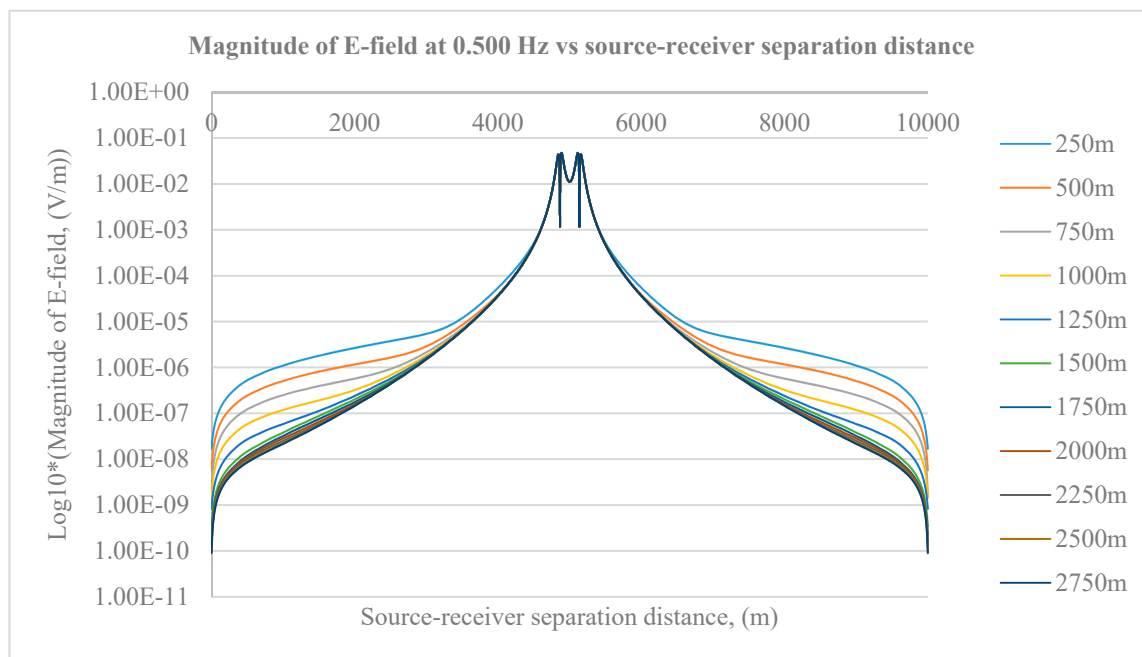

Figure 5. Log10 of magnitude of electric field at 0.25 Hz versus source–receiver separation distance (offset). Different hydrocarbon depths yield different EM responses. The offset is from 0 m (left of the SBL model) to 10,000 m (right of the SBL model).

Figure 6. Log10 of magnitude of electric field at 0.5 Hz versus source–receiver separation distance (offset). Different hydrocarbon depths yield different EM responses. The offset is from 0 m (left of the SBL model) to 10,000 m (right of the SBL model).

From the results, the replicated SBL model is able to reflect the typical synthetic simulation model of EM application. The simulated datasets resulting from the CST software were in a good agreement with the behavior of real-field CSEM data on the effect of the source–receiver separation distance and variations of hydrocarbon depth to the magnitude of the E-field (amplitude). The strength of E-field is inversely proportional to the source–receiver offset and depth of hydrocarbon. If the source–receiver separation is placed further apart and the hydrocarbon layer is located deeper beneath the seabed, the E-field strength significantly decreases. The acquired responses vary in frequency as well. Here, based on figures above, the EM responses are symmetrical. The EM wave was transmitted from the source which was located at the center of the SBL model. The signal travelled equidistant from the source to the boundaries of the model (left and right of the SBL model). Due to this symmetrical setting, only data from 5000 to 10,000 m were considered for processing purposes. Next, from the figures as well, we can see that the magnitudes of the E-field for all hydrocarbon depths are indistinguishable (especially in Figure 6) at source–receiver offset smaller than ~7400 m. This happens because high transmission frequencies have high attenuation, thus the signal is not able to propagate farther than low-frequency EM wave. Thus, we generalized this analysis by utilizing data for the offset from ~7400 to ~9500 m.

From the CST computer output, we developed a 2D forward GP model for every frequency to provide EM profiles at the observed and unobserved depths of the hydrocarbon. Even though the offset distances considered in this paper are from ~7400 m to ~9500 m, the GP models were set to distances from ~2400 to ~4500 m in order to make it easy to interpret, since the EM signal was transmitted from $x = 5000$ (center of the SBL model). The 2D forward GP models for frequencies of 0.125, 0.25, and 0.5 Hz are depicted in Figures 7–9, respectively.
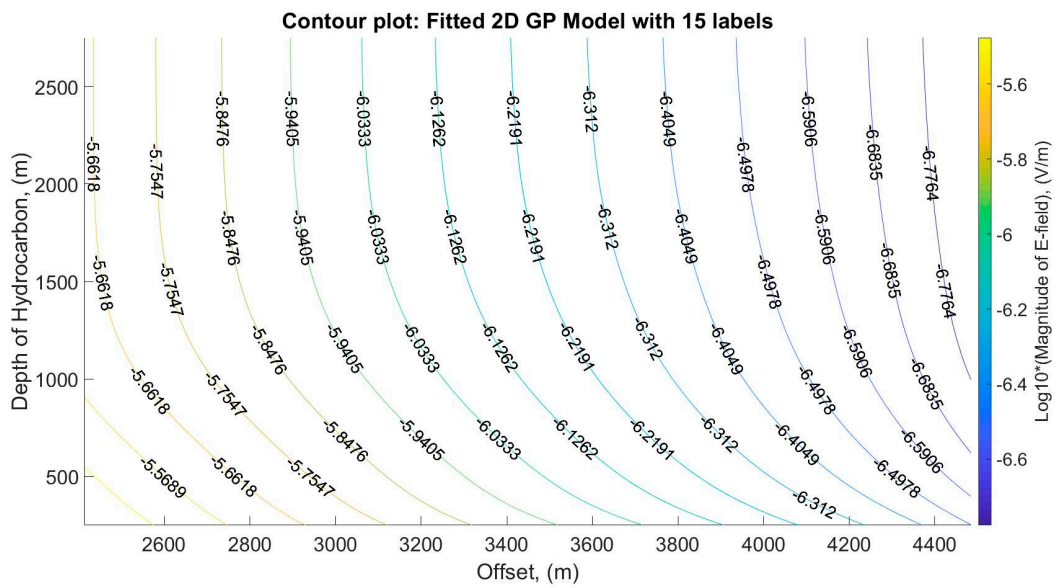
Figure 7. Contour plot of EM profiles (amplitude) for various offset distances (~2400 to ~4500 m) and hydrocarbon depths (250–2750 m) at a frequency of 0.125 Hz with 15 labels.
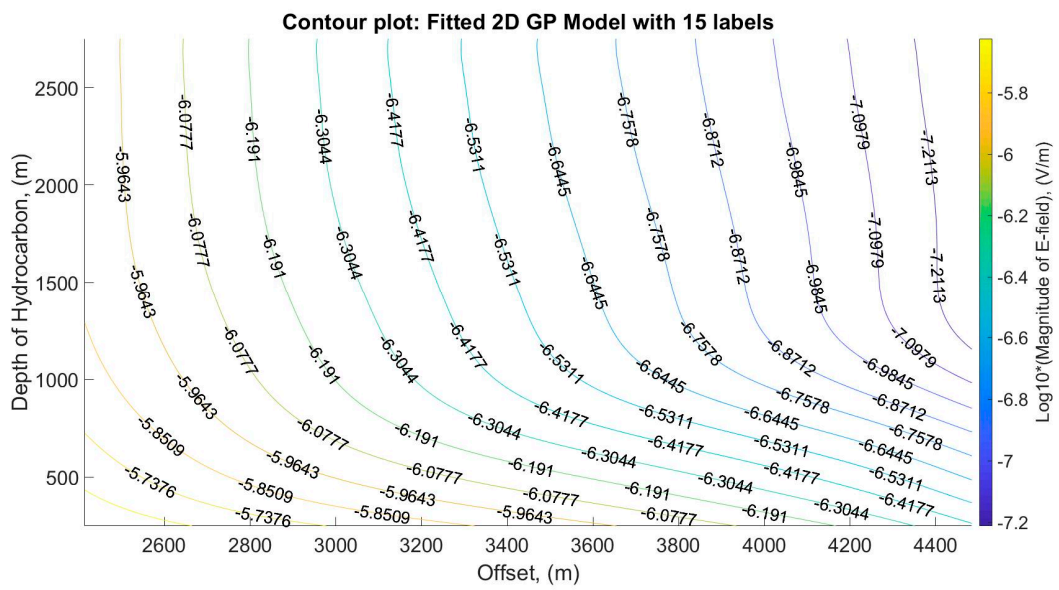

Figure 8. Contour plot of EM profiles (amplitude) for various offset distances (~2400 to ~4500 m) and hydrocarbon depths (250–2750 m) at a frequency of 0.25 Hz with 15 labels.
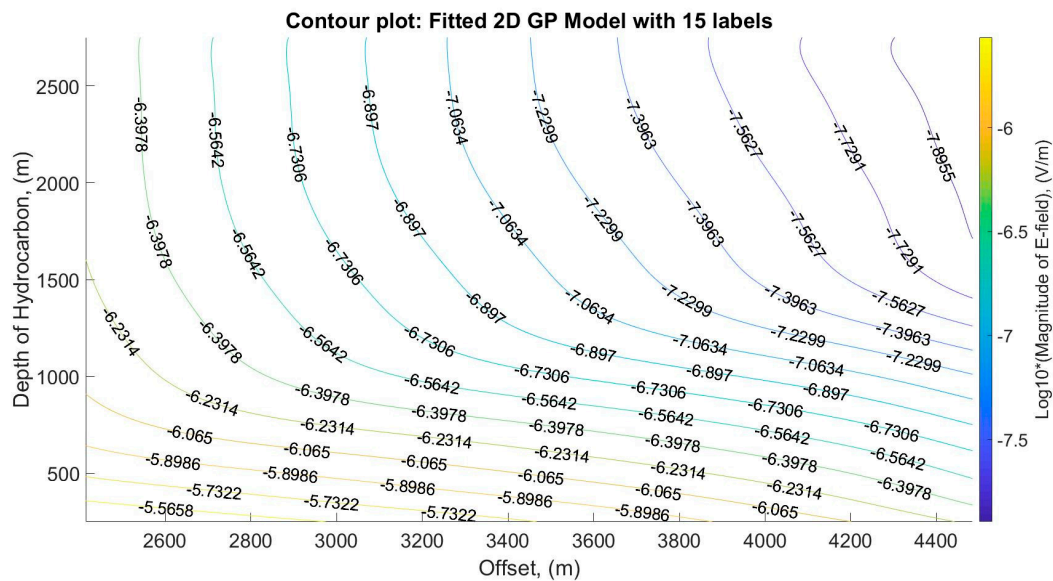
Figure 9. Contour plot of EM profiles (amplitude) for various offset distances (~2400 to ~4500 m) and hydrocarbon depths (250–2750 m) at a frequency of 0.5 Hz with 15 labels.

From the figures, we can see that the GP models are able to provide the information of EM profiles which are the magnitude of the E-field at all desired depths of hydrocarbon (observed and unobserved). Since variance is quantified in GP estimation, we tabulate the average of the variance of EM profiles for all observed depths of hydrocarbon in Table 2 to determine how far the data points are spread out from the mean value.

Table 2. The average of variance of EM responses (frequencies: 0.125, 0.25, and 0.5 Hz) at all tried depths of hydrocarbon (250–2750 m with an increment of 250 m each).

Depth (m)	Average of Variance		
	0.125 Hz	0.25 Hz	0.5 Hz
250	4.6402E–07	7.6034E–07	1.2504E–06
500	4.5931E–07	7.5443E–07	1.2418E–06
750	4.5602E–07	7.5090E–07	1.2372E–06
1000	4.5375E–07	7.4849E–07	1.2343E–06
1250	4.5283E–07	7.4699E–07	1.2326E–06
1500	4.5275E–07	7.4646E–07	1.2321E–06
1750	4.5283E–07	7.4699E–07	1.2326E–06
2000	4.5375E–07	7.4849E–07	1.2343E–06
2250	4.5602E–07	7.5090E–07	1.2372E–06
2500	4.5931E–07	7.5443E–07	1.2418E–06
2750	4.6402E–07	7.6034E–07	1.2504E–06

The confidence interval exercised by this paper is 95% of the data which lies within ± two standard deviations of the mean. Small values of variance indicate that the data points tend to be very close to the mean. Based on Table 2, all values of the average variances are very small. This implies that the 2D forward GP model is capable of fitting the marine CSEM data very well. Next, for better visualization, we depict a combination of 3D surface plots of the developed 2D forward GP models for every frequency in Figure 10.
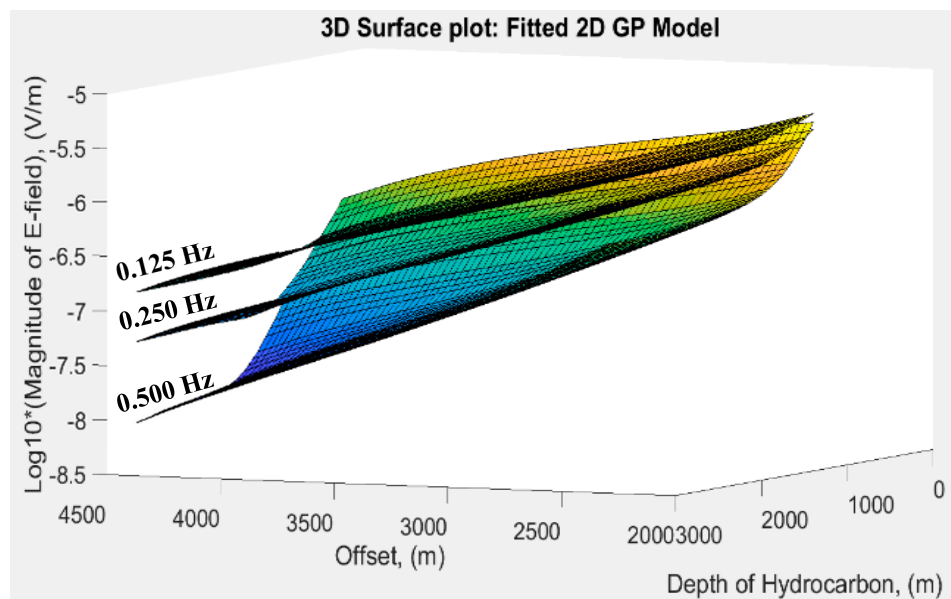
Figure 10. Combination of 3D surface plots of the 2D GP models for frequencies of 0.125, 0.25, and 0.5 Hz. The x-axis is the offset which is the source–receiver separation distance, the y-axis denotes the depth of hydrocarbon (250–2750 m), and the z-axis represents the log10 of magnitude of the electric field.

We determined the reliability of these 2D forward GP models in providing the information of EM profiles by calculating the RMSE and the CV between true (data generated through the CST software) and estimate values (data from the forward model) at unobserved/untried depths of hydrocarbon. In this section, random unobserved hydrocarbon depths (900 m and 2200 m) were selected. The EM profiles from these depths were compared with the CST computer output at the same depth levels. The RMSE and CV of the EM profiles at both depths are tabulated in Table 3.

Table 3. RMSE and CV analyses between EM profiles modelled by GP and EM profiles generated through CST software for all frequencies.

Depth (m)		0.125 Hz	0.25 Hz	0.5 Hz
900	RMSE	8.7419E–04	1.1284E–03	1.2946E–03
	CV (%)	1.4267E–02	1.7604E–02	1.9688E–02
2200	RMSE	5.8993E–04	7.4190E–04	1.3171E–03
	CV (%)	9.4615E–03	1.1241E–02	1.8402E–02

Based on Table 3, the RMSE values obtained are very small and all the CVs are generally less than 1%. This means that the modeling results of the 2D forward GP models are in good agreement with the responses acquired from the CST software even at the unobserved/untried depths of hydrocarbon.

5. Conclusions

We proposed a methodology of processing marine CSEM data using a statistical approach, Gaussian Process (GP). Based on the results, the EM responses estimated by GP are well fitted with the data generated from the CST software. The results (variance) proved that our proposed 2D forward GP model calibrated with computer simulation output is reliable for marine CSEM data-processing. In general, this 2D forward GP model, which contains EM profiles at various hydrocarbon depths, can be compared to surveyed data, and whichever estimate best matches the data measured from the survey will be the more likely case. The importance of this work lies in the application of GP methodology in multiple frequencies marine CSEM technique by developing a data-dependent model with uncertainty quantification to analyze the EM profiles and understand the geological structure

underneath the seabed. It is too risky to directly make a decision in hydrocarbon exploration without any additional analysis. There are too many challenges involved especially when it comes to deeper offshore environments. Therefore, this methodology should be a data-processing tool that provides beneficial information to hydrocarbon exploration using marine CSEM techniques by utilizing the prior information obtained from real-field data before further analysis.

Author Contributions: Conceptualization, H.D. and S.C.D.; methodology, M.N.M.A.; software, H.D.; validation, S.C.D. and K.A.M.N.; writing—original draft preparation, M.N.M.A.; writing—review and editing, M.N.M.A. and K.A.M.N.; supervision, H.D., S.C.D. and K.A.M.N.; funding acquisition, H.D.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ward, S.H.; Hohmann, G.W. *Electromagnetic Theory for Geophysical Applications*; SEG: Oklahoma City, Ok, USA, 1988.
2. Zhdanov, M.S.; Keller, G. *The Geoelectrical Methods in Geophysical Exploration*; Elsevier: Amsterdam, The Niederlande, 1994.
3. Li, Y.; Key, K. 2D marine controlled-source electromagnetic modeling: Part 1—An adaptive finite-element algorithm. *Geophysics* **2007**, *75*, WA51–WA62. [CrossRef]
4. Young, P.D.; Cox, C.S. Electromagnetic active source sounding near the East Pacific Rise. *Geophys. Res. Lett.* **1981**, *8*, 1043–1046. [CrossRef]
5. Evans, R.L.; Sinha, M.C.; Constable, S.C.; Unsworth, M.J. On the electrical nature of the axial melt zone at 13-degrees-N on the East Pacific Rise. *J. Geophys. Res.* **1994**, *99*, 577–588. [CrossRef]
6. Constable, S.; Cox, C.S. Marine controlled-source electromagnetic sounding—The PEGASUS experiment. *J. Geophys. Res.* **1996**, *101*, 5519–5530. [CrossRef]
7. MacGregor, L.M.; Constable, S.; Sinha, M.C. The RAMESSES experiment III: Controlled-source electromagnetic sounding of the Reykjanes Ridge at 57°45′N. *Geophys. J. Int.* **1998**, *135*, 773–789. [CrossRef]
8. MacGregor, L.M.; Sinha, M.; Constable, S. Electrical resistivity structure of the Value Fa Ridge, LauBasin, frommarine controlled-source electromagnetic sounding. *Geophys. J. Int.* **2001**, *146*, 217–236. [CrossRef]
9. Ellingsrud, S.; Eidesmo, T.; Johansen, S. Remote sensing of hydrocarbon layers by seabed logging: Results from a cruise offshore Angola. *Lead. Edge* **2002**, *21*, 972–982. [CrossRef]
10. Eidesmo, T.; Ellingsrud, S.; MacGregor, L.M.; Constable, S.; Sinha, M.C.; Johansen, S.; Kong, F.N.; Westerdahl, H. Sea Bed Logging (SBL), a new method for remote and direct identification of hydrocarbon filled layers in deep water areas. *First Break* **2002**, *20*, 144–152.
11. Hesthammer, J.; Boulaenko, M. The offshore EM challenge. *First Break* **2005**, *23*, 59–66.
12. Carazzone, J.J.; Burtz, O.M.; Green, K.E.; Pavlov, D.A. Three dimensional imaging of marine controlled source EM data. *SEG Expand. Abstr.* **2005**, *24*, 575.
13. Srnka, L.J.; Carazzone, J.J.; Ephron, M.S.; Eriksen, E.A. Remote reservoir resistivity mapping. *Lead Edge* **2006**, *25*, 972–975. [CrossRef]
14. Constable, S.; Srnka, L.J. An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration. *Geophysics* **2007**, *72*, WA3–WA12. [CrossRef]
15. Um, E.S.; Alumbaugh, D.L. On the physics of the marine controlled-source electromagnetic method. *Geophysics* **2007**, *72*, WA13–WA26. [CrossRef]
16. Andréis, D.; MacGregor, L. Controlled-source electromagnetic sounding in shallow water: Principles and applications. *Geophysics* **2008**, *73*, F21–F32. [CrossRef]
17. Zhdanov, M.S. Electromagnetic geophysics: Notes from the past and the road ahead. *Geophysics* **2010**, *75*, A49–A66. [CrossRef]

18. Zaid, H.M.; Yahya, N.B.; Akhtar, M.N.; Kashif, M.; Daud, H.; Brahim, S.; Shafie, A.; Hanif, N.H.H.M.; Zorkepli, A.A.B. 1D EM modeling for onshore hydrocarbon detection using MATLAB. *J. Appl. Sci.* **2011**, *11*, 1136–1142. [CrossRef]

19. Dunham, M.W.; Ansari, S.M.; Farquharson, C.G. *Application of 3D Marine CSEM Finite-Element Forward Modeling to Hydrocarbon Exploration in the Flemish Pass Basin Offshore Newfoundland, Canada*; SEG: Austin, TX, USA, 2016.

20. Cai, H.; Xiong, B.; Han, M.; Zhdanov, M. 3D controlled-source electromagnetic modeling in anisotropic medium using edge-based finite element method. *Comput. Geosci.* **2014**, *73*, 164–176. [CrossRef]

21. Persova, M.G.; Soloveichik, Y.G.; Domnikov, P.A.; Vagin, D.V.; Koshkina, Y.I. Electromagnetic field analysis in the marine CSEM detection of homogeneous and inhomogeneous hydrocarbon 3D reservoirs. *J. Appl. Geophys.* **2015**, *119*, 147–155. [CrossRef]

22. Gelius, L.J. Multi-component processing of sea bed logging data. *PIERS ONLINE* **2006**, *2*, 589–593. [CrossRef]

23. Gehrmann, R.A.S.; Schnabel, C.; Engels, M.; Schnabel, M.; Schwalenberg, K. Combined interpretation of marine controlled source electromagnetic and reflection seismic data in German North Sea: A case study. *Geophys. J. Int.* **2019**, *216*, 218–230. [CrossRef]

24. Reyes, O.C.; de la Puente, J.; Modesto, D.; Puzyrev, V.; Cela, J.M. A parallel tool for numerical approximation of 3D electromagnetic surveys in geophysics. *Comput. Sist.* **2016**, *20*, 29–39.

25. Guo, Z.; Liu, J.; Liao, J.; Xiao, J. Comparison of detection capability by the controlled source electromagnetic method for hydrocarbon exploration. *Energies* **2018**, *11*, 1839. [CrossRef]

26. Loseth, L.O.; Pedersen, H.M.; Schaug-Pettersen, T.; Ellingsrud, S.; Eidesmo, T. A scaled experiment for the verification of the SeaBed Logging method. *J. Appl. Geophys.* **2008**, *64*, 47–55. [CrossRef]

27. Daud, H.; Yahya, N.; Asirvadam, V. Development of EM simulator for sea bed logging applications using MATLAB. *Indian J. Mar. Sci.* **2011**, *40*, 267–274.

28. Chiadikobi, K.C.; Chiaghanam, O.I.; Omoboriowo, A.O.; Etukudoh, M.V.; Okafor, N.A. 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.* **2012**, *1*, 7–18.

29. Weiss, C.J. The fallacy of the 'shallow-water problem' in marine CSEM exploration. *Geophysics* **2007**, *72*, A93–A97. [CrossRef]

30. Booton, R.C. *Computational Methods for Electromagnetic and Microwaves*; John Wiley & Sons: New York, NY, USA, 1992.

31. Puzyrev, V.; Koldan, J.; de la Puente, J.; Houzeaux, G.; Vazquez, M.; Cela, J.M. A parallel finite-element method for three-dimensional controlled-source electromagnetic forward modelling. *Geophys. J. Int.* **2013**, *193*, 678–693. [CrossRef]

32. Key, K.; Weiss, C. Adaptive finite element modeling using unstructured grids: The 2D magnetotelluric example. *Geophysics* **2006**, *71*, G291–G299. [CrossRef]

33. Franke, A.; Börner, R.U.; Spitzer, K. Adaptive unstructured grid finite element simulation of two-dimensional magnetotelluric fields for arbitrary surface and seafloor topography. *Geophys. J. Int.* **2007**, *171*, 71–86. [CrossRef]

34. Li, Y.; Pek, J. Adaptive finite element modelling of two-dimensional magnetotelluric fields in general anisotropic media, *Geophys. J. Int.* **2008**, *175*, 942–954.

35. Bakr, SA.; Pardo, D.; Mannseth, T. Domain decomposition Fourier finite element method for the simulation of 3D marine CSEM measurements. *J. Comput. Phys.* **2013**, *255*, 456–470. [CrossRef]

36. Cox, C.S.; Constable, S.C.; Chave, A.D.; Webb, S.C. Controlled-source electromagnetic sounding of the oceanic lithosphere. *Nature* **1986**, *320*, 52–54. [CrossRef]

37. CST STUDIO SUITE Electromagnetic Field Simulation Software. Available online: https://www.3ds.com/products-services/simulia/products/cst-studio-suite/ (accessed on 15 July 2019).

38. Rasmussen, C.E.; Planck, M. *Gaussian Process in Machine Learning*; Institute of Biological Cybermatics: Tubingen, Germany, 2012.

39. Pal, M.; Dewal, S. Modelling pile capacity using Gaussian Process regression. *Comput. Geotech. J.* **2010**, *37*, 942–947. [CrossRef]

40. Petelin, D.; Grancharova, A.; Kochigan, J. Evolving Gaussian Process models for prediction of Ozone concentration in the air. *Simul. Model. Pract. Theory* **2013**, *33*, 68–80. [CrossRef]

41. Sun, A.Y.; Wang, D.; Xu, X. Monthly Stream flow forecasting using Gaussian Process Regression. *J. Hydrol.* **2013**, *511*, 72–81. [CrossRef]

42. Grbic, R.; Kurtagic, D.; Sliskovic, D. Stream water temperature prediction based on Gaussian Process Regression. *J. Expert Syst. Appl.* **2013**, *40*, 7407–7414. [CrossRef]

43. Liu, D.; Pang, J.; Zhou, J.; Peng, Y.; Pecht, M. Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. *Microelectron. Reliab.* **2013**, *53*, 832–839. [CrossRef]

44. Yin, F.; Zhao, Y.; Gunnarsson, F.; Gustafsson, F. Received-Signal-Strength Threshold Optimization Using Gaussian Processes. *IEEE Trans. Signal. Process.* **2017**, *65*, 2164–2177. [CrossRef]

45. Wang, H.; Gao, X.; Zhang, K.; Li, J. Fast single image super-resolution using sparse Gaussian process regression. *Signal. Process.* **2017**, *134*, 52–62. [CrossRef]

46. Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. Design and Analysis of Computer Experiments. *Stat. Sci.* **2002**, *4*, 409–428. [CrossRef]

47. Harari, O.; Steinberg, D.M. Optimal designs for Gaussian process models |via spectral decomposition. *J. Stat. Plan. Inference* **2014**, *154*, 87–101. [CrossRef]

48. Mukhtar, S.M.; Daud, H.; Dass, S.C. Prediction of hydrocarbon using Gaussian process for seabed logging application. *Procedia Comput. Sci.* **2015**, *72*, 225–232. [CrossRef]

49. Mohd Aris, M.N.; Daud, H.; Dass, S.C. Processing synthetic seabed logging (SBL) data using Gaussian Process regression. *J. Phys. Conf. Ser.* **2018**, *1123*, 012025. [CrossRef]

50. Aris, M.N.M.; Daud, H.; Dass, S.C. Prediction of hydrocarbon depth for seabed logging (SBL) application suing Gaussian process. *J. Phys. Conf. Ser.* **2018**, *1132*, 012075. [CrossRef]

51. Rasmussen, C.E.; Nickisch, H. Gaussian Processes for Machine Learning (GPML) Toolbox. *J. Mach. Learn. Res.* **2010**, *11*, 3011–3015.