Research Article
Infomax Algorithm for Filtering Airwaves in the Field of Seabed Logging

[1]Adeel Ansari, [2]Afza Bt Shafie, [4]Seema Ansari, [1]Abas B. Md Said,
[3]Muhammad Rauf and [1]Elisha Tadiwa Nyamasvisva
[1]Computer Information Sciences Department,
[2]Fundamental and Applied Sciences Department,
[3]Electrical and Electronic Engineering Department, Universiti Teknologi PETRONAS,
Bandar Seri Iskandar, 31750, Tronoh, Perak, Malaysia
[4]Electrical Engineering Department, Institute of Business Management, Pakistan

Abstract: This research focuses on applying Independent Component Analysis (ICA) in the field of Seabed Logging (SBL). ICA is a statistical method for transforming an observed multidimensional or multivariate dataset into its constituent components (sources) that are statistically as independent from each other as possible. ICA-type de-convolution algorithm, Infomax is suitable for mixed signals de-convolution, is proposed and considered convenient depending upon the nature of the source and noise model, in the application of seabed logging. Infomax is applied in the domain of marine Controlled Source Electro Magnetic (CSEM) sensing method used for the detection of hydrocarbons based reservoirs in seabed logging application. The task is to identify the air waves and to filter them out. The infomax algorithm of ICA is considered for filtering the airwaves.

Keywords: ICA, independent component analysis, infomax algorithm, seabed logging

INTRODUCTION

Independent Component Analysis (ICA) is a method for solving the blind source separation problem. It is a way to find a linear coordinate system (the unmixing system) such that the resulting signals are as statistically independent from each other as possible.

In parallel to blind source separation, unsupervised learning rules based on information theory were proposed by Linsker (1992). The goal was to maximize the mutual information between the inputs and outputs of a neural network.

The original infomax learning rule for blind source separation by Bell and Sejnowski (1995a) was suitable for super-Gaussian sources, i.e., sources with probability density functions sharply peaked with heavy tails and positive kurtosis (normalized).

There are two important properties in ICA: the natural gradient and the robustness in ICA against parameter mismatch. The natural gradient or equivalently the relative gradient gives fast convergence. A simple non-linearity used in the ICA learning rule is related to the source density model.

The application domain considered for ICA is Seabed Logging (SBL), which is a technique for finding the resistive Hydrocarbon layers under the oceanic lithosphere. This technique implies a mobile Horizontal Electric Dipole (HED) source that transmits low frequency Electromagnetic wave signals in an Omni-directional manner under water. This source antenna is towed over an array of seafloor receivers that measures the electric and magnetic field responses.

SBL exploits the resistivity contrast as Hydrocarbon reservoirs are 5-100 times more resistive than other host sediments.

In shallow water, signals that have interacted with the air dominate the response. These signals diffuses from the source through the sea water vertically to the air surface and propagates at the speed of light with no attenuation before diffusing back down vertically through water layer to the receiver on the seabed. These signals are called the airwave and carry no information about the seafloor resistivity. This airwave interferes with the signal that comes from sub-surface and due to this, it's considered as an unwanted signal.

Problem statement:

- The presence of airwaves in the signal response effects the detection of Hydrocarbon presence in the seabed logging application.
- A technique is required for filtering the airwaves from the signal response.
- Selection of appropriate algorithm is required for filtering the airwaves.

Corresponding Author: Adeel Ansari, Computer Information Sciences Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750, Tronoh, Perak, Malaysia

- The algorithm must filter out the airwave component from the loaded data.

Solution statement: Independent Component Analysis (ICA) is a popular method for signal de-convolution. The ICA algorithm, Infomax is proposed with respect to the source and noise model, for the identification and filtration of the airwaves in seabed logging application.

Objectives:

- To identify, characterize and filter the air waves within the signal response received from the receiver.
- To identify and apply the Independent Component Analysis (ICA) statistical technique for filtering the airwaves from the signal response.
- To carry out a performance-based evaluation and validation of the suitable ICA algorithm for filtering the airwaves.
- To design and develop a graphical user interface application that will perform ICA to filter out the airwave component by appropriate algorithm from the loaded data.

LITERATURE REVIEW

Seabed logging technique: Sea bed logging uses active source Electro Magnetic (EM) sounding technique in detecting subsurface hydrocarbon. The CSEM method uses a Horizontal Electric Dipole (HED) source to transmit low frequency (typically 0.01-10 Hz) signals to an array of receivers that measure the electromagnetic field at the seafloor (Loseth *et al*., 2010). The method relies on the large resistivity contrast between hydrocarbon-saturated reservoirs and the surrounding sedimentary layers saturated with aqueous saline fluids. Hydrocarbon reservoirs typically have a resistivity of 30-500 Ωm), whereas the resistivity of the over and underlying sediments is typically less than a few Ωm. Both the amplitude and the phase of the received signal depend on the resistivity structure beneath the seabed (Shaw *et al*., 2006).

By studying the variation in the resistivity contrast of the Hydrocarbon layer and thickness, as the transmitting source is towed through the receiver array, the effects of the Electric field at different offsets can be determined at scales of a few tens of meters depths of several kilometers. According to Bell and Sejnowski (1995b), as depicted by Fig. 1, the receivers record the EM responses as a combination of energy pathways including signal transmitted directly through seawater, reflection and refraction via the sea-water interface, refraction and reflection along the sea bed and reflection and refraction via possible high resistivity subsurface layers.
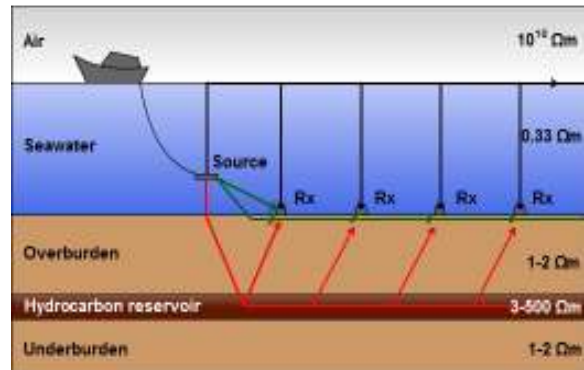

Fig. 1: A schematic diagram of seabed logging application showing the direct waves, air waves, reflected waves and refracted waves
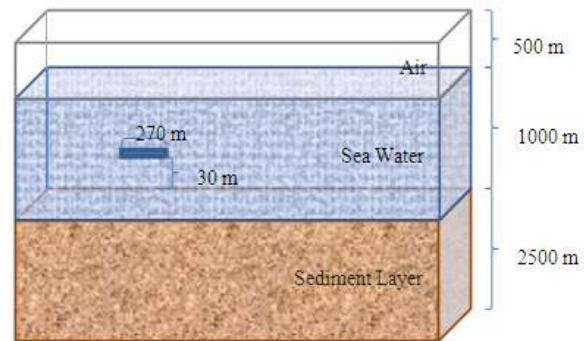

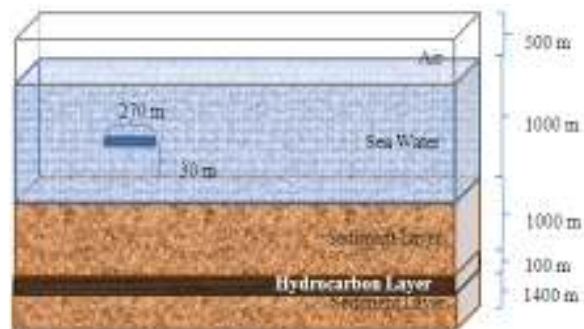Fig. 2: Simulation model 1 showing no presence of hydrocarbon reservoir

Fig. 3: Simulation model 2 showing presence of hydrocarbon reservoir of 100 m thickness, at target depth of 1000 m

In the following sections it will be demonstrated that this resistivity contrast and thickness has a detectable influence on SBL data collected at the sea bed above the reservoir. The effect of the reservoir is detectable in SBL data at an appropriate frequency of 0.1 to 10 Hz (Cardoso, 1998). For this simulation, the frequency is set to 0.125 Hz and the current is 1250 A at the transmitter end.

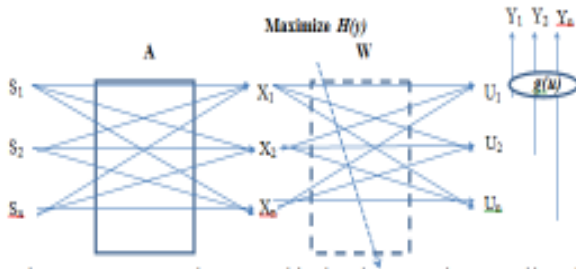**Simulation model 1:** The simulation model proposed here contains no Hydrocarbon reservoir, so as to

Fig. 4: The instantaneous mixing and unmixing model. Independent sources s become mixed by A. The observed sources are x. The goal is to learn W that inverts the mixing A and u are the estimates of the recovered sources. The infomax approach is one way to find the unmixing system W. It requires a nonlinear transfer function g (u)

determine the Electric Field at varying offset during the absence of hydrocarbon reservoir for sea water depth at 100, 500 and 1000 m, respectively. The result obtained from this model will then be used in comparison with the graphs obtained from simulation model 2, which contains the presence of Hydrocarbon reservoirs (Fig. 2).

Simulation model 2: The simulation model proposed here contains a 100 m thickness of Hydrocarbon Layer with a resistivity contrast of 100 Ωm (Fig. 3).

Assumptions:

- The representation of the layer of Hydrocarbon reservoir is considered as a rectangular cuboid.
- For simulation 2, the resistivity of the second Hydrocarbon is with a constant thickness of 100 m at each target depth.
- For simulation 2, the thickness of the second Hydrocarbon is with a constant resistivity of 100 Ωm at each target depth.

Illustration of infomax algorithm: Assume that there is an M-dimensional zero-mean vector $s(t) = [s_1(t),\ldots, s_M(t)]^T$, such that the components $s_i(t)$ are mutually independent. The vector $s(t)$ corresponds to M independent scalar-valued source signals $s_i(t)$.

A data vector $x(t) = [x_1(t),\ldots, x_N(t)]^T$ is observed at each time point t, such that:

$$x(t) = As(t)$$

where, A is a full rank NxM scalar matrix. As the components of the observed vectors are no longer independent, the multivariate p.d.f. will not satisfy the p.d.f. product quality. If the components of $s(t)$ are such that at most one source is normally distributed then it is possible to extract the sources $s(t)$ from the

received mixtures $x(t)$. The mutual information of the observed vector is given by the Kullback-Leibler (KL) divergence of the multivariate density from the product of the marginal densities (Fig. 4).

Kullback-leibler entropy and mutual information: The Kullback-Liebler entropy, also referred to as the Kullback-Leibler divergence, also termed the relative or cross entropy, is a measure of the distortion between the two distributions. The Kullback-Liebler entropy is not a symmetric measure between distributions. The equivalent of this divergence measure is:

$$D(p||q) = \int p_x(x) log \frac{p_x(x)}{q_x(x)} dx$$

The mutual information between variables gives a measure of their statistical independence and is of importance in the field of blind source separation and independent component analysis. For two variables, the mutual information is defined as the Kullback-Liebler divergence between the joint probability density of the variables and the factorized marginal.

Mutual information is a measure of the amount of information that each variable conveys about the other. This can be considered as a measure of the statistical correlation between the variables.

Steps in infomax algorithm: Steps involved in the Information maximization algorithm are as follows:

- Maximize the joint entropy of output Y, which is non-linear.
- If Y is optimized with weights, W, then go to step 3, else go to step 1.
- If the mutual information between the outputs zero:
- If yes then the estimated source density is equivalent to the derivation of the maximum entropy.
- Else, the estimated source density with the errors is equivalent to the derivation of the maximum entropy. The error has to be removed.
- The gradient of the maximum of the joint entropy is evaluated.
- Find Jacobian determinant of transformation from S to X and U to Y.

Flowchart of infomax algorithm: (Fig. 5)

RESULTS AND DISCUSSION

In this research progress, a detailed study and write-up over the infomax algorithm is conducted. The infomax algorithm works in a manner, from which the maximum entropy of the mixed signal is determined
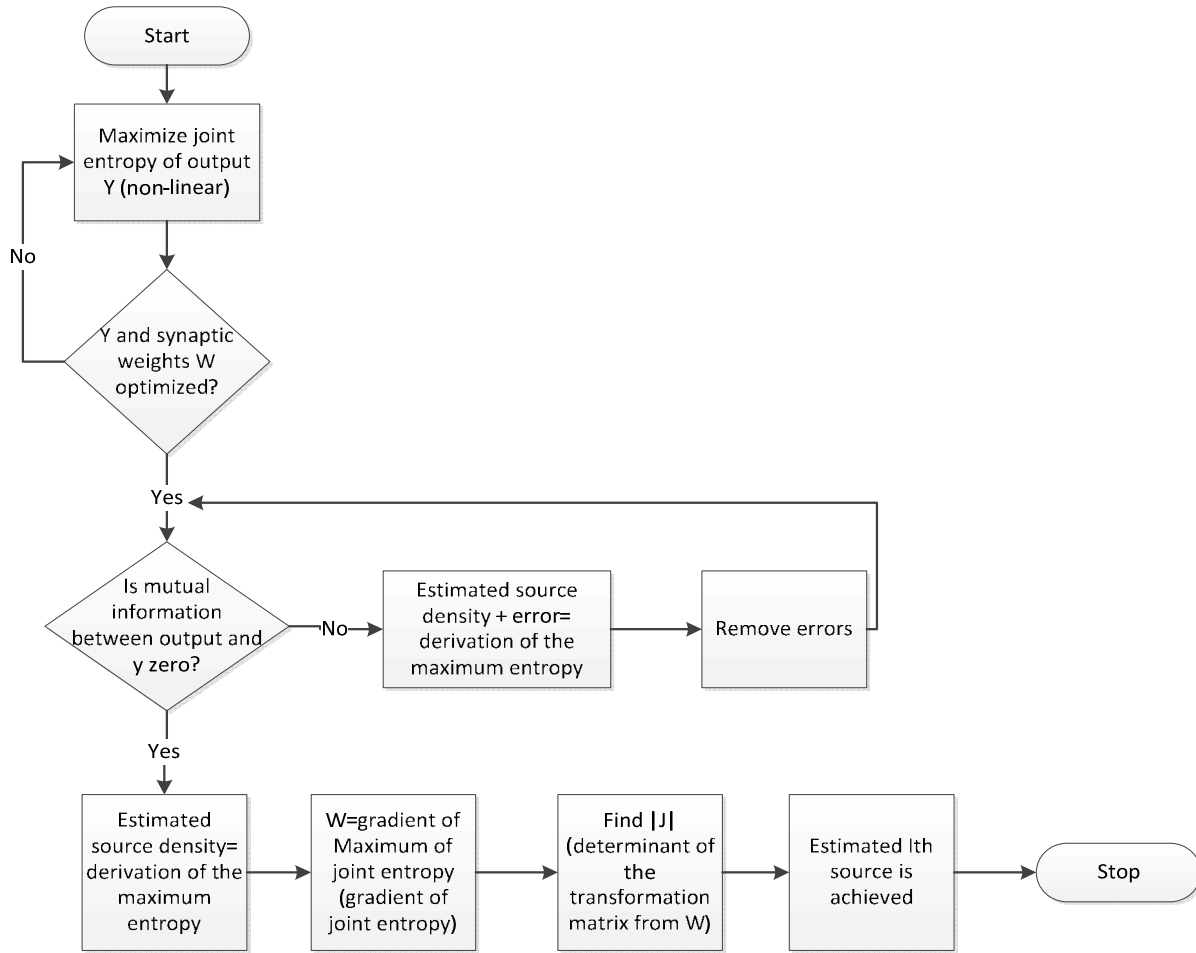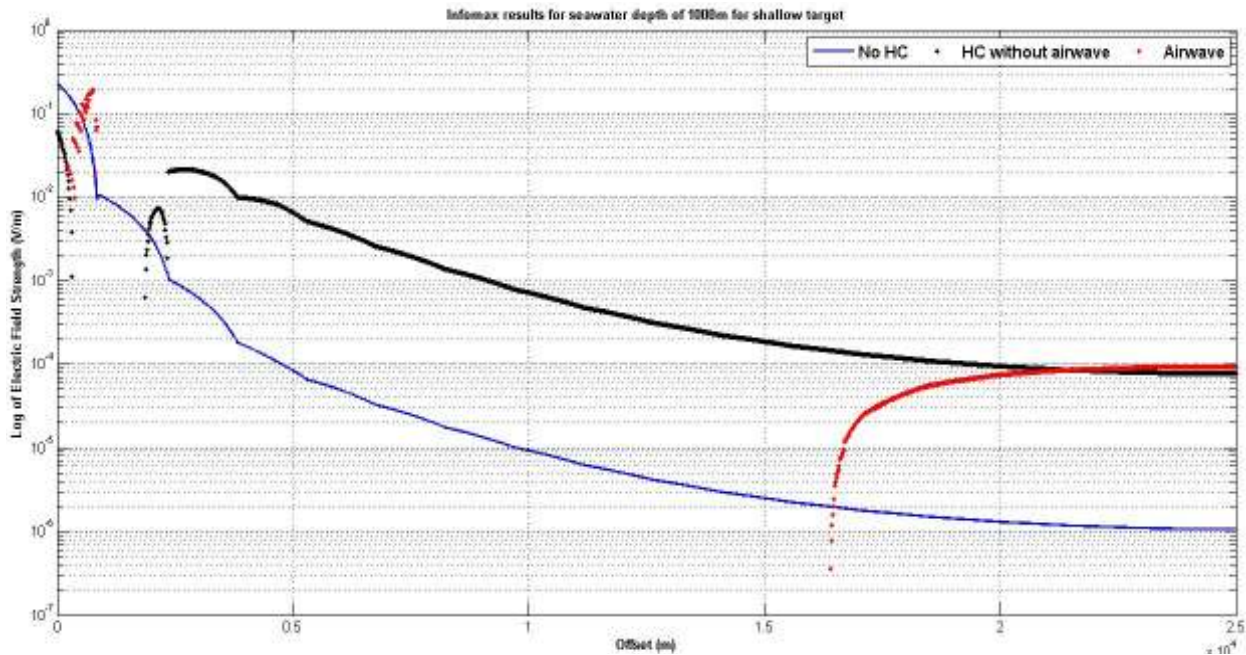
Fig. 5: Flowchart of infomax algorithm

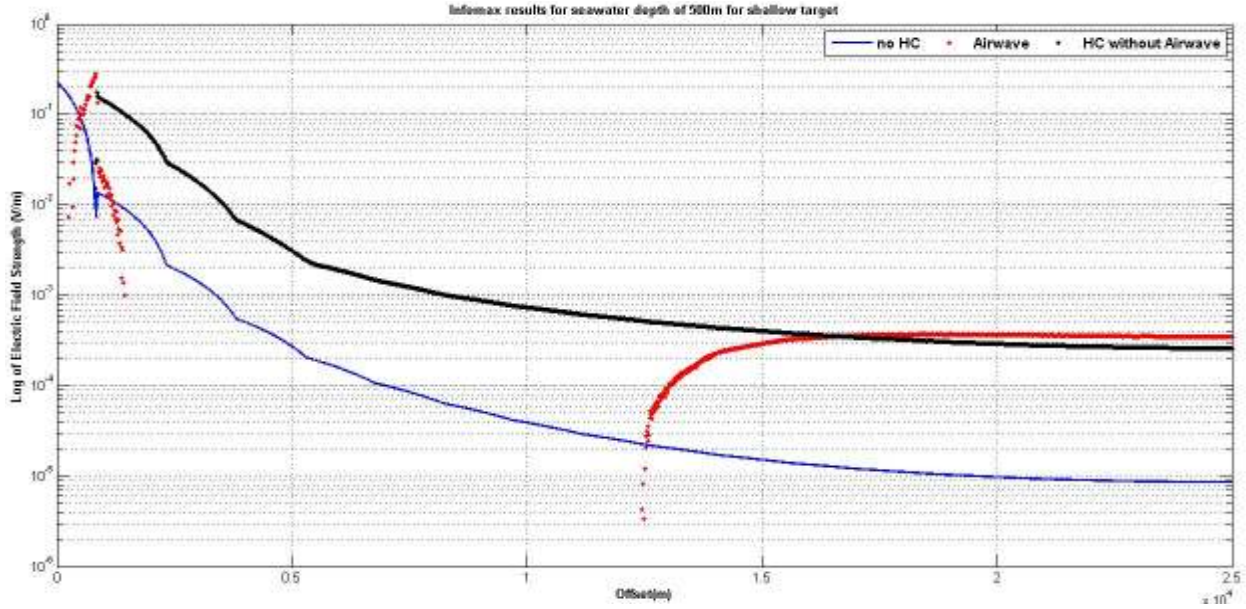Fig. 6: Results from infomax algorithm for 1000 m sea water depth

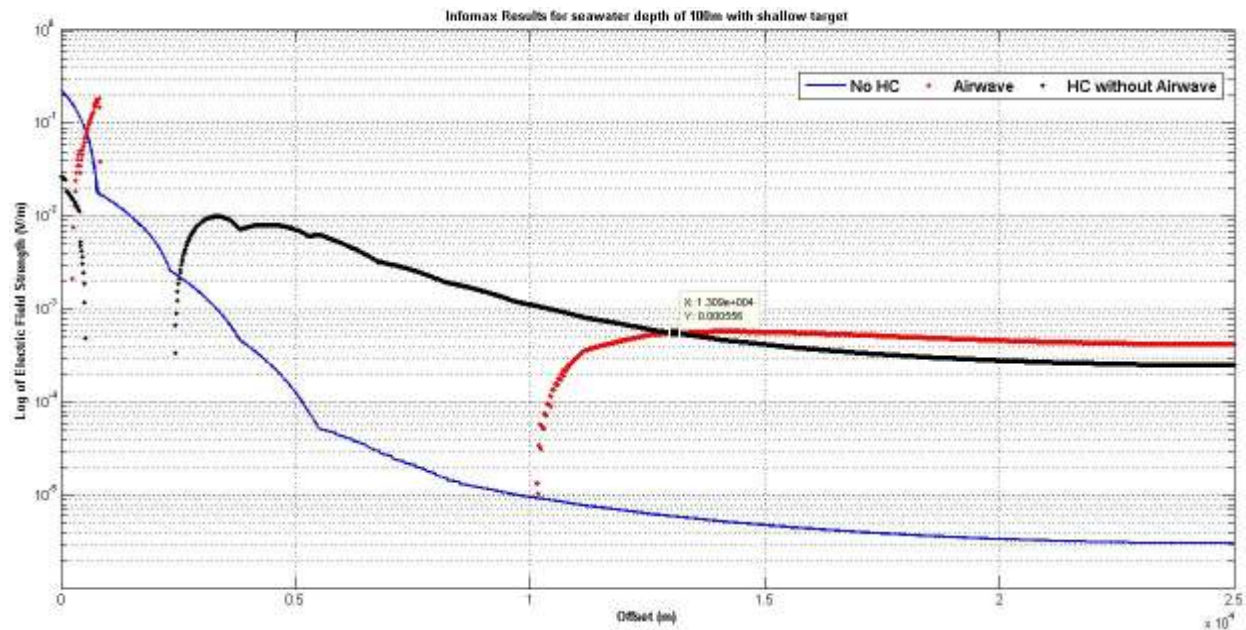Fig. 7: Results from infomax algorithm for 500 m sea water depth

Fig. 8: Results from infomax algorithm for 100 m sea water depth

and deducted from the mixed signal, resulting in only the marginal entropies of the independent signals. The signals fed into the algorithm are the no Hydrocarbon, with Hydrocarbon and the numerical airwave modeled as in Janniche and Lasse (2007).

From Fig. 6, seawater depth of 1000 m is considered for simulation models 1 and 2. The infomax result as shown in Fig. 6 depicts three curves, the curve representing no hydrocarbon, the curve representing the airwave and the curve stating the presence of hydrocarbon without the airwave.

Table 1: Percent difference between the airwave and without the airwave

Seawater depth (m)	Without HC	Airwave	Difference (%)
100	0.0002520	0.000420	-67
500	0.0002614	0.000348	-33
1000	7.72E-05	9.34E-05	-21

From Fig. 7, seawater depth of 500 m is considered for simulation models 1 and 2. The infomax result as shown in Fig. 7 depicts three curves, the curve representing no hydrocarbon, the curve representing the airwave and the curve stating the presence of hydrocarbon without the airwave.
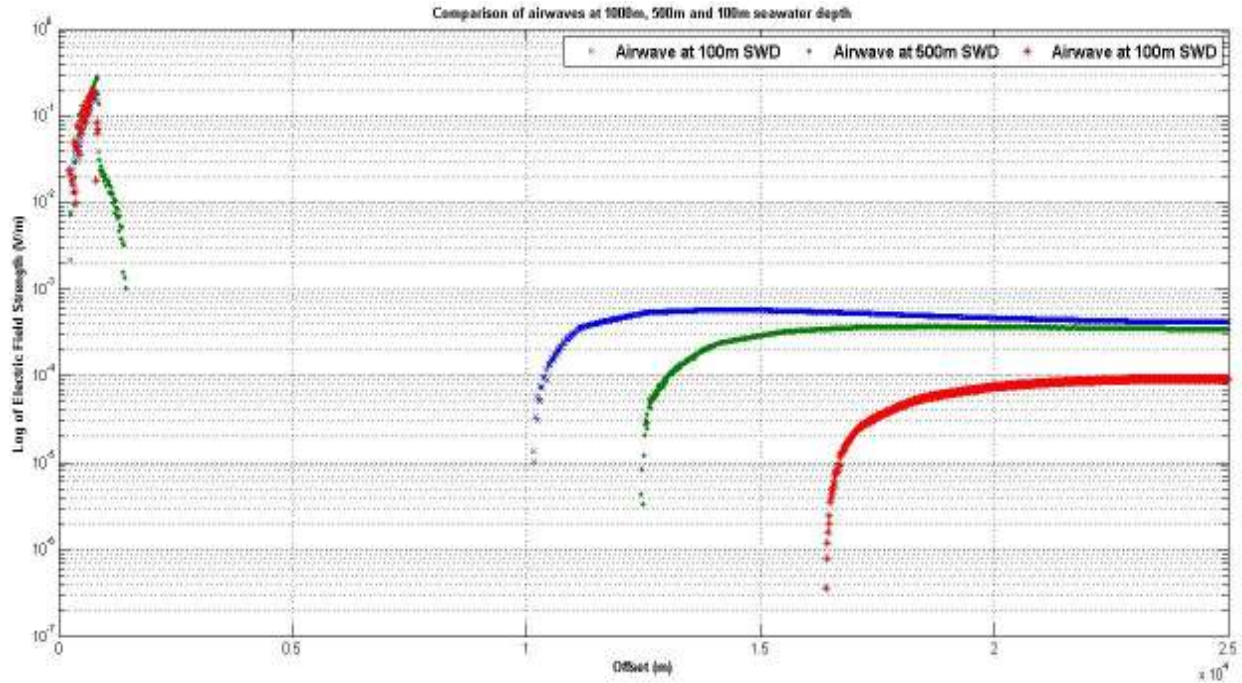
Fig. 9: Comparison of the airwaves from infomax algorithm for seawater depth of 100, 500 and 1000 m, respectively

Table 2: Percent difference between the airwave of varying seawater depth at offset 25 km

Seawater depth	100 m	500 m	100 m	1000 m	500 m	1000 m
Airwave	0.00042	0.000348	0.00042	9.34E-05	0.000348	9.34E-05
Difference	-21%		-350%		-273%	

From Fig. 8, seawater depth of 100 m is considered for simulation models 1 and 2. The infomax result as shown in Fig. 8 depicts three curves, the curve representing no hydrocarbon, the curve representing the airwave and the curve stating the presence of hydrocarbon without the airwave.

From Fig. 6 to 8, we can clearly see that the airwave is discontinuous and is commencing from offset near 10 km (Fig. 8), near 13 km (Fig. 7) and near 16 km (Fig. 6). With increasing depth, the strength of this half or broken curve is decreasing at far offset, see Table 1, at 25 km offset, the electric field strength is decreasing. By reason, direct wave is present at near offsets and the airwaves dominate at far offset. This logic manifests the presence of this broken or half curve being generated by the algorithm, that it is disregarding the portion of the direct wave and is starting from negligible values up to a certain strength and then steadily declining and converging with the curve, that represents the presence of Hydrocarbon without the airwave.

Figure 9, compares the airwaves depicted from the different seawater depths and from conclusion, we can easily see that the intensity of the curve considered as the airwave is losing its strength and dominance with respect to higher seawater depth. Table 1 clearly illustrates at 25 km offset, the strength reduction of the electric field strength (V/m) (Table 2).

Also, from Fig. 6 to 8, the curve representing no airwave but with the presence of the hydrocarbon (based over simulation model 2), the curves dominate longer than compared to the airwave, most likely due to the presence of the guided wave, some presence of noise like the direct and airwaves (to a negligible quantity). This curve is not commencing from the zeroth offset, rather from 1 and 3 km. This sudden gap could be due to the mutual information being eradicated from the mixed signal.

So far, from the graphs and tabular results, it is evident that the results from the Infomax seems reasonable to a certain extent but the results would require more validations so as to confirm the curves represent whether, as the airwaves, the curve without the airwave depicting the presence of the hydrocarbon and other noise presence.

CONCLUSION

In conclusion, we have illustrated in detail the infomax algorithm that exploits the entropy phenomenon to determine the mutual information between the different curves, so as to decipher them as independent sources. The results obtained so far, give us some logical reasoning about the presence of airwave, without the airwave and the negligible noise

being present. For further validation of these results, it is better to perform various simulation models of hydrocarbon at deeper targets, so as to better determine the presence of the airwave easily and the signal without the presence of the airwave.

REFERENCES

Bell, A.J. and T.J. Sejnowski, 1995a. Fast blind separation based on information theory. Neural Comput., 7: 1129-1159.

Bell, A.J. and T.J. Sejnowski, 1995b. An information-maximization approach to blind separation and blind deconvolution. Neural Comput., 7: 1129-1159.

Cardoso, J.F., 1998. Blind signal separation: Statistical principles. Proc. IEEE, 86(10): 2009-2047.

Janniche, I.N. and A. Lasse, 2007. Asymptotic airwave modeling for marine controlled-source electromagnetic surveying. Geophysics, 72(6): F249-F255.

Linsker, R., 1992. Local synaptic learning rules suffice to maximize mutual information in a linear network. Neural Comput., 4: 691-702.

Loseth, L.O., L. Amundsen and A.J.K. Jenssen, 2010. A solution to the airwave-removal problem in shallow-water marine EM. Geophysics, 75(5): A37-A42.

Shaw, A., A.I. Al-Shamma'a, S.R. Wylie and D. Toal, 2006. Experimental investigations of electromagnetic wave propagation in seawater. Proceeding of the 36th European Microwave Conference. Manchester, UK.