

# AUTOMATIC DATA COLLECTION DESIGN FOR REAL-TIME DETECTION OF OIL-SPILLAGE DISASTERS IN NIGERIA

**Bakpo, F. S. and Agu, M. N.**  
Department of Computer Science  
University of Nigeria, Nsukka  
email: fbakpo@yahoo.com

## ABSTRACT

*Oil-spillage disaster is a phenomenon that has thrived and come to stay especially in the Niger Delta region in Nigeria where crude oil is produced both on-shored and off-shored resulting in the loss of billions of crude oils and corresponding loss of income every year. Such disasters can occur at any point within the entire oil-pipeline topology, which is geographically dispersed across the country spanning from the extreme south-south through south-east and south-west to the north central. A major cause of oil spillage disaster is oil-pipeline vandalism, which has recently become one of the primary means of defrauding the Nation by oil bunkers and even angry youths of the host communities. Currently in Nigeria oil spillage disaster is monitored manually under the auspices of the National Emergency Management Agency (NEMA) and involves huge human labor as gathering of useful data has proven to be difficult and untimely using traditional data collection methods given the peculiarity of the operational terrain. An automated data collection design (ADCD) for real-time data collection on oil spillages in the on-shore, mangrove, thick forest and off-shore areas of the Nation offers the much needed solution. This paper examined some current procedures for data collection and highlighted inherent pitfalls. It further presented a robust architecture and model for real-time detection of oil-spillage and discussed incorporated contemporary technological requirements. The major advantages and disadvantages of the proposed system were also discussed.*

**Keywords:** Data collection, wireless networks, detection, neural networks, oil- spills

## 1. INTRODUCTION

Nigeria is the largest oil producer in Africa and the sixth largest oil exporter in the world [1]. The mainstay of Nigeria's economy is the petroleum sector, contributing about 90% of the nation's foreign exchange earnings and about 25% of the Gross Domestic Products [2]. A significant proportion of the nation's oil is produced onshore, although recently oil production has witnessed increased activities in the off-shores. The crude oil

produced from oil fields located at various points in the country including south-south, south- west, and north central are transported by pipelines to four major refineries of Alesa Eleme, Warri, and Kaduna. Over the years, the amount of oil produced and transported between points of production, processing and distribution or export terminals has greatly increased as the demand and dependency on oil increased. Although this increase in oil production level contributes to the national economic

growth, studies have shown that it presents increased potential for environmental pollution and degradation [3, 4]. It has also been observed that thousands of barrels of oil have been spilled into the environment through oil pipelines and storage facilities' failure in Nigeria [5]. The causes of pipeline damage and leakage vary greatly from material defects, pipe corrosion and contact with ship anchors in the offshore operations to vandalism of pipeline onshore.

Oil spillage disaster through pipeline vandalism is a phenomenon that has thrived and come to stay especially in the Niger Delta region in Nigeria, where crude oil is produced both on-shored and off-shored. This has resulted to the loss of billions of crude oils and corresponding loss of income every year. An average of 35,000 barrels of

crude oil is stolen per day in circumstances that threaten lives and the environment [5]. Apart from the lost of lives and property through pipeline fire, the run-off from impacted sites usually degrade the quality of the fresh water sources which serves the domestic rural water supply needs of most communities in Nigeria. The enormous oil installations deployed in the Niger Delta region explains their vulnerability to vandalism. Presently, the Niger Delta region plays host to 606 oil fields of which 355 fields are onshore while 251 are offshore with over 3,000 kilometers of pipelines crisscrossing the region and linking some 275 flow stations to various export terminals. Figure 1, shows the map of the Niger Delta region in Nigeria.



Fig. 1 Map Delta Region

showing Niger in Nigeria

(source, [4])

Oil spillage disasters resulting from pipeline vandalism have continued to be a challenge to spill management in Nigeria and at the moment has defiled all contingency arrangements of oil industry operators with most incidents occurring along major pipelines and manifolds. The major reasons attributed to vandalism of pipeline vary ranging from economic to political. Some of the attacks are carried out by militants who say they are fighting for a fairer share of the natural wealth, others by criminal gangs engaged in a multi-million-dollar trade in stolen oil. It was also discovered in a recent released by security experts that groups of local community youths alias “area youths” frustrated by poverty and a lack of job opportunities, sabotage pipelines in the hope of winning repair contracts [6]. In 2006 Shell Petroleum Development Company- one of the major oil operators in Nigeria recorded 241 oil spill incidents. Of this number sabotage accounted for 165 (69%), while 50 (20%) were controllable incidents (resulting from equipment failure, corrosion or human error). The remaining 26 incidents are yet to be classified due to access restrictions either by communities or the current insecurity in the Niger Delta. Measures taken so far by the Government to manage oil spillages in

Nigeria include the setting up of various legislative frameworks to govern the petroleum industry such as: Federal Environmental Protection Agency Act, 1990; National Oil Spill Detection and Response Agency Act, 2006, National Emergency Management Agency (NEMA), Ministry of Niger Delta, etc. Currently, oil spillage is monitored manually under the provision of the above acts and involves huge human labor as gathering of useful data has proven to be difficult and untimely using traditional data collection method. The rising tide of violence, hostage taking in the Niger Delta region also has significant consequences on oil spill management acts particularly in the area of quick access to spill sites. In this paper, we proposed an automatic data collection design for real-time detection of oil spillage in on-shore, offshore and mangrove forests of the country.

**2. COMPARISON OF CURRENT DATA COLLECTION PROCEDURES**

This section attempts to compare the formal data collection procedures [7]. The strengths and limitations of each method are summarized in Table 1

**Table 1:** Formal Data Collection Procedures

Formal Methods of Data Collection	Strengths	Weaknesses
1. Questionnaires or Structured Interviews	(i) Consistent for everyone (ii) Relatively easy to code (iii) Good when you need data for large numbers of people or need quantitative data (iv) Provides information on “scope” (i.e., how many people do a behavior or are affected by an issue).	(i) Little flexibility for people to raise their own issues (ii) Little opportunity for people to respond in their own words (iii) Little opportunity to go into depth on any issue (iv) Usually limited to quantitative data
2. Unstructured or Semi-Structured Interviews	(i) Permits people to raise issues important to them and speak in their own words (ii) Rapport encourages sharing; (iii) Provides	(i) Time-consuming (ii) Fewer people can be interviewed than surveyed (iii) More difficult to code (iv)

	information on “dept” (why do people feel or act as they do and what does it mean to them) (iv) Open-ended responses can be analyzed qualitatively or quantitatively	Interviewers need in-dept training to interview well
3. Focus Groups	(i) Interaction encourages people to elaborate on their responses (ii) Ideas can be shared (iii) More people can be interviewed in groups than one-to-one (iv) Program staff can get a lot of good suggestions in a short period of time.	(i) Some people do not speak well in groups and others may dominate discussions (ii) Will only work if moderator has good facilitation skills (iii) Not good if you need data from individuals or if topic is sensitive (iv) Scheduling can be problematic for busy people or those with family responsibilities
4. Journals	(i) Provide an ongoing record of people's felt experiences with the program (ii) Focus participants and others on their experiences as they happen (iii) Can provide extensive detailed data to be used in analyses or as ideas for interviews or focus groups	Time-consuming to keep (ii) Some people do not communicate well in written English (iii) Some people may feel uncomfortable writing about themselves or feel they are giving up their privacy.
5. Observation	(i) Provides data other than verbal self-report (ii) Provides an opportunity for in-depth study of behavior, nonverbal communication and the physical and social environment (iii) Allows immersion in the whole context of the program.	(i) Time-consuming and often expensive (ii) May seem intrusive to program staff and participants (iii) Observers need extensive training (iv) Conflict over role may arise (e.g. to what extent can observer be a participant?).
6. Population Statistics	(i) Provide information about change on a broad scale (ii) Useful if program is geared to a large population (iii) Usually are easily available and regularly collected (though government-spending cuts have led to less availability of some stats or higher charges for accessing the data)	(i) Limited to quantitative data (ii) Are influenced by many non-program factors (iii) Short-term changes may be chance fluctuations rather than trends (iv) Only useful if large samples from large populations.

### 3. ARCHITECTURE FOR AUTOMATIC DATA COLLECTION DESIGN

#### • Functions Description

The major functionalities of the automatic data collection design are as follows: (i) to facilitate real-time data collection and (ii) to react to a suspicious pipeline cut or leakage. Consequently, the design of the architecture is based on the following conditions:

- (i) The system provides a non-stop monitoring of pipeline topology within a piloted community of operation.
- (ii) The entire pipeline topology within the area is divided into segments. Each segment is composed of a microcontroller that manages the properties of the segment inscribed by a database known as the wireless sensor control block (WSCB). The WSCB describes the characteristics of

the segment as they relate to a generic set of functions supported such as; sensor identification (SID), sensor location (SL), oil pressure level (OPL), system time, etc.

- (iii) The set of wireless sensors known as wireless sensor network covering the piloted community is monitored by a dedicated satellite, which is connected via GPRS modem to a server in one central office. Real-time data collected by the office are further preprocessed using a technique of artificial neural network known as self-organizing neural network for a better pattern (event) recognition and classification.
- (iv) For a pipeline cut or leakage to occur, a third-party (or an intruder) must apply a certain degree of force to a cross-section of the pipeline. If the degree of forces applied exceeds a preset threshold value, the sensor node of the affected segment sends an alarm to the server, which in turn is detected by the central office. The system is interrupt-driven. This enables appropriate service routines to be deployed for inspection without wasting time. The proposed architecture is shown in Figure 2.

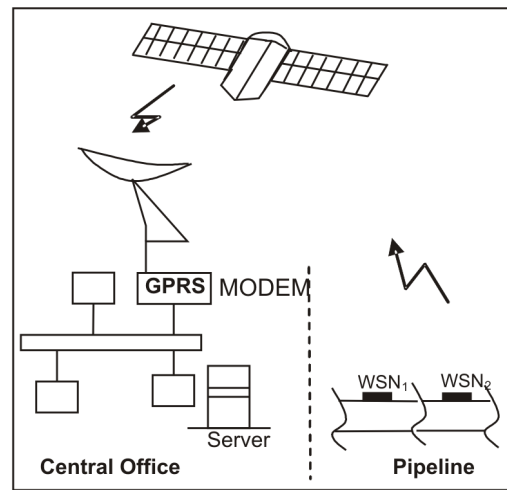


Fig . 2 Oil - Pipeline cut detection system

• **Model Description**

An example of self-organizing map or unsupervised learning algorithm, which solves the well known pattern classification problem, is shown in figure 3. A common self-organizing principle is to search for clusters in the input data distribution. The idea is to find the set of center locations  $w_k$ , or cluster centers, that best describe the data set.

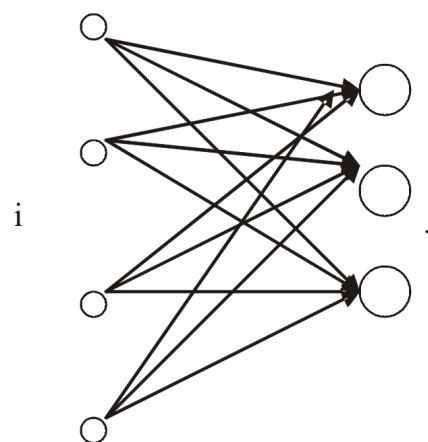


Fig. 3:

The winner-takes-all system

The network is fully connected with weights  $w_{ij}$  connecting neuron  $j$  in the output layer to neuron  $i$  in the input layer. The activation of neuron  $j$  in the output layer is given by

$$a_j = \sum_i w_{ij} x_i = w_j^T x \tag{1}$$

Here,  $w_{ij} = (w_{11}, w_{1j}, w_{21}, w_{ij})$  is the weight vector,  $x_i = (x_1, x_2, \dots, x_i)$  is the input feature vector and each component of  $w$  and  $x$  lies in 0 to 1 (i.e., normalized). The augmented weight vector  $w_{ij}$  is to be estimated based on the available training data. If  $a_i = (a_1, a_2, \dots, a_n)$  is the corresponding class label, then Equation (1) becomes the model:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_i \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1j} \\ w_{21} & w_{22} & w_{23} & \dots & w_{2j} \\ w_{31} & w_{32} & w_{33} & \dots & w_{3j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{i1} & w_{i2} & w_{i3} & \dots & w_{ij} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_i \end{bmatrix} \tag{2}$$

Solution to Equation (2) is possible by searching for clusters in the input data distribution.

Let  $x_1, x_2, \dots, x_n$  be a set of  $n$  samples and  $w_1, w_2$  are some labels representing low and high risk clusters. The goal is to use these training samples to determine the weights in Equation (2) such that

$$a(x_i) = wx \begin{cases} \geq 0, & x_i \in w_1 \\ < 0, & x_i \in w_2 \end{cases} \tag{3}$$

If the conditions in Equation (3) are met, the data set  $x_1, x_2, \dots, x_n$  are called linearly separable. We then find the set of centre locations  $w_k$ , or cluster centres, that best describe the data set. This is done by

minimizing e.g. the reconstruction error or Mean Squared Error (MSE):

$$E = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^K \Lambda_k[X(n)] \|X(n) - w_k\|^2 \tag{4}$$

Subject to the constraint that  $k \ll N$  ( $k$  can be either fixed or adaptive). Here  $\Lambda_k[x(n)]$  is a membership function that takes the values:

$$\Lambda_k[x(n)] = \begin{cases} 1, & \text{if } \|x(n) - w_k\| < \|x(n) - w_j\| \forall j \neq k \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

The MSE solution has the advantage of guaranteeing convergence. Gradient descent on Equation (4) leads to the updating equation.

$$w_k^{(n+1)} = w_k^{(n)} + \eta \sum_{n=1}^N \Lambda_k[x(n)] [x(n) - w_k] \tag{6}$$

Where  $\eta$  - is a small learning constant. It should be noted that from the identity:

$$w_k = w_k + \eta(x - w_k) = (1 - \eta)w_k + \eta x \tag{7}$$

It follows that the updated weight vector is a convex linear combination of the old weight and the pattern vectors. In the end of the training process the final weight vectors point to the centers of gravity of classes. The network is trainable only if classes/clusters of patterns are linearly separable from other classes by hyper-plane passing through origin. To ensure separability of clusters with a priori unknown numbers of training clusters, the unsupervised training is performed with an excessive number of neurons, which

provides a certain separability safety margin. During training neurons whose weights change chaotically are discarded, since their output may not provide any essential clustering information. The weights of remaining neurons should settle at values that are indicative of clusters.

- **Advantages and Disadvantages of ADCD**

The major advantages of the proposed system over existing ones include:

- Automatic data collection
- Real-time detection and localization of pipeline cut
- Large data handling
- Parallel detection
- Good visualization
- Indigenous system design

On the other hand, some of the disadvantages are:

- Requirement of constant power supply at the centre
- Requirement of human experts for 24 hours monitoring
- Equipment failure

#### 4.0 PERFORMANCE ANALYSIS

This section outlined the performance benchmark used for validating the model and compared the performance of ADCD with other oil leak detection models.

#### 4.1 PERFORMANCE BENCHMARK

The proposed architecture relies on a piece of software for real-time detection of oil pipeline cut or leakages. The principles underlying detection software are grounded in classical statistical decision theory. There are two sources that generate inputs to the detection software: pipeline cut ( $H_0$ ) and no cut ( $H_1$ ). The goal of the detection software is to classify each encroachment as cut or

no cut. Two types of errors can occur in this classification:

- Classification of a fraudulent encroachment as legal (false negative); and
- Classification of a legal encroachment as fraudulent (false positive).

We define:

Probability of detection  $P_D = Pr$  (classify into  $H_1|H_1$  is true), or

Probability of false negative =  $1 - P_D$ .

Probability of false positive  $P_F = Pr$  (classify into  $H_1|H_0$  is true).

Let the numerical values for the no cut ( $n$ ) and cut ( $c$ ) encroachments follow exponential distributions with parameters  $\lambda_N$  and  $\lambda_C$ ,  $\lambda_N > \lambda_C$ , respectively. Then we can write the probability of detection  $P_D$  and probability of false positive  $P_F$  as

$$P_D = \int_t^{\infty} \lambda_C e^{-\lambda_C x} dx = e^{-\lambda_C t} \quad (8)$$

$$P_F = \int_t^{\infty} \lambda_N e^{-\lambda_N x} dx = e^{-\lambda_N t} \quad (9)$$

$P_D$  can be expressed as a function of  $P_C$  as

$$P_D = P_C^r, \quad (10)$$

Where  $r = \lambda_C / \lambda_N$  is between 0 and 1. Consequently, the quality profile of most detection software is characterized by a curve that relates its  $P_D$  and  $P_C$ , known as the receiver operating curve (ROC) [8]. ROC curve is a function that summarizes the possible performances of a detector. It visualizes the trade-off between false alarm rates and detection rates, thus facilitating the choice of a decision functions. A comprehensive introduction to ROC curves can be found, for example, in [9, 14]. Figure 4 shows sample ROC curves

for various values of  $r$ . The smaller the value of  $r$  the steeper the ROC curve and

hence the better the performance.

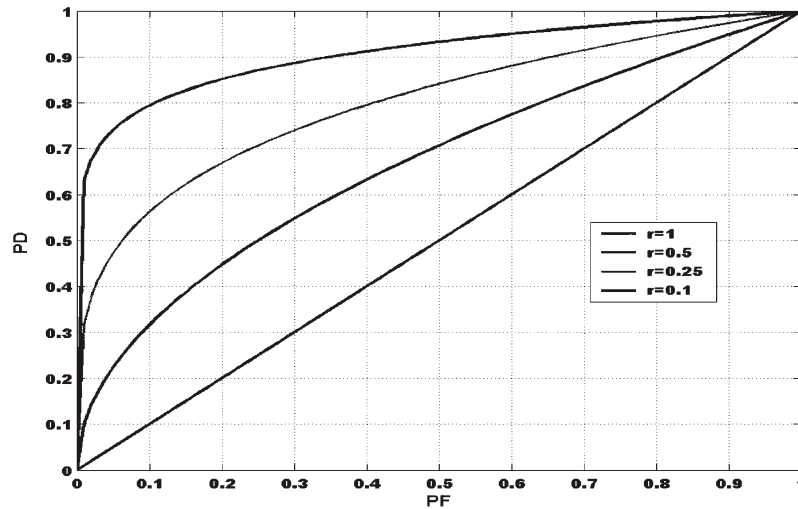


Fig. 4 ROC curves for different  $r$

#### 4. 2 Comparison of ADCD with Two other oil Leak Detection Models

The ROC curve is also used for this comparison. As stated in section 3, if the degree of forces exacted at a cross-section of the pipeline exceeds a certain threshold value, leakage is said to occur. For the analysis, we simulated such degree of

forces as may be exacted by vandal using progress bar in accordance to the following ranges: 0 to 25% represents no cut and 26 to 100% represents cut. This is analogous to the leak characteristic presented in [10] and shown in Figure 5.

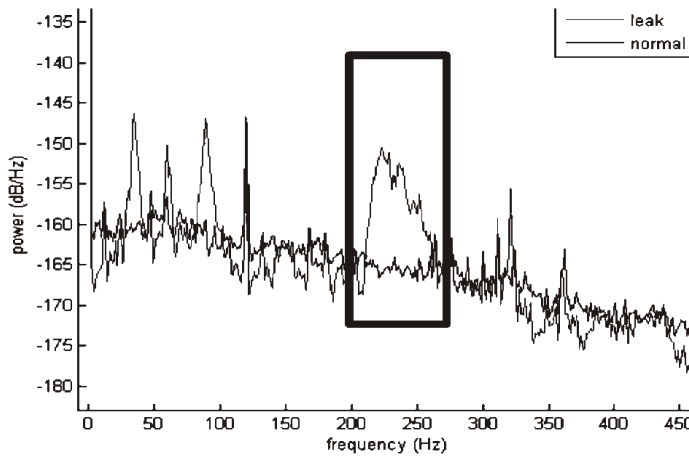




Figure 5 Leak characteristics (Source: [10])

Two different available products such as PIPENET [10] and ATMOS pipe [11, 12] were selected to test the feasibility of the proposed ADCD tool for this purpose. The performance analysis of the respective

detection algorithms was carried out using MATLAB software package [13] and the results compared with the proposed ADCD model as shown in Figure 6.

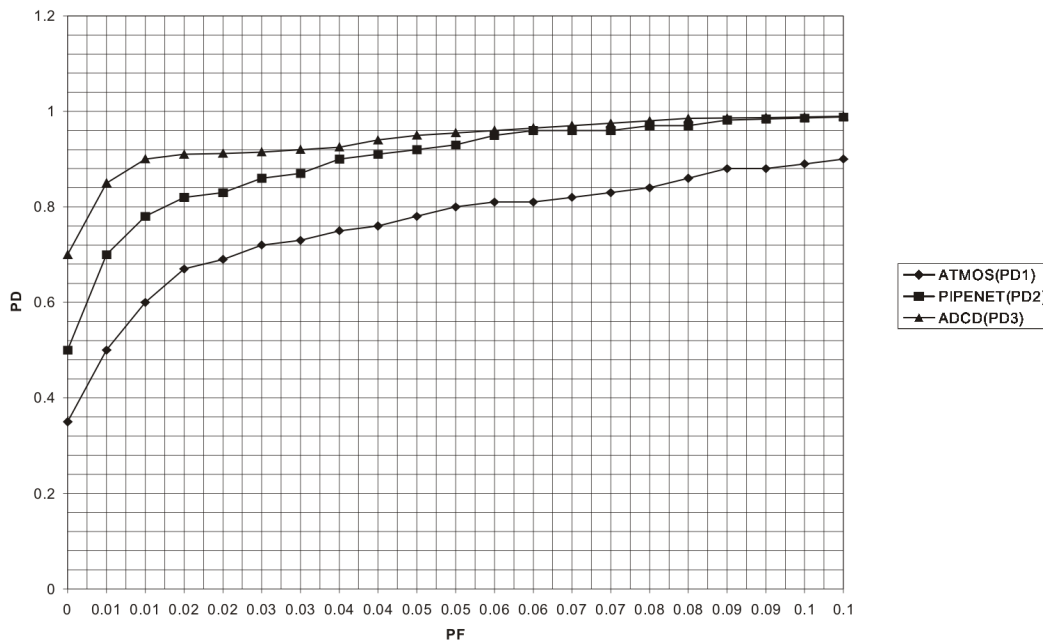


Fig 6 Comparison of ADCD with two other oil-leak detection systems

**Explanations:** From figure 6, the ROC curve for ADCD detects over 70% of pipeline cut cases without causing false alarms. This is followed by the ROC curve for PIPENET with 50% detection with no false alarms. The least is the ROC curve for ATMOS pipe with 35% detection. Figure 6, again shows that the performance of ADCD is in agreement with other detection software, but performs better. This relative advantage of ADCD over others is as a result of combining neural networks with crude flow rules. Using two software products enabled us to illustrate different

vendor user interfaces available, alternative approaches to neural networks models, design for decision making, and the performance metrics of different models. Comparison of performance of neural network model with other statistical methods increased our confidence in the ability of ADCD in successful modeling pipeline cuts.

## 5.0 Conclusion

In a related effort to effectively manage oil spill in Nigeria, the paper reviewed the current situation with exploration of crude

oil in the Niger Delta region of the country and argued that the rising tide of violence, hostage taking in the Niger Delta impaired the use of known formal data collection procedures. As a way forward to provide prompt detection and response to oil spillage, we presented an architecture based on interrupt-driven, which provides communications of wireless sensor network with a satellite connected to a central office to automatically track instances of encroachments or pipeline cut. With the use of a neural network learning capability for preprocessing and pipeline operational rules, the proposed system is able to detect 70% of pipeline violation far above other leak detection systems.

### References

- [1] Daily Times (October, 2003). Nigeria ranked second most corrupt Nation. <http://www.dailytimesofnigeria.com/DailyTimes/2003/October/8/Nigeriaranked.asp/>
- [2] L.P. Eyo-Essien, "Oil spill management in Nigeria: challenges of pipeline vandalism in the Niger Delta region of Nigeria" unpublished.
- [3] P. C. Nwilo, K. O. Peters, and O. T. Bodeji, "Sustainable Management of oil spill incidents along the Nigerian Coastal areas, 2000." unpublished.
- [4] Niger Delta Environmental Survey (NDES), "Environmental and Socioeconomic characteristics, 1997." unpublished.
- [5] B. Ojediran, and J. Ndibe, "Oil Spill Management: SPDC and the Environment, 2005." Unpublished.
- [6] B. A. Ajakaiye, "Welcome address presented at the international workshop on review and update of the National Oil Spill Contingency Plan, 2006." unpublished.
- [7] T. Horne, "Summary of Data Gathering Methods, 1995." <http://www.web.net/tamhorne/data.htm>
- [8] H. V. Trees, *Detection, Estimation and Modulation Theory –Part I*. John Wiley, New York, 2001.
- [9] C. Huseyin, and R. Srinivasan, "Configuration of detection software: a comparison of decision and game theory approaches." *Decision Analysis*, vol. 1, No. 3, 2004, pp131-148.
- [10] I. Stoianov, L. Nachman, and S. Madden, "PIPENET: A wireless sensor network for pipeline monitoring, 2007." <http://db.csail.mit.edu/pubs/ipsn278-nachman.pdf>
- [11] W. Al-Rafai, and R. J. Barnes, "Underlying the performance of real-time software-based pipeline leak-detection systems." *Pipe & Pipelines International*, December, 1999.
- [12] J. Zhang, and L. Xu, "Real-time pipeline leak detection on Shell's North Western Ethylene Pipeline." *IMEchE Conference Transactions 1999-8, [Ageing Pipelines, Optimizing the Management and Operation: Low Pressure-High Pressure.]*
- [13] MATLAB<sup>R</sup> *The MATLAB compiler user's guide. [In Mathworks Handbook; Math works.]*
- [14] Fawcett, T. (2003). ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. Intelligent Enterprise Technologies Laboratory, Palo Alto, Downloaded from: [http://www.purl.org/NET/tfawcett/software/ROC\\_algs.tar.gz](http://www.purl.org/NET/tfawcett/software/ROC_algs.tar.gz)