A statistical analysis on the leak detection performance of underground and overground pipelines with wireless sensor networks through the maximum likelihood ratio test

CHINEDU DURU* and COSMAS ANI

Department of Electronic Engineering, University of Nigeria, Nsukka, Nigeria
e-mail: chinedu.duru@unn.edu.ng; cosmas.ani@unn.edu.ng

Abstract. Leaks in pipelines of the oil and gas industry are an economic and environmental problem that needs to be detected early and effectively. Wireless sensor networks (WSNs) have been researched as one of those technologies to be used in the remote monitoring of pipeline infrastructure. The idea of using tiny sensor nodes on pipelines seemingly provides industries with effective and reliable real-time monitoring, and better coverage density per area. The benefits are apparent in the deployment of WSNs for pipeline monitoring. However, what really lacks is an actual comparison in the detection performance between deployment in overground pipelines and underground pipelines. Extensive research has been going on the use of wireless underground sensor networks for a number of applications. This paper attempts to provide a statistical insight on the concepts of leak detection performance of WSNs when deployed on overground and underground pipelines. The approach in the study employs the hypothesis testing problem to formulate a solution on the detection plan. Through the hypothesis test, the maximum likelihood ratio scheme is used to provide an optimal performance analysis of the detection idea. The test also takes into consideration the signal to noise ratio performance of the two settings of underground and overground and is crucial in bringing up a conjecture on the performance of detection. As would be shown in the paper, thresholds, determined by probability, are the key in ensuring a good detecting performance for the WSN.

Keywords. Wireless sensor networks; underground pipelines; overground pipelines; leak detection; maximum likelihood ratio test; hypothesis test.

1. Introduction

The need to transport the high-demand products of oil and gas from a production point to an area of supply has led to the usage of pipeline infrastructure. These pipelines, which can either be deployed overground or underground, are specifically designed and manufactured to carry these usually toxic and dangerous products through vast areas of a few hundred metres to several thousand kilometres. The areas are often close to hubs of high population or of high environmental sensitivity. With such risks, methods have to be sought to preserve these lines and maintain their condition at a good level [1, 2]. Furthermore, with the intentional tampering of pipeline infrastructure on the increase, through terrorism and vandalism, the need for a reliable pipeline monitoring system is highly significant [3, 4].

Wireless sensor networks (WSNs) are just one of those technologies used in the remote monitoring of pipeline infrastructure [5–7]. With the deployed nodes, a sensor network can check the infrastructure condition for corrosion and most importantly leaks, which are essential occurrences to determine intentional tampering or a natural phenomenon due to degrading or ruptured equipment. Compared with other technologies that use a wired form of communication, the deployment of sensor networks that communicate wirelessly for pipeline monitoring offers several advantages, which include ease of deployment, real-time and reliable data capturing, concealment and better coverage density [8, 9].

This paper focuses on the leak detection performance of WSNs when deployed on an overground pipeline and an underground pipeline. The investigation is through a statistical analysis and evaluation by the maximum likelihood ratio (MLR) technique used in the detection theory. The reason for approaching the MLR scheme in the analytical evaluation of the detection plan is the fact that the MLR is a simpler and more straightforward approach than other estimations such as those using the Bayesian criteria [10]. Through this paper, it would be shown that statistically the MLR criterion needs only a prior distribution model to turn

*For correspondence

the detection or estimation problem into an analytical optimization problem, which is the focus of this paper. Additionally, under the general assumption that leak occurrence comes with a large number of observed measurements, the MLR scheme is seen to achieve the smallest possible variance of any estimator. In other words, it is the best guaranteed estimator that will achieve the most optimum parameter estimation in the detection problem. Therefore, the single focus on the MLR scheme to evaluate detection performance is an apparent approach [11].

The need for such a study is apparent. The occurrence of leaks leads to environmental pollution and a dangerous hazard. Additionally, these leaks are also an economic problem with millions of tonnes of oil and billions of m^3 of gas lost each year, which amount to about 7% of the total transported oil or gas [12]. Early leak detection is therefore of paramount importance and over the years, WSNs are being used for just this in overground and, more recently, underground pipelines [13–17]. This study attempts to provide an insight on the mathematical approach of leak detection performance with WSNs to establish the parameters required for performance. It is important to note that this paper is entirely conceptual and aims to use theory to bring up a conjecture. The approach is for the two settings of overground and underground pipelines, which are distinguished through their wireless communication channel.

2. Related works

Works on pipeline leak detection with WSNs have been pursued at different levels. Magnetic Induction (MI)-based WSN for underground pipeline monitoring, also known as MISE-PIPE, is pursued by Sun *et al* [18]. Through the study it is discovered that the MISE-PIPE scheme involves sensors deployed both inside and outside the pipeline, which include pressure sensors, acoustic sensors and soil property sensors that cooperatively work together to detect and localize a pipeline leak. The MI-based WSN comprises coils connected to the various sensors that are systematically placed along the underground pipeline structure to provide the sensors with efficient and robust wireless communication among themselves. Sun *et al* [18] provide a methodology and theory in deploying the MISE-PIPE idea but do not really provide the detection scheme or detection approach in detecting a leak. To be precise, Sun *et al* [18] focus on the design architecture of the MISE-PIPE rather than on the actual detection performance analysis.

Reference [19] is a research manuscript on WSNs for pipeline water loss detection. Though not entirely focusing on pipelines of the oil and gas industry, Christodoulou *et al* [19] offer a framework on the WATERSENSE project, whose aim is to receive sensor signals of a deployed WSN on a water pipeline for the purpose of a leakage decision support and management system. They focus on the development and implementation of the WATERSENSE project but again do not provide an actual conceptual idea on the detection performance analysis, which can be important in improving the operational performance of the deployed system.

Nawaz and Rehman [20] provide insight into the different leakage problems that may arise in various pipeline industries and how WSN is used as an applied solution to monitor the pipeline leakage. This reference is more of a general paper that puts some of the WSN leakage detection techniques into one manuscript and aims to briefly describe and compare their functionalities. The study provides good insight on the technologies of pipeline monitoring using WSNs but yet again fails to provide insight on the actual detection approach.

Boaz *et al* [21] focus on wireless sensor nodes for gas pipeline leak detection and localization. They discuss the theories and environmental constraints for wireless sensor nodes and its usage in the detection of sound excited by jetting gas indicating a gas pipeline leakage, and about its frequency. This work is precise and focuses on the environmental characteristics of gas in air for sensor detection in depth. An architecture was even proposed and designed using various components such as the XBee RF module and a microcontroller, whose results were produced by simulation. However, just like the previous works, [21] fails to deliver on the performance analysis of detection. This is also the case for references [22–25]; all of them focus on architecture and implementation.

This paper, unlike the others, attempts to provide an actual detection scheme that provides insight on the detection performance analysis of pipeline leakage. This study and application of discussed concepts can enable designing of systems to improve performance. The work also provides a comparison of detection between underground and overground pipelines, which are the two most important scenarios pipelines deployed in industries. The entire work of this paper is based on the MLR scheme, which, as previously mentioned, is the best guaranteed estimator that will achieve the most optimum parameter estimation since, after all, leak detection is a decision estimation process.

3. Overview of study

Consider the scenario of figure 1, where acoustic omnidirectional sensor nodes operating at narrowband frequencies are deployed in cross sectional groups across a pipeline.

Each group makes up a WSN and has a maximum of four sensors covering each cylinder section of the pipeline. A Base Station (BS), also a sensor device, is deployed at the top of the structure to act as a hub where all information from the sensors goes to. The hub can be placed at the very
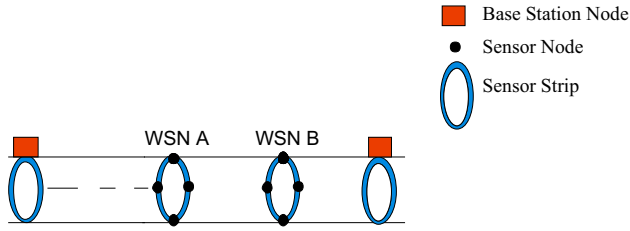
Figure 1. Pipeline structure with sensors.

Figure 3. WSN detection system.

end of the pipeline or at intermediate stops between groups of sensors. The sensors are connected with a strip that also has sensing abilities but with very limited capabilities.

Practically, these WSNs consist of MICAz motes with acoustic and pressure sensing capabilities. The outdoor range of these sensor devices would be about 100 m [26]. Therefore, the optimal distance between the sensor strips would also be 100 m. Hence, for a 1000-km pipeline, for example, the structure would need precisely 10,000 strips each equipped with four motes. Assuming that each mote costs $100, the default deployment costs of the WSNs on the pipeline infrastructure would be $4 million minus the BS sensor nodes. This $4 million, compared with other schemes such as the Fibre optic leak detection cable, which costs around $8 million [27], is a cheaper approach. Fundamentally, the deployment costs of the WSNs would depend on the sensing range capability of the sensor devices. Therefore, an improved sensing range would drastically decrease costs.

When a leak occurs, as shown in figure 2, the sensor nodes of each group detect this leak, make a decision on whether the leak occurred or not and send the information wirelessly to the corresponding nearest BS of the group for a further and final decision on the existence of the leak. A sensor device decides to detect the leak by estimating the approximate distance between it and the occurrence depending on its coverage capability. If within coverage, detection is carried out with the BS making its own decision based on multiple decisions from the detecting sensors. With this scenario played out, the WSN detection system is visualized as shown in figure 3 [28].

In figure 3, K is the number of sensors (S), i represents the individual sensor number (i.e., S_1, S_2, S_3, …, S_i, …, K), A_i represents the acoustic signal from the leak at sensor i (i.e., A_1, A_2, A_3, …, A_i, …, K) and $f(x_i)$ is the function of measurements that is sent to the BS if the sensor decides
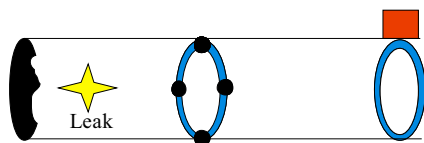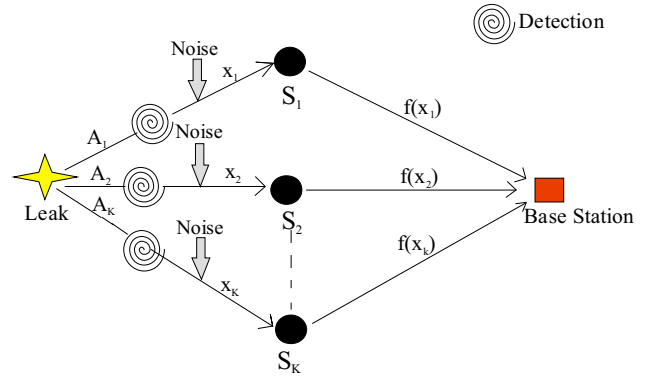
that a leak is present. Fundamentally, a sensor makes this decision based on its noise-degraded measurements whereas the BS makes a decision based on multiple decision functions of measurements from the sensors.

4. Detection testing synthesis

In this section of the paper, a testing problem is formulated to provide insight on the parameters involved with the detection scheme. It is through this preparation that further evaluation is made on the detection performance of WSNs for pipeline leaks.

4.1 *Testing problem*

The signal picked up by the sensor nodes, in accordance with figure 3, is the effective sound pressure produced by the leak under the surrounding background noise. To be precise, the occurred leak emits an acoustic sound signal that alters the normal state of background noise. The aim is to detect the sound signal amidst noise; since this is random and undefined with uncharacterized possibilities, investigating whether the leak is present or not is a hypothesis testing problem [28–30].

Statistically, to carry out the detection testing involves taking measurements and through those measurements a decision is made on the occurrence of the leak. Thus, there are two probable outcomes, which are

- whether a leak is absent amidst the background noise measurements, defined as the zero hypothesis H_0.
- whether a leak is present amidst the background noise measurements, defined as the alternate hypothesis H_1.

To investigate the two possibilities, also referred to as the binary testing procedure, the sound measurements are first considered to be modelled as a continuous random variable x. This random variable is a function that maps the random occurrence of the leak to an infinite number of values and a

Figure 2. Leak occurrence.

good representation in describing the endless unpredictable nature of noise. Under the hypothesis H_0, the testing measurements are modelled as the acoustic background noise that affects the sound pressure measurement of the leak. They consist of sounds of surroundings where the pipelines are deployed, which includes a sum of many sources of sound noise. For simplicity and due to the central limit theorem, which states that the sum of several independent random variables tends to a normal distribution as the number of random variables in the sum grows to infinity [28, 31], the background noise is approximated by a mean square of sets of random variables that follow normal distribution. Thus, without the occurrence of the leak, sound measurements and under the null hypothesis H_0; the acoustic background noise is modelled by a normal distribution N with mean μ_o and variance σ^2:

$$x = N(\mu_o, \sigma^2) \quad Under\, H_0 \tag{1}$$

On the other hand, when a leak occurs this assumption is presumed under the alternative hypothesis H_1. Here, the normal distribution of the background noise is excited to produce a different normal distribution under H_1. This excitation includes the effects of H_0 and an unknown mean shift value A, depending on the loudness of the leak. Therefore, the new and alternate sound measurement under H_1 is modelled as follows:

$$x = N(\mu_o + A, \sigma^2) \quad Under\, H_1 \tag{2}$$

It is important to note that the mean shift here is always a positive value because it is assumed that the A value is an additional parameter to the normal state of noise. In other words, A is an added occurrence and therefore an added shift that can only be positive is included to fit into the intuition of an occurred leak amidst noise. Equations (1) and (2) are considered to be the probability density functions (PDFs) that characterize the probable behaviour of the sound measurements under each hypothesis, hence simplifying the subsequent test.

Given this set-up, the performance characterization of the binary test is approached through different probabilities, namely, the probability of detection, the probability of false alarm and the probability of miss. The following are their definitions [32]:

- The probability of detection (P_D) is where H_1 is true it is decided, denoted as *prob{H_1 true, decide H_1} = P_D*.
- The probability of false alarm (P_{FA}) is where H_1 is decided but H_0 is actually true, expressed as *prob{H_0 true, decide H_1} = P_{FA}*.
- The probability of miss (P_M) is where H_0 is decided but H_1 is actually true, represented as *prob{H_1 true, decide H_0} = P_M*.

Note that the probability of detection can also be expressed as follows: $P_D = 1 - P_M$, and additionally, the outcome

where H_0 is decided and is true can be expressed as follows: *prob{H_0 true, decide H_0} = $1 - P_{FA}$*.

The two most important probabilities that characterize the performance of the testing are P_D and P_{FA}, where the ultimate goal is make the correct decision given the scenario. In deciding which one is true between the two probabilities, the sound measurements are observed to see whether they lie underneath the PDF of P_{FA} or the PDF of P_D. This essentially means the separate integration of the two probabilities with their corresponding PDFs.

Let us assume that R_O indicates the region where the measurements lie under the PDF of P_{FA}, and R_1 describes the region for the PDF of P_D. R_O and R_1 are defined using the function of distribution and a threshold. Assume that $T(x)$ is the function and γ is the threshold. Then, if $T(x)$ is greater than γ, it is decided that R_1 is true and the measurements lie in the region R_1. Alternatively, if $T(x)$ is less than γ, then the measurements lie in R_0. This is intertwiningly linked to the binary test, where the corresponding decisions of R_0 and R_1 also lead to the decisions of H_0 and H_1, respectively, and expressed as follows:

$$T(x) > \gamma = R_1, H_1 \tag{3}$$

$$T(x) < \gamma = R_0, H_0. \tag{4}$$

The entire testing problem is therefore illustrated as shown in figure 4 [32],
where $f_{H_0}(T)$ = PDF of $T(x)$ when H_0 is true and $f_{H_1}(T)$ = PDF of $T(x)$ when H_1 is true. In essence, P_{FA} and P_D are as follows:
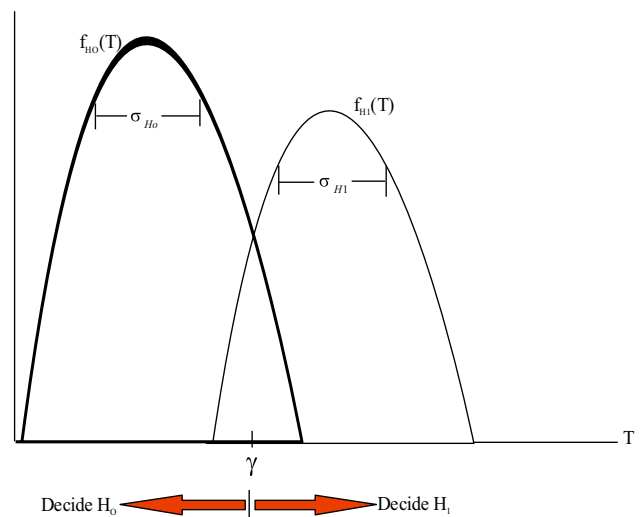
$$P_{FA} = \int_{\gamma}^{\infty} f_{H_0}(T)dT, \tag{5}$$

Figure 4. P_D and P_{FA} illustration for a normal distribution of measurement x.

$$P_D = \int_\gamma^\infty f_{H_1}(T)dT. \tag{6}$$

Putting P_{FA} and P_D together provides a conjecture on the receiver performance of the sensor in detecting the leak. To be precise, what is observed is the variation of P_D and P_{FA} with respect to a changing γ. The next step to the testing is finding the test procedure $T(x)$ that provides optimal performance in the leak detection ability.

4.2 *Likelihood ratio test and maximum likelihood detection*

In providing an optimal test procedure $T(x)$, the known PDFs of H_0 and H_1 are further analysed. It is understood that a good performance detector is one where P_D is as high as possible and P_{FA} is as low as possible. This means the sensor detector has to maximize P_D for a given P_{FA}. To achieve this, the Neyman Pearson (NP) test is used, where the measurements x are observed below a ratio of the PDF of H_1 to that of H_0 and and plugged into Eqs. (3) and (4). The new equations give rise to the Likelihood ratio $L(x)$, which enables the test to maximize P_D for a given P_{FA} expressed as follows [33]:

$$L(x) = \frac{f_{H_1}(x)}{f_{H_0}(x)} > \gamma \quad \text{to decide that H}_1 \text{ is true} \tag{7}$$

$$L(x) = \frac{f_{H_1}(x)}{f_{H_0}(x)} < \gamma \quad \text{to decide that H}_0 \text{ is true} \tag{8}$$

In summary, Eqs. (7) and (8) say that a comparison is made between the likelihood of observing a value x given that H_1 is true to the likelihood of observing a value x assuming H_0 is true. The aim based on Eq. (7) is to get a bigger $f_{H_1}(x)$ for the purpose of choosing H_1. A bigger $f_{H_1}(x)$ will mathematically maximize P_D for a given P_{FA}; hence the MLR test is introduced through $L_{max}(x)$, where all attempts are made to maximize $f_{H_1}(x)$ by estimating unknown parameters [34].

$$L_{max}(x) = \frac{\max(f_{H_1}(x))}{f_{H_0}(x)}. \tag{9}$$

5. Detection testing application with WSNs

Given the scenario of figure 3 in section 3, the detection system follows a decentralized scheme where sensor nodes detect a leak, make a decision on whether a leak is present or not and sends the relevant information to a BS for further decision evaluation on the presence or absence of the leak. Thus, the testing and evaluating scheme has two rules to follow:

1. sensor decision rule
2. BS decision rule

Both rules are tested with the MLR test in evaluating the detection performance of a leak.

5.1 *Sensors' decision rules*

In accordance with figure 3 and in relation to the derived synthesis in section 4, the MLR test is computed for each sensor i. It is assumed that q samples of measurements are taken per distribution for each sensor i, where $q = 1, ..., Q$. Therefore, Eqs. (1) and (2) are modified to the following single equation [28]:

$$x_i[q] = \begin{cases} N(\mu_{o,i}, \sigma_i^2) & Under\ H_0 \\ N(\mu_{o,i} + A_i, \sigma_i^2) & Under\ H_1 \end{cases} \tag{10}$$

where $u_{o,i}$ and σ_i^2 can be calculated as follows:

$$\mu_{o,i} = \frac{1}{Q}\sum_{q=1}^{Q} x[q], \tag{11}$$

$$\sigma_i^2 = \frac{1}{Q-1}\sum_{q=1}^{Q}(x[q] - \mu_{o,i})^2. \tag{12}$$

Since the measurements follow a normal distribution, the PDF equations defined for each hypothesis are expressed as follows:

$$\begin{aligned} f_{H_0}(x_i) &\approx N(\mu_{o,i}, \sigma_i^2) \\ &= \frac{1}{(2\pi\sigma_i^2)^{\frac{Q}{2}}} exp\left\{ -\frac{1}{2\sigma_i^2}\sum_{q=1}^{Q}(x_i[q] - \mu_{o,i})^2 \right\} \end{aligned} \tag{13}$$

$$\begin{aligned} f_{H_1}(x_i) &\approx N(\mu_{o,i} + A_i, \sigma_i^2) \\ &= \frac{1}{(2\pi\sigma_i^2)^{\frac{Q}{2}}} exp\left\{ -\frac{1}{2\sigma_i^2}\sum_{q=1}^{Q}(x_i[q] - (\mu_{o,i} + A_i))^2 \right\} \end{aligned} \tag{14}$$

With the defined PDFs of (13) and (14), the likelihood ratio test for each sensor is expressed as follows:

$$L_i(x) = \frac{f_{H_1}(x_i)}{f_{H_0}(x_i)} > \gamma_i \tag{15}$$

MLR of (15) requires maximizing (14); if $\mu_{1,i} = \mu_{0,i} + A_i, \sigma_i$ then (14) is modified to

$$f_{H_1}(x_i) = \frac{1}{(2\pi\sigma_i^2)^{\frac{Q}{2}}} exp\left\{ -\frac{1}{2\sigma_i^2}\sum_{q=1}^{Q}(x_i[q] - \mu_{1,i})^2 \right\} \tag{16}$$

It is very common with exponential densities to simplify things by taking the natural logarithm. This is because the natural logarithm is a function that essentially does not

change the characteristics for a given order [28]. Thus, the natural log of (16) is

$$\ln(f_{H_1}(x_i)) = -\ln\left(\left(2\pi\sigma_i^2\right)^{\frac{Q}{2}}\right) - \frac{1}{2\sigma_i^2}\sum_{q=1}^{Q}(x_i[q] - \mu_{1,i})^2. \tag{17}$$

The equation to be maximized is (17) and it is known that the extrema of any function is where the derivative equals zero. Therefore, taking the derivative of (17) with respect to $\mu_{1,i}$ gives

$$\frac{\partial\ln(f_{H_1}(x_i))}{\partial\mu_{1,i}} = \frac{1}{\sigma_i^2}\sum_{q=1}^{Q}(x_i[q] - \mu_{1,i}) = \frac{1}{\sigma_i^2}\sum_{q=1}^{Q}(x_i[q] - Q\mu_{1,i}) \tag{18}$$

Equating equation (18) to zero and solving for $\mu_{1,i,\max}$ yields

$$\mu_{1,i,\max} = \frac{1}{Q}\sum_{q=1}^{Q}x_i[q]. \tag{19}$$

Thus, replacing $\mu_{1,i,\max}$ of (19) with its equivalent in (16) maximizes the likelihood ratio of (15), defining the MLR. In perspective, if a sensor decides that a leak is present, it sends its estimated MLR of (15) to the BS.

In addition to maximizing (15), further clarifications are pursued as regards the decision rules of the sensor in detection. A threshold, ρ_i, is introduced to replace the assumption of Eq. (8) given that H_0 is true. Therefore, a combination of this threshold with that of H_1 yields the final decision rule for the sensors given as follows:

$$\Phi_i(x_i) = \begin{cases} L_i(x_i) & \text{if } L_i(x_i) \geq \gamma_i \\ \rho_i & \text{if } L_i(x_i)\gamma_i \end{cases}, \tag{20}$$

where ρ_i and γ_i are thresholds that are set.

5.2 *BS's decision rules*

At the BS, all decisions are received from the sensors depending on Eq. (20). The job of the BS is to decide whether a leak occurred or not depending on the product of the sensors' information. Again, if K is the number of sensors, the decision rule at the BS is formulated as follows:

$$\Phi_0(\Phi_1, \ldots \Phi_K) = \begin{cases} 1 & \text{if } \prod_{i=1}^{k}\Phi_i(x_i) \geq \gamma_0 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

where 1 is when leak is present and 0 is when leak is absent; γ_o is the set threshold.

According to Koskiahde [28], a reasonable γ_o is calculated as follows:

$$\gamma_0 = \alpha\rho^K \tag{22}$$

where σ is a factor that depends on:

- the number of sensors in the network,
- the PFA of the sensors and
- the PFA of the BS.

6. Leak sound and communication channel characteristics

As reiterated, when a leak occurs it emits an acoustic energy that propagates both inside and on the outer surface of the pipeline. This energy gives off a sound pressure, which can be measured and analysed. The sound pressure is converted to sound power and depending on the distance from the leak, the sound pressure level at the sensor node can be calculated as follows [35–37]:

$$L_P = L_W - A_{div} \tag{23}$$

where L_W is the sound power level relative to a 1 pW reference sound power and A_{div} is attenuation due to geometrical divergence given by Eq. [28]

$$A_{div} = 20\log_{10}\left(\frac{d}{d_0}\right) + 11. \tag{24}$$

In Eq. (24), d is the distance the sound propagates and d_o is the reference distance, $d_o = 1$ m. Thus, the sound pressure level L_p [dB] at a sensor is

$$L_P = L_W - 20\log_{10}\left(\frac{d}{d_0}\right) - 11. \tag{25}$$

Equation (25) gives the measure of L_p in terms of distance and sound power level. The relative measure of L_p [dB], however, which is used to determine the pressure level in terms of a reference power level is given by [38]

$$L_P = 20\log_{10}\left(\frac{P_{rms}}{P_{ref}}\right) \tag{26}$$

where P_{rms} is a given sound power in Pascals (Pa) and P_{ref} is the reference level. P_{ref} is usually given as $20 \mu\text{Pa(RMS)} = 2 \times 10^{-5}$ Pa, which is the typical value for hearing in air.

To distinguish between the underground and overground settings for the performance of sound pressure level detection, it is assumed that an underground WSN deployment takes into consideration the presence of soil or rock that hampers performance. It is also assumed that overground there is no such hampering. Therefore, with this consideration, the signal to noise ratio (SNR) would be

different for the two settings, which also affects the probability of detection performance. The SNR is estimated by the following equation [12]:

$$SNR = 10 \log_{10}\left(\frac{L_W}{\sigma}\right) dB \qquad (27)$$

where L_W is the sound power in Watts or numeric representation, and σ is the noise variance. Using Eq. (27), the variance is adjusted to suit the conditions of the settings. Since it is assumed that underground setting has more signal hampering effect than overground, the variance for underground would be significantly higher than the variance of overground ($\sigma_{Under} > \sigma_{Over}$). To elaborate on this notion of variance, in the underground environment, signal propagation would be heavily affected and attenuated by the presence of soil and rock. Hence, measurements coming from the leak to the sensor devices or communication between the sensor devices and the BS would be less precise and less accurate. In terms of the variance of the PDF, a larger variance would indicate a less precise estimator, which consequently would mean a less accurate hypothesis testing [39]. This higher variance flattens the probability distribution curve more than that of the lower variance. Therefore, this decreases the SNR and decreases the probability of detection performance for the underground setting. The SNR percentage difference between the two settings is calculated as follows:

$$SNR = \frac{SNR_{Over}(dB) - SNR_{Under}(dB)}{SNR_{Over}(dB)} \times 100\%. \qquad (28)$$

Linked to the SNR is the packet loss probability for performance. According to [40], the probability of a packet lost in the wireless communication channel is

$$P_\gamma(packet\,lost) = ((1-p) + p(1 - \left(1-p\right)^{B-1}))^S \quad (29)$$

where B is the number of sensors in the network involved in the detection and S is time slots. The optimal probability (p) of sending data in one time slot is

$$p = \frac{2}{K+2} \qquad (30)$$

where K is the total number of sensors in the network.

7. Evaluating leak detection performance with WSN

This evaluation follows the works of Koskiahde [28] and reference [41], where sensors were deployed and a leak occurred 172.5 m away with an acoustic power of 35 dB. Using Eq. (24), P_{rms} was calculated to be 1 mPa. This value was used as the mean of the system, whose variance and SNR values were chosen to fit the settings of an underground and overground environment.

Table 1. Simulation parameters.

Overground parameters	Underground parameters
$\mu = 10^{-3}$	$\mu = 10^{-3}$
$\sigma = 10^{-6}$	$\sigma = 10^{-3}$
$\alpha = 1$	$\alpha = 1$
SNR = 0.02	SNR = 0.58

The leak detection with WSNs uses the MLR scheme, and the parameters presented in to analyse performance for the two settings of underground and overground are presented in table 1. The parameters of table 1 are of a general case and apply to all materials transported through the pipeline. The simulation procedure was as follows.

Sensor nodes and a leak occurrence were first placed in x–y coordinates. It was assumed that both sensor and leak placement were within 1 m in diameter of pipeline structure. Hence, in a two-dimensional setting, the y-axis or height extends from 0 to 1 m, while x-axis extends from 0 to ∞ m. The distance of the leak from all placed sensors and sound pressure were then computed. The sensor range value is presented in table 2; a distance less than the range implies that the associated sensors had detected the leak determined by Eq. (25). Noise was added into the sensor computation using the following equation:

$$x = \mu + \sqrt{\sigma} * (random\,values). \qquad (31)$$

To generate data of the sensors for the H_0 and H_1 criteria, the following condition was used:

$$H_0 = x \quad \text{and} \quad H_0 = x + L_P \qquad (32)$$

where L_p is the calculated sound pressure from leak to all associated sensors determined by range. Once H_0 and H_1 were sought, the maximum likelihood was then calculated using the derivations in section 5.1 (Eqs. (10)–(20)) based on a set threshold ρ. A wireless communication simulation for sensor to BS was then carried out using Eqs. (29) and (30) based on ρ. A centre fusion processing scheme at the BS was then established using Eqs. (21) and (22) with a general simulated result and insight into the number of packets dropped and the number of detections at the BS.

The simulated results and discussion are as follows.

A deployment of two sensors with a varying ρ in the detection of the leak was first simulated for the underground and overground settings. Figure 5 shows the result.

From figure 5, it can be seen that as proven, an increasing ρ (threshold for the probability of false alarm)

Table 2. Sensor parameters

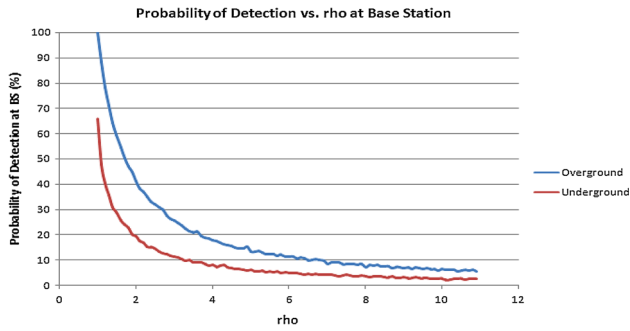| Range	100 m
ρ	1–11

Figure 5. Probability of detection vs. ρ at Base Station for two sensors.
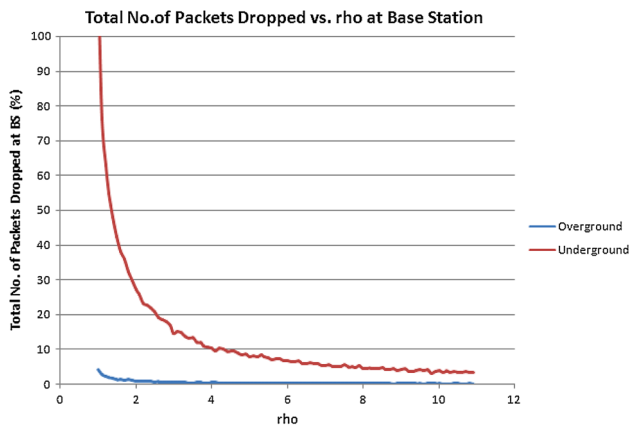
Figure 6. Total number of packets dropped at Base Station vs. ρ for two sensors.

Figure 7. Probability of detection vs. number of sensors for $\rho = 1$.
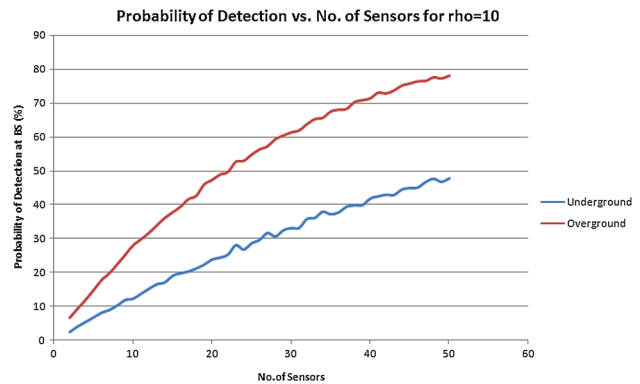
Figure 8. Probability of detection vs. number of sensors for $\rho = 10$.

decreases the probability of detection, therefore hampering performance. However, this degradation is noticeably significant for an underground pipeline than an overground pipeline; in fact, about a 15% and greater difference for a high performing ρ (i.e. $\rho = 1, 2, 3$) and about a 5% and less difference for a low performing ρ (i.e., $\rho = 6, 7, 8$). This just proves that in order to achieve a good detector for the sensor ρ must be as low as possible. However, this low threshold comes with a cost as shown in figure 6.

As observed from figure 6, a low ρ also means more packets dropped during wireless communication between sensor node and the BS. This again is more significant for the underground pipeline than overground, indicating that the presence of soil or rock significantly hampers communication and therefore also affects performance. Consequently, in the underground situation, there needs to be some kind of compromise in setting the threshold to achieve adequate detection and a low rate of dropping of packets.

A further insight is obtained in determining the other factor that can affect the performance of detection. The number of sensors in the WSN was investigated. Figure 7 shows performance through the number of sensors for the best observed threshold of $\rho = 1$.
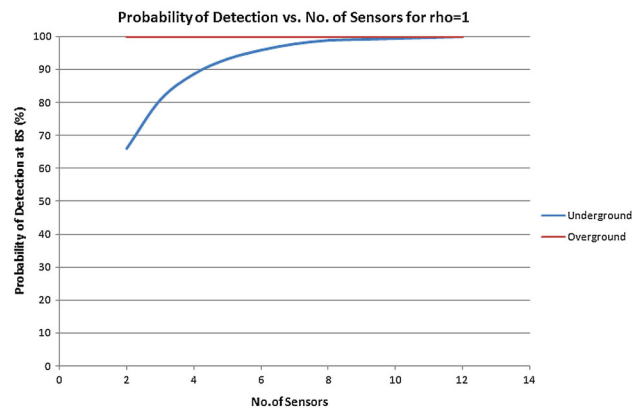
As observed and as proven, the low threshold achieves 100% detection for all number of sensors in the WSN for the overground setting. However, for the underground, a considerable number of sensors will be needed to achieve 100%. This means, according to the default configuration of deployment, a total of two sensors, one on each (opposite) side of the pipeline will be enough to achieve 100% detection for an overground setting; whereas for an underground pipeline, two groups of four sensors, spaced out accordingly, will be required to achieve 100% detection. This indicates that an underground WSN deployment will be a much more expensive endeavour than an overground deployment.

In observing the worst possible scenario, ρ was set high at 10 and figure 8 shows the results.

It can be seen that a high threshold significantly hampers performance and a lot of sensors will be needed to achieve adequate detection. However, this is noticeably worse for an underground setting than an overground one.

A plot of the probability of detection against the total number of packets dropped given the number of sensors and threshold, ρ, was pursued as shown in figures 9–12.
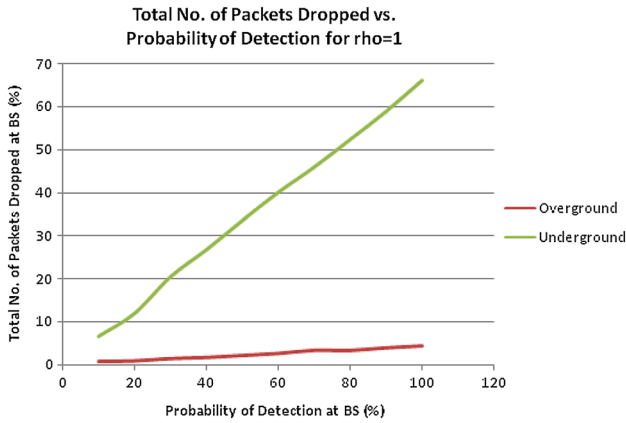
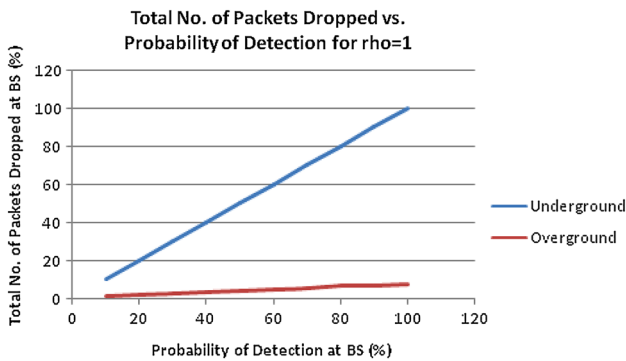**Figure 9.** Total number of packets dropped vs. probability of detection at BS for $\rho = 1$ with two sensors.

Figure 10. Total number of packets dropped vs. probability of detection at BS for $\rho = 1$ with four sensors.
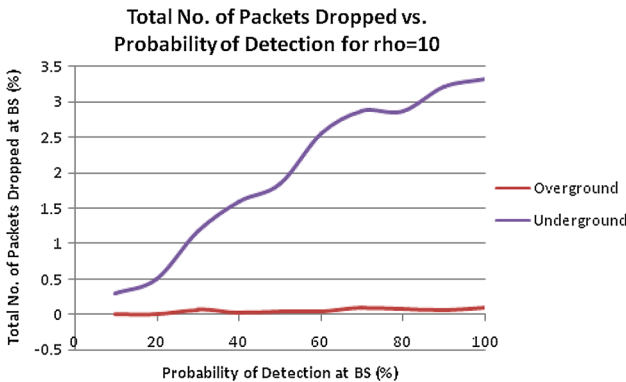

Figure 11. Total number of packets dropped vs. probability of detection at BS for $\rho = 10$ with two sensors.

As observed from the figures, the probability of detection has a linear relationship with the total number of packets dropped. To be precise, as the probability of detection increases so does the total number of packets dropped. Additionally, as the number of the sensors increases in the network so does the probability of detection and
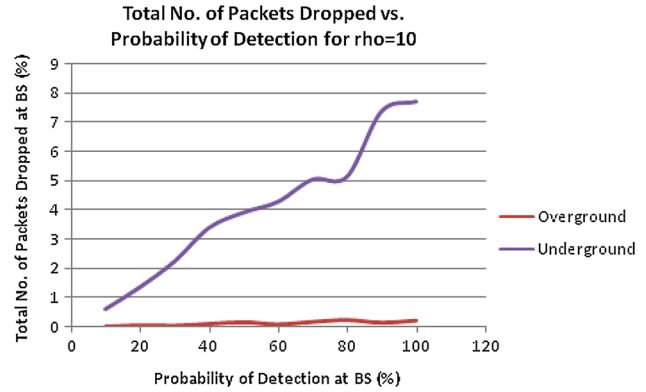

Figure 12. Total number of packets dropped vs. probability of detection at BS for $\rho = 10$ with four sensors.

consequently the total number of packets dropped. In essence, the number of sensors is directly proportional to the detection capability and network communication performance of the WSN.

Applying these observations to the deployment scenario of underground and overground is the fact that the underground setting has more severe packet loss than overground as the probability of detection increases. This literally indicates that the detection performance in underground is weak compared with overground based on the set assumptions. Again, as proven in figure 6, threshold is a significant factor in the total number of packets dropped and hence the detection performance of the WSN. At $\rho = 1$ the total number of packets dropped is more than the packets dropped at a higher set threshold of 10. Equally, also proven in figure 6 is the fact that a higher threshold also means a lower probability of detection. Hence, in perspective, the threshold is inversely proportional to both the total number of packets dropped and the probability of detection.

In summary, to achieve adequate detection for the underground pipeline scenario, ρ must be low enough with a compromise consideration in the total number of packets dropped. Equally, the total number of sensors used for monitoring the pipeline must be more and spaced out at smaller distances between each other than an overground deployment.

8. Conclusion

This paper provides an analysis and evaluation on the leak detection performance for underground and overground pipelines. It was discovered that the threshold for the probability of false alarm is crucial in determining adequate detection of the source. Additionally, the number of deployed sensors is also another factor that affects performance. In relation to the surrounding environment of the underground and overground pipeline, it was discovered that

leak detection performance is severely hampered in the underground situation because of the presence of soil and rock, which affects the noise variance and therefore decreases the probability of detection. For overground pipelines, because of the assumption that there is no presence of any such obstacle, detection performance is pretty much at a high level, with a low noise variance and hence a high probability of detection. From the study, it can be concurred that any kind of underground sensor deployment for detection would require more sensors and as low threshold as possible that also takes into consideration the number of packets dropped for adequate detection performance.

References

[1] Golmohamadi M 2015 *Pipeline leak detection*. Master of Science in Computer Engineering Thesis. USA: Missouri University of Science and Technology

[2] Chis T 2007 Pipeline leak detection techniques. *Ann. Univ. Tibiscus Comput. Sci. Ser.* 5: 25–34

[3] Eboh M 2016 *Nigeria bleeds as pipeline bombings, vandalism wreaks economy*. Nigeria: Vanguard Media Limited

[4] Vanguard News 2016 *FG records N2trn loss to militancy, pipeline vandalism in 2016 – NNPC*. Nigeria: Vanguard Media Limited

[5] Golshan M, Ghavamian A and Abdulshaheed A M A 2016 Pipeline monitoring system by using wireless sensor network. *IOSR J. Mech. Civ. Eng.* 13(3): 43–53

[6] Rashid S, Qaisar S, Saeed H and Felemban E 2014 A method for distributed pipeline burst and leakage detection in wireless sensor networks using transform analysis. *Int. J. Distrib. Sens. Netw.* 10(7): DOI: https://doi.org/10.1155/2014/939657

[7] Du R 2016 *Wireless sensor networks in smart cities: the monitoring of water distribution networks case*. Licentiate Thesis. Sweden: School of Electrical Engineering, KTH Royal Institute of Technology

[8] Akyildiz I F and Stuntebeck E P 2006 Wireless underground sensor networks: research challenges. *Elsevier Ad Hoc Netw. J.* 4: 669–686

[9] Akyildiz I F, Sun Z and Vuran M C 2009 Signal propagation techniques for wireless underground communication networks. *Elsevier Phys. Commun. J.* 2(3): 167–183

[10] Passerini A 2010–2011 *Maximum-likelihood and Bayesian parameter estimation*. Available from http://disi.unitn.it/~passerini/teaching/2010-2011/MachineLearning/slides/06_07_bayesian_learning/talk.pdf

[11] Wikipedia 2017 *Maximum likelihood estimation*. Available from https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

[12] Pnev A B, Zhirnov A A, Stepanov K V, Nesterov E T, Shelestov D A and Karasik V E 2015 Mathematical analysis of marine pipeline leakage monitoring system based on coherent OTDR with improved sensor length and sampling frequency. *J. Phys.: Conf. Ser.* 584

[13] Yu X, Han W and Zhang Z 2015 Survey on a novel wireless sensor network communication at 433 MHz frequency. *Int. J. u- e-Serv. Sci. Technol.* 8(7): 223–234

[14] Li L, Vuran M C and Akyildiz I F 2007 Characteristics of underground channel for wireless underground sensor networks. In: *Proceedings of the Sixth Annual Mediterranean Ad Hoc Networking Workshop*, Corfu, Greece, June 12–15

[15] Can Vuran M and Silva A R 2009 *Communication through soil in wireless underground sensor networks – theory and practice*, Springer Sensor Networks, Signals and Communication Technology, pp. 309–347

[16] Al-Kadi T, Al-Tuwaijri Z and Al-Omran A 2013 Wireless sensor networks for leakage detection in underground pipelines: a survey paper. *Elsevier Proc. Comput. Sci.* 21: 491–498

[17] Vuran M C and Akyildiz I F 2010 Channel model and analysis for wireless underground sensor networks in soil medium. *Elsevier Phys. Commun.* 3: 245–254

[18] Sun Z, Wang P, Vuran M C, Al-Rodhaan M A, Al-Dhelaan A M and Akyildiz I F 2011 MISE-PIPE: magnetic induction-based wireless sensor networks for underground pipeline monitoring. *Elsevier Ad Hoc Netw.* 9: 218–227

[19] Christodoulou S, Agathokleous A, Kounoudes A and Milis M 2010 Wireless sensor networks for water loss detection. *Eur. Water* 30: 41–48

[20] Nawaz F and Rehman K 2016 Remote pipeline monitoring using wireless sensor networks. *J. Netw. Technol.* 7(4): 112–121

[21] Boaz L, Kaijage S and Sinde R 2014 Wireless sensor node for gas pipeline leak detection and location. *Int. J. Comput. Appl.* 100(18): 29–33

[22] Meeravali S and Anusha A 2012 Detection of gas leak and its location using wireless sensors. *Int. J. Eng. Res. Technol.* 1(9): 1–8

[23] Golshan M, Ghavamian A, Mohammed A and Abdulshaheed A 2016 Pipeline monitoring system by using wireless sensor network. *IOSR J. Mech. Civ. Eng.* 13(3): 43–53

[24] Andrey Somov, Alexander Baranov, Denis Spirjakin, Andrey Spirjakin, Vladimir Sleptsov, Roberto Passerone 2013 Deployment and evaluation of a wireless sensor network for methane leak detection. *Elsevier Sens. Actuators A: Phys.* 202(1): 217–225

[25] Stoianov I, Nachman L, Madden S and Tokmouline T 2007 PIPENET: a wireless sensor network for pipeline monitoring. In: *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*, pp. 264–273

[26] Crossbow *MICAz wireless measurement system*. Available from: http://www.openautomation.net/uploadsproductos/micaz_datasheet.pdf

[27] Penty R and Lee M 2013 *Keystone XL pipe shuns infrared sensors to detect leaks*. Available from: https://www.bloomberg.com/news/articles/2013-06-17/keystone-xl-pipeline-shuns-high-tech-oil-spill-detectors

[28] Koskiahde J 2011 *Decentralized detection in realistic sensor networks*. Master of Science in Technology Thesis, School of Electrical Engineering, Aalto University

[29] Bajoivć D, Sinopoli B and Xavier J 2009 Sensor selection for hypothesis testing in wireless sensor networks: a Kullback–Leibler based approach. In: *Proceedings of the 48th IEEE Conference on Decision and Control, 2009, held jointly with the 2009 28th Chinese Control Conference, CDC/CCC 2009*, Shanghai, pp. 1659–1664

[30] Nandi M, Nayak A, Roy B and Sarkar S 2015 *Hypothesis testing and decision theoretic approach for fault detection in wireless sensor networks*. Computing Research Repository

[31] Le Cam L 1986 The central limit theorem around 1935. *Stat. Sci.* 1(1): 78–91

[32] Johnson D *Introduction to detection theory*. Available from: www.ece.iastate.edu/∼namrata/EE527_Spring08/l5.pdf

[33] Penn State Eberly College of Science *Lesson 56: likelihood ratio tests*. Available from: https://onlinecourses.science.psu.edu/stat414/node/309

[34] Statleck Maximum likelihood—hypothesis testing. In: *The digital textbook*. Available from: https://www.statlect.com/fundamentals-of-statistics/maximum-likelihood-hypothesis-testing

[35] Breysse P N and Lees P S J 2006 *Noise*. Johns Hopkins University School of Public Health. Available from: http://ocw.jhsph.edu/courses/PrinciplesIndustrialHygiene/PDFs/Lecture6.pdf

[36] Penn State Courses 2009 *Outdoor sound propagation*. Available from: http://pcfarina.eng.unipr.it/Public/Acoustics-Course/Penn-State-Course/10_osp.pdf

[37] International Standard Organization 1996 *Acoustics—attenuation of sound during propagation outdoors—part 2: general method of calculation*. Available from: https://www.saiglobal.com/PDFTemp/Previews/OSH/ISO/ISO_12345_06-02/T020649E.PDF

[38] UCSB Experimental Cosmology Group 2010 *Acoustics* (Chapter 16). Santa Barbara: University of California. Available from: http://www.deepspace.ucsb.edu/wp-content/uploads/2010/08/033_Chapter-16-lecture-PML.pdf

[39] Wooldridge J M 2015 *Introductory econometrics: a modern approach*, 6th ed. South-Western, Cengage Learning, Ohio, USA, p. 83. Available from: https://books.google.com.ng/books

[40] Rabi M, Stabellini L, Proutiere A and Johansson M 2010 Networked estimation under contention-based medium access. *Int. J. Robust Nonlin. Control* 20(2): 140–155

[41] Ovchinnikov A L and Lapshin B M 2016 Low-frequency acoustic signals propagation in buried pipelines. *J. Phys.: Conf. Ser.* 671