Validation Technique for Pipeline Leakage Localization (VTeLL)

Folorunso Olufemi A. 1), Mohd Shahrizal Sunar 2) , Sarudin Kari 3)

1) Department of Graphics and Multimedia, Universiti Teknologi Malaysia, olufemifolorunso@gmail.com
2) Department of Graphics and Multimedia, Universiti Teknologi Malaysia, shahrizalsunar@gmail.com
3) Department of Graphics and Multimedia, Universiti Teknologi Malaysia, aukauk@itc.utm.my

Keywords: pigging, nugget, predictor, prediction, dataspace

Abstract

Pipelines are essential parts of the infrastructure of modern society; Leak detection systems (LDS) are used to protect the public and the environment from the effects of pipeline failure as well as bringing down operational costs, LDS do this by alerting when a leak occur and solving the problems of leakages most often involves the application of two or more methods which largely depends on several factors. Identification and isolation are important steps in minimizing the consequences of any leak; LDS are used in order to do both. In most cases, it is impossible to pinpoint the exact location of a leak because no single method has become an industry standard due to the limitations of each. We present an alternative validation technique for pipeline leakage prediction through behavioral modeling of the data pattern associated with the various LDS predictions. Our tested data resulted in less than 0.3% localization error.

1.0 Introduction

The major problem with companies using pipelines is the detection and localization of the exact leaking points long their network[1], [2]. Following previous works on visualization of turbulent flow with particles[3], and An improved numerical solution of multiphase flow analysis in soil[4], many solutions have been proposed which includes the pressure transient and sound analysis method [5]; using the genetic-algorithm-enhanced blind system identification method [6]; the application of wavelet method [7]; the cepstrum analysis method [8]; the fuzzy decision-making method [9]; by the double sensors pressure gradient method [10]; using the platelet technology [11]; the spectral analysis of pressure [12]; the acoustic experimental method [13]; the on-line computational technique [14]; and the ATMOSpipe technology developed by Shell[15].

Many leak detection and localisation methods have been used and some are currently being researched to ensure the reliability of pipelines [16],[17],[18]. These methods could generally be divided into three, namely the biological, hardware and the software methods[15].The biological methods employs trained personnel and animals like smell-sentive dogs to locate leaking points along a pipeline, apart from the high risks involved with this method, it is also subjective to individual experience and the time of inspection vis-a-vis the actual leakage. The hardware based methods work based on the principles of their designs. For example, the infrared thermography (a visual devise) was used to detect hot water leaks when there is temperature increase in the surrounding environment after a leak[19]. Acoustic devices based their detection and localisation on the fact that when a leak occurs, noise is generated due to turbulence, their sensors listens to the varying noise signals and triggers an alarm.[18],[20],[21]. The level of atmospheric hydrocarbon vapour is usually done through the gas sampling devices[23], and the negative wavelet method relies on the fact that when a leak occurs, there is a negative wave that travels either ways from the point of leakage, at least two sensors are mounted at strategic locations along the pipelines to detect these and trigger necessary alarms[18]. The software methods on the other hand make use of various data provided by a Supervisory Control And Data Acquisition system (SCADA), as well as the basic flow parameters such as the flow velocities, the pressure, densities and the temperature to localize leakages[16],[18],[24],[25],[26]. The reliability and usability or otherwise of a method or equipment is a function of its sensitivity, localization, availability, alarm rate and costs[27].

When pipelines are left to age without adequate preparation to maintain or replace them, then the rate of
leakages is increased because of aging and other earthen factors. Explosion is therefore imminent especially for those pipelines carrying explosive and highly inflammable substances. This paper proposes a study of how predictions uncertainties can be treated by using a prediction validation formula. VtELL is not a new method but an alternative validation technique for predicted localizations.

Most companies employs at least three different methods and or equipment to effectively localise leakages. The associated costs in terms of pigging, maintenance and use is a source of worry to many pipeline bond companies. The Nigerian National Petroleum Cooperation for example spends about 15 percent of it annual income for the detection and localization of leakages that has claimed thousands of lives in last few years. Our survey showed that not less than 1,800 lives were lost between 1999 and 2008 due to pipeline failures resulting from these leakages.

The Figure 1.0 below shows the summary of lost lives in selected four nations.

![Figure 1.0](image)

Identification and isolation are important steps in minimizing the consequences of any leak, and the pipeline companies are constatly faced with the challenge of balancing between the selection and use of appropriate method and the tradeoffs of their various disadvantages. Some statistical methods have been applied in attempts to perfect these predictions, example is that of ATMOS pipe technology which incorporate the use of advanced pattern recognition function, developed by Shell. This method detects changes in input and output data relationships that could inform the presence of a leak.

The errors associated with each of these methods and equipments limits their applicability in real world [5] and a basic truth in science and engineering is that no analysis should yield results that are more accurate than its inputs[5],[12]. Yet, most LDS predictions are relied upon by their users and this has led to considerable loss of lives and degrading the environments[7]. Every pipeline is subjective to various internal and external factors [7],[9] and in real world measurements, every prediction comes with a degree of errors in pin-pointing exact location where there is/are leakages on a network majorly because the exact state of the underground installation cannot be effectively determined [5]. We present a reduction formula that validates the predictions of any method or equipment through modelling and correlation of the historical behaviour of the associated predicted data patterns.

2.0 Some methods and their error bands

Table1.0 below shows seven commonly used methods and their calculated error bands. The assessment of each method is based on correlated application of each to ten specific predictions and the average error taken as percentages. We looked in detail at the representation of certainty within each model, how model predictions, when correlated with history of observation, could inform the assessment of uncertainty, how that uncertainty can be conveyed to the user and how it can be used to refine the evolving model.

The data were obtained from the company’s database about each of the listed methods applied ten times against the actual values along the 500km lenght pipeline measured over a period of 5 years and the simple averages are as presented for each of the methods as shown below.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>Predicted</th>
<th>Actual</th>
<th>% ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Pressure Transient and Sound Analysis method</td>
<td>10.8</td>
<td>10</td>
<td>8.0</td>
</tr>
<tr>
<td>The Cepstrum analysis method</td>
<td>213.0</td>
<td>200*</td>
<td>6.5</td>
</tr>
<tr>
<td>Double Sensors Pressure Gradient method</td>
<td>223.0</td>
<td>236</td>
<td>5.5</td>
</tr>
<tr>
<td>The Fuzzy decision-making method</td>
<td>295.0</td>
<td>310</td>
<td>5.0</td>
</tr>
<tr>
<td>Using the Genetic-Algorithm-Enhanced blind system identification method</td>
<td>424.2</td>
<td>406</td>
<td>4.5</td>
</tr>
<tr>
<td>The application of Wavelet method</td>
<td>434.1</td>
<td>420*</td>
<td>3.4</td>
</tr>
<tr>
<td>the ATMOS PIPE technology( used by Shell)</td>
<td>473.2</td>
<td>490</td>
<td>3.4</td>
</tr>
</tbody>
</table>

*Will not be included in subsequent tables to eliminate replication within same class boundaries and avoid a bias

3.0 The ideal function

The expectation or the ideal function is of the form:

\[ f(x) \rightarrow y_i = x_i \]  

[1.0]
Where \( y_i \) are the predicted values and \( x_i \) are the actual leakage values. In real world measurements, no equipment and or method could achieve this. In general, the model function for predictions of leakages follows the pattern

\[
f(x) \rightarrow y_i = x_i \pm c_i \quad [2.0]
\]

Where \( c_i \) is a constant. The development of any method or equipment is centred on minimizing \( C \); the lower the \( C \) the more acceptable or reliable the predictor is.

where

\[
C = \sum_{i=1}^{n} \text{mod}(c_i) \quad [3.0]
\]

In effect, for a predictor \( P_i \) its reliability depends on the factor

\[
\lim_{n \to \infty} (C) = 0 \quad \forall \ n \geq 1 \quad [4.0]
\]

When \( n \) (the number of functions involved) is large, and \( C \) is minimal, then the method is reliable. For the ideal function, we expect the following function and table:

\[
f(x) \rightarrow y_i = x_i \pm 0 \quad [5.0]
\]

Note that \( c=0 \);

with \( R^2 = 1^{**} \quad [6.0] \)

from:

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>236</td>
<td>236</td>
</tr>
<tr>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td>406</td>
<td>406</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>490</td>
<td>490</td>
</tr>
</tbody>
</table>

** \( R^2 \) – Is the square of the correlation coefficient between \( y \) and \( x \)

### 5.0 Illustration

If we let \( A \), \( B \) and \( C \) represent our nugget set from the entire methods and equipments observed, the linear functions of a modeled observation from the Biological (Table 3.0), the Hardware (Table 4.0) and the Software methods (Table 5.0) represented as \( A \), \( B \) and \( C \) are:

\[
\begin{align*}
\text{Table 2.0} \\
\text{Predicted} & | \text{Actual} \\
236 & | 236 \\
310 & | 310 \\
406 & | 406 \\
10 & | 10 \\
490 & | 490 \\
\end{align*}
\]

With each pair of \( x \) and \( y \) corresponding to a unique dataspace element, we define the association equation as the linear function

\[
f(x) = y = \sum_{n=1}^{\infty} (a_n x + (b_n \cdot \alpha_n)) \quad [8.0]
\]

Where

\[
C = (b_n \cdot \alpha_n) \quad [9.0]
\]

With \( a_n \) and \( b_n \) as the coefficients or the multiplicities of \( x \) and \( C \) respectively, \( \alpha_n \) is some constant. So, when \( n \) is large,

\[
y = f(x) = x \sum_{n=1}^{\infty} (a_n) + \frac{1}{n} \sum_{n=1}^{\infty} (b_n \cdot \alpha_n) \quad [10.0]
\]

This reduces to:

\[
y = f(x) = x \cdot \bar{a} + \frac{1}{n} \sum_{n=1}^{\infty} (\alpha_n) \quad [12.0]
\]

Or

\[
y = f(x) = x \cdot \bar{a} + \alpha_n \cdot \frac{1}{n} \bar{b} \quad [13.0]
\]

which gives

\[
y = f(x) = x \cdot \bar{a} + \frac{1}{n} \sum_{n=1}^{\infty} (b_n) \quad [14.0]
\]

This means, the higher the number of equations injected into our approximation formula, the better the results, since

\[
\lim_{n \to \infty} \left( \frac{1}{n} \right) = 0 \quad [15.0]
\]

### 4.0 The predictor function

For a predefined dataspace \( U \) with associated elements:

\[
U = \{(x_{11}, y_{11}), (x_{21}, y_{21}), (x_{31}, y_{31}), \ldots (x_{n1}, y_{n1}), (x_{n2}, y_{n2})\} \quad [7.0]
\]
\( y = 0.9893x + 3.5074 \) \hspace{1cm} [16.0]

with \( R^2 = 0.9987 \) \hspace{1cm} [16.1]

from:

**Table 3.0**

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>233</td>
<td>236</td>
</tr>
<tr>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td>400</td>
<td>406</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>501</td>
<td>490</td>
</tr>
</tbody>
</table>

\(^{**} R^2 \) = is the square of the correlation coefficient between \( y \) and \( x \)

\( y = 0.9709x + 3.3912 \) \hspace{1cm} [17.0]

with \( R^2 = 0.9964 \) \hspace{1cm} [17.1]

from:

**Table 4.0**

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>234</td>
<td>236</td>
</tr>
<tr>
<td>312</td>
<td>310</td>
</tr>
<tr>
<td>433</td>
<td>406</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>490</td>
<td>490</td>
</tr>
</tbody>
</table>

And

\( y = 0.9848x + 3.6222 \) \hspace{1cm} [18.0]

with \( R^2 = 0.9984 \) \hspace{1cm} [18.1]

from:

**Table 5.0**

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>234</td>
<td>236</td>
</tr>
<tr>
<td>307</td>
<td>310</td>
</tr>
<tr>
<td>400</td>
<td>406</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>504</td>
<td>490</td>
</tr>
</tbody>
</table>

where \( R \) is the coefficient of correlation between \( y \) and \( x \)

Using \( [16.0] \), \( [17.0] \), \( [18.0] \) in \( [14.0] \) with \( n=3 \) since we only have 3 equations, to obtain:

\[
y = \frac{0.9893+0.9709+0.9848}{3} - \frac{1}{3} (3.5074+3.3912+3.6222)
\]

\[ \ldots \] [19.0]

which gives:

\( y = 0.9817x + 1.1690 \) \hspace{1cm} [20.0]

with \( R^2 = 1 \) \hspace{1cm} [20.1]

a perfect positive correlation!.

**6.0 Localization validation**

Generally, equation \( [14.0] \) perfects all predictions which is interpreted as equation \( [20.0] \) in our last illustration. This equation perfects the predictions of each of the three methods listed in this illustration as shown in **Table 6.0** below. In this experiment, pigging valves were used as temporary artificial taps at the designated positions A, B and C along a 500km length pipe to mimic the presence of a leak. Each representative of the nuggeted methods are then used to predict the leak at each instance. The results obtained are as presented.

**Table 6.0**

<table>
<thead>
<tr>
<th>POSITION</th>
<th>PREDICTED VALUE</th>
<th>VTeLL</th>
<th>ACTUAL VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>302</td>
<td>297.64</td>
<td>298.00</td>
</tr>
<tr>
<td>B</td>
<td>180</td>
<td>177.88</td>
<td>177.57</td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>10.00</td>
<td>10.00</td>
</tr>
</tbody>
</table>

From the table, it could be seen that the localisation prediction is brought closer to the actual values irrespective of the method or technique adopted. The associated errors are as listed in the **Table 7.0** below. The list is indicative but not exhaustive and could be replicated by trying some life data generated from any of the applied method as highlighted earlier. By comparing this result with that obtained by either of the three methods, the predictions are perfect up to, but not limited to 0.1%. By increasing the participating functions by a factor of 1, the error is further reduced by 25% of the initially perfected figures.

**Table 7.0**

<table>
<thead>
<tr>
<th>ACTUAL</th>
<th>PREDICTED Using any method</th>
<th>%E</th>
<th>( %E )</th>
</tr>
</thead>
<tbody>
<tr>
<td>298.00</td>
<td>302</td>
<td>1.34</td>
<td>297.64</td>
</tr>
<tr>
<td>177.57</td>
<td>180</td>
<td>1.37</td>
<td>177.88</td>
</tr>
<tr>
<td>10.00</td>
<td>9</td>
<td>10.00</td>
<td>10.00</td>
</tr>
</tbody>
</table>

**7.0 Discussion and conclusions**

The method described in \( 4.0 \) above is suitable for trusted methods and equipment with linear behavioural data patterns. A method or equipment is trusted when the historical prediction pattern follows a unique trend. Our experiment showed that the method predicts less accurate results if more than 50 percent of the methods involved lacks adequate prediction credibility. We have involved fifteen different methods in our initial selection. Their predictions were correlated with the actuals and any method whose squared correlation falls below 0.9960 is ignored and was not taken as part of our nugget set. The nuggeted set are then filtered for noise and redundancies using the binning system\[17\] after which they were used in the individual predictions. It was observed that methods with less than 0.9980 squared correlation yield more than 4% percentage error.
Although it is absolutely necessary to involve more methods and equipment in the determination of functions and models as illustrated because the larger the n the more reliable the perfected prediction, only one method or equipment is needed to make a near-perfection prediction. This could mean a reduction in cost for the respective company and the simplicity of its determination could interpret to saving a lives and protecting the environment. The selection of appropriateness of method or equipment follows through a filter pattern initially defined to remove noise and redundancies. Our experiments showed that the behaviour of each data pattern could be modelled as illustrated. Still, there are some data patterns that does not follow the traditional expectations in 3.0 and yet produces desirable results. Fitting non-linear models for these sort of data pattern could mean a further reduction in the computational errors and we see a lot of academic challenges in this direction.

Acknowledgements

We are grateful to Mr. Olubunmi Owoso, the rector of Yaba College of Technology, Lagos, Nigeria for the financial assistance through (Study Leave Approval), and M. Eng., Feng Jian of the Northeastern University, China, for his invaluable advice on leakage predictions and the Nigerian National Petroleum Corporation (NNPC) for providing the needed data. Also we want to thank Kalema Kiweewa, UTM, for providing us with useful information on abnormal behaviour of data patterns in information systems and Josephine Okpareke for the secretarial support.

References

[7]. Wu Ming, Wang Wei-qiang (2006): Application of Wavelet to Detect Leak Point, Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA’06), IEEE Xplore 0-7695-2528-8/06.