Leak Detection in Gas Pipelines Using Accurate Hydraulic Models

Miguel Bagajewicz*†‡ and Gary Valtinson‡

†Department of Chemical Biological and Materials Engineering, University of Oklahoma, Norman, Oklahoma 73019, United States
‡Ok-Solutions, Norman, Oklahoma 73072, United States (www.ok-solutions.org)

ABSTRACT: In this paper, we show the implementation of the Generalized Likelihood Ratio (GLR) method to detect and also identify the size and location of leaks in pipelines. We introduce the use of accurate hydraulic models for hypothesis testing and the use of economics to determine the thresholds of detection and identification. We compare the leak detection power and costs to those of other simple leak detection methods. The economic comparison includes computing the losses for not detecting the leaks (false negatives) and detecting leaks that do not exist (false positives). We also illustrate the improvement in the power of our method by using more-accurate instrumentation.

■ INTRODUCTION

Pipelines are used to deliver petroleum products, natural gas, liquid hydrocarbons, and water to consumers. They do so in a cost-effective manner; however, safety and losses due to leaks are the primary concern in pipeline operation. Thus, the challenge is to use leak detection methods capable of accurately detecting leaks and their location in a timely fashion.

There are several hardware-based methods for leak detection and location. They are generally sensitive to small leak sizes and quite accurate, with regard to location of the leak. Typically, instrumentation is run along the entire length of the pipeline, which helps with the detection of both large and small leaks in a timely fashion and allows for the detection of a leak anywhere along the pipeline. Although significant instrumentation provides many of the advantages associated with hardware leak detection, it also provides disadvantages. The high level of instrumentation results in installation and maintenance costs that are significant. Installation is complex, requiring a considerable amount of work below the surface, since many pipelines are buried. These methods are based on several technologies: detection of noise generated by leaks and the measurement of ultrasonic wave speed (acoustic methods), detection of scattered light due to leaks (fiber optics), detection of vapors (typically by running a small pipe with an inert gas permeable to hydrocarbons with an analyzer at certain distances), among others.

Software-based methods make use of existing measurements of flow and pressure. The simplest one is the flow discrepancy method. It assumes steady state and it declares that there is a leak when two flow measurements separated by a certain distance indicate different flow values. This technique only applies to single pipelines with flow moving in one direction, so complex networks such as gas distribution systems in urban areas do not apply. The volume balance does not help in detecting the location of leaks, and it also cannot distinguish between biases and leaks. If one can add the pressure measurements upstream and downstream, one can use the pressure drop equation to look for discrepancies between flow and pressure drop. For liquids, one can use the gradient intersection method to additionally detect the size and location of the leak (see Figure 1). Quite clearly, with some effort, this can be extended to gases, where the pressure gradient is not linear. Threshold values must be set in the gradient intersection method, since normal pressure drop fluctuations occur. Many false alarms are the result of not setting these values, since this method is dependent on the tuning of the model, because measurement errors (along with uncertainty) in fluid properties can cause difficulties.

Transient model-based methods attempt to distinguish the effects of a leak from all other phenomena in a pipeline. While the pressure analysis method cannot distinguish between a leak and anything else that causes a pressure drop, transient models simulate transients in a system in real time. This method is a numerical integration of three different equations: the momentum, continuity, and the energy equation. Generally, an implicit matrix-based solution is used with all three of these equations. One downside to this approach is that many parameters are needed for the method to work accurately.

Special Issue: Jaime Cerda Festschrift

Received: April 20, 2014
Revised: October 3, 2014
Accepted: October 6, 2014
Published: October 6, 2014
Some of these pipeline parameters can be difficult to obtain, such as the inside pipe roughness, the current drift, and calibration of the instruments. Finally, in order to perform calculations in real time, adaptive modeling must be used. This implies that certain parameters in the system will be adjusted when compared to simulation or measured values. A leak is detected if the discrepancy between the actual data and the model data is greater than the determined limits. If no leak is found, the differences between the measured and calculated values are used to adjust parameters, all this just for one pipe segment. Billmann and Isermann\(^3\) showed that detectable leaks were larger than 2\% for liquid and 10\% for gas. One clearly needs to perform better leak detection, especially for gases.

In turn, frequency analysis methods use a steady oscillatory flow produced by periodically opening and closing a valve. Pressure amplitude peaks are developed from this oscillatory flow for a system with leaks, and then the peaks are compared with a system where no leaks are present. This process allows for identification of leaks, as well as their location and magnitude in a given system, but it can be very complex, and normal pipeline operations must be suspended for frequency analysis methods to be implemented.\(^1\)

Finally, the statistically based generalized likelihood methods are based on hypothesis testing and data reconciliation models.\(^{4,5}\) Mukherjee and Narasimhan\(^6\) were the first to propose a leak detection method based on the Generalized Likelihood Ratio (GLR) method for networks (as opposed to just one pipe segment) that are also based on the use of hydraulic equations. Their solution procedure is iterative and only for liquids. What prevented the use of this method for gases in the past is the fact that the pressure drop correlations are rather inaccurate, and, therefore, the error they introduce would lead to a large number of false positives, as well as large errors in prediction when real leaks occur.

We propose to apply the GLR method to gas pipelines, using accurate metamodels for the hydraulics, as well as tuning the thresholds of hypothesis testing using economics. The article is organized as follows: We first review the GLR method, as applied to our particular case. Then, we discuss the power of the method (effectiveness) and the economics of leak detection. Finally, we present a simple illustration.

**GENERALIZED LIKELIHOOD RATIO FOR PIPELINE LEAK DETECTION**

The Generalized Likelihood Ratio (GLR) is a well-known hypothesis testing statistical method based on the comparison of the performance of two alternative models and how they fit the data. One model, represented by the null hypothesis, is that there is no leak, while the alternative model is that there is a leak. There is then a model (a data reconciliation problem for our case) for each hypothesis.

In the case of leaks in material balancing systems, Narasimhan and Mah\(^4\) proposed using the flow measurements at different points and constructing two data reconciliation models: one that assumed no leak and one that assumed a leak in one of the many segments of a pipeline (defined as pipelines between flowmeters). The value of \(T\) has a central \(\chi^2\) distribution with one degree of freedom under Ho. Since the \(T\) values are not independent, the distribution of \(T\) (the maximum value of all \(T\)) cannot be obtained. However, Narasimhan and Mah\(^4\) used arguments similar to those used by Mah and Tamhane\(^6\) and chose the upper \(1−\beta\) quantile of the \(\chi^2\) distribution \((\chi^2_{1,\beta})\) as the critical value. For a given level of significance \(\alpha\) (usually 95\%), \(\beta = 1 − (1 − \alpha)^{\frac{1}{\beta}}\), where \(p\) is the number of gross errors hypothesized (we use a value of 1). This would ensure that the probability of a Type I error is less than or equal to \(\alpha\).

For the case of only flow measurements, Narasimhan and Mah\(^4\) developed analytical expressions. A similar approach was used by Mukherjee and Narasimhan\(^3\) for liquid pipelines. In this latter case, as in ours, they incorporated pressure measurements. They actually used some approximations for the data reconciliation models.

**Our Generalized Likelihood Ratio Approach.** Our proposed method uses pressure measurements at the beginning and end of each pipeline section, in addition to flow measurements. Without any loss of generality, consider a pipeline with \(n\) segments (see Figure 2).

The likelihood of the null hypothesis (no leak) is represented by the following data reconciliation model:

\[
Q_0 = \min \sum_{\gamma \in N} \left[ \frac{(Q_{\gamma} - \bar{Q}_{\gamma})^2}{\sigma_{Q_{\gamma}}^2} + \frac{(P_{\gamma,x} - P_{\gamma,s})^2}{\sigma_{P_{\gamma,x}}^2} + \frac{(P_{\gamma,s} - P_{\gamma,n})^2}{\sigma_{P_{\gamma,n}}^2} \right]
\]

subject to

\[
Q_{\gamma,x} = \sum_{\gamma=1}^n A_{\gamma,x} \left( \bar{P}_{\gamma,n} - \bar{P}_{\gamma,s} \right) - B_{\gamma} \Delta z_n \quad \forall \ n \in N
\]

\[
Q_1 = Q_2 = ... = Q_N
\]

In these equations, \(\bar{Q}_{\gamma}, \bar{P}_{\gamma,s}, \text{ and } \bar{P}_{\gamma,n}\) are the measured values of flow rates and pressures for each pipe segment \(n\), while \(Q_{\gamma}, P_{\gamma,s}, \text{ and } P_{\gamma,n}\) are the estimators. Finally, \(\sigma_{Q_{\gamma}}^2, \sigma_{P_{\gamma,s}}^2, \text{ and } \sigma_{P_{\gamma,n}}^2\) are the variances of the respective flow and pressure measurements. Note that, for the hydraulic equation, which represents the relationship between flow and pressure (eq 2), we use a simplified metamodel\(^7\) that we developed and has a much smaller error than that of existing proposed equations (Phandale, Weymouth, etc.).

We now build the alternative model (a leak hypothesized in segment \(i\)). In the case where there is a leak of magnitude \(l\) at some distance \(x\) in the pipeline, the pressure drop equation for that segment must be written for each interval before and after the leak (see Figure 3).

Thus, we write

\[
Q^2 = \frac{A(L^2 - P_{\gamma,x}^2) - B\Delta z}{x}
\]

Figure 2. Linear pipeline with \(n\) segments and compressors.
Figure 3. Schematic depiction of a segment that has a leak.

\[(Q - l)^2 = \frac{A(P_i^2 - P_{g}^2) - B\Delta z}{L - x}\]  

(5)

When a leak is assumed in pipe segment \(i\), we solve the expression

\[Q = \min \sum_{n \in N} \left[ \frac{(Q_i - Q_n)^2}{\sigma_{Q_i}} + \frac{(P_{n,i} - P_{n,g})^2}{\sigma_{P_{n,i}}} + \frac{(P_{g,i} - P_{g,g})^2}{\sigma_{P_{g,i}}} \right]

\text{subject to}

\[Q_n^2 = \frac{A_n(P_{n,i}^2 - P_{n,g}^2) - B_n\Delta z_n}{L_n} \quad \forall n \in N; n \neq i\]  

(6)

\[Q_i^2 = \frac{A(P_{i,i}^2 - P_{i,g}^2) - B\Delta z_i(x_i)}{x_i}\]  

(7)

\[Q_i - l)^2 = \frac{A_i(P_{i,i}^2 - P_{i,g}^2) - B\Delta z_i(L_i - x_i)}{L_i - x_i}\]  

(8)

\[Q_i = Q_2 = \ldots = Q_i\]  

(9)

\[Q_{i+1} = Q_i - l\]  

(10)

\[Q_{i+2} = \ldots = Q_N\]  

(11)

Here, \(P_{i,g}\) is the pressure at the location of the leak. In turn, \(\Delta z_i(x_i)\) is the change in height between the beginning of the pipe and the location of the leak and \(\Delta z_i(L_i - x_i)\) is the change in height downstream from the leak.

Equations 8 and 9 present the problem that \(\Delta z_i(x_i)\) and \(\Delta z_i(L_i - x_i)\) are functions and not parameters as in the other equations. These equations can be simplified as follows: We multiply 8 by \(x_i\) and 9 by \(L_i - x_i\) and add them together to obtain

\[Q_i^2 x_i + (Q_i - l)^2(L_i - x_i) = A_i(P_{i,i}^2 - P_{i,g}^2) - B_i\Delta z_i(L_i)\]  

(12)

where we used the fact that \(\Delta z_i(L_i) = \Delta z_i(x_i) + \Delta z_i(L_i - x_i)\).

Substituting eqs 8 and 9 by eq 13 eliminates the unmeasured variable \(P_{g,i}\) which is an observable quantity that can be calculated after the problem is solved (although this is rarely needed). The leak detection procedure is as follows:

1. Hypothesize leak in every branch and solve data reconciliation problems
2. Obtain GLR test statistics for each branch \(T_i = \log(O_{i,j}) - \log(O_i)\)
3. Determine the maximum test statistic \(T_i\)
4. Following the GLR method, we compare the max test statistic with the chosen threshold value \((T^*)\). If the statistics is larger than this threshold (i.e., \(T_i > T^*\)), the leak is said to have been identified and located in the branch corresponding to the maximum test statistic

Quite clearly, the method can be extended to multiple branches, by just modifying eq 10 accordingly.

### POWER OF THE METHOD

The overall power of the method is the ratio of the number of leaks correctly detected or identified to the number of leaks simulated. One can add qualifiers to such power. For example, one can define correct identification if the pipeline segment is correctly identified, or if the estimator of the leak is close to the real value within a predetermined error and/or if the location is identified within a certain predetermined error. The key operative word in these definitions is the word “correctly”, which needs to be defined.

More specifically, we distinguish these different types of power:

1. **Detection Power:** This is related to the frequency at which the presence of a leak is detected, regardless of its size and location. The outcome of this test is either “there is a leak” or “there is no leak”. Some methods can only do this.
2. **Identification Power:** We call identification the ability to detect the leak and also give additional information. Of these, there are two varieties:
   a. **Leak Size Identification Power:** In this case, the additional information is the size of the leak.
   b. **Leak Size and Location Identification Power:** In this case, the additional information is the size of the leak, as well as its location.

To obtain the power, several leak sizes were tried at specific locations, sampling 100 different sets of simulated “measurements” normally distributed around the actual flows and pressures.

Our analysis focuses on the ability of our method (ALINA) to detect a leak within a particular segment and locate the leak. As in all GLR-based methods, one must tune the threshold, \(T^*\), using some procedure. We use economics to determine this threshold.

We compare ALINA with other methods:

- The use of GLR based on flow measurements only (as implemented by Narasimhan and Mah)\(^7\). This method uses a threshold (\(T^*\)) similar to that used in ALINA, and its data reconciliation model is given by eqs 6, 10, 11, and 12. This method does not provide location, as it does not regard the hydraulics.
- **Flow discrepancy:** In this case, the difference of two measurements \((\hat{Q} - \hat{Q}_{i,1})\) is compared to the sum of the two standard deviations \((\sigma_{Q_i} + \sigma_{Q_{i,1}})\). A leak is declared if the aforementioned difference is positive and larger than the sum of the standard deviations multiplied by a threshold, \(T^*\). Because this method does not involve the use of reconciliation, one has several choices. Similar to GLR based on flow measurements, the flow discrepancy method does not localize the leak.
- **Hydraulic discrepancy:** This method involves accepting the pressure measurements as true and obtaining a representative flow as follows:

\[Q_i^# = \sqrt{\frac{A_i(P_{i,i}^2 - P_{i,g}^2) - B_i\Delta z_i}{L_i}}\]  

(14)

This flow is now compared to the flow measurement corresponding to the beginning of the same pipe segment. The difference between the measured flow at the beginning of the pipe \(Q_i\) and the flow predicted by hydraulics determines the presence of a leak based on a threshold, \(T^*\). That is, if \(Q_i - Q_i^# > T^*\), then a leak has been identified.

To obtain the location, two pressure profiles are constructed. They obey eqs 8 and 9. As discussed above, these two equations...
can be rewritten, eliminating the pressure at the leak location to obtain eq. 13. Thus, the location of the leak can be obtained manipulating 13 to obtain the following:

\[
x_i = \frac{A_i (\bar{P}_{i+1}^2 - \bar{P}_{i+1}^2) - B_i \Delta z_i (L_i) - \bar{Q}_{i+1}^2 L_i}{\bar{Q}_{i+1}^2 - \bar{Q}_{i+1}^2}
\]  

(15)

Economics. To calculate the cost of each leak of size \(Q\) each time it occurs (\(LC\)), we add the cost of a detected leak (location cost + repair cost + cost of the loss), and the cost of an undetected leak. The cost of a real leak is given in the following expression:

\[
LC(Q) = R(C_{D1} + C_{Q1}(D_1 + D_2)) + (1 - P)Q_i
\]  

\[C_j \times 365\]

(16)

The first term is the cost of the detected leak. This is obtained by multiplying the expected number of successful identifications of a leak \(R\) (incidents per year of the leak of size \(Q\)) with three terms:

- **The cost of physically locating the leak**: This is given by the cost of locating the leak \(C_j\), multiplied by the number of days required to locate the leak \(D_1\). It is assumed that once a leak has been detected, the task of sending crews to physically identify the exact location takes a few days.

- **The cost of repairing the leak**: This is given by the cost per day \(C_{D1}\) multiplied by the number of days needed to repair the leak \(D_1\).

- **The financial loss due to the loss of fluid, during the days needed to locate and repair the leak**: This is given by the price of fluid \(C_{Q1}\) multiplied by the flow rate of the leak \(Q_i\), multiplied by the number of days the fluid is being lost \(D_1 + D_2\); one can also assume that the leak is stopped once it is located, so the number of days of loss reduces to \(D_1\).

We assume that the cost of physical location and repair costs are independent of the size of the leak. The second term is the cost of the undetected leak, which is obtained by multiplying the frequency at which the method does not detect the leak \((1 - R)\) by the size of the leak \((Q_i)\), multiplied by the cost of the fluid \((C_j)\), multiplied by 365 days per year.

There is also a cost of false positives (detecting a leak when there is none); such a cost is given by the number of false positives \(FP\) per year multiplied by the daily cost of locating a leak \((C_j)\), multiplied by the time it takes to give up the search for the nonexistent leak \((D_{FP})\). Thus, the false positive cost (FPC) is

\[
FPC = FPC_{D_{FP}}
\]  

(17)

We consider the value \(D_{FP}\) to be economically equivalent to the time needed to locate a leak without a localizing method. Therefore, FPC is a fixed cost per year. In turn, the number of false positives (FP) is related to the value of the threshold used. The smaller the threshold, compared to the standard deviation of the estimators through data reconciliation, the larger the number of positives. Two ways of ameliorating the impact of these false positives is to increase the threshold, which, in turn, may increase the false negatives, or force the use of more-accurate instrumentation, which will reduce the standard deviation of the estimators. Thus, this is an economic issue.

Something else that can be considered when estimating the cost of a false positive is the use of a threshold related to the leak size. If a leak of size 1% is detected 90% of the time, while leak sizes of <1% are detected significantly less frequently, it may be justifiable to ignore leaks below a size of 1%. This, in turn, reduces the number of false positives. We do not explore this added threshold in the present article.

To evaluate any method, one must compute the expected annual cost. False positives have a fixed cost, so the expected value is already given by FPC. For false negatives, the expected cost can be calculated by integrating the product of the annual cost for each leak size by the probability of such leak size:

\[
ELC = \int_0^\infty P(Q_{ij})LC(Q_{ij}) dQ_{ij}
\]  

(18)

Figure 4 provides the probability of a leak per year for a given size of leak. This was constructed using the significant incidents data from the Pipeline and Hazardous Material Safety Administration. As should be expected, large leaks are not as likely as smaller leaks.

**Optimum Threshold.** The optimum threshold \((T^*)\) can be obtained by calculating the expected cost for each threshold value and selecting the one that provides the smaller annual cost. This applies to three other thresholds as well: \(T^a\) for GLR without flow, \(T^b\) for flow discrepancies, and \(T^c\) for hydraulic discrepancies. The minimum expected cost is the location of the optimum threshold. This point corresponds to the point where false positives have as much economic impact as the false negatives.

**Illustration**

Our example consists of a four-segment pipeline, with the gas being recompressed at the beginning of each segment. Measurements of inlet as well as outlet pressures and flow rates for each segment are made. Using Pro II (SIMSCI, Invensys), we simulated natural gas through a single pipeline 400

Table 1. Economic Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>annual gas price&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>(C_i)</td>
<td>$5.00/MMBTU</td>
</tr>
<tr>
<td>cost of physically locating a leak&lt;sup&gt;c&lt;/sup&gt;</td>
<td>(C_j)</td>
<td>$5,000/day</td>
</tr>
<tr>
<td>cost to repair a leak&lt;sup&gt;c&lt;/sup&gt;</td>
<td>(C_l)</td>
<td>$10,000/day</td>
</tr>
<tr>
<td>time to repair a leak</td>
<td>(D_l)</td>
<td>5 days</td>
</tr>
<tr>
<td>time to physically locate a leak</td>
<td>(D_{FP})</td>
<td>3 days</td>
</tr>
<tr>
<td>all methods except ALINA</td>
<td>(D_{FP})</td>
<td>3 days</td>
</tr>
<tr>
<td>ALINA</td>
<td>(D_{FP})</td>
<td>3 days</td>
</tr>
<tr>
<td>time before giving up search for nonexistent leak (all methods)</td>
<td>(D_{FP})</td>
<td>3 days</td>
</tr>
</tbody>
</table>

<sup>a</sup>Prices and costs given in terms of U.S. dollars (USD).<sup>b</sup>Data taken from ref 9.
km long, with a total of four compressors. The nominal pipe diameter is 20 in.; the inlet temperature and pressure to each segment are 45°C and 1200 psi, respectively. The volumetric flow rate was 346.086 million standard cubic feet (SCF) per day. Once we obtained the pressure into and out of the compressors, we then applied random error to these values to produce 100 points of data.

The economic parameters are provided in Table 1. We assumed it takes 5 work days (8 h per day) to complete the repair. Since ALINA can detect the location of the leak, we used 3 days to locate the leak when using GLR on flows only and 1 day with ALINA. Finally, we assumed that the flow measurements have 1% error, while the pressure measurements have 0.1% error.

We provide ALINA's power curves to detect a leak in the correct segment for various thresholds in Figure 5. As Figure 5 indicates, a leak size of 1% is detected 88% of the time for $T^* = 0.1$ and 1. Below this leak size of 1%, the number of leaks detected decreases by more than 20%, because the leak is more difficult to detect (it may have the same effect in the data reconciliation as noise) without decreasing the threshold, $T^*$. However, decreasing $T^*$ increases the number of false positives. Figure 6 provides the frequency of false positives, compared to the threshold, $T^*$. As can be seen, increasing the threshold decreases the number of false positives. Indeed, at $T^* = 2$ and $T^* = 3$, there is no leak detected when there is no leak (no false positives). When $T^* = 1$, a leak is detected 3% of the time when there is no leak, while for $T^* = 0.1$, ALINA detects a false positive 63% of the time.

Now that we have the power curves at various thresholds for ALINA, we can assess the economics and optimize the method. First, we calculate the annualized cost for each threshold, $T^*$. Figure 7 shows the annualized cost from ALINA using a threshold of $T^* = 1$. At the size of a leak of 0%, the cost is finite due to false positives. At the beginning of the curve, the cost increases because of false negatives. In this case, the false positives have an annual cost of approximately $164 000 per year. Below an existing leak size of 2%, the primary cost associated with the annual cost is strictly false negatives. Eventually, the leak is detected the majority of the time (2% in this case), causing the
cost to be primarily the price of locating it physically, repairing the leak, and the product lost before location and while repairing.

Once the annual cost is determined, we can then determine the expected cost. Figure 8 shows the cost of a leak versus the leak size as a percentage of total flow. It is the product of the annualized cost and the probability (see Figure 4). As the probability decreases with leak size, so does the impact of that particular leak size. As with the annual cost, we used the curve for a threshold of \( T^* = 1 \) using ALINA. The expected cost is obtained by computing the area under the curve.

Figure 9 shows a curve based on the \( T^* \) values from ALINA. The minimum of this curve is the optimum threshold where the false negatives and false positives balance each other economically. As previously indicated, the smaller the threshold, the less false negatives occur; however, the number of false positives will increase. This optimal point corresponds to the point where the false positives and negatives are balanced in the economics.

As Figure 9 indicates, the optimum threshold for ALINA is \( T^* \approx 1 \) with an expected annual cost of \( \$1.09 \) million per year. It is clear
from this figure that this optimum threshold determination is crucial, because cost changes rather rapidly. Knowing the optimal threshold, we can now compare ALINA to other methods. To compare the other methods, we had to perform the same assessment to determine their optimal thresholds.

We obtained the optimal thresholds for the other methods (GLR, flow discrepancy, and hydraulic discrepancy) and compare their power curves to detect a leak in Figure 10. Leak detection through reconciliation of only flows uses the GLR method and thus requires a threshold ($T^*$) similar to that for ALINA. Figure 10 indicates that ALINA has the best performance, achieving 88% detection for a leak size of 1%. We provide the percentage of time that a false positive will occur in Table 2.

Figure 11 depicts the power of ALINA to locate the leak within a specific distance from the true location of the leak. Once again, this was using the optimal threshold found for ALINA, in this case, $T^* = 1$. As the graph indicates, ALINA will have an easier time detecting a leak within 10 km than within 2.5 km. That said, being within 10 km is still significantly better than looking over the full 100-km pipe segment.

We then make a comparison between the ability of ALINA to locate a leak and locating a leak with the hydraulic discrepancy method. Figure 12 shows a comparison between ALINA at its optimal threshold of $T^* = 1$ and the hydraulic discrepancy method with its optimal threshold at $T^* = 6.5$. The other methods

---

Table 3. Expected Cost Comparison between ALINA and Various Leak Detection Methods

<table>
<thead>
<tr>
<th>method</th>
<th>ALINA $T^* = 1$</th>
<th>GLR with only flows $T^* = 1$</th>
<th>flow discrepancy $T^* = 1.25$</th>
<th>hydraulic discrepancy $T^* = 6.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>threshold</td>
<td>$T^* = 1$</td>
<td>$T^* = 1$</td>
<td>$T^* = 1.25$</td>
<td>$T^* = 6.5$</td>
</tr>
<tr>
<td>expected cost (M USD/yr)</td>
<td>$1.09$</td>
<td>$5.52$</td>
<td>$5.63$</td>
<td>$6.96$</td>
</tr>
<tr>
<td>difference from ALINA (M USD/yr)</td>
<td>$0.00$</td>
<td>$4.43$</td>
<td>$4.54$</td>
<td>$5.87$</td>
</tr>
</tbody>
</table>
do not provide an estimator for location, because pressure measurements are required to locate a leak. ALINA locates small leaks of 0.5% of the total flow 5%−15% of the time more often than the hydraulic discrepancy method. On the high end, ALINA achieves >50% localizing within 2.5 km and >90% localizing within 7.5 km for a 5% leak. Meanwhile, the hydraulic discrepancy method remains at <50% localizing for both 2.5 km and 7.5 km for a leak size of 5% of the total flow.

We compiled the optimal thresholds and expected costs for the three other leak detection methods given in Table 3. This includes the use of GLR with flow only, flow measurement discrepancy, and hydraulic discrepancy. As the table indicates, ALINA outperforms these other methods, with $4.43 million USD in savings at the low end and $5.87 million USD in savings at the high end.

We then looked at a case where both instruments have high accuracy (flow measurements with 0.1% error, pressure measurements with 0.1% error). This would be a more-extreme (and more-expensive) case where the accuracy of all instruments is the best possible. Figure 13 provides the power curves for multiple thresholds to identify a leak in the correct pipe segment under these conditions. The frequency of false positives using ALINA for each threshold is provided in Figure 14. As these figures indicate, with a threshold of $T^* = 1$, ALINA can detect a leak equivalent to 0.5% of the total flow rate 98% of the time, with only 4% of the leaks detected being false positives.

Figure 15 provides an overview of ALINA’s ability to locate the leak with these improved flow instruments. Comparing Figure 11 and Figure 15, we see that ALINA experiences a 5%−10% increase in its ability to locate a leak. Thus, as is to be expected, high accuracy equates to better detection and localizing of a leak with ALINA and will have an impact on economics.

**CONCLUSIONS**

We have studied the economics of the implementation of the Generalized Likelihood Ratio (GLR) to leak detection in pipelines. Our method, called ALINA, used both flow and pressure with GLR to detect a leak and locate a leak. As we demonstrated, the threshold has a direct effect on ALINA’s ability to detect a leak and locate a leak, as well as in the economics of leak detection.

We compared our method to the use of leak detection by GLR with only flow measurements, leak detection via flow discrepancy, and leak detection through hydraulic discrepancies. We found that ALINA outperforms all of these methods, by 4.4 million USD or more (Table 3) for our particular small example. We expect this number to increase for larger-sized pipelines with more compression stations.

We also provided an assessment of ALINA’s capabilities if instrumentation is even more accurate (particularly with flow measurements being at 0.1% accurate, versus 1%). This showed a jump in the detection rate for leaks 0.5% in size to 98%. The accuracy of ALINA to locate a leak is also increased by 5%−10%. Thus, ALINA shows a clear capability to detect and locate leaks of at least 2% of the total flow and moderately well below this size with the example considered in this paper.

In future work, we will address the impact of the uncertainty associated with the values of the parameters $A_n$ and $B_n$, a problem that all model-based leak detection formulations share. Other work to be considered is the issue of biases (the method presented in this paper assumes well-calibrated, bias-free instruments), the addition of new instruments, the economics tradeoffs between investment and leak cost reduction, and considering other thresholds (such as size of the leak to be accepted as such) in tandem with the thresholds used in this paper.

**AUTHOR INFORMATION**

Corresponding Author
*Tel.: 1-405-325-5458. E-mail: bagajewicz@ou.edu.

Notes
The authors declare no competing financial interest.
ACKNOWLEDGMENTS

The following undergraduate students worked on portions of the paper and deserve recognition: David Mannel, Andrew S. Negley, Rachel N. Weber, Eric Penner, Josh Stephens, Elijah Odusina, and James Akingbola. Graduate students Mesude Ozturk and Luftife Hacioglu also helped at various points.

REFERENCES