

## Statistical Parameter Estimation to Find Oil Pipeline Leakage Using a Fiber Optic Sensor

N. Liu<sup>1</sup>, R. Paranjape<sup>1\*</sup>, E. Hara<sup>1</sup> and K. Asghari<sup>2</sup>

<sup>1</sup>Electronic Systems Engineering Program, University of Regina, Regina, SK S4S 0A2, Canada

<sup>2</sup>Petroleum Systems Engineering Program, University of Regina, Regina, SK S4S 0A2, Canada

**ABSTRACT.** We present a new method for enhancing the detection of leakage points on an oil pipeline using a fiber optic sensor stretched along the pipeline. The method is based on Maximum Likelihood Estimation (MLE) and is used with an Optical Time Domain Reflectometer (OTDR). The system depends on a series of fuse blocks, which respond to the presence of oil. The fuse blocks bend the fiber, when the fuse blocks are immersed in oil. The location of the bend in the fiber can be detected using an OTDR. This approach, while very effective in general, can become less successful in certain conditions including: low signal-to-noise ratio, long lengths of fiber (> 200 km) and low signal strength. In this work we modify the basic method to add steps including the Maximum Likelihood Estimation of the parameters in the method. This new modified method is tested using various signals and the approach is suggested to be effective for realistic situations in which the fiber is only very slightly bent or in the presence of high noise. The results suggest that the new method is very effective and could be implemented in field trials.

*Keywords:* Distributed fiber sensors, Maximum Likelihood Estimation, oil pipeline leakage, OTDR signal analysis

### 1. Introduction

The detection of slow leaks and of catastrophic failures in oil pipelines is an increasingly important and pressing problem in today's energy dependent world (Chen et al., 1998; Paranjape et al., 2002). Easily available oil reserves are rapidly being exhausted and more and more remote sites are being explored and tapped for oil reserves. The issue of promptly detecting the leakage and rupture of oil pipelines has not been given the attention it warrants given the potential damage to sensitive environments, the cost of clean up, and the loss of revenue to oil producers.

Pinpointing the location of a leak or rupture represents a significant challenge. A distributed sensor with sufficient resolution, which can be stretched along a pipeline, is required.

#### 1.1. Fiber Optic OTDR Sensor

A method, described in Paranjape et al. (2002), is to spiral a long optical fiber around the oil pipeline and wrap it with a plastic sheet. Many fuse devices, which bend or crimp the fiber when they are immersed in oil, are attached to the fiber at equal intervals. These bending devices alter the physical properties of the fiber when the fuse device is exposed to oil or oil products. At one terminal-end of the fiber, an OTDR (Optical Time-Domain Reflectometer) receives backscatter from a narrow optical pulse that has been launched into the fiber. A detailed description of the operation and basic princi-

ples behind optical time-domain reflectometry are available in the literature (TFOM, 2000).

#### 1.2. Relationship to other Oil Pipeline Leakage Detection Technologies

A number of techniques for oil leakage detection have been proposed previously. For example when pipeline leakage occurs parameters describing the flux in the oil pipeline (such as flow rate, accumulated volume, pressure and temperature) will vary at the inlet and the outlet. In this method, a complex mathematical model is constructed with the relationships of flux parameters. Once leakage occurs, these relationships are destroyed. In this statistical method, these flux parameters are monitored continuously, and leakage can be detected by using probability analysis and pattern recognition. The exact location of the leakage, however, is somewhat more difficult to determine in this method. An additional defect of this method is that it requires the use of high accuracy instruments to measure the parameters, and the fact that slight leakage is hard to detect (Liu et al., 2000).

A second method is based on analysis for acoustic signals. Because oil leaking will produce mechanical oscillation, acoustic sensors can be used to obtain the corresponding acoustic signal. Using certain signal processing algorithms the existence and even the location of leakage can be estimated. This method has acceptable performance in case of short pipelines, however in the presence of environmental noise or in the case of long distance pipelines the method is of limited use (Beushausen et al., 2004).

A third method is installation of a system of uniformly

---

\* Corresponding author: Raman.Paranjape@uregina.ca

distributed sensors connected using copper wire placed along the outside of the pipeline. The disadvantage of electric wire is that the resistance of these wires increases with the length of the wire. The Ohmic loss due to the wire becomes significant in most practical situations. In addition, the wire may pick up other ambient electrical signals. Lastly, this system typically fails to detect multi-point leakage.

### 1.3. Challenges for the Fiber Optic OTDR Sensor

The primary challenge with the fiber Optic/OTDR approach is to develop the necessary sensitivity to bending in the fiber. This problem is fundamentally a signal detection problem (Hogg et al., 1977; Kay, 1993; Demirli et al., 2001).

It is relatively straightforward to identify and locate fiber discontinuity when the distance is short, because the energy of light pulse is strong compared with inherent noise in the system. However, in order to develop a practical and useful system, high-resolution (i.e. very short duration) light pulses and long distance fiber (more than 200 km) are required. A high-resolution light pulse has low energy. The long distance means great attenuation of the signal. Both of these factors conspire to reduce the energy of the backscattered light and lead to a low SNR (Signal to Noise Ratio). In addition, if there is only a slight bend in the fiber, the attenuation characteristics of the fiber are changed only slightly and again the discontinuity is not easily identified (TFOM, 2000).

One simple and common approach to improve the SNR is through averaging many waveforms or realizations. This decreases the variance of noise and improves the SNR. There is, however, a statistical limit of how much this approach can improve the signal. In addition, it may be unacceptable to wait for the many hours necessary to collect the necessary amount of data. Another possible solution to the problem is to increase the quality of the OTDR. But again this is not a practical solution as high quality OTDRs are costly and would need to be placed at intervals of about 200 km along the pipeline.

A new approach to this problem is the application of statistical parameter estimation, an advanced method of Digital Signal Processing (DSP). There are many DSP techniques to estimate model parameters in noisy signals. The technique of statistical parameter estimation is chosen because it can help find the location of discontinuity in the recorded signals with little added cost or delay.

## 2. Methods

### 2.1. Algorithm Design

The recorded signal of the backscattered light captured by an OTDR can be modeled as an exponential function. In addition, the waveform is corrupted by noise. It can be expressed as

$$y(n) = \beta e^{-\alpha n} + w(n) \quad n: 0 \sim \infty \quad (1)$$

where  $y(n)$  is the recorded signal,  $\beta$  is a multiplicative attenua-

tion factor,  $-\alpha$  is the exponential term related to the loss of transmission through the fiber,  $n$  is the discrete time variable, and  $w(n)$  is Gaussian white noise with variance  $\sigma^2$ .

We now define a new variable,  $x(n)$ .  $x(n)$  is assumed to be a section of the recorded signal  $y(n)$  spanning the time segment  $[m, m + N - 1]$  which is within the period  $n: 0 \sim \infty$ , given as

$$x(n) = \beta e^{-\alpha(n+m)} + w(n) \quad n: 0 \sim (N+1) \quad (2)$$

where  $\alpha$  and  $\beta$  can be estimated as  $\hat{\alpha}$  and  $\hat{\beta}$  using the Maximum Likelihood Estimation (MLE) method. The segment  $[m, m + N - 1]$  can then be shifted along the  $n$  axis, and the estimates of  $\hat{\alpha}$  and  $\hat{\beta}$  can be recalculated. If the process generating the signal is unchanged we would expect subsequent estimates to be similar. If, on the other hand, the optical fiber has a bend in it due to a fuse block,  $x(n)$  will have a sudden change in the values of  $\alpha$  and  $\beta$  around that point. When the segment  $[m, m + N - 1]$  includes this point, estimated parameters ( $\hat{\alpha}, \hat{\beta}$ ) will be significantly changed from the previous estimates of ( $\hat{\alpha}, \hat{\beta}$ ). The position of the leaking point can be determined based on the value of  $m$ .

In Figure 1 below we see a typical OTDR trace. The leakage point is identified in the trace by the sudden and abrupt increase in the returned signal's attenuation as it passed through a bend in the optical fiber. The horizontal axis in this figure is time or distance along the fiber and the vertical axis is the magnitude of the returned signal. The units of the returned signal magnitude are somewhat as arbitrary as they include gain and losses that are constant within the system. Only the relative difference in the magnitude of the signal is of importance in this analysis. The figure further shows the sliding window  $[m, m + N - 1]$  which is moved along the trace. The values of ( $\hat{\alpha}, \hat{\beta}$ ) are estimated at each value of  $m$ .

### 2.2. Application of the Gaussian-Newton Method

In the MLE method, we attempt to find  $\theta$  which maximize  $p(x, \theta)$ , where

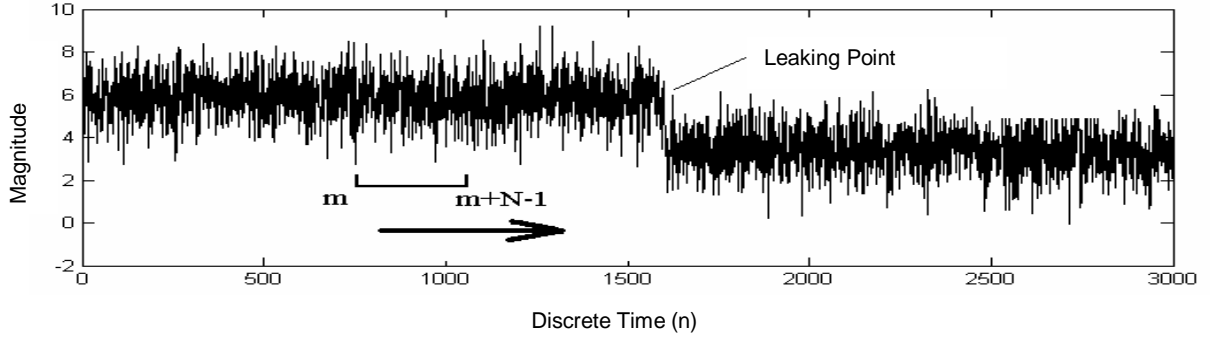
$$p(x, \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [x(n) - \beta e^{-\alpha(n+m)}]^2 \right\} \quad (3)$$

is the joint probability density function obeying the normal distribution.

Because  $w(n)$  is Gaussian white noise, the estimation problem becomes a nonlinear least squares problem. Because of the difficulty in obtaining an analytical solution, an iterative approach must be engaged.

We define, following the textbook by Kay (1993), the following matrices:

$$\theta = (\alpha, \beta)^T, \quad (4)$$



**Note:** this is a typical trace from an OTDR of a section of fiber with a sharp bend in it; the estimation sliding window is also shown in which application of the MLE algorithm allows estimation of model parameters  $(\hat{\alpha}, \hat{\beta})$ .

**Figure 1.** Typical OTDR trace from an abrupt bend in short Fiber.

$$I(\theta) = \begin{bmatrix} -E \left[ \frac{\partial^2 \ln p}{\partial \alpha^2} \right] & -E \left[ \frac{\partial^2 \ln p}{\partial \alpha \partial \beta} \right] \\ -E \left[ \frac{\partial^2 \ln p}{\partial \alpha \partial \beta} \right] & -E \left[ \frac{\partial^2 \ln p}{\partial \beta^2} \right] \end{bmatrix}, \quad (5)$$

$$s(\theta_k) = \beta_k * \exp(\alpha_k * n), \quad (6)$$

$$\Delta(\theta_k) = I^{-1}(\theta_k) \begin{bmatrix} \frac{\partial s}{\partial \alpha_k} & \frac{\partial s}{\partial \beta_k} \end{bmatrix}. \quad (7)$$

An iterative formula for estimation of  $\theta$  is presented by Kay (1993) as:

$$\theta_{k+1} = \theta_k + \Delta(\theta_k) * (x - s(\theta_k)). \quad (8)$$

When the expression of  $p(x, \theta)$  is substituted into  $I(\theta)$ , the result is:

$$I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=0}^{N-1} n^2 \beta^2 e^{-2\alpha(n+m)} & -\sum_{n=0}^{N-1} n \beta e^{-2\alpha(n+m)} \\ -\sum_{n=0}^{N-1} n \beta e^{-2\alpha(n+m)} & \sum_{n=0}^{N-1} e^{-2\alpha(n+m)} \end{bmatrix}. \quad (9)$$

The elements of  $I(\theta)$  are expanded below, using  $r = e^{-2\alpha}$ :

$$I(\theta)_{11} = \frac{1}{\sigma^2} \beta^2 r^m \times \left[ \frac{r(1+r-N^2r^{N-1}+(2N^2-2N-1)r^N+(-N^2+2N-1)r^{N+1})}{(1-r)^3} + \frac{2mr(-Nr^{N-1}+(N-1)r^N+1)}{(1-r)^2} + \frac{m^2(1-r^N)}{1-r} \right], \quad (10)$$

$$I(\theta)_{12} = I(\theta)_{21} = -\frac{1}{\sigma^2} \beta r^m \times \left( \frac{r(-Nr^{N-1}+(N-1)r^N+1)}{(1-r)^2} + \frac{m(1-r^N)}{1-r} \right), \quad (11)$$

$$I(\theta)_{22} = \frac{1}{\sigma^2} r^m \frac{1-r^N}{1-r}. \quad (12)$$

### 2.3. Selection of N

According to Kay (1993), the ML estimated parameter  $\hat{\theta}$  is asymptotically distributed according to  $\hat{\theta} \sim N(\theta, I^{-1}(\theta))$ . That is to say, when the length of  $x(n)$ :  $N \rightarrow \infty$ ,  $E(\hat{\theta}) \rightarrow \theta$ , and  $var(\hat{\theta}) \rightarrow I^{-1}(\theta)$ , which is Cramer Rao Lower Bounds.

We can therefore conclude that if we allow the size of the sliding window,  $N$ , to become very large, we will get a very good estimation of  $\alpha$  and  $\beta$ . However, we would like to keep  $N$  as small as possible because our resolution for detecting the leakage point is dependent on  $N$ . It is important to select  $N$  so that it is of the appropriate length. While the magnitude of  $N$  should be large enough in order that  $var(\hat{\theta})$  is small, it should not be so large as to smooth out the effect of a change in  $\hat{\theta}$  caused by a discontinuity point. Otherwise, the whole point of the analysis, that of finding the leakage point, will be lost.

## 3. Result Analysis

### 3.1. Evaluation of N

In Table 1 below, we present a set of results showing the effect of various lengths of  $N$  and the effectiveness of the method to estimate the values of  $\alpha$  and  $\beta$ .

In the table below we present values of the actual and estimated values of the variance of  $\alpha$  and  $\beta$  for different values of  $N$ . We must select a value of  $N$  which will allow the difference between the actual and estimated values to be greater

than the Signal-to-Noise-Ratio (SNR). This will allow us to readily see the points at which the difference between the actual and the estimate is large (hence due to a leak).

**Table 1.** Theoretical Asymptotic and Actual Variance

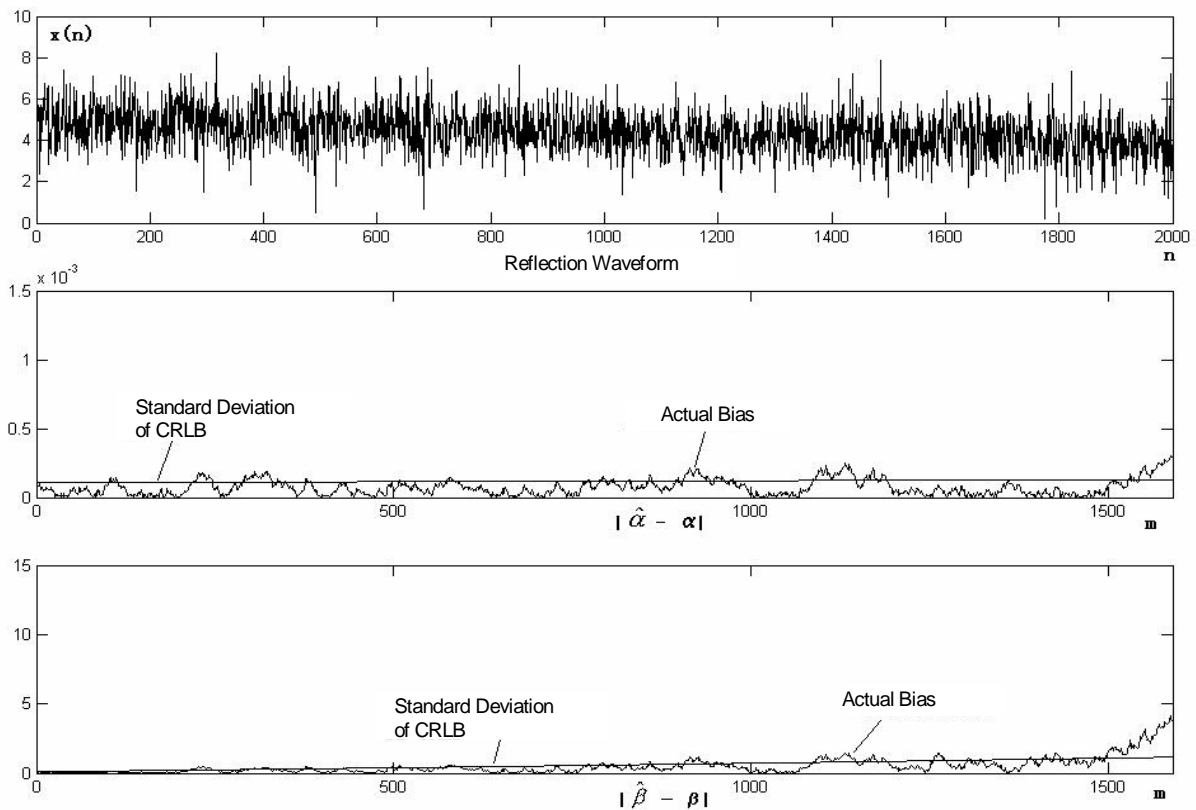
N	Actual Variance of $\alpha$	Theoretical Variance of $\alpha$	Actual Variance of $\beta$	Theoretical Variance of $\beta$
50	1.1385e-05	1.0734e-05	0.0834	0.0779
100	1.1973e-06	1.3494e-06	0.0358	0.0396
150	4.3472e-07	4.0221e-07	0.0273	0.0266
200	1.6633e-07	1.7070e-07	0.0196	0.0201
250	8.9270e-08	8.7926e-08	0.0156	0.0161
300	4.8443e-08	5.1190e-08	0.0135	0.0135
350	3.3818e-08	3.2430e-08	0.0129	0.0116
400	2.1591e-08	2.1857e-08	0.0102	0.0102

When  $N = 350$ ,  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  caused by noise does not exceed those caused by the discontinuity led by leaking point, even when the SNR is low. A larger  $N$  results in smaller variance and higher sensitivity to a leaking point. But if  $N$  is too large, two leaking points with distance less than  $N$  cannot be distinguished. In addition, larger  $N$  means greater computation and therefore the calculation will be more time consuming.

**3.2. Examples of the Application of the Iterative Algorithm**

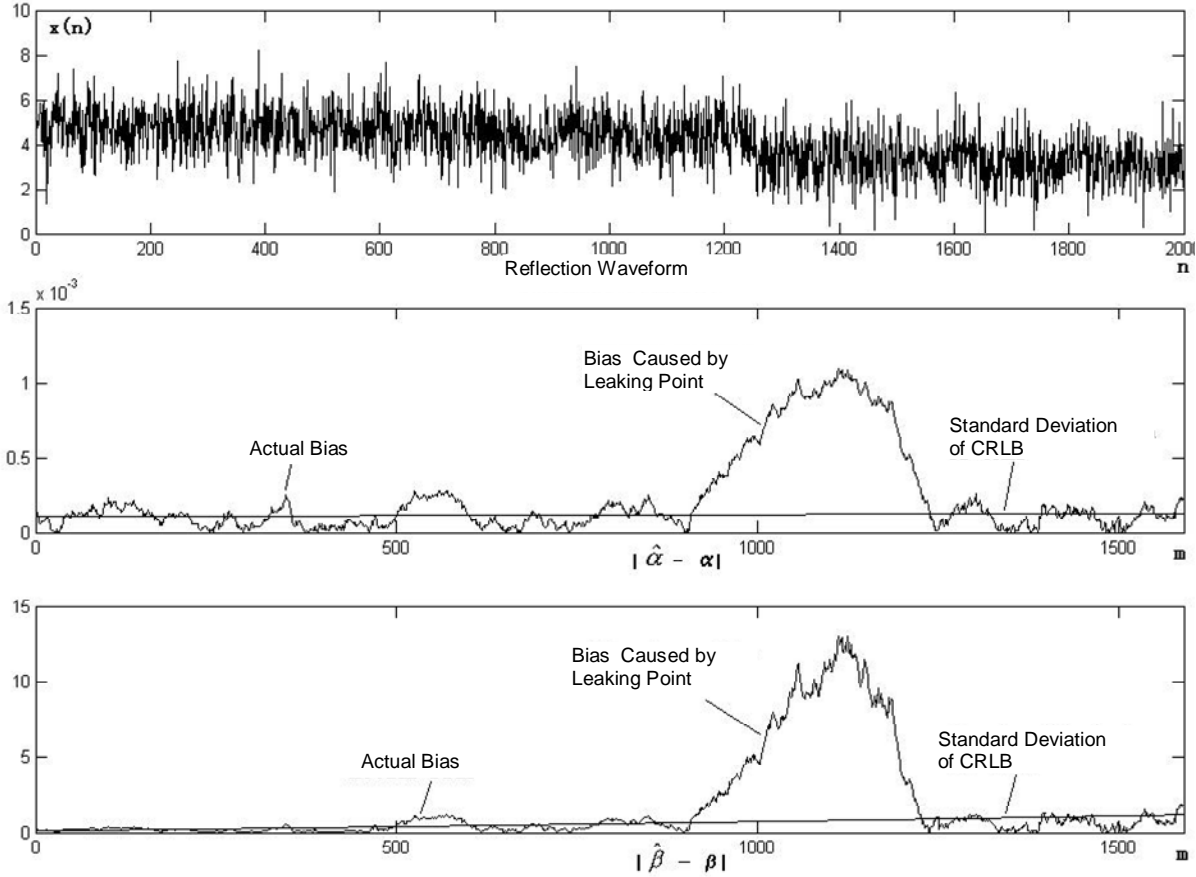
We now present a number of results from the application of the technique to simulated signals. The first example (Figure 2) is of the application of the iterative algorithm to a waveform without a discontinuity point. It is observed that  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  typically has a value which is near the standard deviation of CRLB. The second example (Figure 3) is of the application of the method to an OTDR waveform in which there is a discontinuity point.

Comparing Figure 3 with Figure 2, we can observe that  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  increases abruptly as soon as the do-



**Note:** This figure shows three traces. The first trace shows the returned signal at the OTDR. The second trace shows the result of computing  $|\hat{\alpha} - \alpha|$  (which is labeled the Actual Bias). This trace also has a computed value of the Cramer Rao Lower Bound (CRLB) for the difference between the estimated and the actual value. The final trace is of  $|\hat{\beta} - \beta|$  (which is labeled the Actual Bias) with the CRLB shown again.

**Figure 2.** Application of MLE method to Fiber without bend.



**Note:** this figure shows three traces. The first trace shows the returned signal at the OTDR. The second trace shows the result of computing  $|\hat{\alpha} - \alpha|$  (which is labeled the Actual Bias). This trace also has a computed value of the Cramer Rao Lower Bound (CRLB) for the difference between the estimate and the actual value. The final trace is of  $|\hat{\beta} - \beta|$  (which is labeled the Actual Bias) with the CRLB shown again.

**Figure 3.** Application of MLE method to Fiber with bend.

main  $[m, m + N - 1]$  overlaps with a point at which the fuse block has bent the fiber. As  $m$  is increased by  $N$ , the segment no longer includes the discontinuity and there is a sudden decline in the value of  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$ .

If the SNR of  $x(n)$  is high, or the fuse block bends the fiber sharply, the  $var(\theta)$  caused by noise will not have a significant effect on the ability of the algorithm to locate the leakage point.

In the contrast, when SNR of  $x(n)$  is low, or the fuse block is not able to bend the fiber sharply, the changes in  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  are small, and this can mean that a leakage point is drowned out the signal in the estimation error. In this case, the identification and location of leaking point become difficult.

For extremely small changes in the fiber, the MLE method presented may still not be powerful enough to bring

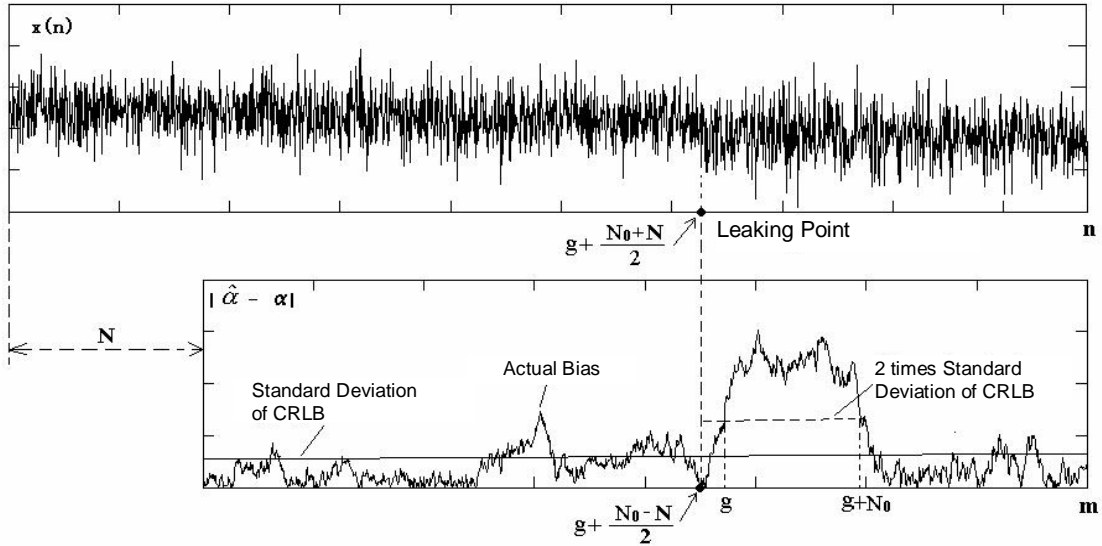
out the signal discontinuity. In this situation an additional step can be taken of using the Chebyshev's Inequality.

The Chebyshev's Inequality states that if the random variable  $t$  has a finite mean  $\mu$  and finite variance  $\delta^2$ , then for any  $d \geq 1$ ,  $p(|t - \mu| \geq d * \delta) \leq 1/d^2$ .

For every  $m$ ,  $\delta$  can be calculated using  $I(\theta)$ . The  $\theta$  will be known after several initial estimations (e.g.  $m: 0 \sim 5$ ). The probability of bias,  $\{(|\hat{\alpha} - \alpha| \text{ or } |\hat{\beta} - \beta|) > \delta_a * d \text{ or } \delta_\beta * d\}$ , is less than  $1/d^2$ . The standard deviation of  $\alpha$  and  $\beta$  is define as  $(\delta_\alpha, \delta_\beta)$ . We see that  $(\delta_\alpha, \delta_\beta)$  for every  $m$  can be calculated with  $I(\theta)$ .

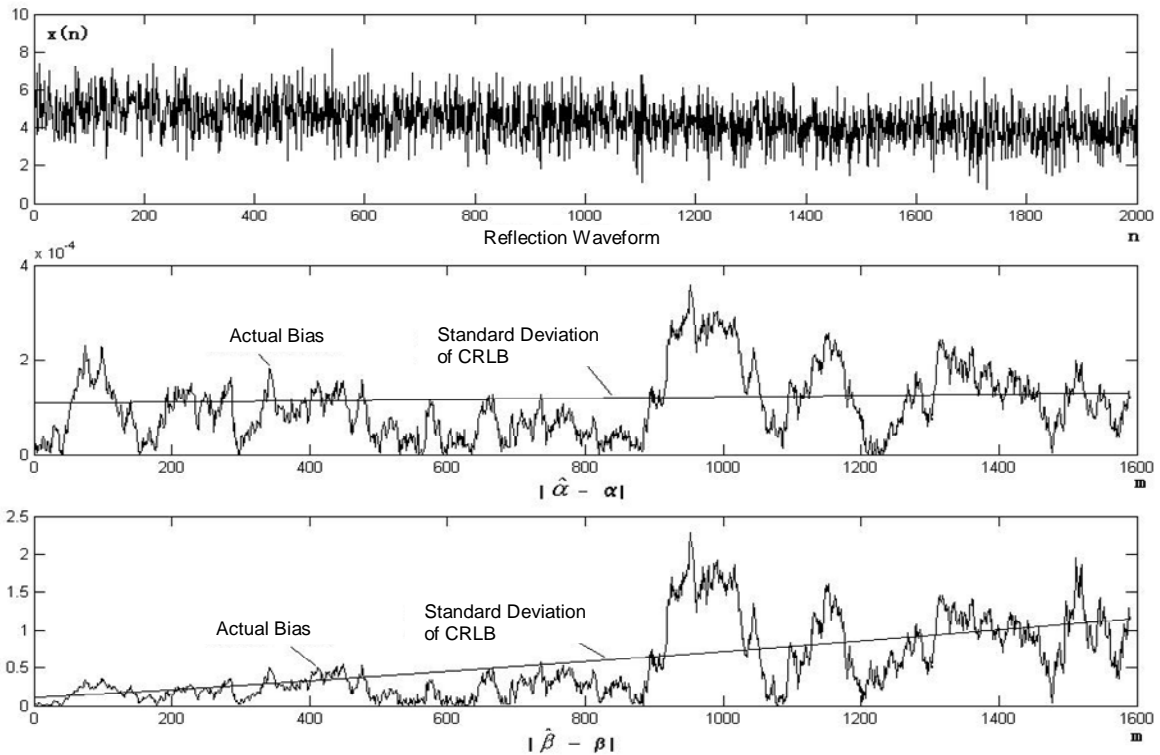
For example, if we let  $d = 2$ , a criteria can be developed as follows:

If  $|\hat{\alpha} - \alpha| > 2 * \delta_\alpha$  and  $|\hat{\beta} - \beta| > 2 * \delta_\beta$  on a segment  $[g, g + N_0]$  on  $m$  axis, in which  $N_0 > N/2$ , there must be a leaking point, whose position is  $g + (N + N_0)/2$  on the  $n$  axis.  $N_0$ ,



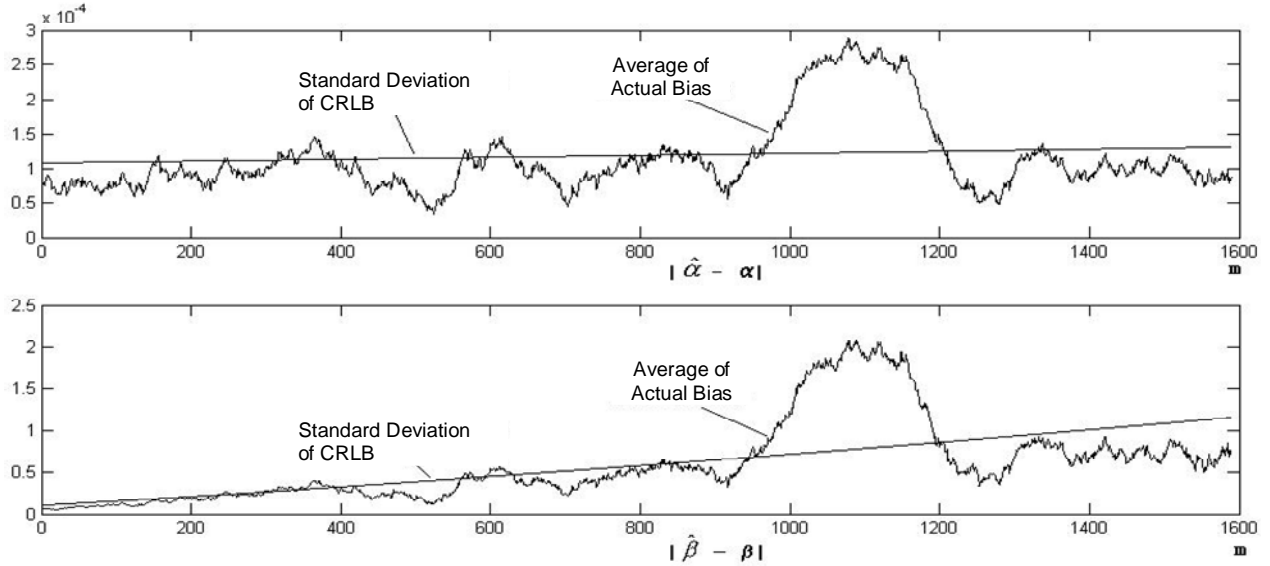
**Note:** An example of  $|\hat{\alpha} - \alpha| > 2 * \delta_a$  from a realization of  $x(n)$ . This example shows that, using the approach suggested above, even small discontinuities in the backscattered signal can be drawn out.

**Figure 4.** Application of Chebyshev's Inequality step to MLE method for a very slight bend in the Fiber.



**Note:** in this example  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  from a realization of  $x(n)$  with weak discontinuity shows a number of potential leakage points. The low SNR gives an unpredictable result.

**Figure 5.** Example of MLE method with out Chebyshev's Inequality step for signal with high SNR.



**Note:** The effect of averaging of  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  from 10 realizations of  $x(n)$  with weak discontinuity using the data for Fig. 5. Clearly the bend in the fiber becomes very evident and the other noise peaks are suppressed.

**Figure 6.** Example of MLE method followed by Chebyshev's Inequality step for signal with high SNR.

rather than  $N$ , is chosen in order to improve the sensitivity.

The method can be summarized in the following bullets:

(1) If  $|\hat{\alpha} - \alpha| < 2 * \delta_\alpha$  and  $|\hat{\beta} - \beta| < 2 * \delta_\beta$  for any  $m$ , there is no leaking point.

(2) If  $|\hat{\alpha} - \alpha| > 2 * \delta_\alpha$  or  $|\hat{\beta} - \beta| > 2 * \delta_\beta$  for some  $m$ , the estimate doesn't meet condition 1, however there could still be a leakage point near  $m$ . In this case, averages of  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  from 10 realizations of  $x(n)$  are calculated. If the  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  after averaging does meet condition 1 the position of leaking point is  $g + (N + N_0) / 2$ .

The method with the Chebyshev's Inequality step added to the MLE method is shown in Figures 5 and 6 also. Figure 5 shows an example of a situation where the SNR is low and the MLE cannot uniquely identify the location of leakage. Following the steps of averaging the traces and using the Chebyshev's Inequality with ten traces it can be observed that the location of the leakage is easily identified.

If  $\sigma^2$  (the noise power) is small and the effect of the leakage point is easily seen and  $|\hat{\alpha} - \alpha|$  and  $|\hat{\beta} - \beta|$  will meet condition 1. On the other hand, if the SNR is low and the fuse block is producing a weakly backscattered signal, condition 2 must be applied.

## 4. Discussion

### 4.1. Sensitivity and Location Accuracy

The method, which we have presented above, is sensitive

to discontinuities in an exponential function, even if the discontinuity in the recorded exponential signal is very small. Because of the variance of  $(\hat{\alpha}, \hat{\beta})$  is affected by noise, the accuracy of identifying the leakage point position can also be affected. There are several ways to solve this problem: using larger  $N$ , using average of more realizations of  $x(n)$ . However, there is tradeoff between accuracy and calculation time.

### 4.2. Variance Increases with $m$

When  $m$  increases, the variance of  $(\hat{\alpha}, \hat{\beta})$  also increase theoretically, which will influence the accuracy of leakage point position. Fortunately, the variance increases slowly when compared with an increase in  $m$  (Figures 1 and 2). An attempt to adjust  $N$  so as to get the same variance for every  $m$  is difficult.

### 4.3. Distinguish Leaking Point from Other Singular Point

The methods we have presented here clearly identify the location of backscatter on the fiber. However, a fuse block responding to oil leakage may not be the only source of backscatter. There may be connectors and splices in a long optical fiber, which will also create back reflections. This will mean that there will be discontinuity in the trace  $x(n)$ . The connector or splice points will have to be distinguished from oil leakage points according to their characteristics. A simple solution to this problem may be to look for changes in the backscatter over time, however a full investigation into this issue is left as a future work.

## 5. Conclusions

A method that uses statistical parameter estimation has been presented and shown to be effective in locating leakage points on an oil pipeline instrumented with a fiber optic sensor. The parameter estimating technique senses irregularities in the backscattered-reflection waveform, and this allows the identification of the leakage points. This method is mainly applicable in noisy situations where the reflected waveform contains a signal from a weak reflector due to a small leak or the initial response of a fuse block to the presence of oil. In order for this method to be useful, it must be able to sense a very slight bend created by a fuse block. The fuse block's force is to some extent proportional to the level of exposure to oil and thus our method may identify small leakage prior to a major rupture of the pipeline.

## References

Paranjape, R., Liu, N., Rumble, C. and Hara, E. (2002). Distributed fiber optic system for oil pipeline leakage detection, Applica-

- tions of photonic technology 5, R. Lessard, G. Lampropoulos, G. Schinin (Eds.), *Proc. of the SPIE*, 4833, pp. 206-213.
- Chen, H., Tu, Y. and Luo, T. (1998). A method for oil pipeline leak detection based on distributed fiber optic technology, *SPIE*, 3555, 77-82.
- TFOM (TFP2 FiberMaster Operator Manual). (2000). Tektronix Inc., Beaverton, Oregon, USA.
- Liu, S., Li, L., Cui, J. and Li, T. (2000). Acoustic emission detection of underground pipeline leakage, *15th World Conference on Non-Destructive Testing*, Rome, Italy, October 15-21.
- Beushausen, R., Tornow, S., Borchers, H., Murphy, K. and Zhang, J. (2004). Transient leak detection in crude oil pipelines, in *Proc. of IPC 2004, International Pipeline Conference*, Calgary, Alberta, Canada, October 4-8.
- Kay, S. (1993). *Fundamentals of Statistical Signal Processing Estimation Theory*, Prentice-Hall Inc., Upper Saddle River, New Jersey, USA.
- Demirli, R. and Saniie, J. (2001). Model-Based Estimation of Ultrasonic Echoes Part I: Analysis and Algorithms. *IEEE Trans. Ultrason., Ferroel. Freq. Control*, 48(3), 787-811.
- Hogg, R.V. and Tanis, E.A. (1997). *Probability and statistical inference*, New York Macmillan.