

Entropy Filter for Anomaly Detection with Eddy Current Remote Field Sensors

Davide Spinello

Department of Mechanical Engineering
University of Ottawa
Ottawa, Ontario, K1N 6N5 Canada
Email: dspinell@uottawa.ca

Wail Gueaieb

School of Electrical Engineering
and Computer Science
University of Ottawa
Ottawa, Ontario, K1N 6N5 Canada
Email: wgueaieb@site.uottawa.ca

Roderick Lee

InvoDane Engineering
30 Mesmill Rd., Unit 2
Toronto, Ontario, M3B 2T6 Canada
Email: rlee@invodane.com

Abstract—We consider the problem of extracting a specific feature from a noisy signal generated by a multi-channels Remote Field Eddy Current Sensor. The sensor is installed on a mobile robot whose mission is the detection of anomalous regions in metal pipelines. Given the presence of noise that characterizes the data series, anomaly signals could be masked by noise and therefore difficult to identify in some instances. In order to enhance signal peaks that potentially identify anomalies we consider an entropy filter built on a posteriori probability density functions associated with data series. Thresholds based on the Neyman-Pearson criterion for hypothesis testing are derived. The algorithmic tool is applied to the analysis of data from a portion of pipeline with a set of anomalies introduced at predetermined locations. Critical areas identifying anomalies capture the set of damaged locations, demonstrating the effectiveness of the filter in detection with Remote Field Eddy Current Sensor.

I. INTRODUCTION

The world is witnessing an increasing demand on water and energy, mainly delivered through large networks of distribution pipelines. Disrupting the flow in these pipelines for maintenance or repairs may lead to a higher stress on other parts of the network and even to economic losses and disturbances in the consumer market in case it yields a shortage of supply. This is in addition to the severe short- and long-term effects an oil spill may have on the environment, for instance. With such a very limited downtime margin, millions of miles of pipelines already operational worldwide, and thousands more either planned or under construction [1], automating their inspection process is becoming a pressing and urgent necessity. This manuscript contributes to the efforts of advancing the state of the art of gas pipeline inspection by fusing Eddy Current Remote Field sensing technology and an entropy filter.

Sensors used in pipeline inspections are generally referred to as conventional pipeline inspection systems (PIGs) or inline-inspection (ILI) tools. Several types of PIGs have been proposed in the literature [2]. Acoustic and ultrasonic sensors were tested in [3]–[6]. A rotating optical geometry and infrared sensors were adopted in [7] and [8], respectively. Despite their promising performance, they remain limited by the shortcomings naturally inherited from image-processing techniques, in general, such as relatively high cost and computational complexity, and dependence on the lighting condition inside

the pipe.

Magnetic sensors, such as magnetic flux leakage [9], Eddy current [10], magnetic particle, and Hall-effect sensors, have been widely adopted in commercial ILI tools, taking advantage of the abundance of metallic utility pipes. This type of sensors are capable of detecting, both, internal and external defects, and are therefore suitable for nondestructive inspection methods [11], [12]. A comparative study between magnetic- and ultrasonic-based inspection techniques is reported in [13].

Magnetic PIGs magnetize the pipe as they travel through it. A magnetic field-related signal is then captured by an array of transducers uniformly distributed around the circumference inside the pipe wall. A difference in the transmitted and received magnetic-dependent signals usually indicates the existence of a flaw within the vicinity of that point [14], [15]. The flaw can be, for instance, due to corrosion, weld, crack, fatigue, or deformation. When noise levels are high, the collected sensory data requires filtering and possibly other signal processing techniques. For instance, artificial neural networks were used in [16], [17] for this purpose. Wavelet-based filtering approaches were adopted in [18]–[20]. The authors in [21] proposed and compared three machine learning approaches, namely, support vector machine, kernelized principal component analysis, and kernelized partial least squares.

In this paper, we investigate the problem of extracting features from data generated by a multi-channel Remote Field Eddy Current Sensor. Sensory data is characterized by low signal to noise ratio, and therefore relevant signals associated with features to be identified may be buried in noise. Several works in the literature have developed the idea of using Shannon entropy [22] to filter noisy signals in order to discriminate structures from background noise [23]–[27]. We extend such idea to the class of data generated by Remote Field Eddy Current Sensors and we delineate a procedure to characterize sensor noise and anomalies signals within the Neyman-Pearson [28] decision making framework. The effectiveness of the algorithm is assessed by detecting known critical regions on a pipeline with lab data. The specific data set used for illustration purposes does not limit the applicability of the analysis, which extends to data sets with comparable characteristics in terms of signal to noise ratio.

The rest of the paper is organized as follows: in Section II we characterize sensory data in terms of related discrete probability density functions, which is the necessary preamble to apply the entropy filter. In Section III we briefly describe the entropy filter, and we determine thresholds for hypothesis testing within the Neyman-Pearson decision making framework with statistical parameters intrinsically related to noise and to anomalies to be detected. The algorithm is illustrated in Section IV by its effectiveness of detecting critical regions associated with known anomalies. Section V is left for conclusions and ongoing and future work.

II. STOCHASTIC CHARACTERIZATION OF THE SENSOR APPARATUS

We consider the problem of extracting signals embedded in noisy data series characterized by low signal to noise ratio. The signal reveals a specific feature to be detected, and therefore its extraction defines the task associated with the system. Specifically, we refer to data series generated by a Remote Field Eddy Current Sensor [29] mounted on a mobile robotic platform which performs non-destructive inspection of gas pipelines. Sensor measurements are correlated to the thickness of the pipeline. Multiple channels distributed along the circumference of the pipe allow for discretized coverage of the surface.

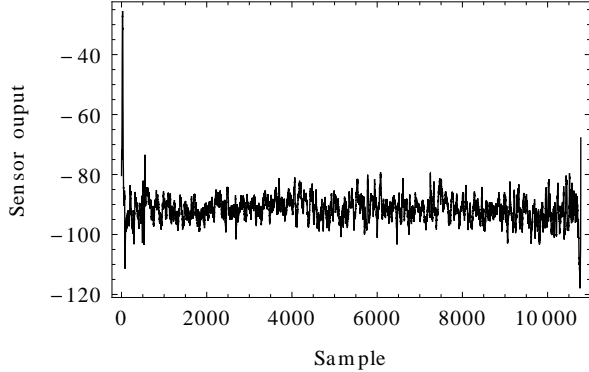


Fig. 1. Time series of the output from a channel of the Eddy Current Sensor.

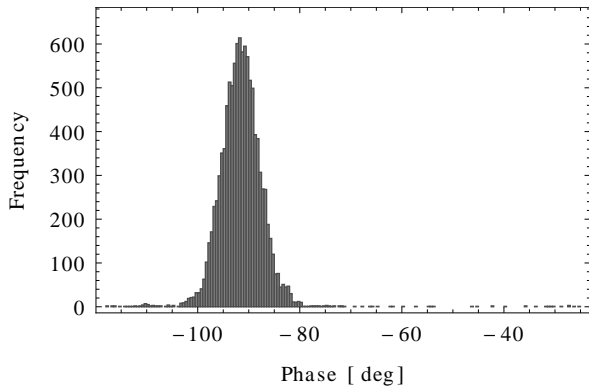


Fig. 2. Histogram of the sensor data in Fig. 1.

An example of a single channel data series is given in Fig. 1 in which it is shown the output of the sensor. The data series refers to a portion of a pipeline with no significant defects, except for the extremities in which peaks associated with joints appear. By considering the data set to be the sample space one can build the corresponding discrete distribution by normalizing the histogram of the data series. For the series in Fig. 1 the histogram is plotted in Fig. 2, showing the expected white noise characteristic behaviour. For a single channel labelled with the integer i the discrete density function is built by dividing the sample space into N_s subintervals $\{\phi_{ij}, \phi_{i,j+1}\}_{j=1}^{N_s}$ and by counting the number of occurrences of elements of ϕ_i normalized with respect to the cardinality of ϕ_i and the size of the bin $\Delta\phi_{ij} = \phi_{i,j+1} - \phi_{ij}$. For the data series in Fig. 1 and $N_s = 100$ the discrete probability density function is plotted in Fig 3. As expected the curve is the normalized histogram of Fig. 2.

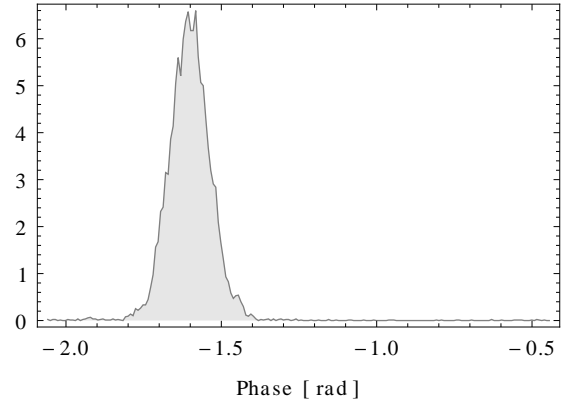


Fig. 3. Discrete probability density function for the data set in Fig. 1.

III. THE ENTROPY FILTER

A. Description of the filter

In the original framework delineated by Shannon [22], entropy is a measure of the quantity of information [30]. Given a scalar data set $\{\phi_{ij}\}_{j=1}^{N_i}$, as for example the series plotted in Fig. 1, let $p_i(\phi)$ be the related discrete probability density function obtained as explained in Section II. For $\phi \in \{\phi_{ij}, \phi_{i,j+1}\}$ the mass probability $P(\phi = \phi_{ij})$, that is the probability of a sample ϕ taking the value ϕ_{ij} is calculated as

$$P(\phi = \phi_{ij}) = p_i(\phi_{ij})\Delta\phi_{ij} \quad (1)$$

For every sample ϕ_{ij} the output of the entropy filter is

$$H(\phi_{ij}, \ell) = - \sum_{k=j-\ell+1}^j P(\phi = \phi_{ik}) \ln P(\phi = \phi_{ik}) \quad (2)$$

where the integer ℓ defines the size of the window used to compute the entropy.

Filtered data of Fig. 1 is plotted in Fig. 4, where the computation in (2) is performed with $\ell = 200$. Since the

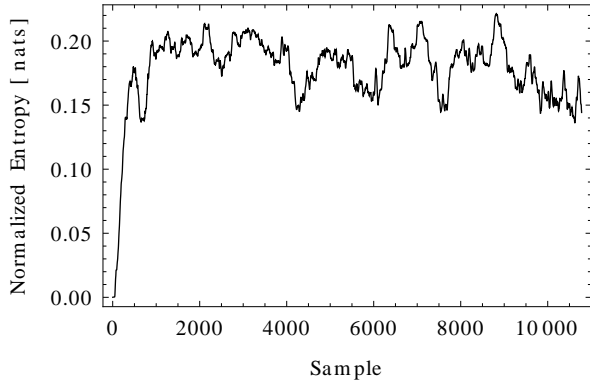


Fig. 4. Filtered data of Fig. 1 with $\ell = 200$, normalized with respect to ℓ .

natural logarithm was employed the unit of entropy is nat. The mean of the normalized entropy

$$\mu_i^H = \frac{1}{N} \sum_{j=1}^N \frac{H(\phi_{ij}, \ell)}{\ell} \quad (3)$$

allows for the characterization of the sensor noise as evinced from the plot in Fig. 5, where it is evident that such value is nearly independent of ℓ and it represents the value of the Shannon entropy of the data set

$$H = - \sum_{j=1}^N P(\phi = \phi_{ij}) \ln P(\phi = \phi_{ij}) \quad (4)$$

which for the series in Fig. 1 equals 0.18 nats.

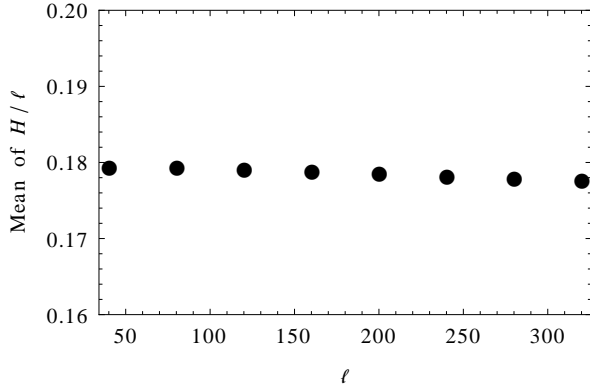


Fig. 5. Mean of the normalized entropy versus ℓ for $\ell \in [40, 320]$.

B. Determination of the threshold

The anomaly detection with the entropy filter is completed with a threshold to discriminate the noise from relevant peaks. The threshold is determined by using the Neyman-Pearson criterion which assigns the threshold value corresponding to the maximum detection probability achievable for a given false detection rate [23], [28].

We consider a simple binary hypothesis test, and define \mathcal{H}_0 and \mathcal{H}_1 to be respectively associated with anomalies and with noise. In probabilistic terms, we introduce the density

functions $f_0(H)$ and $f_1(H)$ defined on the sample space of the output of the entropy filter applied to a data series $\{\phi_{ij}\}$. Let ρ_1 be the decision region for \mathcal{H}_1 , that in the present framework has to be determined given a constraint on the false detection rate. The probability of false detection and the probability of detection are related through ρ_1 as

$$P_F = \int_{\rho_1} f_0(H) dH \quad (5a)$$

$$P_D = \int_{\rho_1} f_1(H) dH \quad (5b)$$

Given P_F , and assuming f_0 and f_1 to be continuous, the detection region ρ_1 is the set of

$$H : \Lambda(H) = \frac{f_1(H)}{f_0(H)} > \eta \quad (6)$$

where $\eta \geq 0$ is determined by solving (5a) with assigned P_F .

In order to characterize the density f_0 we consider the entropy distribution associated with data acquired on a portion of a pipeline with no relevant anomalies. The data in Fig. 2 is a good candidate upon discarding the first portion which is related to a weld. For $\ell = 200$ the plot of the normalized entropy, H/ℓ , of the series in Fig. 1 with borders discarded is shown in Fig. 6. The application of the Neyman-Pearson criterion in the form outlined above is based on the assumption that the probability density functions for the two hypotheses are continuous. For the entropy data set chosen to characterize the sensor noise, the plot in Fig. 7 shows the discrete probability density function (dots) obtained from the normalized histogram. The continuous line in the same plot is the normal probability density function $\mathcal{N}(\mu_0, \sigma_0)$ with

$$\mu_0 = \frac{1}{N} \sum_{j=1}^N \frac{H(\phi_{ij}, \ell)}{\ell}, \quad \sigma_0^2 = \frac{1}{N} \sum_{j=1}^N \left(\frac{H(\phi_{ij}, \ell)}{\ell} - \mu_0 \right)^2 \quad (7)$$

In view of the plot in Fig. 7 we consider the Gaussian density to be an acceptable approximation of the actual discrete density, and therefore assume $f_0 = \mathcal{N}(\mu_0, \sigma_0)$ with $\mu_0 = 0.185$ nats and $\sigma_0 = 0.0179$ nats²

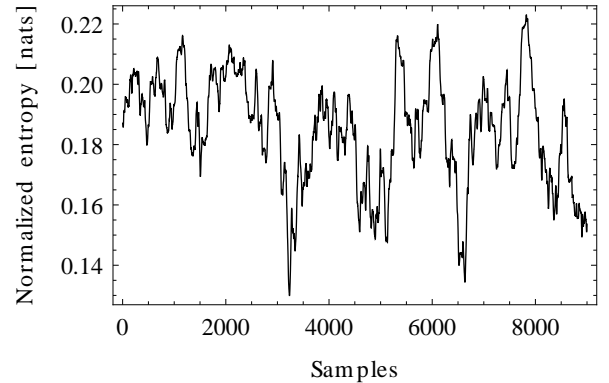


Fig. 6. Normalized entropy of the data set in Fig. 1 with borders discarded in order to derive the probability density function characterizing the entropy of the sensor noise.

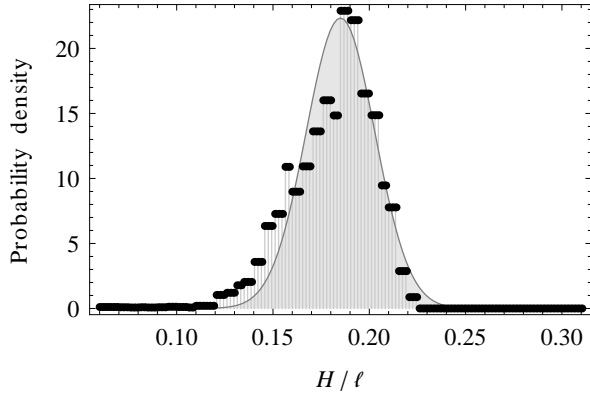


Fig. 7. Discrete (dots) and Gaussian (continuous line) probability density functions for the entropy data set associated with sensor noise.

The characterization of the density f_1 follows the same steps, except that in this case it is considered a data set related to sensor measurements from a portion of pipelines in which there are known anomalies. An example of the filtered data from a single channel output across a region involving several anomalies is given in Fig. 8. We assume that f_1 is normal with estimated parameters $\mu_1 = 0.196$ nats and $\sigma_1 = 0.0738$ nats². The discrete distribution from the normalized histogram of the anomaly data and the approximating Gaussian density are plotted in Fig. 9, validating the Gaussian model for the data. Parameters in f_0 and f_1 are summarized in Table I.

TABLE I
PARAMETERS FOR THE PROBABILITY DENSITY FUNCTIONS f_0 AND f_1 IN HYPOTHESIS TESTING.

μ_0	μ_1	σ_0	σ_1
0.185	0.196	0.0179	0.0738

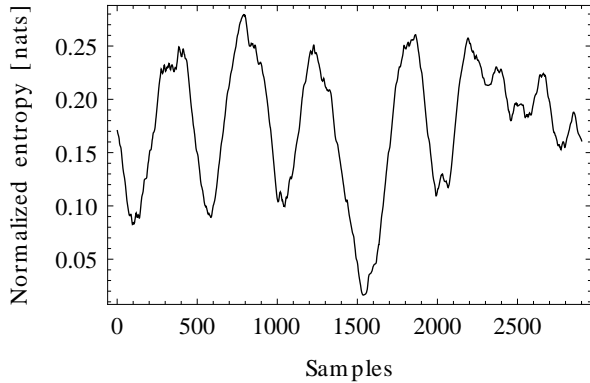


Fig. 8. Normalized entropy of a single channel output across a region containing known anomalies.

Based on the assumptions for the densities f_0 and f_1 , the likelihood ratio for hypothesis testing can be explicitly be

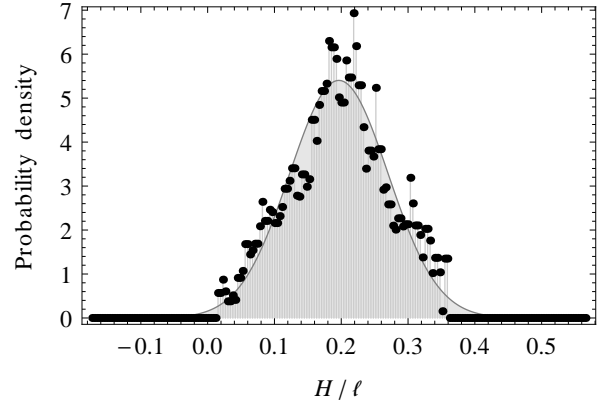


Fig. 9. Discrete (dots) and Gaussian (continuous line) probability density functions for the entropy data set associated with known anomalies.

written as

$$\Lambda(H) = \exp \left(-\frac{1}{2} \left(\frac{H - \mu_1}{\sigma_1} \right)^2 + \left(\frac{1}{2} \frac{H - \mu_0}{\sigma_0} \right)^2 \right) > \eta \frac{\sigma_1}{\sigma_0} \quad (8)$$

Taking the \ln of both sides we obtain a quadratic inequality in H that defines the detection region (critical region) as a function of η

$$\left(\frac{H - \mu_0}{\sigma_0} \right)^2 - \left(\frac{H - \mu_1}{\sigma_1} \right)^2 - 2 \ln \eta \frac{\sigma_1}{\sigma_0} > 0 \quad (9)$$

Let

$$a = \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \quad (10a)$$

$$b = -2 \left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2} \right) \quad (10b)$$

$$c = \frac{\mu_0^2}{\sigma_0^2} - \frac{\mu_1^2}{\sigma_1^2} - 2 \ln \eta \frac{\sigma_1}{\sigma_0} \quad (10c)$$

so that the left-hand side of (9) can be rewritten as $aH^2 + bH + c$. Let $H^-(\eta)$ and $H^+(\eta)$ be the roots of this quadratic function of H , with $H^- \leq H^+$. The detector in (9) therefore dictates the following detection region

$$\rho_1(\eta) = \begin{cases} H < H^-(\eta) \cup H > H^+(\eta) & \text{if } a > 0 \\ H^-(\eta) < H < H^+(\eta) & \text{if } a < 0 \end{cases} \quad (11)$$

The value of the threshold η is found by numerically solving (5a) with the bisection method with assigned P_F . For discrete values of $P_F \in [0.02, 0.8]$ the values of H^- and H^+ are given in Fig. 10, along with the least square fitting second order polynomials:

$$H^- = 0.144P_F^2 + 0.0930P_F - 0.0623 \quad (12a)$$

$$H^+ = 0.225P_F^2 - 0.0930P_F + 0.0623 \quad (12b)$$

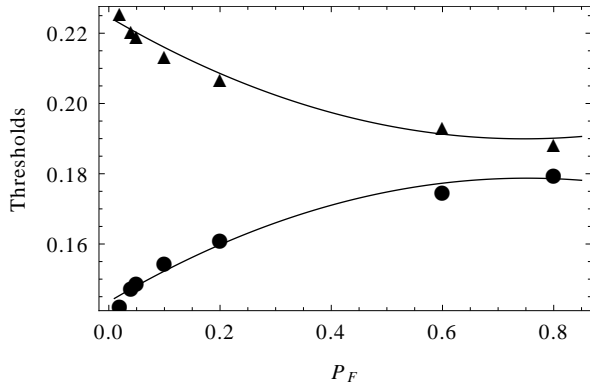


Fig. 10. Thresholds H^- (dots) and H^+ (triangles) and relative fourth order fitting curves in (12) versus P_F .

IV. ILLUSTRATION OF ANOMALY DETECTION

In order to illustrate the application of the entropy filter explained in Section III we apply it to a data set acquired in a portion of a pipeline with known anomalies introduced in a controlled way. For a false alarm rate $P_F = 5\%$ the thresholds are computed from (5a) to be

$$H^- = 0.149, \quad H^+ = 0.219 \quad (13)$$

which correspond to a probability of detection $P_D = 64\%$. Filtered data is considered to belong to a critical region identifying a potential anomaly if $H(\phi_{ij})/\ell \leq H^-$. Gray circles represent approximate locations (centres) of known anomalies which are well captured by the entropy filter (dark regions). The density plot in Fig. 11 is obtained by linearly interpolating data from contiguous channels to reconstruct a two dimensional profile whose support is the surface defining the pipeline. The entropy filter clearly captures the critical regions identified by the dots. This result suggests that the algorithmic tool can be effectively used to post-process data from Eddy Current Sensors characterized by low signal to noise ratio.

For a single channel in the region in Fig. 11 the normalized entropy data is plotted in Fig. 12. The dashed line represents the threshold: values below the threshold are identified as belonging to critical regions.

V. CONCLUSION AND FUTURE WORK

We have applied the entropy filter to the detection of anomalies with a multi-channel Remote Field Eddy Current Sensor. Sensory data and entropy filter outputs are characterized in a probabilistic framework, which allows for thresholding based on Neyman-Pearson decision making criterion. Critical regions for a binary test in a Gaussian framework with the hypotheses characterized by different moments are derived. The effectiveness of the algorithm is illustrated by detecting several anomalies from a multi-channel data set acquired from a portion of a pipeline with damages introduced in a controlled way.

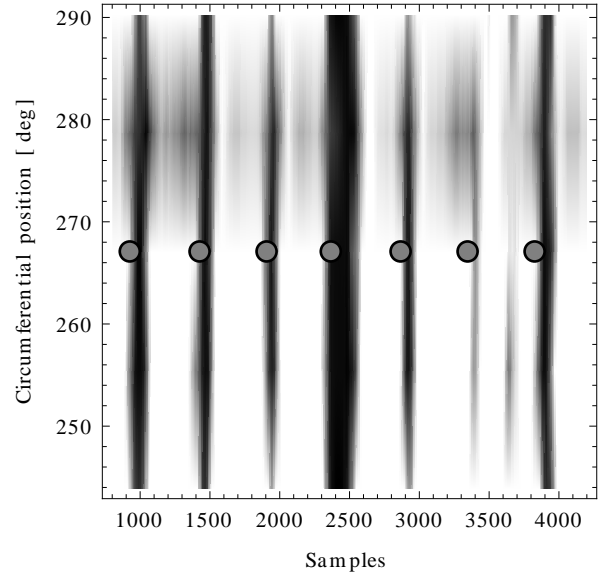


Fig. 11. Application of the entropy filter to a multichannel data set from a portion of a pipeline with known anomalies (gray circles). Dark regions are critical with respect to the entropy filter.

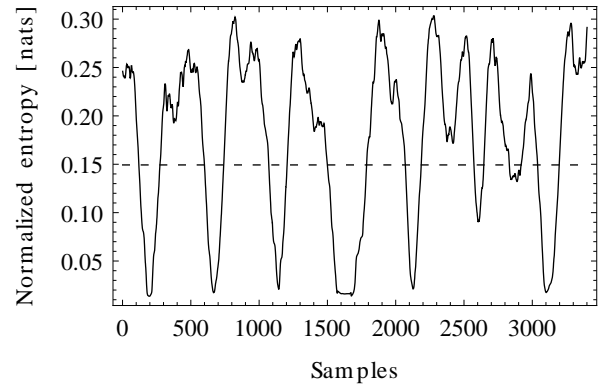


Fig. 12. Normalized entropy data (continuous line) from a single channel across a region with known anomalies, and threshold value used with the entropy filter (dashed line).

Ongoing work include the introduction of multi-channel cross correlation. Characterization of different types of anomalies with the introduction of multiple hypothesis testing is currently being modelled and investigated. Future work will include refined description of a-posteriori probability density functions which generalize the Gaussian framework, and the investigation of the eventual increase of accuracy versus increase of complexity with respect to the Gaussian model.

ACKNOWLEDGMENT

The authors would like to thank Precarn and NSERC whose financial support allowed to start the collaboration between InvoDane Engineering and the University of Ottawa.

The authors would also like to thank the Northeast Gas Association/NYSEARCH for their continued support of this work.

REFERENCES

- [1] R. Tubb, "2011 worldwide pipeline construction report," *Pipeline & Gas Journal*, vol. 238, no. 1, pp. 18–38, 2011.
- [2] W. Y. Du and S. W. Yelich, "Post-earthquake pipeline leak detection technologies," in *Smart Sensors and Sensing Technology*. Springer-Verlag, 2008, ch. 1, pp. 265–283.
- [3] A. Ferrari, "Modelling approaches to acoustic cavitation in transmission pipelines," *International Journal of Heat and Mass Transfer*, vol. 53, no. 19, pp. 4193–203, 2010.
- [4] J. C. Martins and P. S. Jr., "Assessment of the performance of acoustic and mass balance methods for leak detection in pipelines for transporting liquids," *Journal of Fluids Engineering, Transactions of the ASME*, vol. 132, no. 1, pp. 0114011–0114018, 2010.
- [5] D. Bo, Z. Huiping, S. Sha, and T. Jian, "Research on ultrasonic inspection of pipeline corrosion," in *IEEE International Conference on Control and Automation*, 2007, pp. 2572–5.
- [6] B. Dai, S. Sheng, X. Tian, Z. Yang, and Z. Xie, "Research on multi-probe ultrasonic automated in-line inspection system of pipeline corrosion," in *IEEE International Conference on Mechatronics and Automation*, 2007, pp. 3105–10.
- [7] C. W. Frey, "Rotating optical geometry sensor for fresh water pipe inspection," in *IEEE Sensors*, 2008, pp. 337–340.
- [8] K. Kim, K. Kim, H. Jung, and H. Chang, "Measurement of defect thickness of the wall thinning defect pipes by lock-in infrared thermography technique," in *Fourth International Conference on Experimental Mechanics*, vol. 7522, 2009.
- [9] R. K. Amineh, N. K. Nikolova, J. P. Reilly, and J. R. Hare, "Characterization of surface-breaking cracks using one tangential component of magnetic leakage field measurements," *IEEE Transactions on Magnetics*, vol. 44, no. 4, pp. 516–524, 2008.
- [10] P. Laursen, G. C. Vradis, and C. Swiech, "First robotic device to inspect unpiggable gas transmission pipeline," *Pipeline & Gas Journal*, vol. 236, no. 11, 2009.
- [11] W. Du, H. Nguyen, A. Dutt, and K. Scallion, "Design of a GMR sensor array system for robotic pipe inspection," in *IEEE Sensors Conference*, 2010, pp. 2551–2554.
- [12] V. E. Loskutov, A. F. Matvienko, B. V. Patramanskii, and V. E. Shcherbinin, "The magnetic method for in-tube nondestructive testing of gas and oil pipelines: the past and the present," *Russian Journal of Nondestructive Testing*, vol. 42, no. 8, pp. 493–504, 2006.
- [13] H. Goedecke, "Ultrasonic or MFL inspection: Which technology is better for you?" *Pipeline & Gas Journal*, vol. 230, no. 10, pp. 34–41, 2003.
- [14] R. C. Ireland and C. R. Torres, "Finite element modelling of a circumferential magnetiser," *Sensors and Actuators, A: Physical*, vol. 129, no. 1, pp. 197–202, 2006.
- [15] S. O'Connor, L. Clapham, and P. Wild, "Magnetic flux leakage inspection of tailor-welded blanks," *Measurement Science and Technology*, vol. 13, no. 2, pp. 157–162, 2002.
- [16] A. A. Carvalho, J. M. A. Rebello, L. V. S. Sagrilo, C. S. Camerini, and I. V. J. Miranda, "MFL signals and artificial neural networks applied to detection and classification of pipe weld defects," *NDT&E International*, vol. 39, no. 8, pp. 661–7, 2006.
- [17] A. Joshi, L. Udpa, S. Udpa, and A. Tamburrino, "Adaptive wavelets for characterizing magnetic flux leakage signals from pipeline inspection," *IEEE Transactions on Magnetics*, vol. 42, no. 10, pp. 3168–70, 2006.
- [18] W. Han and P. Que, "A modified wavelet transform domain adaptive FIR filtering algorithm for removing the SPN in the MFL data," *Measurement*, vol. 39, no. 7, pp. 621–7, 2006.
- [19] J. Tao, Q. Peiwen, C. Liang, and L. Liang, "Research on a recognition algorithm for offshore-pipeline defects during magnetic-flux inspection," *Russian Journal of Nondestructive Testing*, vol. 41, no. 4, pp. 231–238, 2005.
- [20] R. W. Tucker Jr., S. W. Kerckel, and V. K. Varma, "Characterization of gas pipeline flaws using wavelet analysis," in *Sixth International Conference on Quality Control by Artificial Vision*, 2003, pp. 485–493.
- [21] A. Khodayari-Rostamabad, J. P. Reilly, N. K. Nikolova, J. R. Hare, and S. Pasha, "Machine learning techniques for the analysis of magnetic flux leakage images in pipeline inspection," *IEEE Transactions on Magnetics*, vol. 45, no. 8, pp. 3073–3084, 2009.
- [22] C. E. Shannon, "Prediction and entropy of printed english," *Bell System Technical Journal*, vol. 30, no. 1, pp. 50–64, 1951.
- [23] R. N. McDonough and D. Whalen, A. *Detection of Signals in Noise*. Academic Press, 1995.
- [24] Y. Li, Y. Dong, and G. Sheng, "A signum minimum entropy filter for irregularity detection in optical yarn signals," *Measurement Science & Technology*, vol. 21, no. 3, MAR 2010.
- [25] A. Sheinker, N. Salomonski, B. Ginzburg, L. Frumkis, and B.-Z. Kaplan, "Magnetic anomaly detection using entropy filter," *Measurement Science and Technology*, vol. 19, no. 4, p. 045205, 2008. [Online]. Available: <http://stacks.iop.org/0957-0233/19/i=4/a=045205>
- [26] J. Kaplan and R. Howitt, "Estimating nonpoint source pollution: An application of a sequential entropy filter," *Water Resources Research*, vol. 38, no. 3, MAR 2002.
- [27] J. Zhang, L. Cai, and H. Wang, "Minimum entropy filtering for networked control systems via information theoretic learning approach," in *Proceedings of the 2010 International Conference on Modelling, Identification and Control*, Okayama, Japan, July 17–19 2010, pp. 774–778.
- [28] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London Series A-Containing Papers of a Mathematical or Physical Character*, vol. 231, pp. 289–337, MAR 1933.
- [29] D. J. Hagemaier, *Fundamentals of Eddy Current Testing*. American Society for Nondestructive Testing, 1990.
- [30] J. Shao, *Mathematical Statistics*, ser. Springer texts in statistics. New York: Springer Verlag, 1999.