# A Novel Feature Extraction Method for Monitoring (Vehicular) Fuel Storage System Leaks

Fling Tseng[1], Imad H. Makki[2], Dimitar P. Filev[3], and Ratna Babu Chinnam[4]

[1,2,3]*Research and Advanced Engineering, Ford Motor Company, Dearborn, MI 48121, USA*

*ftseng@ford.com*
*imakki@ford.com*
*dfilev@ford.com*

[4]*Department of Industrial and Systems Engineering, Wayne State University, Detroit, MI 48202, USA*

*ratna.chinnam@wayne.edu*

## ABSTRACT

System state determination with incomplete sensory information set proved to be a technically challenging problem. In this paper, authors tackle a problem of this type associated with vehicle fuel storage systems and proposed a novel feature extraction method. Federal and state regulations require fuel storage leak detection mechanism to be conducted periodically and regulate its execution rate and performance to ensure effective emission controls. Being able to robustly determine a fuel storage system's state in terms of its effectiveness of fuel containment is therefore of great importance to all vehicle original equipment manufacturers (OEM). Prevailing practice in the industry utilizes a method relevant to natural vacuum phenomenon and is loosely associated with ideal gas law. Commonly referred to as "Entry Conditions" in in-vehicle monitoring design literature, major noise factors go through stringent pre-monitoring evaluations before monitoring program execution to ensure ideal test conditions. Differences in ambient conditions compounded with varying customer drive cycle patterns present great challenge to existing monitor designs for the purpose of leak detection. In addition, prevailing practices of evaluation in-tank fuel pressure and temperature information are generally conducted with surrogate or estmiated temperature information due to the absence of in-tank temperature sensor. All this calls for an alternative feature calculation and detection method that are less sensitive to known noise factors, can operate with incomplete sensory information yet being able provide similar or improved detection capability. In this paper, we put the main focus on the derivation of a novel method of feature calculation for the purpose of detecting presence of a leak in a fuel storage tank.



## 1. INTRODUCTION

Murvay (Murvay, 2012) studied state-of-the-art development in terms of hardware (including pressure, acoustic, remote and reflective sensing) and software methods for gas leak detections. It was concluded that a hybrid approach to take advantage of cost effective hardware setup (high localization accuracy) with fast improving software methods (real-time detection capability) would be highly recommended. It also suggests that investment in a hybrid approach may be more cost effective in the long term as software capability enhancements may offset the effect of aging hardware, reducing the need for a complete revamp of leak detection setup, something very cost prohibitive. Zhou (Zhou, 2011) proposed a Bayesian Belief Rule Based (BRB) system where subject expert knowledge and real-time information are incorporated to incrementally improve the performance of the system. Such a combination of human knowledge and data driven refinement to the model is suitable to deal with ever increasingly complex real-world problems. Ghazali's work (Ghazali, 2012) focused on instantaneous frequency analysis (IFA), where comparisons between Hilbert transform (HT), Normalized HT (NHT), Direct Quadrature (DQ), Teager Energy Operator (TEO) and Cepstrum performed on pressure transients (opening a valve or stopping a pump) within a live distribution network were conducted. A detection method that includes multiple modeling techniques was proposed by (Mandal, 2012). They apply rough set theory and artificial bee colony (ABC) trained SVM (Support Vector Machine) to carry out classification tasks in two stages and yielded robust performance when compared with PSO (particle swarm optimization) and

EPSO (enhanced particle swarm optimization) based learning methods.

Leak detection mechanism as part of an overall emission control strategy is gaining importance in recent years. As countries are increasingly pledging reduced carbon footprints, one of the main focuses was to incrementally reduce and eventually eliminate allowable fuel vapors escaped to the ambient air. In the United States, ongoing efforts from Environmental Protection Agency (EPA) and California Air Resources Board (CARB) requires consumer vehicle original equipment manufacturers (OEMs) to equip their products with leak detection monitors to improve monitoring capabilities within a given timeframe (State of California Air Resources Board, 2012). In the meantime, on the field performances are under federal and state level regulations subject to audits. If sampled results are deemed unsatisfactory, fines or even voluntary recalls could be imposed. These penalties are undesirable as they undermine an OEM not only financially but could also negatively affect brand image that take years to even decades to recover if such incidents occur.

Emmission related monitors generally reside in the powertrain control module (PCM) therefore contraints such as A. During calculation memory requirement, B. Computational efficiency and C. Compactness of the code often need to be carefully evaluated due to implications in terms of cost and practicality during implementation phase. In this paper, authors focus on describing a fundamentally different way of extracting information from the in-tank pressure signal stream as it is one of most critical parts of an overall redesign of an in-vehicle monitor. More specifically, we will cover a recursive approach to enable monitor design engineers to have access to physically meaningful probability density function (PDF) type of information continuously in the form of a recursively updated histogram or discretized probability density function (DPDF) from normalization performed on an obtained discretized relative frequency function (DRFF). Feature calculations are performed from evaluation of certain specific bin(s) of the DPDF from which decisions can be made about the fuel tank's status with repect to the presence of a leak. Technique descibed in (Syed, 2009) utilizes a low pass filter (LPF) implementation to extract driver (non-conditional / overall) behavioral information for adaptation of an in-vehicle advisory system. When applied to scenarios where possible alternatives do exist, such calculation produces conditional relative frequency (RF) information which is a precursor of probabilistic information. In (Filev, 2011), organization and conditional updates of trip specific RF values enable the creation of a context senstive predictive system. Proposed feature exraction method strictly operates in the probabilistic space. It represents a significant step forward and a crucial enabling element to improve from prevailing pactice of evaluation of pressure signal (or its manipulated version) alone (Wong, 2003 and Jentz, 2013). Our preliminary analysis suggests proposed feature calculation produces meaningful and promising results. The investigation of promising alternative feature calculations as the one described in this paper is an important first step that shall shed more light on how to redesign a leak detection monitor in the future.

The rest of the paper is organized as the following. In section 2, current prevailing practices in the industry will be discussed where most OEM's approach can be understood as solving a classification problem (leak vs no leak) with a single feature commonly derived from in-tank pressure signal. In section 3, the derivation and computation procedure of obtaining a continuous measure of the content of in-tank pressure signal stream in the form of DPDF. In addition, proposed feature calculation from DPDF vector is desribed in detail. Section 4 covers a simple threshold determination based classification process utilizing the feature calculation described in Section 3 and preliminary results are presented. We conconlude current findings and future work in section 5 followed by cited references.

## 2. INDUSTRY PRACTICE FOR VEHICULAR LEAK DETECTION

Prevailing principle of fuel storage leak detection design relies on well-known "Ideal Gas Equation", which states the governing relationship between system pressure and temperature given certain characterizing constants or a lumped product is known or estimated (Wong, 2003 and Jentz, 2013). Determination of the presence of a leak in the fuel storage system is carried out by evaluation of whether expected pressure change is met within certain threshold (2005, McLain). Due to its evaporative nature, gasoline vapor / liquid state transition activities does not warrant the direct use of the ideal gas equation, therefore, monotor specific "Entry Condition" evaluations have to be carried out before monitoring program execution.

After vehicle key-off, when entry conditions are met, the system is then sealed by operation of certain actuators such as valves. In this phase, in-tank pressure signal is kept alive for evaluation against thresholds that are dynamically adjusted to ambient as well as preceding driving conditions that led to the current stop. During all this time, parallel evaluations of certain run time parameters are common to reduce false state determinations and total engine-off battery draw. When it is deemed an effective determination cannot be reached, execution could self-abort without making a determination as to the system's state. A set of built-in counters are required by law to be in place to keep track of how often a monitor runs against scenarios it is required to do so. The ratio of leak / no leak versus total number of successfully full executions are also being tracked. These values are subjected to insepctions of government agencies and OEM's periodically.

Abovementioned leak detection process can be understood as carrying out a classification procedure with a main feature that is commonly derived from pressure sensor information. The goal of these leak detection monitors is to produce a leak indicator value [0, 1] in which 0 represents no leak state and 1 represents presence of a sizable leak. The original pressure value is subjected to futher common signal procesing methods such as signal smoothing, clipping and flipping. Other common modifications may also include multiple scalers associated with ambient / vehicle conditions. After a series of manipulations, comparison is performed with thresholds resulted from calibrations conducted with a sweep of main noise factors spaces. Different from abovementioned commonly used feature, section 3 describes in detail a recursive procedure continously measure in-tank pressure content in the form of DPDF from which feature(s) will be calculated for the purpose of leak detection.

## 3. FEATURE DERIVATION FROM PROBABILITY DENSITY CURVE FOR CLASSIFICATION PURPOSE

The first step in solving a classification problem generally has to do with identification of effective features. Feature extraction serves at least following purposes: 1) Obtaining informative representation of data, 2) Dimensionality reduction, and 3) Reduction in noise and redundancy. Common feature extraction methods can be grouped into the following categories: 1) Time series based features, 2) Statistics based features, 3) Frequency based features, 4) Mixed domain features, and 5) Model based features. For some applications (e.g., vibration analysis), expert and domain knowledge play important roles in guiding the methodology and techniques involved in the feature extraction process. While certain calculation and data transformation may be common (e.g., Fourier Transform for accelerometer sensing signals), such practice may produce signatures associated with certain frequency range. Depending on subject problem of interest, simple data smoothing, deterministic or moving data window scheme or windowed data overlay techniques may be imposed as part of a feature extraction procedure. Details regarding signal and feature selection process are out of the scope of this paper.

Different from common practice, the authors performed data analysis focused on signatures revealed from the probability density function of in-tank pressure changes. This is one of the signals typically kept "alive" during leak detection monitoring phase after the engine has been turned off and the system has been sealed. More specifically, we developed a non-parametric method to continuously extract signatures indicative of the existence of a leak in a presumably sealed setting. The rationale is that change in overall pressure is a consequence of accumulated pressure (rate) changes. We apply procedures to obtain dprobability distribution function in a discretized form from the frequentist's point of view (of relative frequency). This is procedure is implemneted with a low pass filter (LPF or 1st order exponential smoothing). After initialization phase (where a number of initial signal samples have been observed), proposed method gives a continuous output of the DPDF with predefined partitions. Resolution a DPDF is dependent on pre-determined signal range and number of partitions within that range.

Conceptually, proposed implementation is identical to the creation of a histogram with a moving data windown given some continuously incoming data stream; the counting procedure is carried out by a LPF in which its learning rate controls the size of the moving data window. The crisp partitions within specified signal range act as "competing and possible" scenarios or alternatives where we impose a "winner takes all" rule for relative frequency (RF) updates for all partitions involved. Through this updating rule, the increment of the relative frequency occurs only for one partition at a time while the rest of the competing partitions receive negative updates. At any given time, a DPDF is obtained by normalizing most recent DRFF with the summation of its elements. Details regarding this process are described next.

### 3.1. Recursive Estimation of Discretized Relative Frequency Function (DRFF) as Predecessor of Discretized Probability Density Function (DPDF)

3.1. Recursive Estimation of Discretized Relative Frequency Function (DRFF) as Predecessor of Discretized Probability Density Function (DPDF)

From a frequentist's point of view of probability, probability density function (PDF) comes from obtaining a histogram-like vector (of very fine granulaity or partition), namely a DRFF. After a normalization procedure, a DPDF is obtained and the summation of its content should be 1 (sum of total probability of 1). In the simplest case, the first step in obtaining DRFF vetor is to partition a signal's value space into smaller non-overlapping ones. For example, if a signal X takes values from 0 to 10, an example of such a partition would be to define 10 partitions of the signal space that spans the following consecutive intervals or bins: $0 \leq x < 1$, $1 \leq x < 2$, $2 \leq x < 3$ … $9 \leq x \leq 10$. As a result, they represent mutually exclusive scenarios or value range alternatives regarding numeric content of signal X at any given moment. When a specific component of data stream of signal x is being evaluated, only one of the the alternatives will receive the increment in count from the fact current x's value falls into a corresponding region while other alternatives will receive negative updates. From (Syed, 2009), the construction of a count based histogram can be approximated recursively with an exponentially weighted moving average (EWMA) formulation where counts are replaced with relative frequencies (RF). When such implementation is in place, content captured in an interval in DRFF represents a relative frequency value corresponds to the total number of occurances relative its alternatives (other intervals). For example, if α is 0.05 the moving window is approximately 1/0.05 = 20 meaning that at any given moment the DRFF preserves information from the most recent past 20 observations of signal X. The process of obtaining DRFF can be represented by following equation:

$$DRFF_i(t) = (1-\alpha)\cdot DRFF_i(t-1) + \alpha \cdot Flag_i(t) \qquad (1)$$

where 〖 DRFF〗 _i denotes relative frequency of a partition enclosed by its lower and upper limits, α denotes the learning (0≤α≤1), and 〖 Flag〗 _i denotes a binary flag value of 0 or 1 indicating whether current value of X falls into the regrion defined by the i'th region. All partitions of DRFF go through exactly one update during the evaluation of one incoming signal value with Eq (1) and all but one of the partitions will experience a value increment due to the use of "winner takes all" updating rule.

DPDF is obtained by normalization procedure performed on DRFF with following equation:

$$DPDF_i(t) = {DRFF_i(t)}\Big/{\textstyle\sum_{i=1}^{N} DRFF_i(t)} \qquad (2)$$

With equation (2), DPDF is obtained from updated DRFF from which subsequent feature calculation will be performed.

A numerical example comparing LPF vs actual counts based DPDF is shown in the Figure 1.

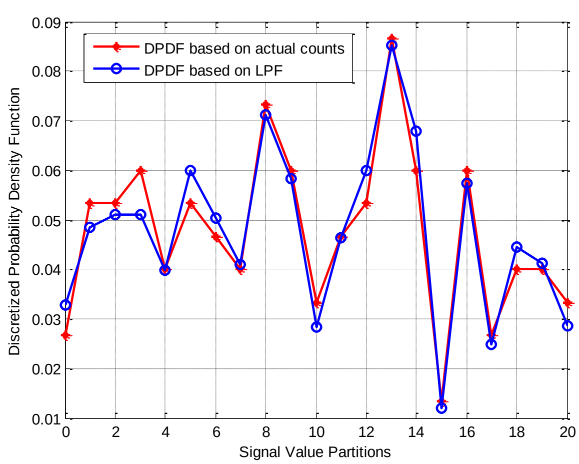


Figure 1: Comparison of recursively obtained DPDF vs Actual Count generated DPDF

In Figure 1, a total of 150 random integers ranging from 0 to 20 were populated.

### 3.2. Extracting Probability Density Content from In-Tank Pressure

#### 3.2.1. Focus of 1st Sealed Stage

During experiments to generate representative datasets, the fuel storage system (fuel tank) goes through a series of state transitions that either expose or seal the system from the atmosphere. The rationale for the transitions contains proprietary information, and hence, will not be discussed here. Our research development focused on the 1st seal stage of all datasets. The reason being that subsequent changes are dependent on information collected during a prior state, making comparison between datasets not realistic. In addition, we identified that the early stage in the 1st sealed phase is much more informative; therefore, we will focus on data collected in the first 300 seconds of each dataset. In addition, we have found that the contrast (separation) between classes reduced for the proposed method very quickly after 300 seconds into the 1st sealed phase.

#### 3.2.2. Pressure Change between Samples vs Pressure Change Rate

The determination that a system has entered its 1st sealed state is conducted by monitoring a set of flags associated with actuators' (valves) states that could be either open or closed. When the system is deemed to have entered its 1st sealed phase, the difference between previous and current in-tank pressures (inch mercury) is calculated continuously. Since our data collection system collects information at a (almost) constant rate of 10 Hz (every 100 milliseconds), pressure change rate in this case is proportional to pressure change between samples, and therefore, we omit the normalization division operation to simplify the calculation.

#### 3.2.3. Obtaining Vector Probability Density Content

First of all, the signal numeric space is defined as 100 equally spaced (0.0003) partitions ranging from -0.015 to 0.015. α is set to be 1/500 or 0.002, which is equivalent of imposing a moving data window containing the last 500 samples as it moves through the data stream. Since the normalization process effectively only scales DRFF through division of its element sum, the overall shape DRFF will be identical to DPDF. A snapshot of DPDF serves as a visual example is shown is Figure 2 according to partitions based on aforementioned definition.

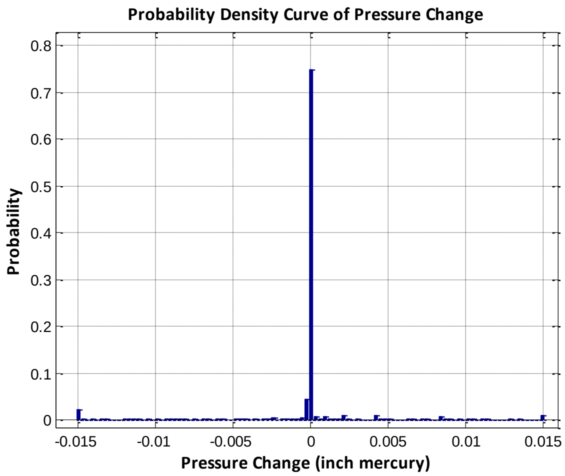


Figure 2: DPDF obtained from normilzation of DRFF covering value range [-0.015, 0.015]. Each partition is of the width of 0.0003.

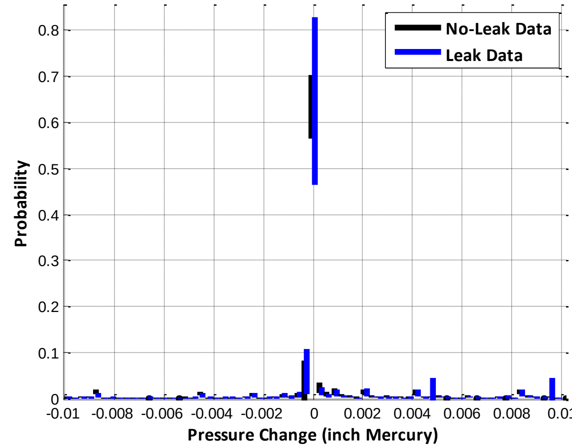


Figure 3: Visualation of DPDF content of Leak (Blue) vs No-Leak (Black) Datasets. For each partition, upper bound / lower bound are obtained with μ+3σ and μ-3σ to visualize the location of the mean value and its spread simultaneously.

### 3.2.4. Identification of Effective Features from DPDF for Classification Purpose

From Figure 2, we noticed an interesting fact that close to 75% of pressure change readings are assigned to the partition centered at 0 for this particular experimental dataset. This is not a coincidence but a result of the sensitivity of the pressure sensor in the existing product.

The next step is to perform the same computational procedures to all datasets. With predefined partitions as described in 3.2.3, resuling DPDF from all datasets are inherently of the same size making it straightforward for us to calculatae the mean and standard deviations separately for two populations: leak vs no leak datasets. As a result, we obtained two sets of means and standard deviations for each partition using following equations:

$$\overline{\mu}_{DPDF_i} = \frac{\sum_{j=1}^{K} DPDF_{i,j}}{K} \tag{3}$$

$$\overline{\sigma}_{DPDF_i} = \sqrt{\frac{\sum_{j=1}^{K} (DPDF_{i,j} - \overline{\mu}_{DPDF_i})^2}{K-1}} \tag{4}$$

i denotes a particular partition, j denotes a dataset and K represents total number of datasets. Since we peforms such calculations for leak and no leak datasets separately, K will take different values if we have an unbalanced datasets where total numbers of leak and no leak datasets are different. From (3) and (4), we obtained population mean and standard deviation of each defined partition. We employ the well-known 6σ definition to show the range spans μ-3σ and μ+3σ for each partition separately for leak (blue line) vs no-leak (black line) datasets as shown in Figure 3.

Selective use of content from DPDF partitions for the purpose of distinguishing between leak and no leak (classification) datasets need to fulfill at least following criteria: 1) Potential content from a partition should exhibit class separation potential and 2) Potential content from a partition should have likelyhood of taking values (non-zero). The first criteria suggests that patterns shown in DPDF should have some class separating capability such as μ_leak ≤ μ_(no-leak) such as the partition around 0.015 as shown in Figure 4. Or, as shown in Figure 3, the partition around zero that the spreads are different between classes, which indicates standard deviations of no-leak datasets may be generally smaller than those of leak datasets. The second criteria has to do with selection of content elements that will take value in the sealed process making sure such content will available to determine the overall system's state in terms of the presence of a leak. This criteria is a basic yet a necessary one to ensure content availability of a partition from DPDF from which subsequent feature calculations are based on.

Following aforementioned criteria, we will mainly focus on the features extracted from DPDF partition near the zero. This is due to the overall low DPDF values of almost all other partitions indicating risks of them to take value on a consistent basis.

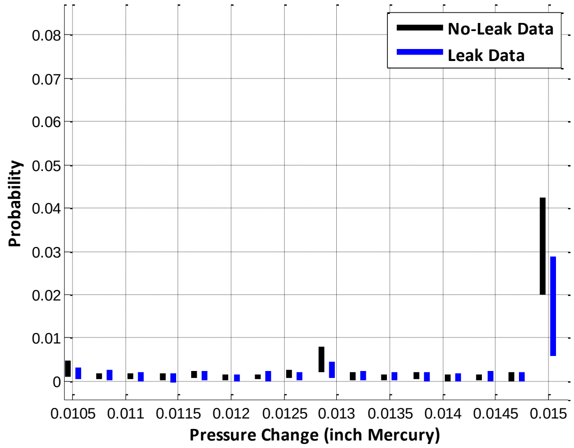


Figure 4: Zoom-in view of Figure 3 focus on partitions on the positive side. For partition centered at 0.015, with some overlapping the means of leak vs no leak populations exhbit certain level of difference.

### 3.2.5. Continuous Evaluation of DPDF Content Derived Features for Leak vs No-Leak System State Determination

One advantage of using recursive equation for feature extraction is the enablement of continuous assessment of the system of interest. In Figure 5, DPDF partition content around zero for multiple leak (upper figure) and no leak (lower feagure) datasets (as described in 3.2.4) are shown in time domain where we can visually validated the continuous class separation capability.

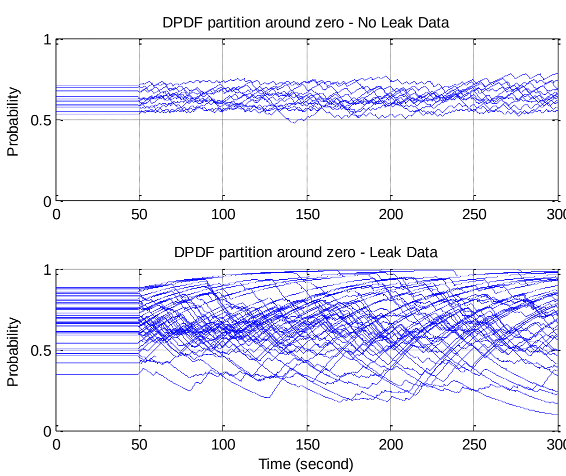


Figure 5: Continuous Evaluation of Content derived from DPDF partition around zero. DPDF content (Y-axis) as shown is presented in terms of probability where 1 equals 100%. Upper figure includes only datasets with no leak. Lower figure includes only datasets with leak.

## 4. CLASSIFICATION WITH A SIMPLE THRESOLD SETTING AND RESULTS

Existing datasets to test out the method contains data streams that are collected for calibration purpose of existing strategies. Due to current monitor's design, datasets collected for this purpose tend to put more focus on datasets with leaks. There are 14 data files labeled as system that has been verified to have no leak and 53 data files that have induced leak. When applied to existing monitor, nearly half of all dataset will be thrown out without being evaluated due to failures to pass one of the entry condtions in place.

For simplication purpose, we will refer to $DPDF_0$ for the probability value obtained from the partition around zero. We employ method described above to calculate $DPDF_0$ continuously at a particular common execution phase of current strategy where the system was commanded to be sealed.

$$\mu_{MAX\ of\ DPDF_0} = \frac{\sum_1^k \max(DPDF_{0,k})}{k} \tag{5}$$

$$\mu_{min\ of\ DPDF_0} = \frac{\sum_1^k \min(DPDF_{0,k})}{k} \tag{6}$$

$$\sigma_{MAX\ of\ DPDF_0} = \sqrt{\left(\frac{\sum_1^k \left(\max(DPDF_{0,k}) - \mu_{MAX\ of\ DPDF_0}\right)^2}{k-1}\right)} \tag{7}$$

$$\sigma_{min\ of\ DPDF_0} = \sqrt{\left(\frac{\sum_1^k \left(\min(DPDF_{0,k}) - \mu_{min\ of\ DPDF_0}\right)^2}{k-1}\right)} \tag{8}$$

The characterization of PDC0 from no leak dataset involves using 10 no leak data files. From these files, means and standard deviations of maximum and minimum values of each PDC0 profiles are obtained. Currently, upper and lower thresholds are estimated separately taking the common form as the following:

$$Threshold_{Upper} = \mu_{MAX\ of\ DPDF_0} + k_1 \cdot \sigma_{MAX\ of\ DPDF_0} \tag{9}$$

$$Threshold_{Lower} = \mu_{min\ of\ DPDF_0} + k_2 \cdot \sigma_{min\ of\ DPDF_0} \tag{10}$$

For each dataset, $DPDF_0$ profiles are evaluated continuously against Threshold_Upperand Threshold_Lower. System is deemed to be leaky if at any given time "either" threshold is exceeded.

Identification of thresholds k1 and k2 are performed with following procedure. We divide both datasets with leak and datasets with no leak into 2 equal sized groups (training and validation). As a result, each group contains 7 no leak datasets. In addition, training group contains 26 leak datasets and validation group contains 27. We enumerate k1 and k2 values between -3 to 3 with 0.1 increments to identify potential pairs of k1 and k2 producing reasonable results. In this case, we define a reasonable performance as being able to at least classify all no leak datasets correctly. After that, passing pairs are ranked based on their detection rate for leak datasets. In this process we found that among 31*31 = 961 pairs there exist 20 pairs of k1 and k2 to have the same results. For these pairs, the overall prediction rates are the same at 100% meaning all leak and no leak datasets were identified correctly. They tend to have k1 around 0.9 ~ 0.18 and k2 to be either -0.7 or -0.8.

Table 1. k1 and k2 pair test sequence and detection rates for leak datasets, no leak datasets and when combined.

| Testing Sequence | K1 | K2 | Detection Rate (%) - No Leak | Detection Rate (%) - Leak | Detection Rate (%) - all |
|---|---|---|---|---|---|
| 1434 | 0.9 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1435 | 1 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1436 | 1.1 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1437 | 1.2 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1438 | 1.3 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1439 | 1.4 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1440 | 1.5 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1441 | 1.6 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1442 | 1.7 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1443 | 1.8 | -0.8 | 100.0% | 100.0% | 100.0% |
| 1495 | 0.9 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1496 | 1 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1497 | 1.1 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1498 | 1.2 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1499 | 1.3 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1500 | 1.4 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1501 | 1.5 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1502 | 1.6 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1503 | 1.7 | -0.7 | 100.0% | 100.0% | 100.0% |
| 1504 | 1.8 | -0.7 | 100.0% | 100.0% | 100.0% |

Using these pairs we obtained best overall detection rate of 88% that is slightly worse yet very similar to the result of the original leak monitor. The two k1 and k2 pairs produced best result during validation have the same k1 to be 0.9 and k2 to be -0.7 and -0.8 respectively at sequence #1434 and #1495. One thing to note is that application of the proposed method does not require a large set of entry conditions before monitoring procedures being executed. In other words, proposed feature calculation with a simple thresholding method result in significantly improved monitor applicability in comparison with current design.

Table 2. k1 and k2 pair validate sequence and detection rates for leak datasets, no leak datasets and when both are combined.

| Validation Sequence | K1 | K2 | Detection Rate (%) - No Leak | Detection Rate (%) - Leak | Detection Rate (%) - all |
|---|---|---|---|---|---|
| 1434 | 0.9 | -0.8 | 85.7% | 88.9% | 88.2% |
| 1435 | 1 | -0.8 | 85.7% | 85.2% | 85.3% |
| 1436 | 1.1 | -0.8 | 85.7% | 85.2% | 85.3% |
| 1437 | 1.2 | -0.8 | 85.7% | 81.5% | 82.4% |
| 1438 | 1.3 | -0.8 | 85.7% | 81.5% | 82.4% |
| 1439 | 1.4 | -0.8 | 85.7% | 81.5% | 82.4% |
| 1440 | 1.5 | -0.8 | 85.7% | 77.8% | 79.4% |
| 1441 | 1.6 | -0.8 | 85.7% | 77.8% | 79.4% |
| 1442 | 1.7 | -0.8 | 85.7% | 77.8% | 79.4% |
| 1443 | 1.8 | -0.8 | 85.7% | 77.8% | 79.4% |
| 1495 | 0.9 | -0.7 | 85.7% | 88.9% | 88.2% |
| 1496 | 1 | -0.7 | 85.7% | 85.2% | 85.3% |
| 1497 | 1.1 | -0.7 | 85.7% | 85.2% | 85.3% |
| 1498 | 1.2 | -0.7 | 85.7% | 81.5% | 82.4% |
| 1499 | 1.3 | -0.7 | 85.7% | 81.5% | 82.4% |
| 1500 | 1.4 | -0.7 | 85.7% | 81.5% | 82.4% |
| 1501 | 1.5 | -0.7 | 85.7% | 77.8% | 79.4% |
| 1502 | 1.6 | -0.7 | 85.7% | 77.8% | 79.4% |
| 1503 | 1.7 | -0.7 | 85.7% | 77.8% | 79.4% |
| 1504 | 1.8 | -0.7 | 85.7% | 77.8% | 79.4% |

## 5. CONCLUSION AND FUTURE WORK

We have proposed a novel method to obtain an effective feature from discretized probabilistic density function continuously. Using a simple threshold mechanism, different thresholds are setup such that exceeding either one indicates the presence of a leak in the system. Compared with existing strategies that use a set of entry conditions to determine whether to execute a test or not, proposed method produced similar detection rate while significantly increases applicability (no entry conditions has to be imposed).

In addition to the simple threshold setting approach presented in this paper, continuing effort will be focused on evaluating the usage of more effective data classification methods such as SVM, Bayesian Classifiers, Fuzzy Classifiers or LVQ with proposed feature. The eventual goal is to redesign computation procedures that minimizes false positives/negatives (robustness), enhances system performance (performance) in real-world settings with broad coverage (applicability). We believe continual effort in this field will ensure future technical advancement in this fundamental yet critical aspect in emission reduction and control.

## REFERENCES

Murvay, Pal-Stefan, and Ioan Silea. (2012). *A survey on gas leak detection and localization techniques.* Journal of Loss Prevention in the Process Industries 25.6 pp.966-973.

Zhou, Z. J., Hu, C. H., Xu, D. L., Yang, J. B., & Zhou, D. H. (2011). *Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection.* Expert Systems with Applications, 38(4), 3937-3943.

Ghazali, M. F., Beck, S. B. M., Shucksmith, J. D., Boxall, J. B., & Staszewski, W. J. (2012). *Comparative study of instantaneous frequency based methods for leak detection in pipeline networks.* Mechanical Systems and Signal Processing, 29, pp. 187-200.

Mandal, S. K., Chan, F. T., & Tiwari, M. K. (2012). *Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM.* Expert Systems with Applications, 39(3), 3071-3080.

State of California Air Resources Board (2012). *California Evaporative Emission Standards and Test Procedures for 2001 and Subsequent Model Motor Vehicles.*

State of California Air Resources Board (2012). *California Refueling Emission Standards and Test Procedures for 2001 and Subsequent Model Motor Vehicles.*

Filev, D., Tseng, F., Kristinsson, J., & McGee, R. (2011). *Contextual on-board learning and prediction of vehicle destinations.* In Computational Intelligence in Vehicles and Transportation Systems (CIVTS), 2011 IEEE Symposium on. pp. 87-91

Wong, Kevin C., and William Western. (2003) *Engine off natural vacuum leakage check for onboard diagnostics.* U.S. Patent No. 6,550,316. 22

Jentz, R. R., Clemens, D. A., Sipes, A. I., Dudar, A. M., & Kluzner, M. I. (2013). *System and method for performing evaporative leak diagnostics in a vehicle.* US 8725347 B2

McLain, Kurt D., and Gregory E. Labus. (2005). *Apparatus and method for fuel vapor leak detection.* US 6880383 B2

Syed, Fazal U., Dimitar Filev, Fling Tseng, and Hao Ying. (2009). *Adaptive real-time advisory system for fuel economy improvement in a hybrid electric vehicle.* In Fuzzy Information Processing Society, NAFIPS 2009. Annual Meeting of the North American, pp. 1-7.