Machine Learning Based Approach for EVAP System Early Anomaly Detection Using Connected Vehicle Data

Ala Eddine Omrani, Pankaj Kumar, Aed Dudar, Michael Casedy, Steven Szwabowski, Brandon Dawson

Ford Motor Company, Dearborn, Michigan, 48126, USA

aomrani@ford.com
pkumar38@ford.com
adudar@ford.com
mcasedy@ford.com
sszwabow@ford.com
bdawso14@ford.com

ABSTRACT

From automobile manufacturers perspective, reduction of warranty cost leads to less expenditures, which then yields higher profits. Thus, it is crucial to leverage the available methods and tools to achieve such outcome. Connected vehicle data is one critical resource that can be a gamechanger, reducing the associated costs and improving the business profitability. This project uses Mode06 connected vehicle data along with contextual data to early detect EVAP and purge monitors anomalies, yielding proactive fix of the detected issue through software (SW) and/or hardware (HW) upgrades (preventive maintenance and system quality improvement). Root cause analysis, which can be developed based on the anomaly detection outcomes and which is not within the scope of this paper, allows diagnostics of HW and/or SW related issues in a timely manner and eventually be prepared ahead of time for system failures. In this paper, statistics-based early anomaly detection models, based on vehicle data and fleet data, are developed. The proposed solution is a generic tool that doesn't make assumptions on data distribution and can be adapted to other systems by tweaking mainly the data cleaning process. It also incorporates specific system definitions of abnormal behavior, which makes it more accurate compared to conventional anomaly detection tools, which are mainly affected by the imbalanced data and the EVAP and purge definition of an anomaly. When deployed with field data, the algorithm showed higher performance, compared to popular anomaly detection techniques, and proved that failures can be prevented through detection of the anomalies several weeks/miles before the actual fail.

---

1. INTRODUCTION

The condition of the EVAP system (Evaporative Emission Control System) is critical not only for gas engine-vehicles' performance, but also for environment. It is also one main source of warranty costs (İbrahim, Altinişik, & Keskin, 2015). The current procedure for repair and improvement of emission control system software and hardware is based on a corrective maintenance. An example is the monitoring of Diagnostic Trouble Code (DTC) rate for a given fleet. Consequently. the SW calibrations may be revised if the rate is alarming (i.e, higher than historical rate). Another example is the review of warranty claims to determine the main failure modes and assess the current performance of the designed system and the need for updates.

The problem with such approaches is the delayed actions aiming to deal with the observed failures or to improve the system performance. In fact, an action based on higher failure rate means that the issue has been already propagated and that it affected a portion of the monitored population, which means higher warranty cost and lower level of customer satisfaction.

In this paper, another approach is proposed to deal with emission control system performance by leveraging connected vehicle data. The approach is based on anomalies early detection, allowing to act proactively to improve the system before failure (preventive maintenance), which will then yield a reduction of the failure propagation rate. For example, a software over the air update (OTA) can be executed for vehicles with no sign of failure, based on early anomalies trends. Such OTA update allows avoidance of customer visits to dealership which will reduce warranty cost and improve customer experience/satisfaction. Another example is hardware improvement involving identification of specific parts as the failure root cause. Subsequently, the identified parts will be replaced during Job 2 phase based on

detected anomalies from Job 1 vehicles (Job 1 refers to the initial production of a given vehicle model year, while Job 2 is related to the mid-model year production with updates).

The designed early anomaly detection method processes emission system's on-board monitoring tests results to develop a data-driven statistical model, which outputs a threshold to determine whether a test result is an anomaly or a healthy outcome. To this end, a fleet data-based model and a vehicle data-based model are introduced. The two models' development flowcharts include data cleaning and pre-processing using test specificities and feedback from the subject matter experts (SMEs). Next, cleaned data is filtered, using the Chebyshev filter, to remove noise which can affect anomaly detection performance and increase false positives and false negatives rates. Lastly, anomaly threshold is calculated based on the pre-processed data distribution and predefined upper and lower threshold limits. An anomaly score is then determined for the test results that exceed the deduced threshold, which consequently helps confirm or reject the anomaly (false positive/True positive). As a continuation of this effort, and for the detected anomalies with higher score, corresponding trip data might be investigated to understand the anomaly context and remove any false positives. Warranty data and software calibration versions might be also analyzed to assess the test performance and whether the anomalies are real or false positives.

One advantage of the above approach for anomaly detection is that it is designed in a generic fashion since it doesn't require a specific data distribution for the anomaly detection task. Algorithm tuning is required though, to adapt the data cleaning step to the monitored system. In addition, one challenge with EVAP and purge data is that data density-based anomaly detection methods (and similar methods) assume that an anomaly is any point "away from the main cluster(s)" which represents the normal response of the system. Such assumption leads to high number of false positives given that the EVAP and purge tests of interest fail in one direction above/below a given test result threshold. Meanwhile, the proposed data-statistics based method is not affected. It is also robust to imbalanced data, which is another main characteristic of EVAP and purge system's tests results. Those observations are detailed in Section 3.1 of this paper.

This paper is divided into 5 sections. In section 2, an overview of EVAP and purge monitors is presented. Section 3 is dedicated to the description of the early anomaly detection methods. Outcomes of the developed methods are then discussed in Section 4. Finally, key findings and future work are discussed in Section 5.

2. EVAP-PURGE MONITORS

The Evaporative Emission Control System (EVAP) is a critical part of vehicles with gas engine. As its name suggests, the main function of this system is to control gases emissions to the atmosphere.

During key-off periods, fuel vapors are contained within the fuel system, and stored in the carbon canister thus preventing them to escape to the atmosphere. During trip time, vapor stored in the charcoal canister are purged through the canister purge valve (CPV) to the intake manifold.

Since gasoline vapors' emissions control function is critical, monitors were developed to make sure that the system is efficient and that the vehicle is not polluting the environment as required by the regulations. One of the deployed monitors is EVAP monitor, which assesses whether the fuel system, illustrated in Figure 1, is leaking fuel vapor above the regulated leak size. Several tests are executed including small leak test (key-off leak size less than 0.02") and medium leak test (key-on leak size is less then 0.04").
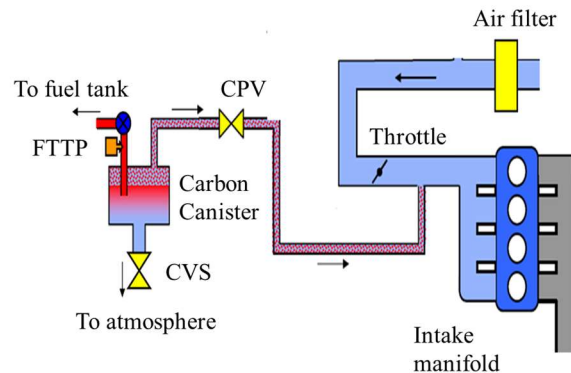

Figure 1: EVAP System Schematics

On the other hand, purge monitor's tests measure the integrity of the different fuel system components that are controlling fuel vapor purging function to the intake manifold. Leaky CPV is one of the purge monitor tests insuring that CPV is not purging when it is closed.

Tests results, comprise test value (depending on the test, it could be pressure measurements, slope or ratio among others). Minimum threshold and maximum threshold are also part of the test results, which are reported as a test output and stored within the mode06 report (On-Board diagnostics reported tests results). Based on the reported data, test is labeled as a pass or a fail and corresponding diagnostic trouble codes (DTC) might be set along with the malfunction indicator light (MIL) depending on the corresponding requirements.

3. ANOMALY EARLY DETECTION PROCEDURE

Anomaly detection is used to capture trends toward failure, lower performance and other key indicators yielding a pro-

active monitoring in terms of applying corrective actions and reducing maintenance and warranty costs.

Several techniques have been introduced for anomaly detection. Statistics-based methods can be used to determine abnormal regions, such as multivariate normal distribution-based analysis, Gaussian mixture model and box plot (Lauer, 2001). Another approach is to use supervised techniques, which require labeled data for model training purposes. Decision trees, Support vector machines (SVM) and naïve Bayes are among the most popular supervised techniques for anomaly detection. The main advantage of decision trees is its simplicity, while it requires large storage. SVM can be generalized to different use cases. However, its interpretability is not straightforward in addition to challenges related to optimal kernel selection. Naïve Bayes is known for its simplicity of implementation and its lower requirement for training samples. Its main disadvantage is that it handles the different features independently, and thus, cannot capture the inter-dependence of the features of concern (Agrawal & Agrawal, 2015) (Al-Garadi, et al., 2020). Semi-supervised methods have been developed to take advantage of both supervised and unsupervised algorithms. Multi-Layered Clustering (MLC) approach and Extreme Learning Machine (ELM) are two well-known semi-supervised techniques. One major challenge with such approaches is that they don't provide the same detection accuracy that is achieved by supervised machine learning (Al-Garadi, et al., 2020). For our specific use case, implementation of the aforementioned supervised and semi-supervised techniques is challenging due to the lack of labeled field data.

With regard to unsupervised techniques, K-means, Principal Component Analysis (PCA) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) are widely used unsupervised tools for data clustering. K-means generates K clusters with no need for labeled data, while is reported to be less effective than supervised techniques for anomaly detection based on previous studies (Zarpelão, Rodrigo, Cláudio, & Sean, 2017). PCA reduces the number of correlated features helping to reduce classification complexity. Meanwhile, it requires additional machine learning techniques to perform the clustering task. As for DBSCAN algorithm, and similar to the other unsupervised clustering methods, a common issue is that it is not optimized to detect the anomalies since the main goal is to define clusters.

In this paper, the retained early anomaly detection method is a data-driven solution deploying statistical analysis to label the test result as either healthy, anomalous or failure. The method details and performance, compared to the above techniques, are illustrated in the subsequent sections.

The first step for anomaly detection, using the proposed method, is data pre-processing to ensure that the consumed data is of good quality. Filters are considered to clean the data ranging from test specific filters, based on expected data range, to generic filters like data redundancy removal. Removing redundancy, and alike issues, is of particular importance to the anomaly detection method since redundancy for instance may skew the data and affect the distribution which consequently affects the model outcomes.

Once the data is cleaned, test output is normalized for the early detection algorithm to be modular and applicable for any use case regardless of the test specificities. The main variable used for anomaly detection is the normalized test result (TestResult), which is the ratio of the test value (TestValue) divided by the corresponding threshold, which could be either the minimum threshold (ThreshMin), maximum threshold (ThreshMax) or both.

Two approaches can be adopted for early anomaly detection, by leveraging either the vehicle data or the fleet data. To take advantage of both methods, a hybrid model is retained. This choice is justified by the fact that each method has strengths which may not be the case for the other approach as explained in Table 1.

Challenge	Analysis using fleet data	Analysis using vehicle data
Low number of tests per vehicle	X	
Detection during training phase	X	
Vehicle/Anomaly specificities		X
Lower number of false positives	X	
Anomaly speed of detection		X

Table 1: Early anomaly detection approaches

The model is hybrid in terms of results analysis and root cause analysis (which is an extension of the present work and not within the scope of this paper). In one example, using both approaches and based on the anomalies' scores, the end user would have a better understanding of one particular vehicle performance through analysis of its own data in addition to the related fleet performance. Such analysis should help confirm an issue specific to the vehicle of concern or rule it out by observing a similar issue on the fleet level for instance. In another example, at early stages of data collection, vehicle data-based model might not be able to generate accurate anomaly thresholds because of lack of data readings. Fleet data should be used instead.

The novelty of the proposed anomaly detection method is the design of a workflow allowing to leverage the EVAP and purge monitors tests' results to assess the condition of the system components ahead of the failure event (preventive maintenance). The workflow is designed in a generic fashion which yields a framework applicable for

other monitors and systems, provided it is tweaked to account for the system specificities.

Moreover, one challenge with Purge and EVAP monitors test results' anomaly detection, is the distribution of the data, which is not normal in many cases. Also, data analysis revealed that the data wrangling and anomaly detection method for fleet data, doesn't yield reliable results at vehicle level. Two different workflows are then considered for each scenario.

Both fleet data driven model and vehicle data driven model use normalized test values (TestResult) to perform the anomaly detection task as illustrated in the below sections.

3.1. Vehicle Data-Based Model

This model aims to develop an anomaly threshold which depends on the vehicle data allowing to take into account the specificities of the vehicle equipment and the inside and outside conditions. It also allows to implement the model in the powertrain control module (PCM) if needed since it only requires vehicle data not fleet data. The first step of the model is to remove noise in order to calculate the anomaly thresholds accurately. Removed data points are reconsidered at a later stage when labeling the data. This first step is executed whenever a new data is collected for both training and test phases as will be discussed later. A Chebyshev filter, as illustrated in (Godwin, 1955), is used for noise removal and it is based on Chebyshev inequality as shown in Eq. (1).

$$P(X > Mean + k * SD \text{ or } X < Mean - k * SD) \leq \frac{1}{k^2} \quad (1)$$

In Eq. (1), *Mean* and *SD* are respectively the average and standard deviation of the considered dataset, k is a calibratable factor which is set to 10 meaning that the likelihood, *P*, of a point to be outside $Mean \pm k*SD$ is less than 1%. An anomaly limit ($Thresh_{SME}$) is used to control the filter outputs. This limit is set based on expected test results and recommendations from the SMEs.

The resultant filtering thresholds are defined as below.

$$Threshold_{Up} = \min(Thresh_{SME}, Mean + k * SD)$$

$$Threshold_{Down} = \max(Thresh_{SME}, Mean - k * SD) \quad (2)$$

In Eq. (2), *min* and *max* refer respectively to the minimum and maximum operators. Any point above $Threshold_{Up}$ or below $Threshold_{Down}$ is considered part of the noisy data and is subsequently removed.

The effect of the Chebyshev filter is shown in Figure 2, where k-means is used for data clustering using data before applying the filter (bottom figure) and after applying the filter (top figure). The shown data represents the small leak test output (i.e TestResult), which is the normalized test value as explained previously. Features scaling is done using the z-score normalization, to ensure that the two

variables have comparable "contributions" when running the classification algorithm.

Knowing that the failure and anomaly regions are dependent mainly on the test result value, data clustering is expected to separate regions horizontally. From Figure 2, it is clear that, with the filter applied (figure on the top side), the clustering makes more sense, through removal of higher test results (noise), which are healthy and don't add much information to the analysis.
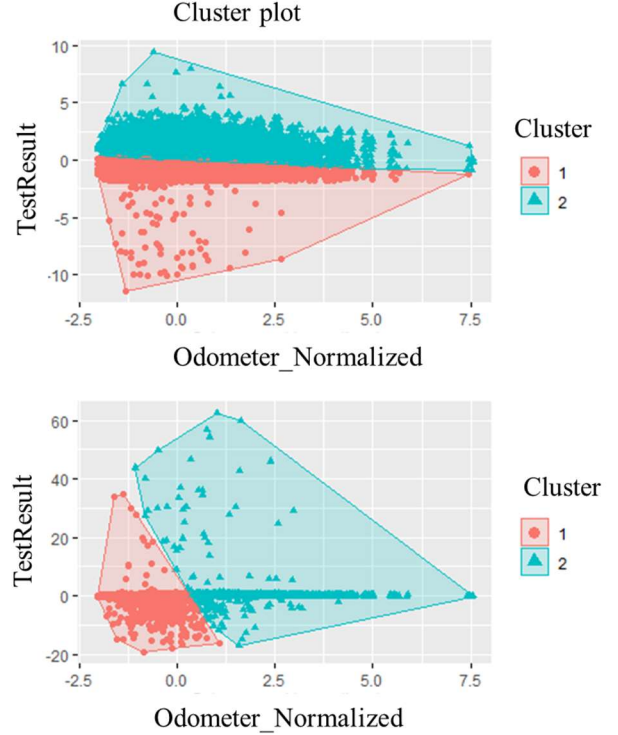

Figure 2: Chebyshev Filter Effect on Data Clustering

Once noise is removed, anomaly thresholds are then determined through a training phase using a calibratable number of initial test points (N_{Train}) and are updated each N miles. Two approaches are considered to calculate the thresholds: The Individual and Moving Range (I-MR) charts and the Box plot technique.

I-MR charts is a Statistical Process Control (SPC) method applied on vehicle data (i.e process data). It is an anomaly detection technique which includes a moving range (MR) chart for data cleaning and an individual I-chart for anomaly detection (Wheeler, 1995). Data normality is not required. The process includes dismissal of data outside a "normal" moving range limit (UCL_{MR}) derived from the training dataset.

$$UCL_{MR} = 3.72 * \frac{\sum_{i=1}^{N_{Train}} |X(i) - X(i-1)|}{N_{Train}} \quad (3)$$

where $X(i)$ the ith point of the train dataset arranged in a chronological order. Any point above UCL_{MR} is considered abnormal and is dismissed from the anomaly threshold calculations in the I-chart as in Eq. (4).

$$UCL_{IMR\_Up} = min(Thresh_{SME}, Mean_{IMR} + 3 * SD_{IMR})$$

$$UCL_{IMR\_Down} = max (Thresh_{SME}, Mean_{IMR} - 3 * SD_{IMR})$$

(4)

where $Mean_{IMR} = \frac{\sum_{i=1}^{N_{Train}} X(i)}{N_{Train}}$ and

$$SD_{IMR} = 0.88 * \frac{\sum_{i=1}^{N_{Train}} |X(i) - X(i-1)|}{N_{Train}}$$

The EVAP and purge tests of interest could be upward or downward tests. An upward test trends upward when failing (meaning that the failure threshold is above the anomaly threshold). In contrast, a downward test trends downward when failing (meaning that the failure threshold is below the anomaly threshold). In Eq. (4), and for an upward test, if *TestResult* is above UCL_{IMR_Up}, then data is labeled as anomaly. As for a downward test, if *TestResult* is below UCL_{IMR_Down}, then data is labeled as anomaly.

As shown in Eq. (4), I-MR chart deploys a non-conventional formula for the standard deviation which is more robust to non-normality of data distribution. The formula is based on absolute moving range rather than the square of the distance from the mean, reducing then the impact of non-normal distributions.

It is worth mentioning that the $3*SD_{IMR}$ term is not derived based on data distribution assumption. It is rather based on experience (i.e empirical limits not probability limits) as shown in references (Shewhart, 1931) and (Wheeler, 1995), where Shewhart and Wheeler suggest that the vast majority of data for any SPC related distribution is within $Mean_{IMR}+/-3* SD_{IMR}$.

Table 2 shows an investigation of the above assumption, using field data to prove the accuracy of this rule.

Test	Vehicle Test Data Count	Points Outside Mean+/-3*SD
Small leak test	1,961	0.48%
Medium leak test	1,449	0.97%

Table 2: IM-R Chart Assumption Verification

I-MR charts is the default method for thresholds calculations since it is a well-established technique for statistical process monitoring. Meanwhile, it is observed that I-MR chart behaves poorly when data is scattered. A Quartile coefficient of dispersion (QCD) is then used to determine the dataset level of dispersion. QCD is defined as in Eq. (5).

$$QCD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

(5)

$$Q_1 = Median [X_{Min} ... X_{Median}]$$
$$Q_3 = Median [X_{Median} ... X_{Max}]$$

Q_1 and Q_3 are the first and third quartiles, defined in Eq. (5) for a given ordered dataset $(X_i)_{i=1 \to M}$, where X_1 is the lowest value and X_M is the highest value.

A large *QCD* means that the data is dispersed. A threshold, λ, is defined in a way that if *QCD* is higher than λ, dataset is dispersed, and box plot is applied. If not, I-MR charts are considered.

The reason for choosing box plot is its robustness to data distribution and data dispersion. The robustness is the result of the percentile calculation, used by box plot to determine the anomalies. In fact, an anomaly is located outside the 25th percentile to the 75th percentile of data, where interquartile (*IQR*) is used to define the anomaly limits as below.

$$IQR = Q_3 - Q_1$$

$$UCL_{Box\_Up} = min (Thresh_{SME}, Q_3 + IQR)$$

(6)

$$UCL_{Box\_Down} = max (Thresh_{SME}, Q_1 - IQR)$$

The box plot technique is chosen based on a comparison with other popular unsupervised anomaly detection techniques. In Table 3 is illustrated the performance metrics for several anomaly detection methods, using fleet data for the EVAP small leak test. The unsupervised models are trained with unlabeled data. Models' metrics are derived using labeled data. The data labeling procedure considers upper and lower thresholds provided by the SMEs based on their experience and expectations. Considering that, for the small leak test, failures are test results that are *below* the DTC threshold, a healthy data is defined as any test result above the lower warning limit and an anomaly is defined as any test result below the upper warning limit. The region between the lower and upper limits is where the early anomaly detection algorithm will help provide more insights as there is no clear information about the correct data condition. Figure 3 illustrates the different thresholds for the small leak test, based on the Test Result value (i.e normalized test value using the associated test thresholds).
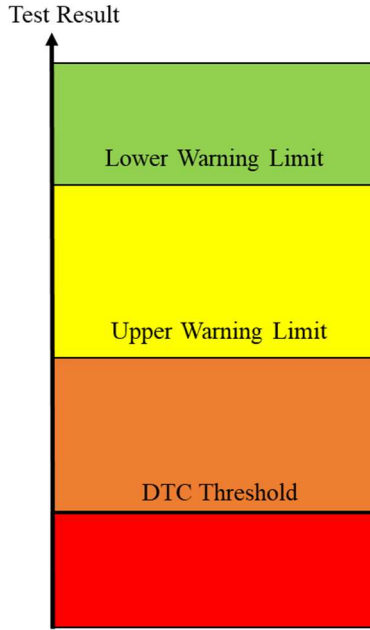
Test Result

Figure 3: Small Leak Test Performance Regions

To conclude, models training is done using data from the different regions in Figure 3. Meanwhile, yellow region data is excluded when performing models testing since data labeling can't be accurately done. Features scaling, using the z-score normalization, is performed on data ingested by DBSCAN, k-means and One-class SVM. This ensures that the two considered features' contributions (test result and odometer) have similar weights when running the classification methods. Data statistics-based methods don't require features scaling since only one feature is considered (i.e Test result).

Method	Accuracy	Precision	Recall
Mean+/-3*SD	0.99	1	0.89
DBSCAN	0.004	0.98	0.003
K-means	0.68	0.19	0.002
One-Class SVM	0.003	0.98	0.003
Box Plot	0.99	1	0.96

Table 3: Anomaly Detection Techniques Performance

For the EVAP and purge tests of interest, the definition of an anomaly is not solely based on data density/clusters. Hence, and considering that DBSCAN and one-Class SVM methods rely essentially on data clustering for anomaly detection, false positive and false negative rates are expected to be higher than other use cases.

One challenge with the EVAP data is that it is imbalanced, meaning that healthy data is the majority class while anomalies and failures is the minority class. Considering the aforementioned labeling strategy for a given fleet of

vehicles, 99.6% of the data is healthy while only 0.4% of the data are anomalies/failures. Such distribution yields a model which is trained heavily on healthy data, while poorly trained on unhealthy data, which would affect the model accuracy to predict anomalies. As shown in Table 3, data imbalance is severely affecting data density-based methods' accuracy, while having no effect on data-statistics based methods. The relatively better performance of k-means method is due to the fact that it requires the number of clusters as an input, which helped improve the classification accuracy.

Another challenge, that data-density based methods struggle with, is the fact that outliers are expected to have a separate cluster. In the present case study, anomalies represent the minority class, which could be then interpreted as noise. A trial-and-error process was required aiming for DBSCAN to define a cluster for the outliers. Meanwhile, the optimal parameters still define a portion of the outliers as noise, which explains the poor performance of the algorithm.

Based on the overall methods performance in Table 3, Box plot and 3*sigma seem to be the best ones to deal with anomaly detection for the EVAP monitor tests where data is imbalanced and unlabeled. Looking at the Recall metric, Box plot does have the best score. An additional advantage of data-statistics methods is the computational cost and the number of tuning parameters. Meanwhile, it is important to mention that data-statistics based methods' performance is related to the specific nature of data distribution for the considered system and how an anomaly is defined. Hence, for other use cases, box plot might not be the best choice, even though its computational cost will always be the lowest.

The above results can be extended to the I-MR charts method since it is also a data statistics-based solution. Another advantage of using the I-MR chart is that it is a well-established method for SPC processes (vehicle test result evolution over time) and it doesn't require normal distribution. It is also suitable for reduced data size unlike data density-based methods which require a fair amount of data to predict the anomaly and healthy clusters accurately.

Finally, an anomaly scoring method is used to prioritize the investigation/root cause analysis (RCA) of detected anomalies for the different vehicles of concern. The lower the score, the higher the false positive likelihood is and vice-versa. The anomaly score is based on four factors:

- Trend: A trending anomaly toward failure increases the likelihood of a real positive (versus singularities).

- Severity: A higher severity leads to higher urgency. Severity is defined as the distance from the lower limit to the anomaly, normalized by the distance from lower limit to Onboard Diagnostics (OBD) threshold.

- Spikes: The likelihood of false positive is higher for a spiky dataset.

- Dispersion: Data dispersion increases the likelihood of having false positives.

A convex combination is used to express the anomaly score for vehicle *Veh* and anomaly *X(n)* at time stamp *n*, as in Eq. (7).

$$Score^{Veh}(X(n))$$
$$= \alpha_1 * Sat\left(\frac{X(n) - LL}{OBD - LL}\right) + \alpha_2 * Sat\left(1 - \frac{\Delta X(n)}{2 * UCL_{MR}}\right) \quad (7)$$
$$+ \alpha_3 * Sat\left(\frac{N_{Trend}}{N_{Norm}}\right) + \alpha_4 * (1 - QCD),$$

where

- $\Delta X(n) = |X(n) - X(n-1)|$
- The operator $Sat(Y)$ is the saturation operator which bounds the variable *Y* between 0 and 1
- $\sum_{i=1}^{4} \alpha_i = 1$
- *LL* is the lower warning limit, as in Figure 3
- *OBD* is the test failure threshold
- The factor $2 * UCL_{MR}$ keeps $\frac{\Delta X(n)}{2 * UC_{MR}}$ lower than 1 for cases where $\Delta X(i) > UCL_{MR}$ (outlier in MR chart)
- N_{Trend} is the number of successive trending points
- N_{Norm} yields an upper limit to the contribution of N_{Trend} to the anomaly score

Table 4 shows the default weighting factors values. The choice of those values is explained by the fact that data gaps affect Spikes and Trend factors. To reduce data gaps' impact, the corresponding factors' weights are reduced.

α_1 Severity	α_2 Spikes	α_3 Trend	α_4 Dispersion
0.4	0.1	0.2	0.3

Table 4 Anomaly Scoring Weighting Factors

3.2. Fleet Data-Based Model

In a similar fashion to the vehicle data-based model, the fleet data model uses the Chebyshev filter to remove noise for mean and standard deviation calculations. Once Nosie is removed, the threshold is calculated as in Eq. (8), assuming that the data has a normal distribution, for a given fleet of vehicles with N_{Fleet} observations ($X_{Fleet}(i)$).

$$UCL_{Fleet\_Up} = min(Thresh_{SME}, Mean_{Fleet} + 3 * SD_{Fleet})$$
$$UCL_{Fleet\_Down} = max(Thresh_{SME}, Mean_{Fleet} - 3 * SD_{Fleet}) \quad (8)$$

where $Mean_{Fleet} = \frac{\sum_{i=1}^{N_{Fleet}} X_{Fleet}(i)}{N_{Fleet}}$ and

$$SD_{Fleet} = \sqrt{\frac{\sum_{i=1}^{N_{Fleet}}(X_{Fleet}(i) - Mean_{Fleet})^2}{N_{Fleet}}}$$

To assess data distribution normality, Anderson-Darling test is then used. If data distribution is not normal, BoxCox transformation is used to attempt to transform the data distribution to be normal.

The one-parameter BoxCox transformation (Box & Cox, 1964), is a power transformation and is defined as in Eq. (9).

$$X_{BoxCox} = \frac{X - 1}{\lambda} \ if \ \lambda \neq 0$$
$$X_{BoxCox} = \ln(X) \ if \ \lambda = 0, \quad (9)$$

where *X* is the original data and X_{BoxCox} is the transformed data.

If BoxCox transformation is successful, *Mean+/-3*SD* of the transformed data is then used. Otherwise, Box plot method is used to determine the anomaly threshold. Data statistics-based methods are chosen for fleet-data model based on similar performance reasons as illustrated in section 3.1 for the vehicle-data model.

Anomaly scoring is then determined similar to the vehicle-based model.

Flowcharts in Figure 4 and Figure 5 summarize the main blocks that yield early anomaly detection using the two discussed approaches.
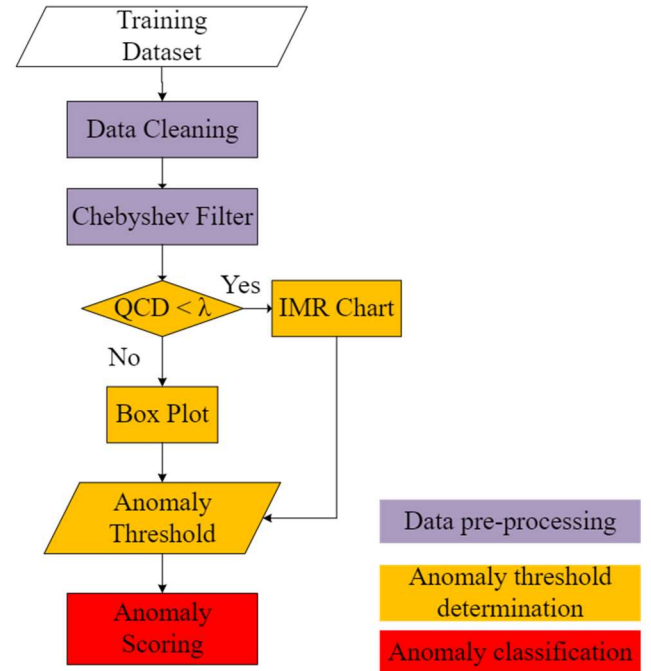

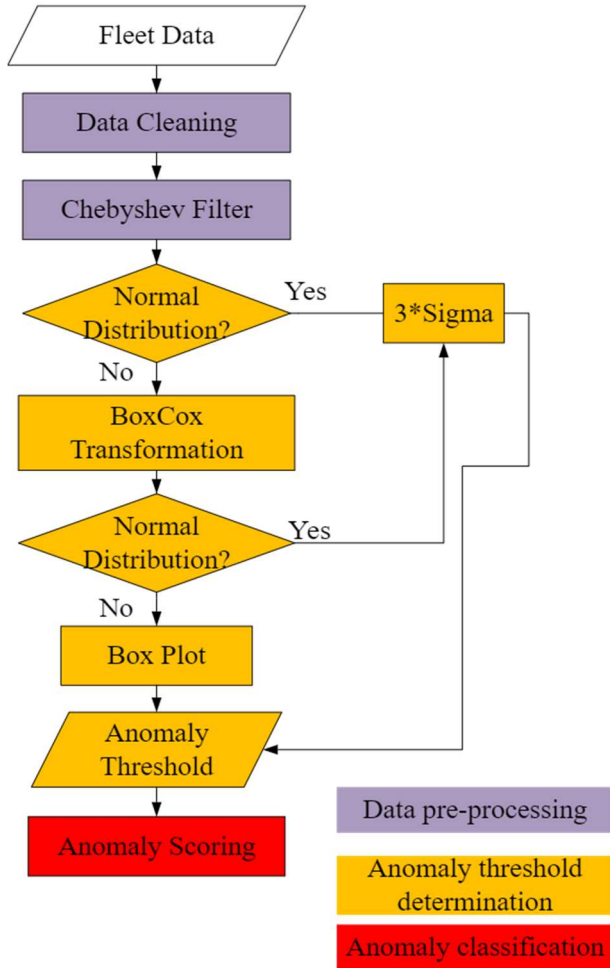Figure 4: Vehicle Data-Based Early Anomaly Detection Flowchart

Figure 5: Fleet Data-Based Early Anomaly Detection Flowchart

4. EXECUTION

Shown below in Figure 6 is an example of anomaly detection algorithm execution for a given fleet of vehicles using the fleet data-based model, applied to the medium leak test normalized test value (i.e TestResult).

The x axis represents the mileage at which the test was executed, and the y axis represents the test result (TestResult). In this example, three performance regions are identified, namely:

- Healthy region: Test passes and test result is far from the OBD/failure limit. The upper limit of this region is dependent on the data distribution and on SMEs inputted limit for healthy region.

- Warning region: Test passes with test result closer to the OBD/failure limit. The lower limit is same as healthy region's upper limit while the upper limit is dependent on the data distribution and on SMEs

inputted limit for warning region. At this level, the system starts to behave abnormally.

- Alert region: Test passes and test result is at an alarming level compared to the OBD/failure limit. The lower limit is same as warning region's upper limit while the upper limit is dependent on the data distribution and on SMEs inputted limit for the alert region. At this level, the system is at a near fail stage.

- Failure region: Test fails. The lower limit is same as the OBD threshold.
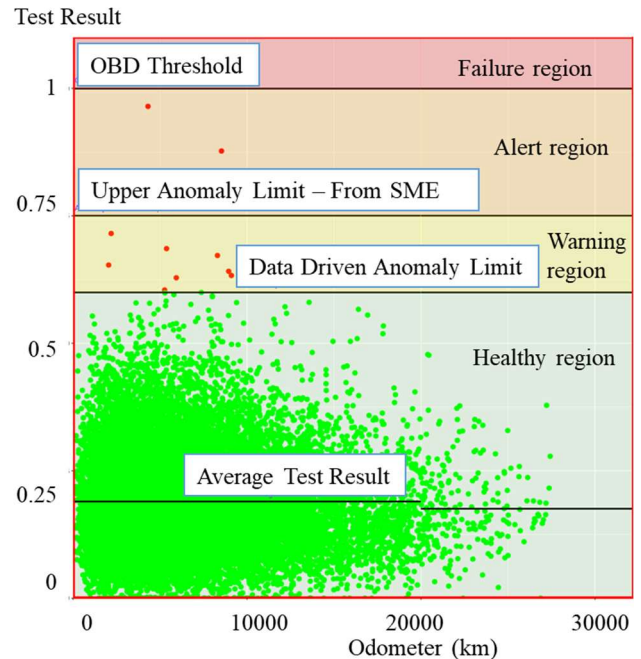

Figure 6: Fleet Data Based Model Execution

Based on the test count in each performance region, relative to the total test count, fleet performance for the selected test can be assessed and preventive actions can be taken. Root cause analysis can then be carried out if for instance a specific fleet is showing higher anomaly rate, or a given feature causes test to have more anomalies. Accordingly, preventive actions can then be taken such as Over the Air (OTA) updates to fix the identified root cause for the fleet of concern.

As for the failure region information, only corrective actions can be considered, which shows the advantage of deploying early anomaly detection methods for quality improvement.

Similarly, test performance and root cause analysis can be performed using vehicle data-based anomaly detection model by looking at the test result trends in a chronological order. Figure 7 illustrates medium leak normalized test data for a given vehicle, part of the fleet in Figure 6.
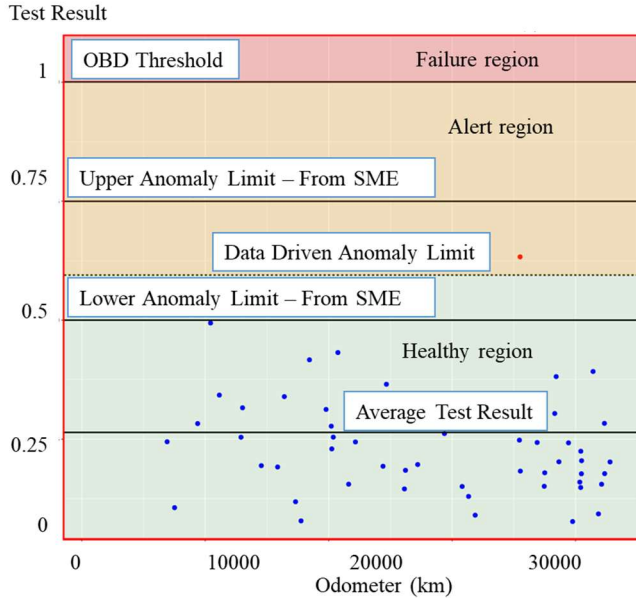
Figure 7: Vehicle Data-Based Model Execution

Similar to fleet data-based model, three regions are considered for vehicle data-based model outputs labeling, namely healthy, warning, alert and failure regions. The advantage of the vehicle data-based model, over the fleet data-based model, is that the trends can be analyzed by considering an evolving process with time, allowing to apply SPC methods for trends monitoring as explained in the previous section. For instance, the anomaly, captured in Figure 7, can be investigated using contextual data for the test results and model performance before and after the anomaly event. One can analyze the internal and external variables of concern, like atmospheric temperature and barometric pressure, for correlations with the observed anomaly. Also, trends versus spikes can be investigated, as shown in the anomaly score calculation in Eq. (7).

In another example shown in Figure 8, vehicle data-based anomaly detection algorithm allowed to capture first abnormal behavior of the EVAP medium leak test before around 40 days or 3,300 miles of test failure, by leveraging the normalized test value (i.e TestResult).
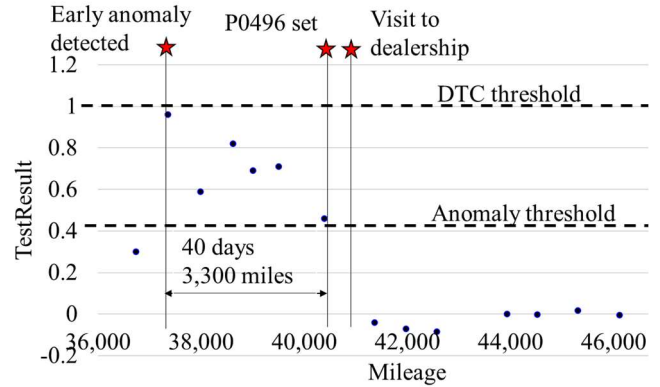

Figure 8: Anomaly Detection Use Case

It is worth mentioning that the vehicle diagnostic algorithm outputs either a test pass or fail, while the algorithm, presented in this paper, outputs three states (healthy, anomaly, failure). Anomaly often appears before the failure event, considering that failure is a continuous event where it is expected to detect a trend toward failure. Moreover, the medium leak test's DTC is of type B, meaning that even if a failure takes place, a "pending" DTC is set, with no malfunction indicator light (MIL). If the test fails in the next diagnostic session, "confirmed DTC" and MIL will get set. Such logic results in an additional delay to the HW/SW issue detection process and consequently a delayed customer visit to the dealership. Meanwhile, the presented algorithm in this paper, and by leveraging vehicle diagnostic's algorithms outputs, labels a test result as an anomaly once it gets above the anomaly threshold which is, as shown in Figure 8, is below the DTC threshold, enabling then an earlier detection. Finally, it is also important to stress on the fact that there are other reasons for the EVAP diagnostic algorithm not to show a failure. One of those reasons is test calibration (minimum and maximum thresholds). Also, missed failures, due to data collection/decoding process related issues, might be another reason, which is unlikely to happen based on extended experience with the process. Finally, the system might be well calibrated, and all data might be properly collected and decoded and only healthy response and/or anomalies and near failures are shown.

In a real-world implementation, the algorithm can be deployed onboard in the powertrain control module (PCM) to analyze system performance in a real time fashion and to provide feedback either to the customer or to the engineering teams. The solution can also be implemented in cloud. Considering this approach, data would be analyzed as received based on a scheduled collection frequency.

5. CONCLUSION

The presented work targeted improvement of EVAP and purge monitors performance monitoring by leveraging mode06 reports, which provide a "continuous" and chronological understanding of system performance progress, from an individual vehicle perspective, but also from a fleet overall performance perspective, when fleet data is considered. The developed solutions are generic and can be extended to other systems with a minimum level of calibration to adapt to the monitored system specificities. Meanwhile, one main challenge with the presented method is the false positives rate estimation. In fact, even though an anomaly is detected (below the failure limit), it may not necessarily turn into a failure in the near or long term, and it might be solely due to specific external conditions for instance. Such anomaly can then be justified since it is an abnormal system response. However, it shouldn't be treated as a potential failure. It is then crucial to perform a root cause analysis based on the observed anomaly rate, which is a natural extension of this work. Such analysis should help provide more context to the observed anomalies' trends and rates. Ultimately, the outcomes should be able to provide guidance to the end user with regard to the next steps to fix the HW/SW related issues.

NOMENCLATURE

CPV	Canister Purge Valve
DTC	Diagnostic Trouble Code
EVAP	Evaporative Emission Control System
HW	Hardware
I-MR	Individual - Moving Range
IQR	Interquartile
Job 1	Initial production of a given vehicle model year
Job 2	Mid-model year production with updates
MIL	Malfunction Indicator Light
MR	Moving Range
Mode06	On-Board diagnostics reported tests results
OBD	Onboard Diagnostics
OTA	Over the Air Updates
PCM	Power Train Control Module
QCD	Quartile Coefficient of Dispersion
RCA	Root Cause Analysis
SD	Standard Deviation
SME	Subject Matter Expert
SPC	Statistical Process Control
SW	Software
UCL	Upper Control Limit

REFERENCES

Agrawal, S., & Agrawal, J. (2015). Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, 708 – 713.

Al-Garadi, M. A., Mohamed, A., Khalid, A., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE Communications Surveys & Tutorials*, 1646-1685.

Bonett, D. (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis, 50*(11), 2953-2957.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*.

Godwin, H. J. (1955). On generalizations of Tchebychef's inequality. *Journal of the American Statistical Association*.

İbrahim, R., Altinişik, K., & Keskin, A. (2015). The pollutant emissions from diesel-engine vehicles and exhaust aftertreatment systems. *Clean Technologies and Environmental Policy*, 15-27.

Lauer, M. (2001). A mixture approach to novelty detection using training data with outliers. *Proceedings of the 12th European Conference on Machine Learning*, (pp. 300–311).

Montgomery, D. (2005). *Introduction to statistical quality control.* New Jersey: John Wiley & Sons.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product.* London: Macmillan And Co Ltd.

Tukey, J. (1977). *Exploratory data analysis.* Reading, Mass. : Addison-Wesley Pub. Co.

Wheeler, D. J. (1995). *Advanced topics in statistical process control.* Knoxville, TN: SPC press.

Zarpelão, B., Rodrigo, S., Cláudio, T., & Sean, C. (2017). A survey of intrusion detection in Internet of Things. *Journal of Network and Computer Applications*, 25-37.