A survey on diagnostic methods for automotive engines

J Mohammadpour*, **M Franchek,** and **K Grigoriadis**
Department of Mechanical Engineering, University of Houston, USA

The manuscript was received on 29 December 2010 and was accepted after revision for publication on 11 August 2011.

DOI: 10.1177/1468087411422851

Abstract: Faults affecting automotive engines can potentially lead to increased emissions, increased fuel consumption, or engine damage. These negative impacts may be prevented or at least alleviated if faults can be detected and isolated in advance of a failure. United States Federal and State regulations dictate that automotive engines operate with high-precision onboard diagnosis (OBD) systems that enable the detection of faults, resulting in higher emissions that exceed standard thresholds. In this paper, we survey and discuss the different aspects of fault detection and diagnosis in automotive engine systems. The paper collects some of the efforts made in academia and industry on fault detection and isolation for a variety of component faults, actuator faults, and sensor faults using various data-driven and model-based methods.

1 INTRODUCTION

The basic concept of automotive onboard diagnosis (OBD) systems is to result in malfunction indicator light (MIL) illumination after a fault has been detected on two consecutive driving cycles. Pending fault codes are stored on the first detection and matured to 'active' or 'confirmed' codes once the MIL comes on. A defection is considered to progress to a fault when it leads to produced emissions that exceed prespecified thresholds.

The introduction of diagnostic systems in vehicles has not been facilitated by customers, as they do not necessarily consider this feature to be an important one when they are purchasing a vehicle. Instead, the installation of OBD systems has been enforced by regulatory requirements. By law, an on-road vehicle system must monitor for the deterioration of its emissions control system and issue a warning to the driver when necessary. In 1970, the US Congress passed the Clean Air Act (CAA) as law, which was designed to curb the impact of automotive emissions on the environment. At the same time, it became mandatory for car manufacturers to equip their vehicles with OBD features to detect emissions control performance deterioration. Around the same time, the Environmental Protection Agency (EPA) and California Air Resources Board (CARB) were also established [**1**]. In 1988, OBD standards were re-organized to align with the Society of Automotive Engineers (SAE) standards. In 1996, a new standard (called OBD II) was introduced. The new regulation mandated the automotive manufacturers to monitor a higher number of items. New regulations were first introduced in California and then spread elsewhere in the US. Similar regulations are in place in Japan and European countries. Listed in Table 1 [**2**] are the items that OBD II mandates to be monitored.

There have been a number of survey papers on diagnostics of automotive systems, including references [**3–5**], with a limited scope of topics and focus on engine subsystems. The present paper is the first comprehensive attempt to survey the work in the area of fault detection and diagnostics for automotive engines and after-treatment systems. The aim is to classify the most relevant research articles from an academic perspective. It should be noted that there are many patents issued or pending in this area that are not surveyed in the present paper.

The paper is organized as follows. In section 2, we present a review of the different methods (both model based and data driven) to diagnose a dynamic system. Section 3 presents a description of

**Corresponding author: Department of Mechanical Engineering, University of Houston, TX 77204-4006, USA.*
email: jmohammadpour@uh.edu

Table 1 The OBD II requirements [**3**]

Item	Requirements
Misfire monitoring	Multiple cylinder misfire, catalyst damage
Evaporative system monitoring	Purge flow, leak detection
Fuel system monitoring	Exceeds 1.5 times the standard
Air-conditioning system monitoring	Exceeds 1.5 times the standard
Secondary air system monitoring	Exceeds 1.5 times the standard
Oxygen sensor monitoring	Exceeds 1.5 times the standard, detection of a lack in circuit continuity
EGR system monitoring	Increase or decrease from the specified EGR flowrate, causing emissions to exceed 1.5 times the standard
PCV system monitoring	Monitors for disconnection of system tubing/hoses
Engine cooling system monitoring	Thermostat, engine coolant temperature (ECT) sensor
Cold start emission monitoring	Key control or feedback parameters
VVT system monitoring	Exceeds 1.5 times the standard
Catalyst monitoring	Non-methane organic gas (NMOG) $>$ 2.5 times the standard, conversion efficiency drops to 50% or lower

CARB regulations for vehicle OBDs and the recent efforts in both academia and industry to address the ever-growing diagnostic requirements. In section 4, we discuss different methods for detection and diagnosis of sensor faults and leaks in automotive engines. Section 5 reviews some of the emerging topics of interest, including engine combustion diagnostics, remote diagnostics, and integration of diagnostics and closed-loop control to improve engine performance and reliability.

2 APPROACHES TO FAULT DETECTION AND DIAGNOSIS OF ENGINEERED SYSTEMS

In this section, we review the various methods that are available for diagnostics of engineered systems classified under the two categories: data-driven methods and model-based methods. It is noted that this review is not, by any means, comprehensive. The interested reader is referred to references [**6**] and [**7**] for details on the methods described below.

2.1 Data-driven methods for fault detection

The effectiveness of any data-driven method depends heavily on the characterization of the process data variations. Since variations in the process data are inevitable, statistical theory plays a key role in most system monitoring and fault-detection schemes. Application of statistical theory to monitor processes depends on the assumption that the characteristics of the data variations are relatively unchanged unless a fault occurs in the system. This implies that the properties of the data variations, such as mean and variance, are repeatable for the same operating conditions, even though the actual sequences of the data might not be predictable. The repeatability of the statistical properties allows thresholds for certain measures that effectively define out-of-control status. A common approach for this purpose is using statistical methods for monitoring processes that employ the multivariate T^2-statistics. Let the data in the training set, consisting of m observation variables and n observations for each variable, be stacked into a matrix $\mathbf{X} \in \Re^{n\times m}$, given by $\mathbf{X} = [x_{ij}]_{i=1,\ldots,n\,\&\,j=1,\ldots,m}$. Then, the sample covariance matrix of the set is equal to $\mathbf{S} = \frac{1}{n-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$. An eigenvalue decomposition (EVD) of the matrix $\mathbf{S}$ given by $\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^{\mathrm{T}}$ shows the correlation structure for the matrix $\mathbf{S}$, where Λ includes the covariance matrix eigenvalues, and the eigenvector matrix $\mathbf{U}$ is orthogonal. The projection $\boldsymbol{y} = \mathbf{U}^{\mathrm{T}}\boldsymbol{x}$ of an observation vector $\boldsymbol{x}$ decouples the observation space into a set of uncorrelated variables corresponding to the elements of $\boldsymbol{y}$. Assuming that $\mathbf{S}$ is non-singular and defining $\boldsymbol{z} = \Lambda^{-1/2}\mathbf{U}^{\mathrm{T}}\boldsymbol{x}$, Hotelling's T^2-statistic is given by $T^2 = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}$ [**8**]. The T^2-statistic is a scaled squared 2-norm of the variation of an observation vector $\boldsymbol{x}$ from its mean. The scaling in the direction of the eigenvectors allows a scalar threshold to characterize the variability of the data in the m-dimensional observation space. Appropriate threshold values for the T^2-statistic can be determined by employing the probability distributions. Discussions on how to determine thresholds for the T^2-statistic based on a level of significance with application to fault detection and prognosis can be found in reference [**9**]. Various approaches for data-driven fault detection have been employed for engine diagnostics purposes that will be described next.

2.1.1 Principal component analysis (PCA)

Principal component analysis (PCA) is a linear dimensionality reduction technique which is optimal in the sense that it captures the variability of the data [**10**]. It determines a set of orthogonal vectors, called leading vectors, ordered by the amount of variance in the leading vector directions. In order to optimally capture the variations of the data while minimizing the effect of random noise capturing the PCA representation, the loading vectors corresponding to the k largest singular values of $\mathbf{X}$ are typically retained [**10**]. The motivation for reducing the dimensionality of the PCA representation is analogous to similar concepts used for pattern classification. Selecting the columns of the loading matrix $\mathbf{P} \in \Re^{m \times k}$ to correspond to the loading vectors associated with the first k singular values, the projections of the observations in $\mathbf{X}$ into a lower dimensional space are contained in the space matrix $\mathbf{T} = \mathbf{XP}$ and the projection of $\mathbf{T}$ back into the m-dimensional observation space $\hat{\mathbf{X}} = \mathbf{TP}^{\mathrm{T}}$. The difference between $\mathbf{X}$ and $\hat{\mathbf{X}}$ is the residual matrix $\mathbf{E}$, (i.e. $\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}}$). The residual matrix captures the variations in the observation space spanned by the loading vectors associated with the $m - k$ smallest singular values. The space contained in the matrix $\mathbf{E}$ has a small signal-to-noise ratio, and the removal of this space from $\mathbf{E}$ produces a more accurate representation of the process $\hat{\mathbf{X}}$. Defining $\boldsymbol{t}_i$ to be the i^{th} column of $\mathbf{T}$ in the training set, a new observation vector in the set $\boldsymbol{x} \in \mathbf{R}^m$ can be projected into the lower dimensional score space (spanned by $\hat{\mathbf{X}}$) by $t_i = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{p}_i$, where $\boldsymbol{p}_i$ is the i^{th} loading vector. The transformed variable t_i is called the i^{th} principal component of $\boldsymbol{x}$. With the vectors projected into the lower dimensional space using PCA, only k variables need be monitored, as opposed to the m variables without employing PCA. When enough data are gathered in the testing set, the score vectors $\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_k$ can be formed. If these score vectors do not satisfy the specific properties, the testing set is most likely collected during different operating conditions than for the training set [**11**]. This abstraction of structure from multidimensional data is a key factor of the score contribution method for fault identification and prognosis. In fact, for detecting different types of faults, T^2- and Q-statistics, along with their appropriate thresholds, are employed. The Q-statistic, also known as the squared prediction error (SPE), is a squared 2-norm that measures the deviation of the observations with respect to the lower dimensional PCA representation. In recent years, PCA has been successfully used for fault identification and diagnosis in automotive engines. Antory *et al.* [**12**] used a non-linear PCA technique to diagnose a sensor fault relating to a drift in the fuel flow reading, as well as a process fault relating to a partial blockage of the intercooler. King and Burnham [**13**] used PCA for the misfire detection purposes and in reference [**14**] for detecting small air leaks in diesel engines. In reference [**15**], non-linear PCA was used to illustrate the potential improvements compared to the linear PCA in detecting leaks (as small as 2 mm) in the inlet manifold plenum chamber of a diesel engine.

2.1.2 Partial least squares (PLS)

Partial least squares (PLS) can be successfully applied in supervised classification tasks for different applications (e.g. fault diagnosis and text classification). It is a dimensionality reduction technique for maximizing the covariance between the $n \times m$ independent training data matrix $\mathbf{X}$ and the $n \times p$ dependent matrix $\mathbf{Y}$ (corresponding to p fault classes) for each component of the reduced space. Here, n represents the number of patterns from p classes and m is the number of features. It builds a regression model between $\mathbf{X}$ and $\mathbf{Y}$. While forming the datasets, in each row of $\mathbf{Y}$, the fault class label corresponding to the pattern in $\mathbf{X}$ is represented by one in the corresponding column of $\mathbf{Y}$ [**16**]. The PLS generates uncorrelated latent variables, which are linear combinations of the original features. The basic idea is to select the weights of the linear combination to be proportional to the covariance between the features and pattern classes. Once the latent variables are extracted, a least-squares regression is performed to estimate the fault class. Both matrices $\mathbf{X}$ and $\mathbf{Y}$ are decomposed into a number of components, which is known as the model reduction order, plus residuals. Each component captures a certain amount of variation in the data. The reduction order is determined by cross-validation. Partial least squares can be viewed as a two-phase optimization problem [**16**], and its application in fault identification and prediction is discussed in detail in reference [**17**]. The use of PLS for oil monitoring and prediction of wear in diesel engines is described in reference [**18**].

2.1.3 Canonical variate analysis (CVA)

Canonical variate analysis (CVA) is a linear dimensionality reduction technique that is optimal in terms of maximizing the correlation between two sets of variables. The CVA theorem states that given a vector of variables $\boldsymbol{x} \in \mathbf{R}^m$ and another vector $\boldsymbol{y} \in \mathbf{R}^m$ with covariance matrices $\sum_{xx}$ and $\sum_{yy}$, respectively, and cross covariance matrix $\sum_{xy}$, there exist matrices $\mathbf{J} \in \mathbf{R}^{m \times n}$ and $\mathbf{L} \in \mathbf{R}^{n \times n}$ such that $\mathbf{J}\sum_{xx}\mathbf{J}^{\mathrm{T}} = \mathbf{I}_{\bar{m}}$, $\mathbf{L}\sum_{yy}\mathbf{L}^{\mathrm{T}} = \mathbf{I}_{\bar{n}}$, and $\mathbf{J}\sum_{xy}\mathbf{L}^{\mathrm{T}} = \mathbf{D} = \mathrm{diag}(\gamma_1, \ldots, \gamma_r, 0, \ldots 0)$ where $\gamma_1 \geqslant \ldots \geqslant \gamma_r$, and $\bar{m} = \mathrm{rank}(\sum_{xx})$, $\bar{n} = \mathrm{rank}(\sum_{yy})$. $\mathbf{I}_{\bar{n}} \in \mathbf{R}^{n \times n}$ is a diagonal matrix containing the first $\bar{n}$ elements as 1s and the rest of the elements as 0s [**8**]. The CVA is equivalent to generalized singular value decomposition (GSVD). A CVA-related approach in the multivariate statistics literature is known as the canonical correlation approach, which can be generalized into the CVA theorem. The CVA method has been extensively used for fault detection of engineered systems [**19**]. This method has also been used for automotive engine fault diagnostics. For instance, Larimore and Javaherian [**20**] refined the non-linear CVA method developed in the literature for engine airpath system identification and for detection of an air-to-fuel ratio sensor bias fault.

2.1.4 Black-box methods based on neural networks

Signal processing tools have been applied to many fault-diagnosis and prognosis problems. In many applications, it would be appealing to be able to predict the signs of occurred malfunctions based on physics or experience. For example, it may be expected that the appearance of the j^{th} harmonic of the primary rotational frequency in a rotating system is symptomatic of a specific kind of damage. In that case, the signal processing toolbox is well equipped to render the appropriate fault signal. However, sometimes, there does not seem to be such a well-defined body of knowledge available. What might actually exist is only a database of fault and no-fault signals. Under such circumstances, what is needed is a capability of pattern classification based on training data sets. Neural networks excel at recognition and classification types of problems. These techniques isolate faults by creating classifiers to identify faulty components or failure modes within components. The classifiers are trained using data obtained under both normal and faulty operating conditions. In this framework, a virtual test bench (VTB) is used as the data extraction tool to help extract the fault-test dependency information. The fault-test dependency measurements are used as input signature patterns to the various classifiers employed in the neural network toolbox. Most parts of the input data are used for training the neural networks, and the rest of the data are used to validate the trained network models. Cross-validation is also used in order to evaluate robustness of the network models. In the following, the classifiers implemented in the neural network toolbox will be briefly discussed. An excellent survey that appeared in 2000 [**21**] discussed the properties of several methods of combining quantitative and qualitative system information and their practical values for fault diagnosis of real engineered systems. In particular, the use of neural networks for automotive engines fault diagnosis has been studied extensively; for example, in sensor fault detection [**22**], intake manifold leakage, and exhaust gas recirculation (EGR) valve fault detection [**23**], crankshaft speed sensor, intake pressure sensor and throttle valve position sensor faults [**24**, **25**], air path sensors and valves [**26**], and intake manifold air leak during transient operation [**27**]. Some additional applications of the use of neural networks for automotive engine system monitoring and fault detection will be described in the next sections.

2.1.5 Restricted Coulomb energy (RCE) classification

The restricted Coulomb energy (RCE) classifiers rely on the approximation of a decision region via a union of hypersphere cells [**28**]. An RCE consists of three layers of cells with a full set of connections between the first and second layers, and a partial set of connections between the second and the third layers. The use of RCE for engine system fault detection has been investigated in reference [**29**].

2.1.6 Decision trees

We have described two pattern classification methods based on feature vectors of real-valued numbers in which there has been a natural measure of distance between vectors. However, there is another common case in which we specify the values of a fixed number of properties by a *d-tuple* and describe the pattern by a variable-length string of nominal attributes. It is natural and intuitive to classify a pattern through a sequence of questions, in which the next question depends on the answer to the previous question. A decision tree classifier partitions the space of possible observations into subregions through the answers (which must be one of a discrete set of alternatives) of a series of questions. In a wide range of applications, decision trees have resulted in classifiers with accuracy comparable to other methods, such as neural networks and nearest neighbour classifiers, particularly when specific prior information about the appropriate form of classifier is lacking [**28**]. The use of decision trees for fault diagnosis of automotive engines was studied in reference [**30**], where the authors proposed an approach based on qualitative deviation models for the automatic derivation of OBDs using decision trees. In reference [**30**], the common-rail fuel delivery system was investigated as a case study.

2.1.7 Fuzzy adaptive resonance theory mapping (fuzzy ARTMAP)

Adaptive resonance theory mapping (ARTMAP) is a class of network architectures that perform incremental supervised learning of recognition categories and multidimensional maps in response to input vectors presented in arbitrary order. Fuzzy ARTMAP is a more general ARTMAP system that learns to classify inputs by a fuzzy set of features, or a pattern of fuzzy membership values between 0 and 1, indicating the extent to which each feature is present. This method has been applied in misfire detection in automotive engines [**31**], air flow sensor fault detection [**32**], diesel engines oil analysis diagnosis [**33**], sensor and actuator calibration errors [**34**], as well as detecting a variety of faults including lean and rich fuel mixture, fuel system leakage, and low compression [**35**].

2.1.8 Single and multiple layer perceptrons

Perceptrons can be either single or multilayer networks with threshold, logistic, or hyperbolic tangent function activation units. A perceptron network usually uses a fixed set of processing elements followed by a layer of adaptive weights and a threshold activation function. The processing elements typically have threshold activation functions. The multiple layer perceptron implemented for fault-diagnosis purposes uses back propagation for training the network [**36**]. Some of the applications of these methods to engine fault diagnostics have appeared in references [**12**] and [**37–39**].

2.1.9 Radial basis function (RBF) networks

Radial basis function (RBF) networks are a major class of neural network models, where activation of a hidden unit is determined by the distance between the input vector and a prototype vector. The RBFs have their origins in the techniques for performing exact interpolation of a set of data points in a multidimensional space. The use of RBFs for fault diagnosis of automotive engines has been investigated in references [**40–42**].

2.2 Model-based fault-detection and diagnosis methods

The main advantage of model-based approaches is the ability to incorporate physical understanding of the process for system monitoring. Another advantage is that, in many cases, the changes in the feature vector are closely related to model parameters. Therefore, it can also establish a functional mapping between the drifting parameters and the selected diagnostic features. Moreover, based on such understanding of the system degradation, the model can be adapted to increase its accuracy and to address delicate performance issues. Consequently, in general, model-based methods can outperform data-driven approaches. A literature survey shows that most applications of the model-based approaches have been on systems with a relatively small number of inputs, outputs, and states. It is indeed difficult to apply this approach to large-scale systems since it requires a detailed process model in order to be effective. Detailed models for large-scale systems are cumbersome to obtain due to the present coupling between different actuator and sensor signals in a multivariable system. For model-based fault detection, various methods have been employed, including parity equations, parameter estimation, state variable estimation, or neural networks [**6**, **43**, **44**]. The basic concepts of some of these methods are summarized below.

2.2.1 Parameter estimation

In parameter estimation, the residuals are the difference between the nominal model parameters and the estimated model parameters (based on the measurements and the least-squares method results). Such deviations in the model parameters serve as the basis for detecting and isolating faults. The parameter estimation-based methods not only detect and isolate a fault, but can also estimate its size. A key requirement for this method is that a mathematical model should be identified and validated to express the physical behaviour of the system reasonably well. If the nominal parameters are not exactly known, they are required to be estimated from observed data. Two different parameter identification approaches that are available for this purpose are described below [**6**].

(a) *Equation error methods.* Equation error methods use the fact that faults in dynamic systems are reflected in physical parameters. A five-step procedure has been proposed in the literature for parameter estimation as follows [**45**]:

(a) obtain a mathematical model of the system relating the measurable input and output variables $y(t) = f\{u(t), \boldsymbol{\theta}_0\}$, where $\boldsymbol{\theta}_0$ is the nominal parameter vector;
(b) determine the relationship between the model parameters $\boldsymbol{\theta}$ and the physical system coefficients p as $\boldsymbol{\theta} = g(p)$;
(c) identify the model parameter vector $\boldsymbol{\theta}$ from the measured data $\boldsymbol{Y}^N = \{y(k)\colon\ 0 \leqslant k \leqslant N\}$ and $\boldsymbol{U}^N = \{u(k)\colon 0 \leqslant k \leqslant N\}$;
(d) calculate the system parameters $p = g^{-1}(\boldsymbol{\theta})$ and the deviations from the normal coefficients $p_0 = g^{-1}(\boldsymbol{\theta}_0)$ and compute the residual $\Delta p = p - p_0$;

(e) diagnose faults by using the relationship between system faults and the deviations in the coefficients Δp.

For a linearized model with lumped parameters, the estimate of the parameter vector can be computed via the recursive least-squares (RLS) algorithm. The least-squares parameter estimation requires the time derivatives of the noisy input and output variables, and estimation of these time derivatives poses a considerable numerical challenge. The equation error method has been used for fault diagnosis in an electromechanical automotive throttle valve actuator in reference [**46**].

(b) *Output error (prediction-error) methods*. For a multiple input–multiple output system, assume that we have collected data from the system $Z_N = [y(1), u(1), y(2), u(2), \ldots, y(N), u(N)]$. Let the output error be provided by a certain model, and parameterized by θ, given by $e(k,\theta) = y(k) - \hat{y}(k|\theta)$. Let the above output-error sequence be filtered through a stable filter L and assume the filtered output to be $\boldsymbol{e}_\mathrm{F}(k,\theta)$. The estimate $\hat{\theta}_N$ is then computed via $\hat{\theta}_N = \arg\min_\theta V_N(\theta, Z_N)$, where $V_N(\theta, Z_N) = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{e}_\mathrm{F}^\mathrm{T}(k,\theta)\Sigma^{-1}\boldsymbol{e}_\mathrm{F}(k,\theta)$, and $\sum$ is the covariance of the error vector. The above minimization problem can be solved iteratively, and the updated parameter estimates at the i^{th} iteration are

$$\hat{\theta}_N^{i+1} = \arg\min_\theta \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{e}_\mathrm{F}^\mathrm{T}(k,\theta)\left[\hat{\sum}_N^i\right]^{-1} \boldsymbol{e}_\mathrm{F}(k,\theta)$$

$$\hat{\sum}_N^i = \frac{1}{N-1}\sum_{k=1}^{N} \boldsymbol{e}_\mathrm{F}(k,\theta_N^i)\boldsymbol{e}_\mathrm{F}^\mathrm{T}(k,\theta_N^i)$$

Analogous to the RLS algorithm, one can also derive a recursive version for the output-error methods. In general, the function $V_N(\theta, Z_N)$ cannot be minimized by analytical methods and hence the solution should be obtained numerically. The computational effort of this method is significantly large and real-time applications may not be practical. However, this approach provides more accurate parameter estimates than the equation error method [**45**]. Some applications of these approaches for engine diagnostics purposes will be described later in section 3.

2.2.2 Observer-based methods

The observer-based methods reconstruct a prediction of the system outputs from the measurements or a subset of the measurements with the aid of observers. The difference between the measured outputs and the estimated outputs is then used as the vector of residuals to detect and isolate faults. These methods are appropriate if the faults are associated with changes in actuators, sensors, or unmeasurable states; that is, they are especially suitable for detecting and isolating additive faults. The observer-based methods are in contrast to the CVA-based method for process monitoring, in which the states are directly constructed from the process data, rather than through the use of a known process model and an observer. The design equations for an unknown input observer (UIO) are given in reference [**6**], where a general method for the design of an observer-based fault detection and isolation (FDI) system is presented. In the UIO, the observer matrices are designed such that the residual and estimation error are independent of the plant input, plant states, and the disturbance signals. Chen and Saif [**47**] proposed a method of fault detection for a class of time-delay non-linear systems based on an iterative learning observer. The method was shown to be able to provide accurate disturbance estimation to the control system when no fault is present. In the presence of a fault, it showed the ability to detect and estimate the faults. The method was used to diagnose the faults in a spark ignited (SI) engine intake manifold system. A non-linear observer design approach was also taken in reference [**48**] to detect misfire events in internal combustion (IC) engines. Hsu *et al.* [**49**] proposed a method for engine fault detection when both sensor and actuator faults take place at the same time, where the challenge is their mutual effects on residuals. In reference [**49**], a decision logic was proposed by relating all the possible failure patterns to the residual code. The use of observer-based design methods for fault detection in automotive engine applications will be discussed in detail in the following sections.

Another commonly used method for observer design is the Kalman filtering technique, in which the optimal state estimator is derived for a set of linear system governing equations under the assumption of measurement white noise. However, there are situations where the system dynamics is largely altered following a standard linearization procedure. In this case, the use of an extended Kalman filter (EKF) has been proposed to derive solutions for non-linear systems. The use of the EKF for diagnostics of a variety of sensors, actuators, and engine component faults will be described in the next sections.

2.2.3 Parity relations

This method checks the consistency of the mathematical equations of the system model with the measured data. The parity relations are subjected to a linear dynamic transformation, with the transformed residuals used for detecting and isolating faults.

Parity relations are rearranged forms of the input–output or state-space equations of the system. The essential characteristic of this approach is to check for consistency of the inputs and outputs. Under normal operating conditions, the magnitudes of residuals or the values of parity relations are small [**50**]. To improve residual-based fault isolation, directional, diagonal, and structured residual design schemes are proposed. In the directional residual scheme, the response to each single fault is confined to a straight line in the residual space. If the response directions are independent, directional residuals support fault isolation. In the diagonal scheme, each element of the residual vector responds to only one fault. Diagonal residuals are ideal for the isolation of multiple faults, but the method can only handle r faults, where r equals the number of outputs. Structured residuals are designed to respond to different subsets of faults, and they are insensitive to process variations that are not in these subsets. For a detailed discussion on the application of the parity relation approach to identify and diagnose the presence of faults, the reader is referred to reference [**50**].

The application of the directional residual design has been demonstrated on a linear discrete-time model of an automotive engine [**51**]. In the context of diagnosing faults in an automotive engine, Krishnaswami *et al.* [**52**] employed a non-linear auto-regressive moving average modelling with exogenous inputs (NARMAX) technique for system identification in order to implement a non-linear parity equation residual generation (NPERG) scheme. The non-linear parity generation method [**53**] and parameter estimation methods (based on an auto regressive moving average (ARMA) model) were used to detect and isolate process coefficient faults in electronic throttle control systems. Monte and Pisu [**54**] presented a fault-detection method for the idle speed control of an IC engine considering throttle faults and manifold pressure sensor fault, where the proposed approach is based on the residual generators using the parity equation and an active threshold calculation. It is noted that parity equations require less computational effort, but do not provide as much insight into the process as parameter estimation schemes.

To summarize this section, fault detection using model-based signal analysis and residual processing is performed as follows. The residuals are generated from the comparison between the observed features of the system and the nominal behaviour of the process. Detectable deviations of the residuals result in symptoms, which are then processed in the fault-diagnosis module by means of fault–symptom–causalities [**55**].

2.3 Structured hypothesis testing for fault detection

In the structured hypothesis test (SHT) framework initially proposed in reference [**56**], it is possible to use detailed models for each fault of interest. When using SHT, the different faults are classified into different behavioural modes, where one of the modes is the healthy (no-fault) case. The modes list is chosen according to how often the faults occur, how difficult they are to diagnose with non-model-based algorithms, and how significant an effect they have on the system. An illustrative example (taken from [**57**]) is presented next to describe the principle of the fault-diagnosis method using SHT.

Let F_p denote the present behavioural mode (healthy or faulty). To describe the i^th hypothesis test, introduce the set M_i, which is a set of behavioural modes. The so-called null hypothesis and the alternative hypothesis can then be represented as

$$H_i^0 : F_\mathrm{p} \in M_i$$
$$H_i^1 : F_\mathrm{p} \in M_i^\mathrm{C}$$

This classification corresponds to the two cases where 'some behavioural mode in M_i can explain the measurement data' and 'no behavioural mode in M_i can explain the data', respectively. The convention is that when H_i^0 is rejected, H_i^1 holds true. However, when H_i^0 is not rejected, no conclusion can be drawn most of the times. Overall, each hypothesis test contributes a piece of information eventually leading to a decision on the behavioural modes that can be matched to the data. The following example shows how a set of different hypothesis tests can be used to diagnose and isolate faults.

Example 1. Assume that the diagnosis system contains the following set of three hypothesis tests

$$\begin{array}{ll} H_1^0 : F_\mathrm{p} \in M_1 = \{NF, F_1\} & H_1^1 : F_\mathrm{p} \in M_1^\mathrm{C} = \{F_2, F_3\} \\ H_2^0 : F_\mathrm{p} \in M_2 = \{NF, F_2\} & H_2^1 : F_\mathrm{p} \in M_2^\mathrm{C} = \{F_1, F_3\} \\ H_3^0 : F_\mathrm{p} \in M_3 = \{NF, F_3\} & H_3^1 : F_\mathrm{p} \in M_3^\mathrm{C} = \{F_1, F_2\} \end{array}$$

Then, if only H_1^0 is rejected, we can imply that $F_\mathrm{p} \in M_1^\mathrm{C} = \{F_2, F_3\}$, which means that the present behavioural mode is either F_2 or F_3. If both H_1^0 and H_2^0 are rejected, we can imply that $F_\mathrm{p} \in M_1^\mathrm{C} \cap M_2^\mathrm{C} = \{F_3\}$, meaning that the current mode is F_3. With the behavioural mode defined initially corresponding to the healthy or faulty cases (multiple faults can be easily handled simultaneously in the SHT framework), the diagnosis problem is then stated as follows: 'Given measurement data, which different models, defined by the different behavioural modes, can best represent the measured data?' [**57**]. Nyberg and colleagues have extensively used the SHT for fault detection and diagnostics in both gasoline and diesel engines [**56–62**].

3 MONITORING REQUIREMENTS FOR AUTOMOTIVE ENGINES AND RELEVANT WORK

This section details the major monitoring requirements for modern automotive engines. For each monitor, we describe its purpose, what quantity needs to be detected (i.e. 'malfunction criteria'), and some of the recent work that describes its usage. It is important to note that OBD regulations only require the system to be designed and calibrated to detect a *single component* failure at the required malfunction criteria rather than having to detect every combination of multiple component degradations that can cause emissions to exceed the malfunction threshold (e.g. 1.5 times the standards). In other words, OBD is not required to take into account synergistic effects of multiple component failures. For example, when calibrating an EGR low flow fault that would exceed the threshold, manufacturers would be required to implant only a low flow fault in the EGR system and leave other emission control components and subsystems (e.g. catalysts, particulate matter (PM), filters) in the nominal condition. It is noted that research on some areas of engine fault diagnostics including three-way catalyst (TWC) monitoring, misfire monitoring, leak detection, and universal exhaust gas oxygen (UEGO) sensor fault detection has matured enough to allow implementation in production vehicles. Other areas including nitrogen oxide (NO_x) sensor fault-detection, EGR, lean NO_x trap (LNT), and diesel particulate filter (DPF) monitoring and emerging topics, including advanced combustion and after-treatment on diesel low-temperature combustion (LTC), and premixed charge compression ignition (PCCI) / homogeneous charge compression ignition (HCCI), are still in a developing research phase.

3.1 Fuel system

Manufacturers are required to detect fuel system faults that cause emissions to increase. The faults are often associated with the fuel system pressure control (e.g. common-rail fuel pressure control or hydraulic pressure control), and the focus is on detecting faults when the feedback system can no longer deliver the desired pressure. Given the critical importance of proper fuelling for emission control, monitoring for appropriately injected fuel quantity and injection timing are also required.

A fuzzy-based pattern recognition method has been investigated in reference [**63**] for real-time detection of fuel injection system faults in a diesel engine. The fuel system health diagnosis system consists of a piezoelectric pressure sensor to measure fuel injection pressure patterns and a fault-diagnosis algorithm to detect abnormal injection pressure patterns and identify the causes contributed to the abnormal patterns. A multinet (artificial neural network (ANN)-based) diagnosis algorithm was proposed in references [**39**] and [**64**] to detect a leaking fuel injector nozzle in a diesel engine, where it only used a pressure transducer. The use of an ANN is justified here due to the viability of an online neural net based fault-diagnosis system using the information from cylinder pressure only. Non-linear estimators using a sufficiently accurate model of the powertrain system of an SI direct injection engine were designed by Lee *et al.* [**65**] to detect different actuator faults, including high-pressure fuel injectors. Since the coupled breathing and fuelling system is highly non-linear in nature, the fault-diagnosis method developed in reference [**65**] works by generating residuals based on multiple non-linear observers. Recently, Schilling *et al.* [**66**, **67**] developed a system to detect and isolate faults due to ageing of the air and fuel path of common-rail direct injected diesel engines using an algorithm based on the information obtained from lambda (air-to-fuel ratio) and NO_x emissions sensors. Faults corresponding to the quantity of the injected fuel, the mass air flow (MAF), and the manifold air pressure (MAP) sensors are taken into account to explain discrepancies in the expected lambda and/or NO_x measurements. The fault-detection method of the latter work is model based and uses a bank of EKFs. This work is the first reported to use emission sensors for fault-detection purposes in diesel engines. The reason to use an emission sensor, as proposed by reference [**67**], is the fact that faults in the air or fuel path have a direct impact on emissions.

Payri *et al.* [**68**] proposed a diagnosis method for the injection process using rail pressure measurements. The authors explored and evaluated different data-driven techniques to detect faults in common-rail injection systems. The reason to use rail pressure for diagnostics purposes is the fact that the injection event leads to an instantaneous drop in the rail pressure, whose variations are also affected by the dynamics of the high-pressure pump that supplies fuel to the rail. Chandroth *et al.* [**69**] proposed to use cylinder pressure and vibration data to detect the presence of a blockage in the fuel injector and poor fuel atomization. The method involved training two sets of ANNs and usage of the features extracted from the cylinder pressure measurements and vibration amplitudes. The goal was to demonstrate the advantages of using sensor fusion or decision fusion compared to a single-sensor system.

3.2 Misfire and knock monitoring

For the 2010 to 2012 model year, manufacturers are required to detect malfunctions that cause a complete single-cylinder misfire (e.g. one cylinder completely dead). For 2013 and subsequent years, misfire monitoring will be required to be performed continuously (under all loads and speeds) and to look for lower levels of misfire (i.e. a cylinder or combination of cylinders that are intermittently misfiring).

3.2.1 Misfire detection

Engine misfire detection is an important element of OBD systems since engine misfire can induce an increasing level of exhaust emission and potentially damage the catalytic converters. Many methods have been proposed in the literature to address this problem, including algorithms based on variation in engine shaft angular speed (also acceleration and torque), spark-plug voltage [**70**], oxygen sensor signal [**71**], knowledge-based expert system, and neural networks.

Methods using variations in crankshaft velocity, acceleration, and torque variations share the same basic principle described next [**72**]. There are n firing events during one engine cycle for an n-cylinder engine. Each firing produces a power pulse to the engine, and these pulses cause shaft fluctuations in velocity, acceleration, or torque. If the effects of the inertia torque, load torque, and friction and pumping torques are ignored, each fluctuation in velocity, acceleration, or torque is directly related to one wave of combustion power contribution. Theoretically, any abnormality in power contribution will be reflected in the variations in velocity, acceleration, or torque. Therefore, examining these variations can provide a measure of cylinder misfire. Methods in this category are classified into two main approaches.

1. The first approach is based on the evaluation of the instantaneous angular velocity signal *without* using an engine model. These methods evaluate the characteristics of the time-domain, angular-domain, or frequency-domain signal engine speed. The extracted features are then used to detect misfire through simple threshold checks or more complex decision-making algorithms (e.g. [**73–77**]). Most of these algorithms give satisfactory results at low speeds, but due to the lack of a proper engine model, it is difficult to correct the influence of the inertia torque at higher engine speeds.
2. The second approach is based on the use of model-based techniques, where a dynamic engine model to estimate the indicated torque or in-cylinder pressure is used. Rizzoni and Connolly and Rizzoni [**78**, **79**] proposed an algorithm to estimate the effective torque based on the deconvolution of the engine speed signals in the frequency domain. As the inertia torque depends on the mean angular speed, this term is added to obtain the indicated torque. Connolly and Rizzoni and Rizzoni *et al.* [**80–83**] proposed the use of sliding-mode observers to estimate the indicated torque, while Kiencke [**84**] proposed a Kalman-filter-based algorithm. To tackle the issue of high complexity of the torque-estimation methods discussed above, researchers have examined methods to improve real-time implementability and its use for online misfire detection [**85**, **86**].

A large number of papers have appeared in the literature on misfire detection using neural networks and knowledge-based expert systems. The methods in this category are computationally less intensive, but they suffer from low flexibility due to the need for retuning a specific engine model. Some of these efforts are based on pattern recognition and classification (e.g. [**76**, **87**]) or the evaluation of the instantaneous angular speed through neural networks (e.g. [**88**, **89**]). Energy models have been used for misfire detection by Tinaut *et al.* [**90**], where they defined two dimensionless energy indices for each cylinder, with the first index evaluating the change in kinetic energy during the compression stroke, and the second one evaluating the change in kinetic energy during the expansion stroke. These two indices collectively provided a tool to detect the fault condition of each cylinder. The algorithm developed in reference [**90**] is based on the model proposed by Lim *et al.* [**91**] and Schmidt *et al.* [**92**], which estimated the indicated torque of the engine from instantaneous engine speed through an energy model. To this end, they considered the difference in kinetic and potential energy between two samples and derived the corresponding indicated torque. The rationale behind the use of energy models for misfire detection is that these models are conceptually easy to understand and are based on straightforward calculations.

The methods proposed in references [**77**] and [**93**, **94**] were aimed at: (i) detecting a missing combustion event and (ii) classifying the event into either a misfuel event (i.e. missing injection) or misfire (i.e. missing ignition) using feedback from an appropriate signal. A different approach to misfire detection was proposed in references [**95–97**], where the authors described new analysis techniques based on a wavelet approach allowing for the extraction of the frequency components related to a misfire event and its localization in the time domain. The use of wavelet for misfire detection was justified as follows: the high-frequency components of engine-block vibration signals during misfire and normal combustion show clear differences in duration and scale when a continuous wavelet transform (CWT) of the output signal is taken.

3.2.2 Knock monitoring

Engine knock is caused by spontaneous ignition of a portion of the air–fuel mixture during the combustion cycle. The very fast release of chemical energy in the mixture results in high local pressure and produces a shock wave. This shock leads to the resonance of the combustion chamber, thus producing the knocking sound. Excessive knock could lead to engine damage, and hence knock detection is a very important requirement for the engine control unit.

A number of methods have been proposed to detect the knock phenomenon in SI and IC engines. Samimy and Rizzoni [**98**] used joint time–frequency signal processing methods to detect knock in IC engines. The idea is based on the use of the relationship between the engine excitation frequency, taking into account the combustion chamber geometry, and the speed of sound in the cylinder charge. The frequency can be estimated using an acoustic model of the combustion chamber given by Draper's equation

$$f_{m,n} = \frac{c_0\sqrt{T}\eta_{m,n}}{\pi B} \tag{1}$$

where f is the resonance frequency, $\eta_{m,n}$ is a non-dimensional parameter, and the integers m and n refer to the radial and circumferential mode numbers. The parameter c_0 is the phase velocity constant, T is the gas temperature, and B is the cylinder bore diameter. From equation (1), another equation was extracted in reference [**98**] to enable the prediction of the existence of frequency shifts in the knock signals, suggesting that conventional knock-detection methods employing a stationary signal model can be improved by applying a time-varying signal-detection method.

In laboratory applications, the in-cylinder pressure can be used to identify the knock characteristics. A number of statistical-based methods using the in-cylinder pressure signal have been proposed in the literature as a means to determine the knock intensity (KI). The most commonly used KI is the absolute value of the peak magnitude of the filtered in-cylinder pressure defined as $KI = \max|p_{\mathrm{ic}}|$, where p_{ic} is the band-pass filtered pressure signal. The filter cut-off frequencies are selected depending on the engine resonance frequency characteristics given by equation (1) [**99**]. Borg *et al.* [**99**] presented a method to determine the knocking condition of an SI engine using the discrete wavelet transform as a means of analysing the engine-block vibration signal and a fuzzy inference scheme to generate an estimate of the knock intensity introduced above. The block vibration sensor responds to knock-induced pressure waves produced when a portion of the gas mixture in the cylinder combusts spontaneously. Previous efforts had been also made in detecting knock using the CWT [**95**, **100**]. In addition to wavelet transform, Fourier analysis has also been used for knock detection [**101**, **102**], where the spectral intensity of the knock resonance frequencies is used as the statistical test for knock determination. Other methods to detect knock (among many parametric and non-parametric methods) are: (i) the use of non-parametric statistical methods [**103**], (ii) the use of parametric methods based on neural networks [**104**], and (iii) the use of an ARMA parametric model, where the change in one of the model parameters estimated in real time can detect the knock, even in the initial stage of knock [**105**].

3.3 Exhaust gas recirculation system monitoring

The EGR system is required to be monitored for three primary failure modes: low flow, high flow, and slow response to achieve the desired flow. Exhaust gas recirculation is one of the primary NO_x emission control mechanisms for the majority of engine manufacturers, and it is critical that the desired flowrate is being delivered. Accordingly, most manufacturers use feedback control systems to modulate the EGR valve to achieve a desired flowrate. The feedback system usually uses an MAF sensor, and the system compensates for small errors to achieve the desired flowrate. As long as the system can provide the desired flowrate, NO_x emissions stay relatively low.

However, when the system can no longer achieve the flow it needs, or it takes too long to reach the desired recirculation flow, NO_x emissions can increase dramatically. For a system that is feedback controlled to an actual flowrate, this emission increase should not occur until the system is close to its control limits (e.g. cannot compensate and deliver the desired flowrate). In addition, the performance of the EGR cooler would also need to be monitored to ensure it has sufficient cooling capacity.

Both data-driven and model-based methods have been proposed to detect faults in the EGR system. Gaussian RBF neural networks with adaptive classifiers were employed in reference [**106**] to detect a stuck EGR valve failure in SI engines. A neural-network-based method using self organizing maps (SOMs) was employed in reference [**107**] to detect malfunctions of the EGR system of a passenger car diesel engine. The SOM outputs a measure of similarity to 'typical system behaviour patterns', and for the OBD system, this value can be used as a metric for system anomaly detection. Semi-physical models (identified with local linear neural networks) were used in reference [**57**] to detect a leaky or stuck EGR valve in combustion engines, where residuals were generated using signal models and filters. Another diagnosis system for the diesel engines proposed in the SHT framework (discussed in detail previously) was developed to detect an EGR valve stuck in a closed position in reference [**62**], where it was shown that this framework is a useful engineering tool to systematically design model-based diagnosis systems. Another model-based fault-detection method to identify and isolate EGR valve actuator faults was proposed by making use of non-linear estimators for a model of an SI direct injection engine in reference [**65**]. A more recent attempt to address the EGR system diagnostics in diesel engines was made by Mohammadpour *et al.* [**108**], where their model-based method (based on the standard orifice flow equation representing the EGR flow back to the intake manifold) was used to detect low-flow and high-flow faults in the EGR system. The proposed fault-detection scheme used a recursive total-least-squares (RTLS) method to estimate two parameters, whose changes were shown to be indicative of the fault type and its severity. The use of RTLS compared to RLS was appealing due to the robustness of the former against model parametric uncertainties.

3.4 Boost pressure

Most manufacturers use some form of feedback control involving the manifold pressure sensor, as well as increasing the use of variable geometry turbo (VGT) with vane position and/or turbine speed. While boost pressure control problems may not have as much impact on emissions as EGR or fuel system problems, operating at incorrect boost levels can increase emissions over time. In addition to the boost pressure, intercooler performance would also be required to be monitored to make sure it provides sufficient cooling of the charge air. Different methods to diagnose a boost pressure sensor will be described later in section 4.1.

3.5 After treatment subsystems

In *diesel engines*, oxidation catalysts located upstream of the PM filter should be monitored to help the regeneration. The requirement also covers monitoring of the other hydrocarbon (HC) converting catalysts, such as NO_x adsorbers and selective catalytic reduction (SCR) catalysts. The NO_x catalysts, including LNT catalysts and SCR catalyst systems, should be monitored. In general, the catalyst itself would be monitored to make sure it has sufficient NO_x conversion to keep emissions below a threshold, while additional components such as the SCR injection system components (urea or ammonia) are monitored for proper functioning. For the 2010 to 2012 model years, the catalysts would need to be monitored, and a fault needs to be detected when NO_x emissions exceed the standards by an additional 0.3 gram per brake-horsepower per hour (g/bhp-hr). For instance, for engines certified to a 0.2 g/bhp-hr standard, a fault would need to be detected when NO_x emissions reach 0.5 g/bhp-hr. In 2013 models, however, the threshold will drop down to the standard plus a 0.2 g/bhp-hr value jump. For diagnostics purposes, the same NO_x sensor for feedback control must be also used for monitoring. For non-feedback SCR systems or passive LNTs, manufacturers are expected to be reaching lower conversion efficiencies. In the following sections, we discuss the efforts recently made on fault diagnosis of engine after treatment system to improve the emission control system reliability. For PM emissions, the regulations require manufacturers to detect faults leading to the PM threshold of 0.05 g/bhp-hr for 2010 to 2012 model years and 0.025 g/bhp-hr for 2013 and subsequent years. For HC emissions, monitoring would be required to verify that the HC catalyst has sufficient conversion efficiency to maintain emission below 2.0 times the HC standards for 2010 to 2012 model year engines and below 1.5 times the standards for 2013 and subsequent year engines. In gasoline engines, the OBD system monitors the catalyst system for proper conversion capability. The fault needs to be detected when the catalyst system's conversion capability drops to the point that emissions roughly exceed 1.75 times the federal test procedure (FTP) full useful life standards to which the vehicle has been certified.

3.5.1 Diesel particulate filter system monitoring

As described in detail in reference [**109**], the only technology available to meet the DPF leakage monitoring requirement in 2007 was a pressure sensor combined with a flow measurement. This was generally found to be of limited capability [**110**] due to little or no separation between healthy and damaged filters and far-reaching implications on the monitor frequency. The main contributing factor in the limited performance of the DPF diagnostics methods based on pressure sensors is the high tolerance caused by the sensor due to noise factors not measured by the engine control system, as well as the driving conditions under which the monitoring occurs. It is suggested by CARB that model-based methods may result in a more accurate detection rather than merely a pressure sensor-based monitoring.

Most DPF leakage monitors currently in production are based on pressure drop. The Darcy–Forchheimer equation modelling the pressure drop across the DPF with a constant soot loading is given by reference [**111**]

$$\Delta p = a_0 + a_1 \upsilon F + a_2 \rho F^2 \tag{2}$$

where υ, F, and ρ represent the kinematic viscosity of the exhaust gas, volumetric flow, and gas density, respectively. For a DPF loaded with soot, the above equation is rewritten as

$$\Delta p = a_0 + a_1 R(\text{soot}) \upsilon F + a_2 \rho F^2 \tag{3}$$

with R being the (normalized) restriction, which is a function of soot. The monitor detects a leakage in the DPF system when the pressure drop Δp is much lower than what the right-hand side of equation (2) predicts. Note that this threshold depends on temperature and exhaust flow. There are many reasons contributing to the discrepancy between the Darcy–Forchheimer model output and reality. These, along with an inaccurate approximation of the model coefficients a_0, a_1, and a_2 estimated from experimental data, taken with imperfect measurement equipment, lead to a far-from-accurate leakage-detection method. The work by Cunningham *et al.* [**112**] provides a promising extension to the mean value pressure drop correlation to particulate load through Darcy's law, which is expected to be useful for DPF monitoring and control. van Nieuwstadt and Brahma [**109**] investigated the ability of model-based DPF leakage detection over the pressure-sensor-based DPF leakage monitor. They presented the noise factors entering the relevant models and a numerical evaluation to assess the capability of the model-based leakage monitor under typical ranges of the noise factors. A recent work by Surve [**113**] proposed to correlate the pre- and post-DPF temperature and pressure signals to define the transfer function characteristics for the baseline DPF behaviour. Assessment of how these characteristics change as a result of a fault in DPF forms the basis of the proposed fault-detection algorithm in reference [**113**]. The method achieved a fault detection of lightly failed DPF not possible by current algorithms based on mean value pressure drop. In fact, the main contribution of reference [**113**] was the extension of dynamic pressure signal analysis from steady-state engine operation (proposed in [**112**]) to transient operating conditions. Gheorghiu *et al.* [**114**] proposed to use a *spark discharge soot sensor* to detect the presence of a crack in the filter, causing the filter to become no longer airtight. The sensor could be installed before the DPF or the exhaust gas turbo charger, which, similar to the lambda sensor of the TWC, regulates the soot emissions in a feedback loop.

The DPF diagnostics have also been investigated from the soot loading perspective. Sappok *et al.* [**115**] studied the use of radio frequency (RF)-based sensing techniques to directly monitor soot accumulation in DPFs, where the RF system was configured to generate and transmit RF signals across the DPF over a wide frequency range. The results of the steady-state testing reported in reference [**115**] reveal the feasibility of using RF sensing to monitor soot loading in DPFs. The RF sensor measurements showed quite a good sensitivity for monitoring even low-level PM accumulation on the DPF. Compared to the pressure drop measurements, RF sensing has the advantage of being unaffected by flowrate variations, and demonstrates a stable performance, even at relatively high exhaust temperatures. More recent efforts for soot loading diagnostics have also been made using the developed PM sensors built from multilayer ceramic sensor technology [**116**, **117**]. The use of these sensors makes the detection of lower soot concentrations possible due to the fact that the sensing method is less influenced by particulate conductivity [**116**].

3.5.2 Lean NO_x trap system monitoring

The LNT technique is one of the promising technologies for controlling NO_x. It can be used in two modes [**118**]. One is a lean-to-rich switching mode in which NO_x in the exhaust is adsorbed when the engine runs under normal lean conditions and released in the form of nitrogen when the engine runs under rich conditions. The other mode is an active LNT technology, in which engine fuel is injected upstream of the catalyst to provide a reducing agent for the nitrate regeneration. There are two main challenges associated with LNT technology. One is the control of the transition between lean and rich phase operation, which is critical to achieve the optimal tradeoff among competing requirements such as fuel economy, drivability, and emissions. A second challenge is the detection of LNT malfunction. The LNT can experience performance deterioration and malfunctions that may go undetected. A simplified storage model that can be integrated into the existing control strategy for real-time LNT control and diagnosis was developed in reference [**119**], which captures the dynamics of NO_x adsorption, reaction rate, and physical mass transfer process. Deactivation of the LNT catalysts is one type of fault that can compromise the NO_x conversion efficiency if it is not properly monitored and compensated for. Thermal exposure during high-load operating conditions or filter regenerations could also lead to the loss of activity, which is irreversible. Another cause of deactivation is related to the presence of sulphur in the fuel and lubricated oil due to the formation of sulphates on the catalyst surface, which reduces the LNT storage capacity. Therefore, the LNT catalyst degradation must be monitored. There have been very limited contributions to fault detection and isolation of LNT after treatment systems [**120–123**]. Recently, Canova *et al.* [**120**] used a time-varying non-linear ordinary differential equation model of the LNT system to generate fault residuals using the system model, through comparison of the predicted and measured values of selected variables. The fault-diagnosis method in reference [**120**] was designed to detect and isolate critical faults of the LNT after treatment system, including sulphur poisoning, deactivation of the catalyst storage sites due to thermal ageing, regeneration controller faults, and faults in the sensors (including outlet NO_x sensor and temperature sensor). The proposed diagnostic method is based on the parity equation approach and on the analytical redundancy. It is noted that the model developed in reference [**120**] is a detailed one consisting of oxygen storage dynamics, NO_x storage dynamics, and catalyst temperature dynamics.

3.5.3 Diesel SCR exhaust after treatment system monitoring

Selective catalytic reduction is a well-proven NO_x reduction technology used in power generation for more than 30 years. In the past decade, SCR has received a lot of attention for NO_x reduction in automotive diesel engines. Selective catalytic reduction catalysts are considered the technology of choice for future heavy-duty applications, while LNTs appear to be more promising for passenger cars and light-duty trucks. This is due to the conversion efficiency, reliability, and cost-effective approach for regenerating the system using the onboard fuel. The control problem in the SCR system consists of achieving the appropriate regulation of the injected urea to minimize NO_x emissions without significant ammonia slip.

Even though there have been some attempts on model-based control design for urea injection in SCR systems, there has not been much progress on the development of monitoring methods for SCR systems. It is, however, noted that all the after-treatment systems share a common type of failure that is to be monitored for: *catalyst ageing*. Ammonia storage capacity is an important parameter directly reflecting the SCR catalyst ageing. Estimation of this parameter is then helpful in monitoring the SCR system physical condition. In reference [**124**], a simple yet effective, method was proposed to estimate SCR catalyst ammonia coverage ratio and storage capacity based on an EKF. In reference [**125**], a sliding-mode observer was designed to estimate the SCR catalyst ammonia storage based on the measurements of NO_x, ammonia, and temperature, and an EKF was used to eliminate the NO_x sensor cross-sensitivity to ammonia. The real-time monitoring of the catalyst storage capacity ensures that the SCR system is in a healthy status and also allows the control gains to adapt to any unexpected changes in the storage capacity to keep the NO_x emission level below the threshold.

3.5.4 Diesel engine oxidation catalyst monitoring

The results of a feasibility study for the diagnostics of diesel oxidation catalysts (DOCs) were reported in reference [**126**], with the focus on the effect of noise factors on the ability of the diagnostics method to distinguish threshold from marginally healthy catalysts. The study revealed that, given the current sensing and catalyst technology, the separation between the two was poor.

3.5.5 Gasoline engine TWC exhaust system monitoring

Modelling and control of TWCs has been a widely discussed research topic [**42**]. In addition, a variety of diagnostics methods based on the detailed thermodynamics-based modelling of the TWC have been developed [**127**]. Recently, simplified models capturing TWC dynamics as an oxygen storage/release process have been employed for catalyst monitoring purposes. The goal of the oxygen storage model is to capture this property in a sufficiently accurate manner. Similar to other catalysts discussed so far, during its life, the TWC loses its storage properties, which can be considered as an indirect index of ageing and deterioration of the component. The proposed online diagnosis methods in the literature based on the oxygen storage model aim to monitor the oxygen storage mechanism in order to detect the difference between a healthy and a deteriorated one. Let $0 < \theta < 1$ be the fraction of oxygen sites occupied in the catalyst, also known as the relative oxygen level. The oxygen storage capacity can be modelled as a limited integrator as

$$\dot{\theta} = \frac{1}{C(\mathrm{MAF})} \times 0.23 \times \mathrm{MAF} \times \rho(\lambda_{\mathrm{FG}}, \theta) \times \left(1 - \frac{1}{\lambda_{\mathrm{FG}}}\right) \quad (4)$$

where MAF, $C(.)$, ρ, and λ denote the MAF, effective catalyst capacity, oxygen exchange between the exhaust gas and the catalyst, and relative air–fuel ratio downstream of the catalyst, respectively. Brandt and Grizzle reference [**128**] employed the above model to develop a diagnostic algorithm and analysed that in the context of a *hypothesis test* based on the oxygen storage capacity of the TWC. The Neyman–Pearson criterion was used in reference [**128**] as a basis for the hypothesis test. The method was applied to a multiple-sample case through the use of *Student's t-test*, and the improved fidelity of the *t-test* was demonstrated by showing that larger sample sizes provide further improvement in the quality of the hypothesis test. A slightly modified version of the model (equation (4)) proposed by Fiengo *et al.* was used in reference [**129**] to present a model-based *stochastic approach* for TWC fault detection. The method first generated a 'residual signal', which is the difference between the measured quantity of the oxygen storage capacity and the estimated one using the simplified phenomenological model. The diagnostic algorithm then worked on the generated signals and implemented a stochastic analysis in order to provide a statistical confidence in the TWC's condition. The real-time decision-making procedure was based on the so-called cumulative sum (CUSUM) algorithm that uses observations from the system and the output of the model.

Using even further simplified versions of the TWC dynamics, where the catalyst is modelled as a simple limited integrator with an adaptive integral gain, has been the focus of recent work by Muske *et al.* [**130**, **131**] and Dawson *et al.* [**1**, **132**]. The gain identified in real time is used as a diagnostic metric since it indicates the oxygen storage capacity, and hence the health of the system.

There have been additional efforts on catalyst health monitoring using statistical approaches. Arsie *et al.* [**133**] used the comparison between upstream and downstream catalyst lambda sensors to extract information about the catalyst actual conversion efficiency. They calculated different indices by means of both deterministic and statistical approaches that allow the performance of a *confidence analysis* concerning the estimation of catalyst efficiency. The proposed approaches were applied to estimate the probability of fault occurrence during intermediate aged catalyst status. The methods were also compared in terms of the risk levels associated with both undetected catalyst faults and false alarms. A simplified TWC model relating a TWC health measure K_{TWC} (as the output) to the air mass flowrate (as the input) was identified in references [**1**] and [**131**], and was shown to be a linear transfer function with one zero and two poles. The preliminary experimental results showed that a model direct current (DC) gain in an identified range represents a 'threshold' catalyst, while a DC gain of another identified range would be a *healthy* catalyst. In addition, an increase in the model bandwidth represented increased ageing. The results showed that analysing the DC gain of the identified models could serve as a diagnostic metric since a good separation exists between a healthy catalyst and faulty ones with different ageing levels.

3.6 Exhaust gas sensors

Wide-band oxygen sensors and NO_x sensors are required to be monitored. Oxygen sensors located upstream of any after-treatment system are typically used for EGR control. For NO_x sensors and oxygen sensors located downstream of the after-treatment system components, monitoring becomes complicated due to the variable effect of the after-treatment components on the exhaust gas. There are a number of efforts to diagnose oxygen sensors, which will be described in detail in section 4.1.

3.7 Engine cooling system

Proper warm up of an engine can be crucial to emission controls, and many other component control strategies are linked to the engine warm up. This includes items like EGR control, fuel injection quantities, or timing restrictions to reduce smoke coming out of the exhaust pipe, and may also be used for after-treatment control strategies. Manufacturers also often use coolant temperature as an enabling condition for other monitors, where the engine cooling system is monitored for thermostat malfunctions (e.g. too slow to warm up) and coolant temperature sensor malfunctions (e.g. irrational, biased, offset, or stuck sensors and circuit faults).

Wu and Chen [**134**] used a CWT technique for fault detection in an IC engine and its cooling system by using both vibration and acoustic emission signals. The acoustic emission signal was used in reference [**134**] to evaluate the CWT for fault diagnosis in defective engine cooling fan blades under fixed revolution, acceleration, and deceleration conditions.

3.8 Positive crankcase ventilation system

In general, manufacturers would be required to identify a fault if part of the system becomes disconnected or starts venting crankcase emissions directly to the atmosphere. Unlike other monitors, much of the positive crankcase ventilation (PCV) monitoring requirements can be satisfied by meeting design criteria regarding the types of materials, connectors, hose-routing, and repair procedures, rather than diagnostics that sense when disconnections occur.

3.9 Comprehensive components

Manufacturers would be required to monitor input and output components used for other OBD monitors (e.g. pressure sensors used to monitor the PM filter). Manufacturers would also be required to monitor input and output components that can, by themselves, increase emissions when they malfunction. It is important to note that for comprehensive components, there would be no emission thresholds, as monitoring only consists of circuit checks (opens/shorts), rationality checks (does the sensor reading make sense?), and functional checks (does the valve open and close when commanded to do so?). Rationality faults are those where the sensor is reading within the sensor's operating range, but is indicating a value that does not make sense based on other available information. An example is a MAP sensor that indicates a high boost condition, while other sensors indicate the engine is operating at an idle speed and load. A functional fault is defined as a component that does not properly respond to computer commands.

4 DIAGNOSIS OF SENSOR FAULTS AND LEAKS IN AUTOMOTIVE ENGINES

In the previous section, we described the monitoring requirements for automotive engines as mandated by EPA and CARB. We also cited a number of references corresponding to each requirement. There exist additional types of faults that are of equal (if not of more) importance than those already discussed due to their significant impact on emissions. In this section, we describe two categories of these faults, including sensor faults and leaks. It is noted that some of the sensor faults described in the previous section are discussed in more detail in this section.

4.1 Detection and diagnosis of engine sensor faults

Sensor systems are critical components in all modern engineering systems. These measuring systems are extensively used not only to obtain system operational information, but also to determine control actions. A sensor fault is typically characterized by a change in the sensor parameters or in its operational characteristics. The detection and diagnosis of these undesired changes plays a critical role in the operation of many engineering systems, and automotive systems are no exception to this. The design of sensor fault-diagnosis schemes using the *hardware redundancy* and *analytical redundancy* approaches has been addressed in the literature [**135**]. In the hardware redundancy approach, redundant sensor systems are incorporated into the control system to improve the reliability of sensor measurement and enable sensor fault detection; however, cost and space make this approach unattractive. In contrast, the analytical redundancy-based fault-diagnosis architectures use system physics-based models and information processing methods to achieve the necessary redundancy.

In the automotive systems literature, data-driven and model-based approaches have been proposed to diagnose different sensor faults. An early work using analytical redundancy method in reference [**136**] described the applicability of the model-based detection filters to diagnose a variety of sensor failures in automotive engines. The work in reference [**136**] showed that the applicability of the FDI is only limited by the complexity of the mathematical model used for the engine, upon which the FDI algorithms are based. The work opened up new directions to further focus on developing the real-time implementation and improving the robustness of detection filters to account for ageing and vehicle-to-vehicle variability. The use of adaptive strategies was proposed in reference [**137**] to diagnose the occurrence of an *unknown bias* in a *noisy* sensor. The non-linear model of the system was assumed to contain modelling uncertainties, where the modelling uncertainties and the sensor noise were unstructured and bounded with *a priori* known bounds. The method of reference [**137**] was successfully demonstrated to detect and diagnose a bias in a UEGO sensor. Kim *et al.* [**138**, **139**] used the information from an integral sliding-mode control and a designed observer with hypothesis tests to detect UEGO sensor faults. Other efforts to detect UEGO sensor faults include the work reported in references [**140–142**].

In addition to UEGO sensors, throttle position (TP) [**57**, **138**, **140**, **142–144**], intake mass flow (MAF) [**32**, **34**, **62**, **140**, **141**, **143–150**], engine speed (ES) [**35**, **141**, **143**], and MAP sensors [**57**, **62**, **141–143**, **146**, **148**, **150**, **151**] are among those extensively studied and experimented sensors to be monitored. Model-based and data-driven methods are also proposed to diagnose the ambient pressure sensors (both intake and exhaust) [**141**, **151**, **152**]. The sensor faults are treated as either additive (bias term) or multiplicative uncertainties in model-based approaches.

A method for residual generation based on non-linear parity equations was proposed in reference [**52**] to detect sensor faults in IC engines, where a NARMAX model was used to develop the engine model. The sensor fault-detection algorithms in reference [**150**] were based on the use of EKFs, where a non-linear model of a diesel engine, combined with a sensor fault model, was used for fault parameter estimation purposes. The use of EKFs was justified here because of the non-linear dynamic model of the system.

4.2 Detection of leaks in automotive engines

As described in detail in reference [**14**], the detection of a leak, and particularly an air leak, in the intake manifold can be difficult since, under a range of operating conditions, the turbocharger wastegate inherently tries to counteract the fault and maintain the manifold boost pressure at a prespecified level. Therefore, depending on the magnitude of the leak, the fault may not be noticeable by the driver. If there is an intake manifold air leak, then the overall air–fuel ratio in the cylinder will be lower than the value assumed by the engine management system. This discrepancy could lead to an increase in the levels of CO, unburned HC, and PM released into the atmosphere. Depending on the location of the leak within the intake manifold and the control method applied, the EGR process may be also be affected, leading to an increase in NO_x emissions [**14**].

A large number of efforts have been made to detect and diagnose different types of leaks in automotive engine systems. The use of non-linear model-based adaptive observers to detect intake leaks in diesel engines was proposed by Ceccarelli *et al.* [**153**, **154**], where the authors designed observers with fixed and variable gains. Vinsonneau *et al.* [**148**] also designed a non-linear observer to detect the manifold leakage in SI engines in real time. The idea in reference [**148**] was to model the leak effect on the air flow as

$$\dot{m}_{\mathrm{ai}} = (f_0(\theta, N) + f_1(\theta, p_{\mathrm{m}}))\Psi\left(\frac{p_{\mathrm{m}}}{p_{\mathrm{atm}}}\right)$$

where the term f_1 represents the flow perturbation due to the leak. The model was obtained following the approach described next. Structured hypothesis tests were employed to detect a manifold leak (a leak in the intake manifold) and a boost leak (a leak between air mass flow and throttle right after the intercooler) in SI engines [**57**, **155**]. The aforementioned model-based leak-detection methods used the flow equation through a restriction; that is, the model for flow past the throttle. For instance, the boost leak can be described by

$$W_{\mathrm{BL}} = \frac{A_{\mathrm{b}} p_{\mathrm{b}}}{\sqrt{T}} \Psi\left(\frac{p_{\mathrm{amb}}}{p_{\mathrm{b}}}\right)$$

where A_{b}, p_{b}, p_{amb}, T, and Ψ are leak effective area, boost pressure, ambient pressure, intake manifold temperature, and correction factor, respectively. It is noted that due to the pressure difference direction, the air flow through a boost leak is in the direction out of the air tube. The manifold leak can be also described by

$$W_{\mathrm{ML}} = \frac{A_{\mathrm{m}} p_{\mathrm{amb}}}{\sqrt{T_{\mathrm{amb}}}} \Psi\left(\frac{p_{\mathrm{m}}}{p_{\mathrm{amb}}}\right)$$

where A_{m}, p_{m}, p_{amb}, and T_{amb} are manifold leak effective area, manifold pressure, ambient pressure, and ambient temperature, respectively. It is noted that in this case, the leak air flow is in the direction into the intake manifold. Nyberg and Stutte [**57**, **62**, **155**] used the above model along with SHT to diagnose the leaks, whereas a simple parameter identification based on the above leak models was employed in references [**61**] and [**156**] to detect the faults by tracking the changes in the estimated parameters of interest. The manifold leak model was added to a non-linear state-space representation of a diesel engine's air path system model in reference [**150**], and an EKF was designed to estimate the flow corresponding to the leak and to detect possible intake manifold leakage.

Model-based techniques have also been used in detecting a leak in the exhaust manifold. The exhaust leakage detection is critical since it can cause an offset in oxygen sensor measurement, leading to increased emissions. Andersson and Eriksson [**157**] have developed a model-based method combining an observer-based virtual exhaust pressure sensor with estimated air-to-fuel ratio to detect the leakages in the exhaust manifold. The diagnosis method was based on SHT with only two behavioural modes, and the sensors used for the leak-detection purpose included heated exhaust gas oxygen (HEGO), MAP, MAF, and temperature sensors. The leak model used was based on the aforedescribed one using the flow equation through the restriction.

In addition to model-based methods, data-driven techniques have also been used to detect and isolate air leaks. Antory demonstrated in reference [**14**] a diagnostic model capable of exploring underlying *hidden* information from the experimental data using the PCA method. The model was shown to capture the interdependency of the original signals and transform into a new and smaller number of independent signals. The variation (using the T^2-statistics) and the residual generator (using Q-statistics) of the uncaptured signals were used as the back-bone of the fault-detection and fault-diagnosis processes. The method was implemented to detect air leaks at the intake manifold plenum chamber of a diesel engine. Further analysis in reference [**14**] using contributions to T^2- and Q-statistics showed the effect of the air leaks on fuel consumption, and indicated that they may contribute to increased emissions. The use of cylinder pressure sensors and vibration signals was proposed in reference [**69**] to detect leaks in air intake and exhaust valves, where a set of ANNs was trained using features extracted from cylinder pressure and vibration measurements.

5 EMERGING AREAS OF RESEARCH IN FAULT DIAGNOSIS OF POWER TRAIN SYSTEMS

In this section, we discuss some of the recent efforts on emerging topics in diagnostics of automotive engines. A fairly well investigated topic relevant to what will be described in section 5.3 is the prognostics for automotive engines that has been briefly discussed in reference [**3**].

5.1 Diagnostics for advanced combustion

A highly attractive solution for vehicle emission reduction is to alter the combustion process such that engine exhaust emissions are at levels to remove or reduce the requirements for auxiliary devices while maintaining or improving engine efficiency. This is the concept behind advanced combustion processes such as HCCI, PCCI, and other LTC modes, which exhibit high efficiency with significant reductions in NO_x and PM emissions. Significant progress continues to be made for these advanced combustion systems, and the operational range continues to be expanded to better cover the speed and load combinations consistent with light-duty and heavy-duty driving cycles [**158**]. In HCCI, a homogeneous (i.e. uniformly mixed) fuel–air mixture is created in the combustion cylinder by injecting fuel during the intake stroke. This is in contrast to the heterogeneous (i.e. not uniformly mixed) mixture used in traditional diesel fuel injection. The HCCI is similar to the fuel injection method used in gasoline engines, but it relies on the pressure and heat of the compressed gas to initiate combustion. The homogeneous charge property in HCCI mode eliminates the fuel-rich regions in the cylinder, and hence combustion occurs almost simultaneously throughout the cylinder. This leads to a more complete, lower-temperature combustion with reduced NO_x and PM formation compared to conventional diesel combustion. In parallel, research on the corresponding emission control systems is underway to increase their efficiency and durability for overall emissions compliance at an acceptable cost. There has not been much effort on diagnostics for engines running in advanced combustion modes other than those by the Sandia National Laboratory, Argonne National Laboratory, and the University of Wisconsin using in-cylinder laser diagnostics [**158**, **159**]. The authors expect that the research in this area will draw significant attention in the years to come, mainly due to the importance of these modes in a large number of applications including on-road and off-road vehicles, as well as military ground vehicles.

5.2 Remote diagnostics

Data-driven methods have been proposed to diagnose a variety of faults in the vehicles equipped with telematic capabilities that provide enhanced services to customers by allowing *remote diagnosis* [**160**, **161**]. This capability provides a means of collecting a large amount of data and perform data mining to monitor the health of system components remotely. Namburu *et al.* [**140**, **162**] used data-driven methods to achieve high diagnostic accuracy by detecting the faults in a reasonable time. The method was based on wavelet-based preprocessing of the data, statistical hypothesis tests to detect faults, and pattern recognition techniques to classify various faults in engines. To classify faults, five different methods for pattern recognition were tried including support vector machines (SVMs), k-nearest neighbours (KNN), Gaussian mixture model (GMM), linear discriminant (LD) analysis, and probabilistic neural network (PNN), and the trained weights corresponding to each classifier were imported by the online module for real-time fault detection.

Although basic research in model-based diagnosis has matured, there is still a lack of sufficient knowledge on how to integrate different diagnostic modelling techniques, especially those that combine mathematical and graph-based dependency models, for an intelligent diagnosis. Luo *et al.* [**163**] presented a hybrid model-based diagnostic method to improve the telematic diagnostics, the diagnostic system's accuracy, and the consistency of those solely based on graph-based models. Luo *et al.* [**163**] developed a fault-diagnosis toolset, comprised of both model-based and data-driven techniques to provide a 'sand box' for test engineers to experiment with, and to systematically select relevant algorithms/techniques to detect and isolate their specific fault problems. This integrated process is implemented on a hardware-in-the-loop simulation platform.

5.3 Integration of fault detection and engine control design

Recent efforts have been made on making use of the available control structure for the implementation of model-based fault diagnosis *without using additional models* [**138**, **139**, **141**, **164**]. This potentially leads to a diagnosis system realization with low demands on engineering costs, computational power, and memory. A variety of SI engine component faults (including manifold leaks and throttle blockages) and sensor faults (including biased ambient pressure sensors, manifold pressure sensors, and lambda sensor) were investigated in references [**141**] and [**164**] using control-oriented models (the same models used for control design purposes).

Kim *et al.* [**138**] considered a throttle position sensor fault and a fuel injector fault to demonstrate their integrated control/diagnostic method. They used the NPERG diagnostic scheme developed by Krishnaswami *et al.* [**52**], where the desired observers were designed based on the sliding-mode approach. The basic concept behind the NPERG method is the use of input and state observers to provide fault detection and isolation using dynamic models of a system.

To this end, the observers are configured such that sensor faults are detected and isolated using non-linear output estimators, while input and plant parameter faults are isolated using non-linear input estimators. Kim *et al.* [**139**] used the powertrain model and the integrated design scheme of control and diagnostics in reference [**138**] by combining the integral sliding-mode control method and observers with the hypothesis testing method. In their approach, information obtained from sliding-mode control and observers with hypothesis testing were used so that a fault can be detected, isolated, and compensated for in the air and fuel dynamics of an IC engine.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1 **Dawson, B., Franchek, M. A.,** and **Grigoriadis, K. M.** A systematic method for automotive three-way catalyst diagnostics: experimental results. In 1st *ASME Annual Dynamic Systems and Control Conference*, Ann Arbor, MI, 2008.
2 **Kido, K., Togai, K.,** and **Yamaura, H.** Knowledge-based diagnosis failure detection and isolation by state comparison with model behavior. *Mitsubishi Motors Tech. Rev.*, 2007, **19**, 33–39.
3 **Rizzoni, G., Onori, S.,** and **Rubagotti, M.** Diagnosis and prognosis of automotive systems: motivations, history and some results. In 7th IFAC Symposium on *Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS)*, Barcelona, Spain, 2009.
4 **Rizzoni, G., Soliman, A.,** and **Passino, K.** A survey of automotive diagnostic equipment and procedures. In *SAE International Congress & Exposition*, Detroit, MI, 1993.
5 **Jones, N. B.** and **Li, Y. H.** A review of condition monitoring and fault diagnosis for diesel engines. *Tribotest*, 2000, **6**(3), 267–291.
6 **Chen, J.** and **Patton, R. J.** *Robust model-based fault diagnosis for dynamic systems*, 1999 (Kluwer Academic Publishers, Norwell, MA).
7 **Chiang, L., Russell, E.,** and **Braatz, R.** *Fault detection and diagnosis in industrial systems*, 2001 (Springer Verlag, London, UK).
8 **Russell, E., Chiang, L.,** and **Braatz, R.** *Data-driven methods for fault detection and diagnosis in chemical processes*, 2000 (Springer Verlag, London, UK).
9 **MacGregor, J.** and **Kourti, T.** Statistical process control of multivariate processes. *Contr. Engng Pract.*, 1995, **3**(3), 403–414.

10 **Jolliffe, I.** *Principal component analysis*, 2002 (Springer Verlag, New York, UK).
11 **Piovoso, M. J.** and **Kosanovich, K. A.** Applications of multivariate statistical methods to process monitoring and controller design. *Int. J. Contr.*, 1994, **59**(3), 743–765.
12 **Antory, D., Kruger, U., Irwin, G.,** and **McCullough, G.** Fault diagnosis in internal combustion engines using non-linear multivariate statistics. *Proc. IMechE, Part I: J. Systems and Control Engineering*, 2005, **219**(4), 243–258. DOI: 10.1243/095965105X9614.
13 **King, P. J.** and **Burnham, K. J.** Application of a PCA model approach for misfire monitoring. In *UKACC Control Conference*, Manchester, UK, 2008.
14 **Antory, D.** Application of a data-driven monitoring technique to diagnose air leaks in an automotive diesel engine: a case study. *Mech. Sys. Sig. Process.*, 2007, **21**(2), 795–808.
15 **Wang, X., Kruger, U., Irwin, G. W., McCullough, G.,** and **McDowell, N.** Nonlinear PCA with the local approach for diesel engine fault detection and diagnosis. *IEEE Trans. Contr. Sys. Tech.*, 2008, **16**(1), 122–129.
16 **Rosipal, R.** and **Krämer, N.** Overview and recent advances in partial least squares. In *Subspace, Latent Structure and Feature Selection Techniques*, 2006, pp. 34–51 (Springer, Berlin, Germany).
17 **Chiang, L., Russell, E.,** and **Braatz, R.** Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemomet. Intell. Lab. Sys.*, 2000, **50**(2), 243–252.
18 **Liu, Y., Tian, H., Sun, Y.,** and **Pei, W.** Application of partial least-squares regression to oil atomic emitting spectrum data of a type diesel engine. In IEEE International Conference *on Industrial Engineering and Engineering Management*, 2010, pp. 1657–1660.
19 **Juricek, B. C., Seborg, D. E.,** and **Larimore, W. E.** Fault detection using canonical variate analysis. *Ind. Engng Chem. Res.*, 2004, **43**(2), 458–474.
20 **Larimore, W. E.** and **Javaherian, H.** Identification and monitoring of automotive engines. In *American Control Conference*, St. Louis, MO, 2009, pp. 1800–1807.
21 **Patton, R., Clark, R.,** and **Frank, P. M.** *Issues of fault diagnosis for dynamic systems*, 2000 (Springer Verlag, London, UK).
22 **Capriglione, D., Liguori, C., Pianese, C.,** and **Pietrosanto, A.** On-line sensor fault detection, isolation, and accommodation in automotive engines. *IEEE Trans. Instrument. Measure.*, 2003, **52**(4), 1182–1189.
23 **Sangha, M. S., Gomm, J. B.,** and **Yu, D.** Neural network fault classification of transient data in an automotive engine air path. *Int. J. Modell., Identif. Contr.*, 2008, **3**(2), 148–155.
24 **Capriglione, D., Liguori, C.,** and **Pietrosanto, A.** Real-time implementation of IFDIA scheme in automotive systems. *IEEE Trans. Instrument. Measure.*, 2007, **56**(3), 824–830.
25 **Desbazeille, M., Randall, R. B., Guillet, F., El Badaoui, M.,** and **Hoisnard, C.** Model-based diagnosis of large diesel engines based on angular speed variations of the crankshaft. *Mech. Sys. Sig. Process.*, 2010, **24**(5), 1529–1541.
26 **Schwarte, A.** and **Isermann, R.** Neural network applications for model based fault detection with parity equations. In *IFAC 15th World Congress*, Barcelona, Spain, 2002.
27 **Wang, X., McDowell, N., Kruger, U., McCullough, G.,** and **Irwin, G. W.** Semi-physical neural network model in detecting engine transient faults using the local approach. In *IFAC 17th World Congress*, Seoul, Korea, 2008, pp. 7086–7090.
28 **Tu, F., Wen, F., Willett, P. K., Pattipati, K. R.,** and **Jordan, E. H.** Signal processing and neural network toolbox and its application to failure diagnosis and prognosis. In SPIE Conference on *Component and Systems Diagnostics, Prognosis, and Health Management*, Orlando, FL, 2001.
29 **Marko, K. A., Bryant, B.,** and **Soderborg, N.** Neural network application to comprehensive engine diagnostics. In IEEE International Conference on *Systems, Man and Cybernetics*, Chicago, IL, 1992, pp. 1016–1022.
30 **Cascio, F., Console, L., Guagliumi, M., Osella, M., Panati, A., Sottano, S.,** and **Theseider Dupré, D.** On-board diagnosis of automotive systems: from dynamic qualitative diagnosis to decision trees. In Proceedings of the 16th International Joint Conference on *Artificial Intelligence, Stockholm, Sweden*, 1999, pp. 1–14.
31 **Murphey, Y., Chen, Z.,** and **Feldkamp, L.** An incremental neural learning framework and its application to vehicle diagnostics. *Appl. Intell.*, 2008, **28**(1), 29–49.
32 **Chen, P. C.** Neuro-fuzzy-based fault detection of the air flow sensor of an idling gasoline engine. *Proc. IMechE, Part D: J. Automobile Engineering*, 2005, **219**(4), 511–524. DOI: 10.1243/095440705X11022.
33 **Macian, V., Tormos, B., Sala, A.,** and **Ramirez, J.** Fuzzy logic-based expert system for diesel engine oil analysis diagnosis. *Insight – Non-dest. Test. Cond. Monitor.*, 2006, **48**(8), 462–469.
34 **Laukonen, E. G., Passino, K. M., Krishnaswami, V., Luh, G. C.,** and **Rizzoni, G.** Fault detection and isolation for an experimental internal combustion engine via fuzzy identification. *IEEE Trans. Contr. Sys. Tech.*, 1995, **3**(3), 347–355.
35 **Çelik, M.** and **Bayir, R.** Fault detection in internal combustion engines using fuzzy logic. *Proc. IMechE, Part D: J. Automobile Engineering*, 2007, **221**(5), 579–587. DOI: 10.1243/09544070JAUTO366.
36 **Korbicz, J., Koscielny, J.,** and **Kowalczuk, Z.** *Fault diagnosis: models, artificial intelligence, applications*, 2004 (Springer Verlag, Berlin, Germany).
37 **Parikh, C. R., Pont, M. J.,** and **Barrie Jones, N.** Application of Dempster–Shafer theory in condition monitoring applications: a case study. *Patt. Recog. Lett.*, 2001, **22**(6–7), 777–785.
38 **Scaife, M., Charlton, S. J.,** and **Mobley, C.** A neural network for fault recognition. In *SAE International*

Congress & Exposition, Society of Automotive Engineers, Detroit, MI, 1993.

39 **Sharkey, A. J. C., Chandroth, G. O.,** and **Sharkey, N. E.** A multi-net system for the fault diagnosis of a diesel engine. *Neural Comput. Appl.*, 2000, **9**(2), 152–160.

40 **Hamad, A., Yu, D., Gomm, J. B.,** and **Sangha, M. S.** Radial basis function neural network in fault detection of automotive engines. *Int. J. Engng Sci. Tech.*, 2010, **2**(10), 1–8.

41 **Sangha, M., Gomm, J. B., Yu, D. L.,** and **Page, G. F.** Fault detection and identification of automotive engines using neural networks. In *IFAC World Congress*, Prague, Czech Republic, 2005.

42 **Balenovic, M.** *Modelling and model-based control of a three-way catalytic converter*, 2002 (Technische Universiteit Eindhoven, Eindhoven).

43 **Isermann, R.** and **Balle, P.** Trends in the application of model-based fault detection and diagnosis of technical processes. *Contr. Engng Pract.*, 1997, **5**(5), 709–719.

44 **Isermann, R.** Model-based fault-detection and diagnosis-status and applications. *Ann. Rev. Contr.*, 2005, **29**(1), 71–85.

45 **Luo, J. F. T., Azam, M., Pattipati, K., Qiao, L.,** and **Kawamoto, M.** Intelligent model-based diagnostics for vehicle health management. In *SPIE Conference Proceedings*, Orlando, FL, 2003.

46 **Pfeufer, T.** Application of model-based fault detection and diagnosis to the quality assurance of an automotive actuator. *Contr. Engng Pract.*, 1997, **5**(5), 703–708.

47 **Chen, W.** and **Saif, M.** An iterative learning observer for fault detection and accommodation in nonlinear time delay systems. *Int. J. Robust Nonlin. Contr.*, 2006, **16**(1), 1–19.

48 **Molinar-Monterrubio, J. M.** and **Castro-Linares, R.** Sliding mode observer for internal combustion engine misfire detection: experimental results. In IEEE International Conference on *Mechatronics*, Malaga, Spain, 2009, pp. 1–6.

49 **Hsu, P. L., Lin, K. L.,** and **Shen, L. C.** Diagnosis of multiple sensor and actuator failures in automotive engines. *IEEE Trans. Vehic. Tech.*, 1995, **44**(4), 779–789.

50 **Gertler, J. J.** Fault detection and isolation using parity relations. *Contr. Engng Pract.*, 1997, **5**(5), 653–661.

51 **Gertler, J. J.** and **Monajmey, R.** Generating directional residuals with dynamic parity relations. *Automatica*, 1995, **33**(4), 627–635.

52 **Krishnaswami, V., Luh, G.,** and **Rizzoni, G.** Nonlinear parity equation based residual generation for diagnosis of automotive engine faults. *Contr. Engng Pract.*, 1995, **3**(10), 1385–1392.

53 **Conatser, R., Wagner, J., Ganta, S.,** and **Walker, I.** Diagnosis of automotive electronic throttle control systems. *Contr. Engng Pract.*, 2004, **12**(1), 23–30.

54 **Montes, C.** and **Pisu, P.** Fault detection in idle speed control of IC engines. In *SAE World Congress & Exhibition*, SAE International, Detroit, MI, 2009, p. 2009-01-0729.

55 **Kimmich, F., Schwarte, A.,** and **Isermann, R.** Fault detection for modern diesel engines using signal- and process model-based methods. *Contr. Engng Pract.*, 2005, **13**(2), 189–203.

56 **Nyberg, M.** *Model based fault diagnosis: methods, theory, and automotive engine applications*, 1999 (Department of Electrical Engineering, Linköping University, Linköping, Sweden).

57 **Nyberg, M.** Model-based diagnosis of an automotive engine using several types of fault models. *IEEE Trans. Contr. Sys. Tech.*, 2002, **10**(5), 679–689.

58 **Berton, A., Frisk, E.,** and **Nyberg, M.** Improving diagnosis performances on a truck engine making use of statistical charts. In IFAC Symposium on *Advances in Automotive Control*, 2004.

59 **Biteus, J., Nyberg, M., Frisk, E.,** and **Slund, J.** Determining the fault status of a component and its readiness, with a distributed automotive application. *Engng Applic. Art. Intell.*, 2009, **22**(3), 363–373.

60 **Biteus, J., Jensen, M.,** and **Nyberg, M.** *Decentralized diagnosis in heavy duty vehicles*, 2004 (CCSSE, Norrkoping, Sweden).

61 **Nyberg, M.** and **Nielsen, L.** Model-based diagnosis for the air intake system of the SI engine. *SAE Trans.*, 1997, **106**, 9–20.

62 **Nyberg, M.** and **Stutte, T.** Model based diagnosis of the air path of an automotive diesel engine. *Contr. Engng Pract.*, 2004, **12**(5), 513–525.

63 **He, Y.** and **Feng, L.** Diesel fuel injection system faults diagnosis based on fuzzy injection pressure pattern recognition. In *5th World Congress on Intelligent Control and Automation*, 2004, pp. 1654–1657.

64 **Chandroth, G. O., Sharkey, A. J. C.,** and **Sharkey, N. E.** Artificial neural nets and cylinder pressures in diesel engine fault diagnosis. In Proceedings of *INMARCO Shipping Trends for the Next Millennium*, 1998, vol. 1, pp. 9.1–9.2.

65 **Lee, B., Guezennec, Y.,** and **Rizzoni, G.** Model-based fault diagnosis of spark-ignition direct-injection engine using nonlinear estimations. In SAE *World Congress & Exhibition*, Detroit, MI, 2005.

66 **Schilling, A.** *Model-based detection and isolation of faults in the air and fuel paths of common-rail DI diesel engines equipped with a lambda and a nitrogen oxides sensor*, 2008 (ETH Zurich, Zurich, Switzerland).

67 **Schilling, A., Amstutz, A.,** and **Guzzella, L.** Model based detection and isolation of faults due to ageing in the air and fuel paths of common rail direct injection diesel engines equipped with a and a nitrogen oxides sensor. *Proc. IMechE, Part D: J. Automobile Engineering*, 2008, **222**(1), 101–117. DOI: 10.1243/09544070JAUTO659.

68 **Payri, F., Luján, J. M., Guardiola, C.,** and **Rizzoni, G.** Injection diagnosis through common-rail pressure measurement. *Proc. IMechE, Part D: J. Automobile Engineering*, 2006, **220**(3), 347–357. DOI: 10.1243/09544070JAUTO34.

69 **Chandroth, G. O., Sharkey, A. J. C.,** and **Sharkey, N. E.** Cylinder pressures and vibration in

internal combustion engine condition monitoring. In Proceedings of *Comadem 99*, Sunderland, UK, 1999.
70 **Shimasaki, Y., Kanehiro, M., Baba, S., Maruyama, S., Hisaki, T.,** and **Miyata, S.** Spark plug voltage analysis for monitoring combustion in an internal combustion engine. In *SAE International Congress & Exposition*, Detroit, MI, 1993.
71 **Plapp, G.** and **Klenk, M.** Methods of on-board misfire detection. In *SAE International Congress & Exposition*, Detroit, MI, 1990.
72 **Shiao, Y.** and **Moskwa, J. J.** Cylinder pressure and combustion heat release estimation for SI engine diagnostics using nonlinear sliding observers. *IEEE Trans. Contr. Sys. Tech.*, 1995, **3**(1), 70–78.
73 **Azzoni, P. M., Moro, D., Porceddu-Cilione, C. M.,** and **Rizzoni, G.** Misfire detection in a high-performance engine by the principal component analysis approach. In *SAE International Congress & Exposition*, Detroit, MI, 1996.
74 **Geveci, M., Osburn, A. W.,** and **Franchek, M. A.** An investigation of crankshaft oscillations for cylinder health diagnostics. *Mech. Sys. Sig. Process.*, 2005, **19**(5), 1107–1134.
75 **Tinaut, F. V., Melgar, A., Horrillo, A., Fernandez, L.,** and **Montero, V.** Faults detection in a reciprocating internal combustion engine from instantaneous engine speed. In CIMAC International Congress on *Combustion Engines*, Copenhagen, Denmark, 1998.
76 **Osburn, A. W., Kostek, T. M.,** and **Franchek, M. A.** Residual generation and statistical pattern recognition for engine misfire diagnostics. *Mech. Sys. Sig. Process.*, 2006, **20**(8), 2232–2258.
77 **Gani, E.** and **Manzie, C.** Misfire-misfuel classification using support vector machines. *Proc. IMechE, Part D: J. Automobile Engineering*, 2007, **221**(9), 1183–1195. DOI: 10.1243/09544070JAUTO301.
78 **Rizzoni, G.** Estimate of indicated torque from crankshaft speed fluctuations: a model for the dynamics of the IC engine. *IEEE Trans. Vehic. Tech.*, 1989, **38**(3), 168–179.
79 **Connolly, F. T.** and **Rizzoni, G.** Real time estimation of engine torque for the detection of engine misfires. *J. Dyn. Sys., Measure., Contr.*, 1994, **116**, 675.
80 **Rizzoni, G., Drakunov, S.,** and **Wang, Y. Y.** On-line estimation of indicated torque in IC engines via sliding mode observers. In *American Control Conference*, Seattle, WA, 1995, pp. 2123–2127.
81 **Kao, M.** and **Moskwa, J. J.** Model-based engine fault detection using cylinder pressure estimates from nonlinear observers. In *IEEE Conference on Decision and Control*, Lake Buena Vista, FL, 1994.
82 **Chen, S. X.** and **Moskwa, J. J.** Application of nonlinear sliding-mode observers for cylinder pressure reconstruction. *Contr. Engng Pract.*, 1997, **5**(8), 1115–1121.
83 **Wang, Y.** and **Chu, F.** Real-time misfire detection via sliding mode observer. *Mech. Sys. Sig. Process.*, 2005, **19**(4), 900–912.
84 **Kiencke, U.** Engine misfire detection. *Contr. Engng Pract.*, 1999, **7**(2), 203–208.
85 **Liu, S., Gu, F.,** and **Ball, A.** The on-line detection of engine misfire at low speed using multiple feature fusion with fuzzy pattern recognition. *Proc. IMechE, Part D: J. Automobile Engineering*, 2002, **216**(5), 391–402. DOI: 10.1243/0954407021529200.
86 **Gyan, P., Ginoux, S., Champoussin, J. C.,** and **Guezennec, Y. G.** Crankangle-based torque estimation: mechanistic/stochastic. In *SAE World Congress*, Detroit, MI, 2000.
87 **Brown, T. S.** and **Neill, W. S.** Determination of engine cylinder pressures from crankshaft speed fluctuations. In *SAE International Congress & Exposition*, Detroit, MI, 1992.
88 **Gu, F., Jacob, P. J.,** and **Ball, A. D.** Non-parametric models in the monitoring of engine performance and condition: part 2, non-intrusive estimation of diesel engine cylinder pressure and its use in fault detection. *Proc. IMechE, Part D: J. Automobile Engineering*, 1999, **213**(2), 135–143. DOI: 10.1243/0954407991526757.
89 **Ali, A., Magnor, O.,** and **Schultalbers, M.** *Misfire detection using a neural network based pattern recognition technique*,In Proceedings of the IEEE International Conference on *Electrical Engineering*, Lahore, Pakistan, pp. 1–6, 2007.
90 **Tinaut, F. V., Melgar, A., Laget, H.,** and **Domínguez, J. I.** Misfire and compression fault detection through the energy model. *Mech. Sys. Sig. Process.*, 2007, **21**(3), 1521–1535.
91 **Lim, B., Lim, I., Park, J., Pae, S., Yoon, Y. S.,** and **Kim, E.** SI engine misfire detection through the energy model. In *International Fuels & Lubricants Meeting & Exposition*, Baltimore, MD, 1994.
92 **Schmidt, M., Kimmich, F., Straky, H.,** and **Isermann, R.** Combustion supervision by evaluating the crankshaft speed and acceleration. In *SAE World Congress*, Detroit, MI, 2000.
93 **Corti, E.** *Modelling and diagnostic constraints of missing combustion phenomena*, SAE Technical Paper 2004-01-0896, Detroit, MI, 2004, DOI: 10.4271/2004-01-0896.
94 **Lambert, C., Westerman, D., Manzie, C.,** and **Gani, E.** Intelligent classification of engine misfire and classification of engine misfire and misfuel using support vector machines. In *Intelligent Vehicles and Road Infrastructure*, Melbourne, Australia, 2005.
95 **Chang, J., Kim, M.,** and **Min, K.** Detection of misfire and knock in spark ignition engines by wavelet transform of engine block vibration signals. *Measure. Sci. Tech.*, 2002, **13**, 1108.
96 **Aihua, J., Xiaoyu, L., Xiuchang, H., Zhenhua, Z.,** and **HongXing, H.** Detection of engine misfire by wavelet analysis of cylinder-head vibration signals. *Int. J. Agric. Biol. Engng*, 2008, **1**(2), 1–7.
97 **Montani, M.** and **Speciale, N.** Multiple misfire identification by a wavelet-based analysis of crankshaft speed fluctuation. In IEEE International Symposium on *Signal Processing and Information Technology*, Vancouver, BC, 2006, pp. 144–148.
98 **Samimy, B.** and **Rizzoni, G.** Mechanical signature analysis using time-frequency signal processing:

application to internal combustion engine knock detection. *Proc. IEEE*, 1996, **84**(9), 1330–1343.
99 **Borg, J. M. K. C. C., Saikalis, G.,** and **Oho, S.** Wavelet-based knock detection with fuzzy logic. In IEEE International Conference on *Computational Intelligence for Measurement Systems and Applications*, Giardini Naxos, Italy, 2005, pp. 26–31.
100 **Zhang, Z.** and **Tomita, E.** Knocking detection using wavelet instantaneous correlation method. *Japan Soc. Auto. Engng Rev.*, 2002, **23**, 443–449.
101 **Jonghwa Lee, S. H., Lim, J.,** and **Cho, Y. –S.** A new knock-detection method using cylinder pressure, block vibration and sound pressure signals from an SI engine. In *SAE International Fuels & Lubricants Meeting & Exposition*, Detroit, MI, 1998.
102 **Kaneyasu, M. N. K., Katogi, K.,** and **Tokuda, H.** Engine knock detection using multi-spectrum method. In *SAE International Congress & Exposition*, Detroit, MI, 1992.
103 **Wu, G.** A real time statistical method for engine knock detection. In *SAE World Congress & Exhibition*, Detroit, MI, 2007.
104 **Thomas, J. -H., Dubuisson, B.,** and **Dillies-Peltier, M. -A.** Engine knock detection from vibration signals using pattern recognition. *Meccanica*, 1997, **32**(5), 431–439.
105 **Ettefagh, M. H. S., Pirouzpanaha, V.,** and **Arjmandi Tash, H.** Knock detection in spark ignition engines by vibration analysis of cylinder block: a parametric modeling approach. *Mech. Sys. Sig. Process.*, 2008, **22**(6), 1495–1514.
106 **Sangha, M. S., Yu, D. L.,** and **Gomm, J. B.** Robustness assessment of adaptive FDI system for engine air path. *Int. J. Inf. Sys. Sci.*, 2007, **3**(1), 180–205.
107 **Fischer, M. J. B., Kirkham, C.,** and **Georgi, R.** OBD of diesel EGR using artificial neural networks. In *SAE World Congress & Exhibition*, Detroit, MI, 2009.
108 **Mohammadpour, J., Grigoriadis, K., Franchek, M.,** and **Zwissler, B. J.** Real-time diagnosis of the exhaust recirculation in diesel engines using least-squares parameter estimation. *ASME J. Dyn. Sys., Measure., Contr.*, 2010, **132**(1).
109 **van Nieuwstadt, M.** and **Brahma, A.** Uncertainty analysis of model based diesel particulate filter diagnostics. *SAE Int. J. Comm. Veh.*, 2008, **1**(1), 356–362.
110 **van Nieuwstadt, M. J.** and **Trudell, D. F.** Diagnostics for diesel particulate filters. In *SAE World Congress & Exhibition*, Detroit, MI, 2004.
111 **Konstandopoulos, A. G., Kostoglou, M., Skaperdas, E., Papaioannou, E., Zarvalis, D.,** and **Kladopoulou, E.** Fundamental studies of diesel particulate filters: transient loading, regeneration and aging. In *SAE World Congress*, Detroit, MI, 2000.
112 **Cunningham, P., Meckl, P.,** and **Shah, C.** Correlating dynamic pressure signal features to diesel particulate filter load. In *SAE World Congress & Exhibition*, Detroit, MI, 2007.
113 **Surve, P. R.** Diesel particulate filter diagnostics using correlation and spectral analysis. In *ECE*, Purdue, 2008.
114 **Gheorghiu, V., Ueberschär, D.,** and **Zikoridse, G.** Soot sensor for diesel emission onboard control systems. In *FISITA Congress*, Jokohama, Japan, 2006.
115 **Sappok, A., Bromberg, L., Parks, J.,** and **Prikhodko, V.** Loading and regeneration analysis of a diesel particulate filter with a radio frequency-based sensor. In *SAE Powertrains Fuels & Lubricants Meeting*, San Diego, CA, 2010.
116 **Kondo, A., Yokoi, S., Sakurai, T., Nishikawa, S., Egami, T., Tokuda, M.,** and **Sakuma, T.** New particulate matter sensor for on board diagnosis. In *SAE World Congress & Exhibition*, SAE International, Detroit, MA, 2011.
117 **Ochs, T., Schittenhelm, H., Genssle, A.,** and **Kamp, B.** Particulate matter sensor for on board diagnostics (OBD) of diesel particulate filters (DPF). *SAE Int. J. Fuels Lub.*, 2010, **3**(1), p. 61–69.
118 **Kabin, K. S., Muncrief, R. L.,** and **Harold, M. P.** NO_X storage and reduction on a Pt/BaO/alumina monolithic storage catalyst. *Catalysis Today*, 2004, **96**(1–2), 79–89.
119 **Wang, X., Wu, H., Zhang, Q., Hansen, A. C.,** and **Malinga, B.** Lean NO_X trap storage model for diesel engine aftertreatment control and diagnosis. *Trans. ASABE*, 2006, **49**(6), 1687–1693.
120 **Canova, M., Midlam-Mohler, S., Pisu, P.,** and **Soliman, A.** Model-based fault detection and isolation for a diesel lean NO_X trap aftertreatment system. *Contr. Engng Pract.*, 1(11), 1307–1317, 2010.
121 **Pezzini, A., Canova, M., Onori, S., Rizzoni, G.,** and **Soliman, A.** A methodology for fault diagnosis of diesel NO_X aftertreatment systems. In *SAFEPROCESS*, Barcelona, Spain, 2009.
122 **Pisu, P., Canova, M.,** and **Soliman, A.** *Model-based fault diagnosis of a NO_X aftertreatment system*, In Proceedings of the 17th IFAC World Congress, Seoul, Korea, pp. 7072–7078, 2008.
123 **Siebenbrunner, P., Fischperer, R., Bachler, J., Vitale, G.,** and **Hulser, H.** OBD of de-NO_X systems – requirements for software development and calibration for 2010, and beyond. In *SAE World Congress & Exhibition*, Detroit, MI, 2008.
124 **Hsieh, M. F.** and **Wang, J.** An extended Kalman filter for ammonia coverage ratio and capacity estimations in the application of diesel engine SCR control and onboard diagnosis. In *American Control Conference*, Baltimore, MD, 2010.
125 **Hsieh, M. F.** and **Wang, J.** Observer-based estimation of selective catalytic reduction catalyst ammonia storage. *Proc. IMechE, Part D: J. Automobile Engineering*, 224(9), 1199–1211, 2010. DOI: 10.1243/09544070JAUTO1482.
126 **van Nieuwstadt, M.** and **Upadhyay, D.** Diagnostics for diesel oxidation catalysts. In *SAE Commercial Vehicle Engineering Conference*, Rosemont, IL, 2005.

127 **Theis, J.** Catalytic converter diagnosis using the catalyst exotherm. In *International Fuels & Lubricants Meeting & Exposition*, Baltimore, MD, 1994, SAE.

128 **Brandt, E. P.** and **Grizzle, J. W.** Three-way catalyst diagnostics for advanced emissions control systems. In *American Control Conference*, Arlington, VA, 2001.

129 **Fiengo, G., Glielmo, L.,** and **Santini, S.** On-board diagnosis for three-way catalytic converters. *Int. J. Rob. Nonlin. Contr.*, 2001, **11**(11), 1073–1094.

130 **Muske, K. R.** and **Jones, J. C. P.** Model-based control and diagnostic monitoring for automotive catalyst systems. *Contr. Intell. Sys.*, 2008, **36**(3), 296–302.

131 **Muske, K. R., Peyton Jones, J. C.,** and **Howse, J. W.** A model-based approach to automotive three-way catalyst on-board monitoring. *J. Proc. Contr.*, 2008, **18**(2), 163–172.

132 **Dawson, B. M., Franchek, M. A.,** and **Grigoriadis, K. M.** Data driven simplified three-way catalyst health diagnostic models: experimental results. In *ASME International Mechanical Engineering Congress and Exposition*, Seattle, WA, 2007.

133 **Arsie, I., Flauti, G., Pianese, C., Rizzo, G., Barberio, C., Flora, R., Serra, G.,** and **Siviero, C.** On-board diagnosis of SI engine catalyst efficiency: a confidence level analysis. In *International Workshop on Diagnostics in Automotive Engines and Vehicles*, Fisciano, Italy, 2002.

134 **Wu, J. D.** and **Chen, J. C.** Continuous wavelet transform technique for fault signal diagnosis of internal combustion engines. *NDT and E Int.*, 2006, **39**(4), 304–311.

135 **Frank, P. M.** Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – a survey and some new results. *Automatica*, 1990, **26**, 459–474.

136 **Rizzoni, G.** and **Min, P. S.** Detection of sensor failures in automotive engines. *IEEE Trans. Vehic. Tech.*, 1991, **40**(2), 487–500.

137 **Vemuri, A. T.** *Diagnosis of sensor bias faults*, In Proceedings of the American Control Conference, San Diego, CA, vol. 1, pp. 460–464, 1999.

138 **Kim, Y. W., Rizzoni, G.,** and **Utkin, V.** Automotive engine diagnosis and control via nonlinear estimation. *IEEE Contr. Sys. Mag.*, 1998, **18**(5), 84–99.

139 **Kim, Y. W., Rizzoni, G.,** and **Utkin, V. I.** Developing a fault tolerant power-train control system by integrating design of control and diagnostics. *Int. J. Rob. Nonlin. Contr.*, 2001, **11**(11), 1095–1114

140 **Namburu, S. M., Chigusa, S., Prokhorov, D., Qiao, L., Choi, K., Pattipati, K., Eng, T. M., America, M. N.,** and **Arbor, A.** *Application of an effective data-driven approach to real-time time fault diagnosis in automotive engines*, In Proceedings of the IEEE Aerospace Conference, Big Sky, MT, pp. 1–9, 2007.

141 **Weinhold, N., Ding, S. X., Jeinsch, T.,** and **Schultalbers, M.** Embedded model-based fault diagnosis for on-board diagnosis of engine control systems. In *IEEE Conference on Control Applications*, Toronto, Canada, 2005, pp. 1206–1211.

142 **Gertler, J. J., Costin, M., Fang, X., Hira, R., Kowalczuk, Z.,** and **Luo, Q.** Model-based on-board fault detection and diagnosis for automotive engines. *Contr. Engng Pract.*, 1993, **1**(1), 3–17.

143 **Soliman, A., Rizzoni, G.,** and **Kim, Y. W.** Diagnosis of an automotive emission control system using fuzzy inference. *Contr. Engng Pract.*, 1999, **7**(2), 209–216.

144 **Franchek, M. A., Buehler, P. J.,** and **Makki, I.** Intake air path diagnostics for internal combustion engines. *J. Dyn. Sys., Measure., Contr.*, 2007, **129**, 32.

145 **Chen, P. C.** Fault correction of an airflow signal in a gasoline engine system using a neural fuzzy scheme and genetic algorithm. *Proc. IMechE, Part D: J. Automobile Engineering*, 2009, **223**(4), 533–547, DOI: 10.1243/09544070JAUTO1064.

146 **Olofsson, N.** *Evaluation of observers for fault diagnosis on an automotive engine*. MSc Thesis, Department of Electrical Engineering, Linkoping University, Linkoping, Sweden, 2000.

147 **Vinsonneau, J. A. F., Shields, D. N., King, P.,** and **Burnham, K. J.** *Use of neural networks for modelling and fault detection for the intake manifold of a SI engine*, 2003 (Springer Verlag, Wien).

148 **Vinsonneau, J. A. F., Shields, D. N., King, P. J.,** and **Burnham, K. J.** *Improved SI engine modelling techniques with application to fault detection*, In Proc. IEEE International Conference on Control Applications, vol. 2, pp. 719-724, Glasgow, Scotland, 2002.

149 **Clever, S.** and **Isermann, R.** *Model-based fault detection and diagnosis with special input excitation applied to a modern diesel engine*. In Proceedings of 17th IFAC World Congress, Seoul, Korea, 2008, vol. 17, pp. 2031–2036.

150 **Gunnarsson, M.** *Parameter estimation for fault diagnosis of an automotive engine using extended Kalman filters*, 2001 (Linköping University, Linköping, Sweden).

151 **Nilsson, F.** *Diagnosis of a truck engine using nonlinear filtering techniques*, 2007 (Department of Electrical Engineering, University of Linköping, Linköping, Sweden).

152 **Wang, Y. Y.** and **Haskara, I.** Exhaust pressure estimation and its application to variable geometry turbine and wastegate diagnostics. In *American Control Conference*, Baltimore, MD, 2010.

153 **Ceccarelli, R., De Wit, C., Moulin, P.,** and **Sciarretta, A.** Model-based adaptive observers for intake leakage detection in diesel engines. In *American Control Conference*, St. Louis, MO, 2009, pp. 1128–1135.

154 **Ceccarelli, R., Moulin, P.,** and **De Wit, C.** Robust strategy for intake leakage detection in diesel engines. In *IEEE Conference on Control Applications*, Saint Petersbourg, Russia, 2009, pp. 340–345.

155 **Nyberg, M.** Using hypothesis testing theory to evaluate principles for leakage diagnosis of

automotive engines. *Contr. Engng Pract.*, 2003, **11**(11), 1263–1272.
156 **Nyberg, M., Perkovic, A.,** and **Nielsen, L.** Model based diagnosis of leaks in the air intake system of an SI-engine. In *SAE International Congress & Exposition*, Detroit, MI, 1998.
157 **Andersson, P.** and **Eriksson, L.** *Detection of exhaust manifold leaks on a turbocharged SI-engine with wastegate*, In Proceedings of SAE World Congress and Exhibition, Detroit, MI, 2002.
158 **US Department of Energy**. Annual progress report for advanced combustion engine research and development, DOE Energy Efficiency and Renewable Energy Vehicle Technologies Program, 2010.
159 **Lachaux, T., Musculus, M. P. B., Singh, S.,** and **Reitz, R. D.** Optical diagnostics of late-injection low-temperature combustion in a heavy-duty diesel engine. *J. Engng Gas Turbine Power*, 2008, **130**, 032808.
160 **Baltusis, P.** On-board vehicle diagnostics. In *SAE Convergence International Congress & Exposition on Transportation Electronics*, Detroit, MI, 2004.
161 **Carr, B.** Practical application of remote diagnostics. *SAE Trans.*, 2005, **114**(7), 376–382.
162 **Namburu, S. M., Chigusa, S., Qiao, L., Azam, M.,** and **Pattipati, K. R.** Application of signal analysis and data-driven approaches to fault detection and diagnosis in automotive engines. In *IEEE SMC conference*, Taipei, Taiwan, 2006.
163 **Luo, J., Pattipati, K. R., Qiao, L.,** and **Chigusa, S.** An integrated diagnostic development process for automotive engine control systems. *IEEE Trans. Sys., Man, Cyber., Part C: App. Rev.*, 2007, **37**(6), 1163–1173.
164 **Ding, S. X., Weinhold, N., Zhang, P., Ding, E. L., Jeinsch, T.,** and **Schultalbers, M.** Integration of FDI functional units into embedded tracking control loops and its application to FDI in engine control systems. In *IEEE Conference on Control Applications*, Toronto, Canada, 2005.

APPENDIX

Notation

ANN	artificial neural network
ARMA	auto regressive moving average
DOC	diesel oxidation catalyst
DPF	diesel particulate filter
ECT	engine coolant temperature
EGR	exhaust gas recirculation
EKF	extended Kalman filter
FDI	fault detection and isolation
FTP	federal test procedure
g/bhp-hr	gram per brake-horsepower per hour
HC	hydroscarbon
HCCI	homogeneous charge compression ignition
HEGO	heated exhaust gas oxygen
IC	internal combustion
LNT	lean NO_x trap
LTC	low-temperature combustion
MAF	mass air flow
MAP	manifold air pressure
MIL	malfunction indicator light
NMOG	non-methane organic gas
NO_x	nitrogen oxides
NPERG	non-linear parity equation residual generation
OBD	on-board diagnostics
PCA	principle component analysis
PCCI	premixed charge compression ignition
PCV	positive crankcase ventilation
PLS	partial least squares
PM	particulate matter
RBF	radial basis function
SCR	selective catalytic reduction
SHT	structured hypothesis test
SI	spark ignited
TWC	three-way catalyst
UEGO	universal exhaust gas oxygen
VGT	variable geometry turbo
VVT	variable valve timing