2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

A Data-Driven and Probabilistic Approach to Residual Evaluation for Fault Diagnosis

Carl Svärd, Mattias Nyberg, Erik Frisk, and Mattias Krysander

***Abstract*— An important step in fault detection and isolation is residual evaluation where residuals, signals ideally zero in the no-fault case, are evaluated with the aim to detect changes in their behavior caused by faults. Generally, residuals deviate from zero even in the no-fault case and their probability distributions exhibit non-stationary features due to, e.g., modeling errors, measurement noise, and different operating conditions. To handle these issues, this paper proposes a data-driven approach to residual evaluation based on an explicit comparison of the residual distribution estimated on-line and a no-fault distribution, estimated off-line using training data. The comparison is done within the framework of statistical hypothesis testing. With the Generalized Likelihood Ratio test statistic as starting point, a more powerful and computational efficient test statistic is derived by a properly chosen approximation to one of the emerging likelihood maximization problems. The proposed approach is evaluated with measurement data on a residual for diagnosis of the gas-flow system of a Scania truck diesel engine. The proposed test statistic performs well, small faults can for example be reliable detected in cases where regular methods based on constant thresholding fail.**

I. INTRODUCTION

Fault Detection and Isolation (FDI) typically contains three essential steps: residual generation, residual evaluation, and fault isolation, see e.g. [1]. In the first step, a model of the system is used together with measurements to generate residuals. In the second step, the residuals are evaluated with the aim to detect changes in the residual behavior caused by faults, and in the third step the detected faults are isolated. This paper addresses the second step, residual evaluation.

Ideally, residuals are signals that are zero when the system is fault-free, and non-zero otherwise. However, in practice residuals deviate from zero even in the no-fault case due to uncertainties such as modeling errors and measurement noise. Furthermore, the magnitude of uncertainties is time-varying because of changes in operating conditions, which causes the probability distributions of residuals to be non-stationary. For a real-world illustration, consider Figure 1 in which a model-based residual for diagnosis of the gas-flow system in a truck engine is shown. Clearly, the residual is not zero in the no-fault case and it is obvious that the distribution of the residual exhibit non-stationary features. Moreover, it can be noted that the difference between the residual in the no-fault and fault cases is time-varying. Nevertheless, the incidence of a difference implies that the fault is potentially detectable.

Apparently, when facing the problem of evaluating a residual as the one depicted in Figure 1, approaches based on solely detecting changes in the mean or variance of the residual by means of constant thresholding may not be successful. A potential solution is to consider adaptive thresholds [2], where additional system knowledge, either qualitative [3], [4], [5] or quantitative [6], [7], are exploited to derive non-constant thresholds that take time-varying uncertainties into account. The viewpoint taken in this work is however that all information regarding the system that can be used for modeling has been used and incorporated in the model. Thus, no additional system knowledge is available and uncertainties must be handled in some other way. In this paper, a data-driven and probabilistic approach is proposed.

The main contribution is to base the residual evaluation on an explicit comparison of the probability distribution of the residual estimated on-line using current data, and the distribution of the no-fault residual, estimated off-line using no-fault training data. To handle the non-stationarity, the distribution of the no-fault residual is continuously adapted by characterizing it as a mixture of different no-fault distributions. The comparison is done in the framework of statistical hypothesis testing by application of the Generalized Likelihood Ratio (GLR). The approach is data-driven and no assumptions regarding the properties of the residual distributions nor the faults to be detected are made.

In Section II, the problem formulation is stated and cast as a statistical hypothesis test. In Section III the GLR is utilized to design a preliminary test statistic for the hypothesis test, and the likelihood function together with the emerging likelihood maximizations are discussed. Section IV explores how the likelihood maximizations can be computed. In Section V, the preliminary test statistic is improved, in terms of test power and computational efficiency, by instead considering a properly chosen approximation to one of the likelihood maximization problems. In Section VI, the proposed residual evaluation method is applied to a residual for diagnosis of the gas-flow system of a Scania truck diesel engine. The paper is concluded in Section VII.

II. PROBLEM FORMULATION

Let the discrete random variable R with range $\mathcal{R} = \{x_1, x_2, \ldots, x_m\}$ represent the discretized and sampled value of a residual. It is assumed that the probability distribution of R is described by the probability mass function (pmf) $f_R\left(r|\theta\right)$, which is fully parameterized by $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ and given by

$$f_R\left(r|\theta\right) = \theta_j, \quad \text{if } r = x_j, \tag{1}$$

for $j = 1, 2, \ldots, m$, where $\theta_j \geq 0$ and $\sum_{j=1}^{m} \theta_j = 1$.

This work was supported by Scania CV AB

All authors are affiliated with Department of Electrical Engineering, University of Linköping, SE-58183 Linkoping, Sweden carl@isy.liu.se

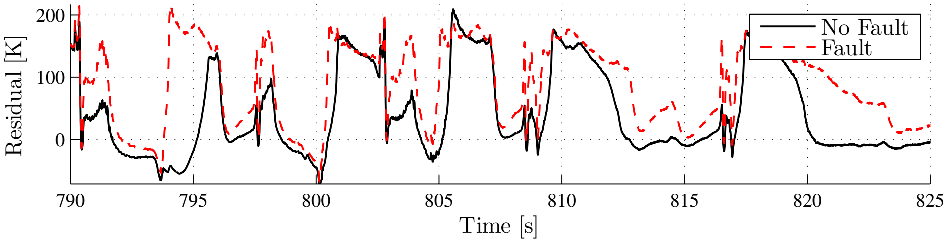

Fig. 1. A model-based residual for diagnosis of the gas-flow in a truck engine in the no-fault (solid) and fault (dashed) cases.

As a prerequisite, it is assumed that representative no-fault data, quantified in terms of estimated probability distributions describing the residual in the no-fault case, are available. Consequently, let Θ_{NF} represent a set of no-fault distributions parameterized according to (1). Typically, a certain $\theta \in \Theta_{NF}$ characterizes the distribution of the residual under some specific operating conditions.

Consider now a set of N observed residual values $r_1, r_2, \ldots, r_N$, i.e., a sample $R_1, R_2, \ldots, R_N$ of the joint residual distribution $f_R(r_1, r_2, \ldots, r_N|\theta)$ where the outcome of R_i is r_i. Throughout this work, it is assumed that the sample $R_1, R_2, \ldots, R_N$ is independent and identically distributed (iid) with pmf $f_R(r_i|\theta)$ according to (1).

Note that the iid assumption may not be valid since residuals often are obtained as output from dynamic systems and thereby exhibit Markovian properties. The assumption is still necessary to make subsequent derivations tractable, or even possible. It can however often be fulfilled in practice by sampling the residual at a sufficiently low rate, and the approach has nevertheless shown to be applicable in practical settings, for example in the experimental evaluation presented in Section VI.

Given the iid sample $R_1, R_2, \ldots, R_N$ of the distribution (1), the problem of residual evaluation is to determine if $\theta \in \Theta_{NF}$ or not. This problem can be formulated as the hypothesis test

$$H^0 : \theta \in \Theta_{NF}, \quad H^1 : \theta \notin \Theta_{NF}, \tag{2}$$

where the null hypothesis H^0 corresponds to the no-fault case and the alternative hypothesis H^1 to the faulty case. Next section deals with the problem of designing a test statistic for the hypothesis test (2).

III. DESIGN OF TEST STATISTIC

A standard approach when encountering a hypothesis test with non-simple hypotheses is to utilize the Generalized Likelihood Ratio (GLR), see e.g. [8], [9]. For testing hypothesis H^0 versus H^1 in (2) the GLR test statistic is

$$\Lambda(\mathbf{r}) = \frac{\max_{\theta\in\Theta} L(\theta|\mathbf{r})}{\max_{\theta\in\Theta_{NF}} L(\theta|\mathbf{r})}, \tag{3}$$

where $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ and $L(\theta|\mathbf{r})$ is the *likelihood function*. The sets Θ and Θ_{NF} denote the entire parameter space and the no-fault parameter space, respectively. The hypothesis H^0 is rejected in favor of H^1, i.e., a fault is present in the system, if the test statistic $\Lambda(\mathbf{r}) > J$, where J is a (constant) threshold. Note that for on-line use, the residual observations in $\mathbf{r}$ may be taken by using a sliding window, i.e., at time instant t the observations are given by $\mathbf{r}_t = (r_{t-N+1}, r_{t-N+2}, \ldots, r_t)$.

In order to employ the test statistic $\Lambda(\mathbf{r})$, the maximizations in the denominator and numerator of (3) must be performed. Before considering these maximization problems, the objective function, i.e., the likelihood function $L(\theta|\mathbf{r})$, will be studied in more detail.

A. The Likelihood Function

Consider now the iid sample $R_1, R_2, \ldots, R_N$ of the distribution (1), where the outcome of R_i is r_i. Let n_j denote how many of the residual values $r_1, r_2, \ldots, r_N$ that have value x_j, i.e.,

$$n_j = |\{r_k \in \{r_1, r_2, \ldots, r_N\} : r_k = x_j, x_j \in \mathcal{R}\}|, \tag{4}$$

for $j = 1, 2, \ldots, m$. Since the sample $R_1, R_2, \ldots, R_N$ is iid, the random variables R_i are mutually independent and thus the joint residual distribution can be written as $f_R(\mathbf{r}|\theta) = f_R(r_1, r_2, \ldots, r_N|\theta) = \prod_{i=1}^{N} f_R(r_i|\theta)$, where $f_R(r_i|\theta)$ is given by (1). Consequently, the likelihood function $L(\theta|\mathbf{r}) = f_R(\mathbf{r}|\theta)$ takes the form

$$L(\theta|\mathbf{r}) = \prod_{i=1}^{N} f_R(r_i|\theta). \tag{5}$$

Using (1) and (4), the likelihood function (5) reduces to

$$L(\theta|\mathbf{r}) = \prod_{i=1}^{N} f_R(r_i|\theta) = \prod_{j=1}^{m} \theta_j^{n_j}. \tag{6}$$

In order to simplify calculations, the log-likelihood function

$$l(\theta|\mathbf{r}) = \log L(\theta|\mathbf{r}) = \sum_{j=1}^{m} n_j \log \theta_j, \tag{7}$$

will instead be considered, where it, for $j = 1, 2, \ldots, m$, is assumed that $n_j > 0$ and $\theta_j > 0$.

Note that the assumption $n_j > 0$ can be done without loss of generality. If $n_k = 0$, i.e., there are no observations with value x_k, the corresponding term in (7) is $0 \cdot \log \theta_k \equiv 0$, independent of θ_k. Thus, this term can be neglected and the log-likelihood function written as $l(\theta|\mathbf{r}) = \sum_{j=1, j\neq k}^{m} n_j \log \theta_j$. The assumption $\theta_j > 0$ is just a technicality due to the usage of $\log(\cdot)$. It however turns out, see Section IV-A, that $n_j > 0$ implies $\theta_j > 0$.

B. The Maximization Problems

Since $\log(\cdot)$ is a strictly increasing function, it is equivalent to maximize the likelihood and log-likelihood functions. By utilizing (7), the two maximizations in (3) can be re-stated as

$$\max_{\theta\in\Theta} \; l(\theta|\mathbf{r}) = \max_{\theta\in\Theta}\sum_{j=1}^{m} n_j \log\theta_j \tag{8}$$

and

$$\max_{\theta\in\Theta_{NF}} \; l(\theta|\mathbf{r}) = \max_{\theta\in\Theta_{NF}}\sum_{j=1}^{m} n_j \log\theta_j. \tag{9}$$

Both (8) and (9) involves finding a θ that (globally) maximizes the log-likelihood function $l(\theta|\mathbf{r})$ given the observations in $\mathbf{r}$, i.e., finding the Maximum Likelihood Estimator (MLE) of θ based on $\mathbf{r}$, see e.g. [8]. The two problems differ by the space in which the maximizing θ should be contained. In (8) the maximization should be performed over the entire parameter space defined by

$$\Theta = \{\theta \in \mathbb{R}^m : \theta_j \geq 0, \sum_{j=1}^{m}\theta_j = 1\}, \tag{10}$$

that is, any θ such that $f_R(r|\theta)$ is a pmf is valid. In (9), however, the maximization should be performed over the space $\Theta_{NF} \subset \Theta$, containing all θ that describes the behavior of the residual in the no-fault case. The space Θ_{NF} will be properly defined in Section IV-B.

IV. LIKELIHOOD MAXIMIZATIONS

This section is devoted to explore in detail how to find the distribution estimates that solve the two MLE problems (8) and (9), corresponding to the numerator and denominator of the GLR test statistic (3), respectively.

A. MLE for the Numerator of the GLR Test Statistic

Consider first the maximization problem (8). With Θ given by (10), (8) can be equivalently stated as

$$\max_{\theta\in\mathbb{R}^m}\sum_{j=1}^{m} n_j \log\theta_j, \quad \text{s.t.} \quad \theta_j \geq 0, \; \sum_{j=1}^{m}\theta_j = 1, \tag{11}$$

which is a general non-linear constrained maximization problem. An important property of (11) is given by the following result.

Lemma 1: The maximization problem (11) is a concave maximization problem.

Proof: The problem (11) is a concave maximization problem if $f(\theta) \triangleq \sum_{j=1}^{m} n_j \log\theta_j$ is a concave function, $g_j(\theta) \triangleq -\theta_j$, $j = 1, 2, \ldots, m$, are convex functions, and $h(\theta) \triangleq \sum_{j=1}^{m}\theta_j - 1$ is affine. The last two conditions are trivially satisfied since both $g_j(\theta)$ and $h(\theta)$ are linear functions. Since $\frac{\partial}{\partial\theta_k}f(\theta) = \frac{n_k}{\theta_k}$, for $k = 1, 2, \ldots, m$, it follows that

$$\frac{\partial^2}{\partial\theta_k\partial\theta_l}f(\theta) = \begin{cases} -\frac{n_k}{\theta_k^2}, & l = k \\ 0, & l \neq k. \end{cases} \tag{12}$$

Thus the eigenvalues of the Hessian of $f(\theta)$, whose (k, l)-element is given by (12), are strictly negative, since by assumption $n_j > 0$, and consequently the Hessian of $f(\theta)$ is negative definite and $f(\theta)$ is a concave function. ■

It turns out that (11), and equivalently (8), can be solved explicitly.

Proposition 1: The global solution to (11) is given by

$$\theta^* = \frac{1}{N}(n_1, n_2, \ldots, n_m). \tag{13}$$

Proof: Since, according to Lemma 1, the maximization problem (11) is concave it is sufficient to show that there exist constants μ_j, $j = 1, 2, \ldots, m$, and λ such that θ^* satisfies the Karuhn-Kuhn-Tucker (KKT) conditions, see e.g. [10], [11],

$$\nabla f(\theta^*) + \sum_{j=1}^{m}\mu_j\nabla g_j(\theta^*) + \lambda\nabla h(\theta^*) = 0, \tag{14a}$$

$$g_j(\theta^*) \leq 0, \quad j = 1, 2, \ldots, m \tag{14b}$$

$$h(\theta^*) = 0, \tag{14c}$$

$$\mu_j \geq 0, \quad j = 1, 2, \ldots, m \tag{14d}$$

$$\mu_j g_j(\theta^*) = 0, \quad j = 1, 2, \ldots, m, \tag{14e}$$

where $f(\theta) \triangleq \sum_{j=1}^{m} n_j \log\theta_j$, $g_j(\theta) \triangleq -\theta_j$, and $h(\theta) \triangleq \sum_{j=1}^{m}\theta_j - 1$. Consider first condition (14c) which is trivially satisfied since $\sum_{j=1}^{m}\frac{n_j}{N} = 1$ by definition of n_j in (4). By assumption and without loss of generality, see Section III-A, it holds that $n_j > 0$ and thus condition (14b) is satisfied with strict inequality since $\theta_j^* = \frac{n_j}{N} > 0$ for $j = 1, 2, \ldots, m$. This implies, due to condition (14e), that $\mu_j = 0$ for $j = 1, 2, \ldots, m$ and thereby also (14d) is satisfied. Differentiation of $f(\theta) = \sum_{j=1}^{m} n_j \log\theta_j$, $g_j(\theta) = -\theta_j$, and $h(\theta) = \sum_{j=1}^{m}\theta_j - 1$, gives that $\nabla f(\theta) = \left(\frac{n_1}{\theta_1}, \frac{n_2}{\theta_2}, \ldots, \frac{n_m}{\theta_m}\right)^T$, $\nabla g_j(\theta) = -e_j^T$, $\quad j = 1, 2, \ldots, m$, and $\nabla h(\theta) = (1, 1, \ldots, 1)^T$, where e_j denotes the j:th unit vector. With $\theta = \theta^*$ and $\mu_j = 0$, for $j = 1, 2, \ldots, m$, condition (14a) reduces to $N + \lambda = 0$ and hence (14a) is satisfied by $\lambda = -N$ and the proof is complete. ■

Since (11) is equivalent to (8), Proposition 1 states that θ^* given by (13) is the MLE of θ, under the assumption that the sample $R_1, R_2, \ldots, R_N$ is iid so that the derivations in Section III-A are valid.

Note that if the MLE of θ as given by (13) is considered, $n_j > 0$ implies $\theta_j > 0$, which justifies the technical assumption $\theta_j > 0$ posed in Section III-A.

Also note that the MLE of θ given the observations in $\mathbf{r}$, can be obtained from the normalized histogram, with m bins, calculated from $\mathbf{r}$.

B. MLE for the Denominator of the GLR Test Statistic

Consider now the maximization problem (9) and let $\theta^1, \theta^2, \ldots, \theta^p$, where $\theta^j \in \mathbb{R}^m$ is a column vector, characterize distributions that describe the residual in the no-fault case. In the ideal case, one set of observations can be described by exactly one of the given distributions. In this ideal case it would be natural to define $\Theta_{NF} = \{\theta^1, \theta^2, \ldots, \theta^p\}$ and the problem (9) would be reduced to simply picking one θ^j from the set Θ_{NF} that maximizes the log-likelihood in $l(\theta|\mathbf{r})$.

In the general case, the observations however origins from more than one of the distributions in Θ_{NF} as defined above.

A typical situation is illustrated in Figure 2, where a sudden change of the operating conditions after 30 seconds causes the distribution of the residual to change. Note that this change should not result in detection of a fault. If the set of observations contains samples from both before and after the change, the approach described above will not yield a satisfactory result.

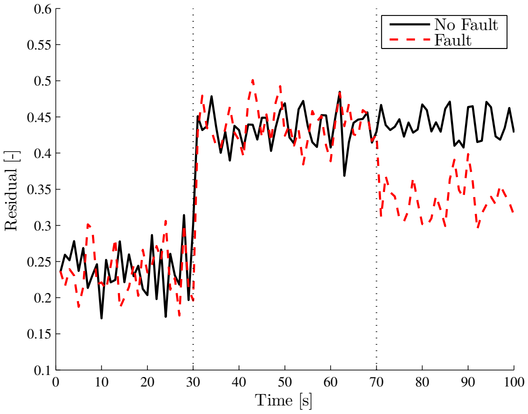

Fig. 2. A typical fault situation where the residual is distributed in one way the first 30 seconds, in another way the last 70 seconds, and the fault occurs after 70 seconds.

Note however that if the distribution of the residual in Figure 2 is described by θ^1 the first 30 seconds, and by θ^2 the last 70 seconds, the complete distribution can be described, in the no-fault case, by the *mixture distribution* characterized by $\theta^1\gamma_1 + \theta^2\gamma_2$, where $\gamma_1 = 0.3$ and $\gamma_2 = 0.7$. Under the assumption that the time spent between two distributions is short, this observation suggests that the set Θ_{NF} of no-fault distributions is extended to also include convex combinations of the original distributions characterized by the parameters $\theta^1, \theta^2, \dots, \theta^p$, i.e.,

$$\Theta_{NF} = \{\theta \in \Theta : \exists \gamma \in \Gamma; \theta = \Delta\gamma\}, \tag{15}$$

where $\Gamma = \{\gamma \in \mathbb{R}^p : \sum_{i=1}^{p} \gamma_i = 1, \gamma \geq 0\}$ and

$$\Delta = \begin{bmatrix} \theta^1 & \theta^2 & \cdots & \theta^p \end{bmatrix} \tag{16}$$

is a $m \times p$ matrix whose columns consist of the distribution parameters $\theta^1, \theta^2, \dots, \theta^p$.

With this extension of Θ_{NF}, the original problem (9) can be equivalently formulated as

$$\max_{\gamma \in \mathbb{R}^p} \sum_{j=1}^{m} n_j \log\left(\Delta_j \gamma\right), \quad \text{s.t.} \quad \gamma_i \geq 0, \ \sum_{i=1}^{p} \gamma_i = 1, \tag{17}$$

where $\Delta_j = \left[\theta_j^1, \theta_j^2, \dots, \theta_j^p\right]$ denotes the j:th row of the matrix Δ. As well as (11), problem (17) exhibits the following important property.

Lemma 2: The maximization problem (17) is a concave maximization problem.

Proof: The proof can be carried out in the same manner as the proof to Lemma 1 with $f(\gamma) \triangleq \sum_{j=1}^{m} n_j \log\left(\Delta_j \gamma\right)$, $g(\gamma) \triangleq -\gamma$, and $h(\gamma) \triangleq \sum_{i=1}^{p} \gamma_i - 1$. As an alternative to differentiate $f(\gamma)$, which may be cumbersome, it suffices to note that the $\log(\cdot)$ function is a concave function. ■

In the general case, it is unfortunately not possible to obtain an explicit solution to (17), as was the the case for (11). There are however several efficient numerical approaches, see e.g. [11]. The concavity property of (17) facilitates the solving since it implies that if a local maximum can be found, then it is also a global maximum.

V. IMPROVED TEST STATISTIC

Even though the problem (17) can be solved, at least to some specified tolerance degree, it is in fact better to use an approximate solution. It is in this section shown that the considered approximation to (17), and equivalently (9), improves the test statistic (3) in terms of test power, and may also reduce the required computational effort.

A. The Approximative Problem

Consider now instead of (17) the minimization problem

$$\min_{\gamma \in \mathbb{R}^p} \|\Delta\gamma - \theta^*\|_2^2, \quad \text{s.t.} \quad \gamma_i \geq 0, \ \sum_{i=1}^{p} \gamma_i = 1, \tag{18}$$

where $\theta^* = \frac{1}{N}(n_1, n_2, \dots, n_m)$, i.e., the solution to (11) and equivalently (8). Intuitively, it makes sense to chose γ so that $\theta = \Delta\gamma$ is as close as possible to θ^*, in terms of the Euclidean norm, but still is contained in Θ_{NF}, since θ^* is the solution to (11) which is equivalent to (17) if $\theta = \Delta\gamma$. Usage of (18) as an approximation to (17) is formally justified by the following result.

Lemma 3: Let $\theta^* = \frac{1}{N}(n_1, n_2, \dots, n_m)$ be an element in Θ_{NF}, then γ^* is the solution to (17) if and only if γ^* is the solution to (18).

Proof: First note that from the assumptions regarding θ^*, Proposition 1, and Lemma 2, it holds that there always exists a unique γ so that $\theta^* = \Delta\gamma$ and γ is a solution to (17). Hence, if γ^* is the solution to (17) it follows that $\Delta\gamma^* - \theta^* = 0$ and γ^* is also a solution to (18). Conversely, let γ^* be a solution to (18). Since the Euclidean norm $\|\cdot\|_2$ defines a convex function and the conditions in (18) are linear the minimization problem (18) is a convex optimization problem, see the proof to Lemma 1 for more details. Due to the convexity, γ^* is a unique solution. From the assumption $\theta^* \in \Theta_{NF}$ and the uniqueness of γ^* it follows that $\theta^* = \Delta\gamma^*$ and hence θ^* is the solution to (17). ■

B. Improving the Test Statistic

Recall the GLR test statistic

$$\Lambda(\mathbf{r}) = \frac{\max_{\theta \in \Theta} L(\theta|\mathbf{r})}{\max_{\theta \in \Theta_{NF}} L(\theta|\mathbf{r})} = \frac{L(\theta^*|\mathbf{r})}{L(\theta_{NF}^*|\mathbf{r})}, \tag{19}$$

where θ^* is the solution to (11), $\theta_{NF}^* = \Delta\gamma^*$ and γ^* is the solution to (17), and the likelihood $L(\theta|\mathbf{r})$ is given by (6). Consider now instead the test statistic

$$\hat{\Lambda}(\mathbf{r}) = \frac{L(\theta^*|\mathbf{r})}{L(\hat{\theta}|\mathbf{r})}, \tag{20}$$

where $\hat{\theta} = \Delta\hat{\gamma}$ and $\hat{\gamma}$ is the solution to (18).

The important implication of Lemma 3 is that if $\theta^* \in \Theta_{NF}$, i.e., under H^0, the solution $\hat{\theta}$ coincides with θ_{NF}^* and hence $\hat{\Lambda}(\mathbf{r}) \equiv \Lambda(\mathbf{r})$. If instead $\theta^* \notin \Theta_{NF}$, i.e., under H^1, it holds that $L(\hat{\theta}|\mathbf{r}) \leq L(\theta^*_{NF}|\mathbf{r})$, due to the concavity property of $L(\theta|\mathbf{r})$, or equivalently $l(\theta|\mathbf{r}) = \log L(\theta|\mathbf{r})$, see Lemma 1. Therefore, under H^1, it holds that $\hat{\Lambda}(\mathbf{r}) \geq \Lambda(\mathbf{r})$. Thus, by using the test statistic $\hat{\Lambda}(\mathbf{r})$ instead of $\Lambda(\mathbf{r})$ in the test $\Lambda(\mathbf{r}) > J$, the probability of rejecting H^0 in general increases, that is, the test $\hat{\Lambda}(\mathbf{r}) > J$ is more *powerful* than $\Lambda(\mathbf{r}) > J$.

The power of a statistical test $\Lambda(\mathbf{r}) > J$ is measured by the *power function*

$$\beta(\theta) = \Pr(\text{reject } H^0|\theta) = \Pr(\Lambda(\mathbf{r}) > J|\theta), \qquad (21)$$

which gives the probability of rejecting the hypothesis H^0 given a fixed θ and a fixed threshold J, see e.g. [8]. With this notion, the discussion above can be compiled into the following result.

Proposition 2: Let $\Lambda(\mathbf{r})$ be given by (19), $\hat{\Lambda}(\mathbf{r})$ by (20), and let $\beta(\theta)$ and $\hat{\beta}(\theta)$ denote the power functions of the tests $\Lambda(\mathbf{r}) > J$ and $\hat{\Lambda}(\mathbf{r}) > J$, respectively. Then $\hat{\beta}(\theta) \geq \beta(\theta)$, with equality if $\theta \in \Theta_{NF}$.

C. Implementation Issues

The optimization problem (18) is equivalent to a linear least squares problem with equality and non-negative constraints for which efficient algorithms exist, see e.g. [12]. A further approximation can be obtained by relaxation of the constraint $\sum_{i=1}^{p} \gamma_i = 1$ in (18), which then may be incorporated in the objective function. This relaxed problem can be stated as a Non-Negative Least Square (NNLS) problem, see e.g. [13], [14]. In MATLAB, these two approximations can be solved with the commands `lsqlin` and `lsqnonneq`, respectively.

D. Parameters

The parameters involved in the design of the test $\hat{\Lambda}(\mathbf{r}) > J$ in (20) are the size N of the residual sample, the discretization m of the residual, the number p of no-fault distribution parameters in (16), and the detection threshold J.

1) Sample Size: The sample size N is a trade-off between detection performance and complexity. A large N will give the test statistic smoothed, low-pass, characteristics. This makes it possible to detect small changes in the residual, but on the other hand a large N may increase the detection time. Moreover, a large N requires more memory and is more computationally demanding.

2) Distribution Resolution: Choosing the discretization m of the residual, or likewise the resolution m of the residual distribution (1), in fact corresponds to the well-studied, but nevertheless difficult, problem of choosing the number of bins in a regular histogram given a sample of data. Numerous approaches for solving this problem exist, see for example [15] and references therein. Regardless of the method used to solve the problem, the choice of m is a trade-off between accuracy and complexity, in terms of computational load and memory. A larger m results in a more accurate discretization of the residual and higher resolution of the probability distributions. On the other hand, a large m requires more memory and involves more computations. The choice of m is also related to the choice of N, since a small N together with a large m will result in an inadequate estimation of the distribution, i.e., a sparse histogram.

3) Number of No-Fault Distributions: The value of p determines the number of no-fault distribution parameters in the set Δ defined according to (16). Typically, a $\theta^i \in \Delta$ characterizes the distribution of the no-fault residual under some specific operating condition. Thus, the value of p is determined by the total number of considered operating conditions of the studied system. In general, determination of p requires expert knowledge regarding the system. For a systematic approach, which is out of the scope of this work, machine learning approaches such as clustering, see e.g., [16], may also be exploited. Using a clustering approach, a given set of no-fault data can be automatically partitioned into an appropriate number of clusters, where each cluster contains data of the same distribution.

4) Detection Threshold: The choice of detection threshold J, is a trade-off between detection time and probability of false detections. The higher the threshold, the longer the detection time and the lower the probability of false detections. The value of J is also related to the choice of N. A larger N implies in general that the sampled residual origins from several distributions, e.g., the system may be used under several different operating conditions during the sampling period, c.f. Figure 2. If the time spent between two distributions is long, this may result in an increase of the test statistic. Therefore, a larger N may imply usage of a larger threshold J in order to avoid false detections.

VI. EXPERIMENTAL EVALUATION

The proposed residual evaluation method has been applied to the problem of fault detection in the gas-flow system of a Scania 6 cylinder, 13 liter, truck diesel engine equipped with Exhaust Gas Recirculation (EGR), Variable Geometry Turbine (VGT) and intake throttle. The purpose of the evaluation is to analyze the performance of the proposed test statistic (20) using measurement data, and also to see how the performance is influenced by different choices of the involved parameters, mainly the window size N.

A. Gas-Flow Diagnosis

The gas-flow system, or rather the truck itself, is a complex system that operates under a variety of conditions, for example high-way and city drive, different ambient conditions, different loads and truck drivers, etc. Diagnosis of the gas-flow system consists of detecting and isolating faults in sensors (pressure, temperature, mass-flow), actuators, as well as detection of, e.g., manifold leakages and clogged air filters. The main incentives for gas-flow diagnosis are fault tolerant control, and On-Board Diagnosis (OBD) regulations.

The model of the system, which is described in [17], relies on both fundamental first principle physics and gray-box modeling. For diagnosis of the gas-flow system, a set of residual generators based on the engine model were designed with the method described in [18]. Naturally, the model does not describe all aspects of the system, i.e., under all operating conditions, leading to that all residuals exhibit properties similar to those illustrated in Figure 1.

The residual considered in this study is sensitive to 10 faults: 3 leakages, 6 sensor faults, and 1 actuator fault. The residual is the output from a non-linear residual generator using both integral and derivative causality, and which has 9 input signals. The value of the residual is based on a comparison of two modeled values of the temperature before the cylinders.

B. Estimation of Residual Distributions

For the gas-flow system, the magnitude of uncertainties and therefore also the properties of the residual distributions, depend on numerous factors. The approach used in this case study relies on expert knowledge regarding the considered system and utilizes that operating conditions for the gas-flow system, and thereby the residual distributions, can be approximately parameterized by the engine torque. This approximation enables the use of a basic and straightforward approach for estimating the distributions, but nevertheless shows the potential of the method.

The engine torque is not measured directly but the boost pressure is, which is approximately proportional to the engine torque. By partitioning the range of the boost pressure into intervals and partition the residual accordingly, all residual values originating from the same distribution, i.e., operating condition, can be picked out. Having partitioned the residual data according to distribution, the parameters θ^i in (16) must be estimated. One option, which has been used in this paper, is to use the MLE of θ^i, for which an explicit expression is given in Proposition 1.

Two data sets were used to estimate the distributions. The first data set is about half an hour long and contains engine test bed measurements from a World Harmonized Transient Cycle (WHTC) test cycle. The second data set is approximately 3 hours long and contains measurements from a part of a test drive, including both city and high-way driving, from Södertälje to Arvidsjaur in Sweden. The two data sets were concatenated and then split into 25 partitions, i.e., 25 distributions. This number is a good trade-off between the coverage of all different features of the residual data and complexity, i.e., memory requirements and computational load.

For each of the 25 partitions of the residual data, the MLE of the parameter θ^i was computed using normalized histograms. The resolution m of the distributions, i.e. the discretization of the residual and likewise the number of bins in the histograms, was chosen as $m = 30$. For this application, this is a good trade-off between complexity, in terms of required memory and computational effort, and accuracy. In Figure 3, the resulting distributions are shown. Note that the characteristics of the estimated distributions are different, some are multi-modal and some have only one single mode.

C. Evaluation Setup

The data set used in the evaluation contains road measurements from a four hour drive that includes both high-way and city parts. This data set is a different data set than the two sets used for estimating the residual distributions.

The fault considered in this case study is a fault in the boost pressure sensor. The relation between the boost pressure sensor signal $y_{p_{im}}$ and the considered residual is dynamic and highly non-linear. The residual value r depends on the derivative of the boost pressure sensor signal, as well as the actual sensor signal, i.e., $r = F(y_{p_{im}}, \dot{y}_{p_{im}}, \ldots)$, where $F(\cdot)$ is a non-linear function. The considered fault scenario is a gain fault in the sensor, that is, the sensor signal $y_{p_{im}}$ fed to the residual generator is $y_{p_{im}} = \delta \cdot p_{im}$, where p_{im} is the actual boost pressure, and $\delta \neq 1$ indicates a gain fault. Gain faults in the range $\delta \in [0.2, 1.8]$ were implemented off-line by modification of the sensor signal.

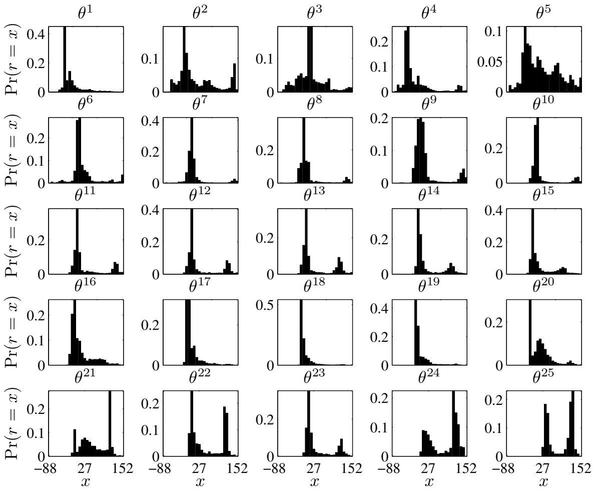

Fig. 3. The 25 estimated residual distributions used in the evaluation.

1) Evaluation Metrics: The main metric considered in this study is the power function, in this context defined as

$$\beta(\delta) = \Pr(\text{detection}|\delta) = \Pr(\lambda(\mathbf{r}) > J|\delta). \qquad (22)$$

If $\delta \neq 1$ the power function (22) gives the *probability of fault detection* and also the *probability of missed detection*, as $1 - \beta(\delta)$, given a fixed δ. If $\delta = 1$, the power function gives the *probability of false detection*, see for example [8]. Note that $\delta = 1$ in (22) corresponds to $\theta \in \Theta_{NF}$ in (21). To study another important aspect of the detection performance, the *Mean Time to Detection* (MTD) will also be considered. Note that the choices of the values of the parameters N and J, i.e. the sample size and detection threshold, are a trade-off between the metrics measured by the power function and the MTD, see Section V-D.

In order to be able to say something about the relative performance of the proposed test statistic, it will be compared to the often in practice used test statistic $s(\mathbf{r}) = \frac{1}{N}\sum_{i=1}^{N} \bar{r}_i^2$ where $\bar{\mathbf{r}} = (\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_N)$ is a low-passed filtered version of the sample $\mathbf{r}$, and N is the sample size. Note that the purpose of this comparison is merely to give a feeling of the relative performance of the proposed test statistic, and the comparison is not claimed to be exhaustive. The low-pass filtering was in this study performed with a first-order Butterworth filter and for comparison, four different cut-off frequencies, $f_1 = 0.005$ Hz, $f_2 = 0.05$ Hz, $f_3 = 0.5$ Hz, and $f_4 = 4.5$ Hz, were used. The corresponding test statistics are denoted $s_1(\mathbf{r})$, $s_2(\mathbf{r})$, $s_3(\mathbf{r})$, and $s_4(\mathbf{r})$. The residual is sampled with a frequency of $f_s = 10$ Hz.

2) Implementation Details: All test statistics were implemented and run off-line in a MATLAB environment. The test statistic $\lambda(\mathbf{r}) = \log \hat{\Lambda}(\mathbf{r})$, where $\hat{\Lambda}(\mathbf{r})$ is given by (20) was utilized, and the optimization problem (18) solved with the command `lsqnonneq`, see Section V-A. The residual observations were taken by using a sliding window, see Section III. The thresholds for all test statistics were computed based on the two data sets used in the estimation of the residual distributions and chosen in order to give a probability of false detection of 5%.

D. Results

To illustrate how the power of the test $\lambda(\mathbf{r}) > J$ varies with the window size N, Figure 4 shows the power function for the test for different values of N. Figure 4 clearly shows that the power of the test increases with N. Moreover, it can be seen that as small faults as $\delta \approx 0.95$ and $\delta \approx 1.05$, corresponding to gain faults in the boost pressure sensor of about ± 5%, may be reliably detected if N is sufficiency large.

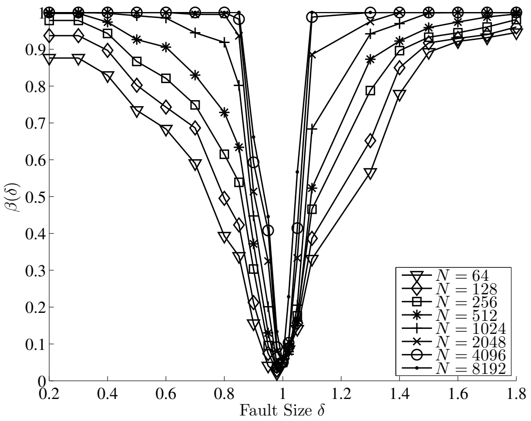

Fig. 4. Power function for the test $\lambda(\mathbf{r}) > J$ for different window sizes N. The power increases with N.

Figure 5 shows the residual and the test statistics $\lambda(\mathbf{r})$ and $s_1(\mathbf{r})$, corresponding to $f_1 = 0.005$ Hz, with $N = 2048$ for a data sequence where $\delta = 0.85$. First of all, it may be noted that as in Figure 1, the residual is non-zero in the no-fault case and its distribution is non-stationary in both the no-fault and fault cases. Moreover, the difference between the no-fault and fault residuals is also time-varying and the residuals even coincide at some occasions. It can also be seen that there is a significant difference between the test statistic $\lambda(\mathbf{r})$ in the no-fault and fault cases, and also that $\lambda(\mathbf{r})$ is above the threshold in the fault case and thus the present fault can be detected. The fault can however not be detected with the test statistic $s_1(\mathbf{r})$, which in this case performed better than each of $s_2(\mathbf{r})$, $s_3(\mathbf{r})$, and $s_4(\mathbf{r})$. It may be noted that $s_1(\mathbf{r})$ is larger in the no-fault than the fault case.

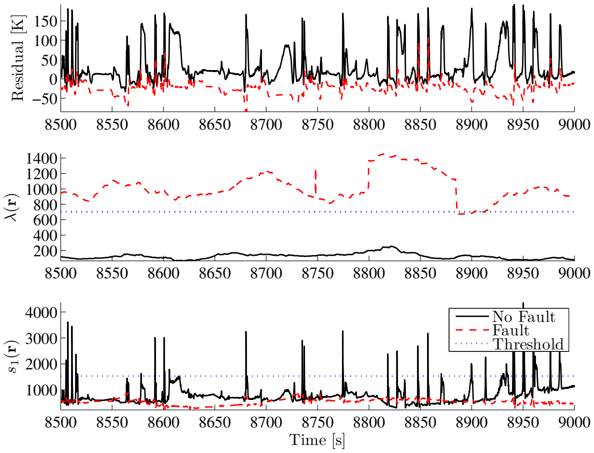

Fig. 5. Residual (top) and test statistics $\lambda(\mathbf{r})$ (middle) and $s_1(\mathbf{r})$ (bottom) with $N = 2048$ during a test sequence where $\delta = 0.85$. Quantities are solid in the no-fault case, and dashed in the fault case.

Figure 6 shows a comparison of the power functions for the tests based on the test statistics $\lambda(\mathbf{r})$, $s_1(\mathbf{r})$, $s_2(\mathbf{r})$, $s_3(\mathbf{r})$, and $s_4(\mathbf{r})$, for different values of the parameter N. First, note that the power of all tests increases with N and that the differences between the power of the tests seem to decrease with an increasing N. Second, it can be seen that the power function for the $\lambda(\mathbf{r})$ test is near symmetric for all N, while the power functions for the other tests are asymmetric and tend to be less powerful for faults sizes $\delta < 1$. For e.g. $N = 128$, the difference in power for $\delta < 1$ is significant.

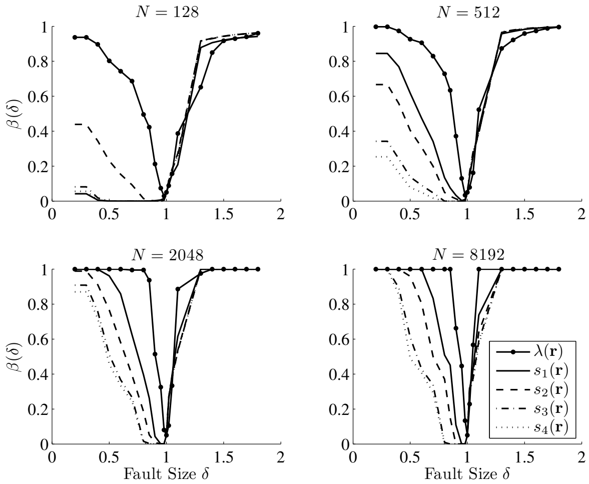

Fig. 6. Comparison of power functions for the tests based on $\lambda(\mathbf{r})$ (solid with dot markers), $s_1(\mathbf{r})$ (solid), $s_2(\mathbf{r})$ (dashed), $s_3(\mathbf{r})$ (dash-dotted), and $s_4(\mathbf{r})$ (dotted), for different window sizes N.

The mean time to detection (MTD) for the tests based on $\lambda(\mathbf{r})$, $s_1(\mathbf{r})$, $s_2(\mathbf{r})$, $s_3(\mathbf{r})$, and $s_4(\mathbf{r})$, for different values of N and fixed false detection probability, is shown in Figure 7. The MTD was computed as the mean of the detection time for all considered fault sizes $\delta \in [0.2, 0.8]$, where each fault of fixed size was injected in the test sequence at 10 time instances. It can be seen that the MTD increases with N for $\lambda(\mathbf{r})$ but tend to decrease with N for $s_1(\mathbf{r})$, $s_2(\mathbf{r})$, $s_3(\mathbf{r})$, and $s_4(\mathbf{r})$. It is also worth noting that for smaller values of N, the MTD for $\lambda(\mathbf{r})$ is shorter than for $s_1(\mathbf{r})$, $s_2(\mathbf{r})$, $s_3(\mathbf{r})$, and for larger N the MTD for all test statistics are almost of the same magnitude.

When the MTD shown in Figure 7 was computed, missed detections were not taken into account. Although the number of missed detections can be deduced from the power functions in Figures 4 and 6, this metric is explicitly shown in Figure 8, in the same format and as a contrast to the results in Figures 4, 6, and 7. In Figure 8 it can be seen that for all test statistics, the number of missed detection decreases with N and also that $\lambda(\mathbf{r})$ has less missed detections than $s_1(\mathbf{r})$, $s_2(\mathbf{r})$, $s_3(\mathbf{r})$, and $s_4(\mathbf{r})$, independent of N.

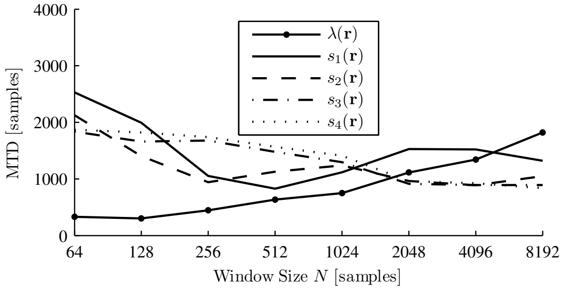

Fig. 7. Comparison of the mean time to detection (MTD) for test statistics $\lambda(\mathbf{r})$ (solid with dot markers), $s_1(\mathbf{r})$ (solid), $s_2(\mathbf{r})$ (dashed), $s_3(\mathbf{r})$ (dash-dotted), and $s_4(\mathbf{r})$ (dotted), for different window sizes N.

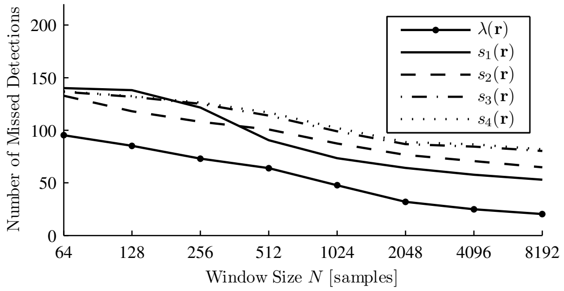

Fig. 8. Comparison of the number of missed detection for test statistics $\lambda(\mathbf{r})$ (solid with dot markers), $s_1(\mathbf{r})$ (solid), $s_2(\mathbf{r})$ (dashed), $s_3(\mathbf{r})$ (dash-dotted), and $s_4(\mathbf{r})$ (dotted), for different window sizes N.

VII. CONCLUSIONS

As clearly illustrated in Figure 1, residuals in practice often deviate from zero even in the no-fault case due to uncertainties such as modeling errors and measurement noise. Furthermore, since uncertainties are time-varying the probability distributions of residuals exhibit non-stationary features, and the difference between residuals in the no-fault and fault cases are time-varying.

To handle these issues, this paper has proposed a novel approach to residual evaluation based on an explicit comparison of the residual distribution estimated on-line using current data, and the distribution of the no-fault residual, estimated off-line using no-fault training data. In the approach, the non-stationarity is handled through a continuous adaption of the no-fault distribution to current residual data, which is done by characterizing the distribution as a mixture of the a priori known no-fault distributions. With the GLR test statistic as starting point, a more powerful and computational efficient test statistic has been derived by considering a properly chosen approximation to one of the emerging likelihood maximization problems.

The proposed test statistic has been evaluated with measurement data on a residual for diagnosis of the gas-flow system of a Scania truck diesel engine. The results are promising, and the proposed test statistic performs well despite non-conventional properties of the considered residual. For example, small faults can be reliable detected in cases where regular methods based on constant thresholding fail.

Future work includes an analysis of how the properties of the given no-fault distributions influence the performance of the method and how these distributions should be chosen and obtained, in order to maximize performance. A more thorough, theoretical, study of how different parameters influence the detection performance and an investigation regarding the implementation issues and computational complexity of the proposed approach is also desirable, as well as a more exhaustive comparison of the proposed approach with other state-of-the-art methods.

REFERENCES

[1] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*, 2nd ed. Springer, 2006.

[2] J. Chen and R. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Boston: MA: Kluwer, 1999.

[3] A. Ingimundarson, A. Stefanopoulou, and D. McKay, "Model-based detection of hydrogen leaks in a fuel cell stack," *Control Systems Technology, IEEE Transactions on*, vol. 16, no. 5, pp. 1004 –1012, sept. 2008.

[4] X. Zhang, M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems," *Automatic Control, IEEE Transactions on*, vol. 47, no. 4, pp. 576 –593, Apr. 2002.

[5] A. Emami-Naeini, M. Akhter, and S. Rock, "Effect of model uncertainty on failure detection: the threshold selector," *Automatic Control, IEEE Transactions on*, vol. 33, no. 12, pp. 1106 –1115, Dec. 1988.

[6] H. Sneider and P. Frank, "Observer-based supervision and fault detection in robots using nonlinear and fuzzy logic residual evaluation," *Control Systems Technology, IEEE Transactions on*, vol. 4, no. 3, pp. 274 –282, May 1996.

[7] P. Frank, "Residual evaluation for fault diagnosis based on adaptive fuzzy thresholds," in *Qualitative and Quantitative Modelling Methods for Fault Diagnosis, IEE Colloquium on*, Apr. 1995, pp. 4/1 –411.

[8] G. Casella and R. Berger, *Statistical Inference*, 2nd ed. Duxbury Press, 2001.

[9] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes - Theory and Application*. Prentice-Hall, 1993.

[10] H. Kuhn and A. Tucker, "Nonlinear programming," in *Proceedings of 2nd Berkeley Symposium*. Berkeley: University of California Press, 1951, pp. 481–492.

[11] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[12] K. Haskell and R. Hanson, "An algorithm for linear least squares problems with equality and nonnegativity constraints," *Mathematical Programming*, vol. 21, no. 1, pp. 98–118, 1981.

[13] C. Lawson and R. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[14] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *Journal of Chemometrics*, vol. 11, no. 5, 1997.

[15] L. Davies, U. Gather, D. Nordman, and H. Weinert, "A comparison of automatic histogram constructions," *ESAIM: Probability and Statistics*, vol. 13, pp. 181–196, 2009.

[16] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2000.

[17] J. Wahlström and L. Eriksson, "Modeling diesel engines with a variable-geometry turbocharger and exhaust gas recirculation by optimization of model parameters for capturing non-linear system dynamics," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 225, no. 7, July 2011.

[18] C. Svärd and M. Nyberg, "Residual generators for fault diagnosis using computation sequences with mixed causality applied to automotive systems," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 40, no. 6, pp. 1310–1328, 2010.