

Model Based Diagnosis for the Air Intake System of the SI-Engine

Mattias Nyberg and Lars Nielsen

Vehicular Systems, ISY, Linköping University
S-581 83 Linköping, Sweden
email: matny, lars@isy.liu.se
phone: +46 13 282369
fax: +46 13 282035

ABSTRACT

Because of legislative regulations like OBDII, on-board diagnosis has gained much interest lately. A model based approach is suggested for the diagnosis of the air intake system of an SI-engine. Important research issues are modeling concepts, residual generation and evaluation, overall performance, and limiting factors. The diagnosis system is based on a non-linear semi-physical model and uses a combination of different residual generation methods. It is capable of detecting and isolating faults in the throttle actuator, throttle sensor, air mass flow sensor and manifold pressure sensor. The scheme is experimentally validated on a real production engine.

1. INTRODUCTION

On-board diagnosis of car engines has become increasingly important because of environmentally based legislative regulations such as OBDII (On-Board Diagnostics) [1]. Other reasons for incorporating diagnosis in vehicles are reparability, availability and vehicle protection. Today, up to 50% of the engine management systems are dedicated to diagnosis. The techniques used in production vehicles [2][3] are mainly based on limit checking and active diagnosis. Also basic model based diagnosis is used to some extent. Active diagnosis means that during special operating conditions, the engine is manipulated in such a way that possible faults will be revealed. It is for example used heavily during idle. These techniques may be insufficient to fulfill the upcoming more restrictive regulations. Also the lack of good diagnosis schemes, have forced engineers to chose suboptimal engine control solutions to accomodate the diagnostic requirements. It is therefore desirable to find new diagnosis techniques which perform better and do not rely on special operating conditions and active diagnosis.

One way to increase the performance of the engine diagnosis system may be to increase the use of model based diagnosis in which more process knowledge is uti-

lized in the form of a mathematical process model. The diagnosis can to a larger extent be performed passively and over a wider operating range if a proper model is available. However the problem is challenging because of the many conflicting requirements on such a diagnosis system:

- Computing power is limited due to cost constraints.
- Good performance is required by the legislative regulation; erroneous diagnosis may lead to fines for the car manufacturer [1].
- The diagnosis system must be able to handle vehicle to vehicle variations and aging. Also the vehicle operates in an uncontrolled environment which makes the process different from day to day, for example because of changes in atmospheric conditions.

Model based diagnosis has received much attention during the last decade [4][5][6][7]. The studies have been mainly general in character, and few concerned with automotive engine diagnosis.

The SI-engine in its basic configuration can be divided into three subsystems, each with its special diagnostic requirements:

- air intake system: air mass flow meter, throttle, manifold pressure sensor, engine speed sensor
- fuel/combustion system: fuel injector, spark plug, misfire
- exhaust after-treatment system: lambda sensors, catalyst

Common are also evaporation systems and EGR. All these components need to be diagnosed due to OBDII. However it is not sure that the same diagnosis technique is suitable for all systems or components. The combustion chamber is difficult to model and for example a promising ion-current based algorithm has been used to diagnose misfire [8]. The only known solution for catalyst monitoring so far, is to use a post catalyst

lambda sensor in addition to the pre catalyst lambda sensor which is used primarily for closed loop air-fuel control. Therefore such a system has two lambda sensors and this hardware redundancy is suitable to utilize for diagnosis.

DIAGNOSIS OF THE AIR INTAKE SYSTEM

The air intake system contains a number of components and sensors related by reasonably well understood thermodynamics and physics of gases. However, a detailed physical model would be very complicated. For this subsystem of the SI-engine, it is still appealing to investigate the possible use of model based diagnosis. This is because of several reasons. One is that a semi-physical model can be found with reasonable effort. Another reason is that many engine variables need not to be considered because of subsystem decoupling. Such variables are for example road load, lubrication, friction, wear, ignition timing, fuel quality, lambda, cooling water and oil temperatures. These variables are present in the other subsystems. Variables that however may affect the air intake system are ambient air temperature, pressure, and humidity. Finally the air intake system do not contain any large disturbances that must be taken into account when designing the diagnosis system.

The undertaking is thus to see if a diagnosis system based on a semi-physical model can be developed, and to investigate overall performance and limiting factors. A review of model based diagnosis is presented in Section 2, and the application to automotive engines is reviewed in Section 3. A semi-physical model based diagnosis system is developed. In Section 4 and 5, a model of the air intake system of a production SAAB 2.3 liter engine is built. From this model, a diagnosis system capable of diagnosing faults in throttle actuator, throttle sensor, air mass flow sensor, and manifold pressure sensor, is constructed in Section 6 and validated in Section 7.

2. MODEL BASED DIAGNOSIS

Model based diagnosis is based on *analytical redundancy*. A process contains analytical redundancy if an input or output can be calculated by using only other inputs or outputs. In the simplest case, the analytical redundancy is utilized by comparing the outputs from the real process and outputs from a process model, which is fed by the same inputs as the real process. By the term *diagnosis*, both fault detection and *isolation* are considered. Isolation means to determine the location of the fault. The primary benefits of model based diagnosis compared to limit- and trend-checking are that it has large potential to:

- be performed passively
- work over a wide operating range

- have better performance; smaller faults can be detected, faults are isolated
- for the same level of diagnosis performance, be less computationally intensive

An important property of a diagnosis system is robustness, i.e. to withstand model uncertainties, noise, and disturbances and still be able to detect and isolate small faults. This implies that we may need a good, possibly complex, model, complex design procedures, and complex algorithms. The goal of the design of a diagnosis system can therefore be summarized as to maximize robustness while keeping reasonable complexity.

It has been reported [9] that the main effort in constructing a model based diagnosis system is to develop a sufficiently good model. However this model can be used also for other purposes, e.g. control algorithms.

The main structure of a diagnosis system is shown in Figure 1. The process is affected by three kinds of faults: *actuator faults*, *component faults*, and *sensor faults*. It is usually assumed that only one fault can occur at the same time. Inputs to the diagnosis system are the process inputs and outputs. The diagnosis is performed in two stages: *residual generation* and *residual evaluation*.

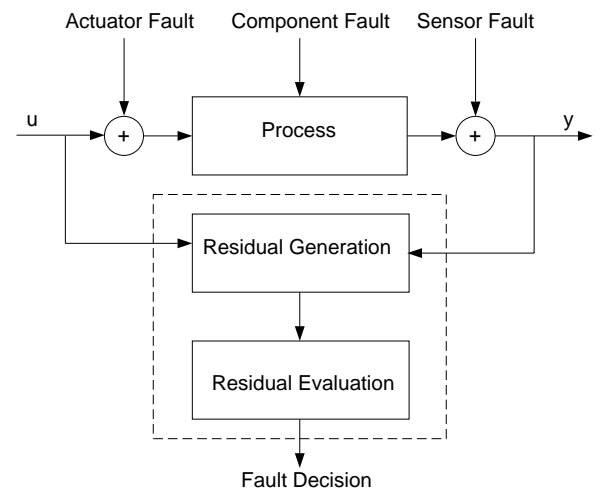


Fig. 1. The process with its faults and the diagnosis system. The actuator and sensor faults are here modeled as additive faults.

A *residual* is a signal that reacts on inconsistencies between the model and the real process. In the case of no fault the residual should be close to zero and in the case of a fault the residual should be significantly non zero if it is sensitive to that particular fault. A number of residuals are used and they are made sensitive to different faults in such a way that isolation can be achieved. This is called *structured residuals*. An important property of the residual structure is that it should be *strongly isolating*. A strongly isolating structure protects small faults from being incorrectly isolated in the

case when some residuals are not excited enough to reach the thresholds. Some processes have a considerable disturbance that can not be ignored in the model. In these cases the disturbance must be *decoupled* so that it does not affect the residuals. There is often a compromise between decoupling and isolation so the diagnosis performance may be degraded.

The second stage is the residual evaluation. The purpose is to determine if the residuals are zero or not. A common and simple approach is to filter the residuals and then test them against thresholds. This stage can also contain means to improve the robustness, e.g. statistical detection, threshold crossing counting or fuzzy thresholds [10].

RESIDUAL GENERATION Many schemes for designing residual generators has been proposed, e.g. parity equations [11], dedicated observer schemes [12], detection filter [13], unknown input observer [14], frequency domain approaches [15], Kalman filters [16]. Most of these schemes are primarily for linear systems but parity equation and some non-linear observer schemes have been used with non-linear systems.

The most intuitive residual generator that can be formed is when static relationships exist between inputs and outputs. This is called *direct redundancy* and an example of this is

$$r_1(t) = y_1(t) - \hat{y}_1(t) = y_1(t) - h(u(t), y_2(t)) \quad (1)$$

where $r_1(t)$ is the residual, $u(t)$ the input, $y_1(t)$ and $y_2(t)$ outputs, and h a function that defines the static relationship $y_1(t) = h(u(t), y_2(t))$.

When dynamics is present, *temporal redundancy* can be utilized. This means that the residual is also a function of earlier inputs or outputs, or contains states. An example of this is, if we have an observer that estimates the output $y_2(t)$, then a residual can be formed as

$$\begin{aligned} x(t+1) &= g(x(t), u(t)) + K(y - \hat{y}) \\ r_2(t) &= y_2(t) - \hat{y}_2(t) \end{aligned} \quad (2)$$

An observer used for diagnosis is called *diagnostic observer* and does not necessarily estimate any state.

A common principle for residual generation is *parity equations* [11]. They are in the discrete domain defined as consistency relations of the inputs and outputs of the process. Inputs and outputs from different times can be used. For the linear case, consider the system

$$y(t) = \frac{B(q)}{A(q)}u(t) + f(t) \quad (3)$$

where u is the input, y the output and f the fault. This can be rewritten as

$$A(q)y(t) = B(q)u(t) + A(q)f(t) \quad (4)$$

This is an example of a linear parity equation and a residual can be formed as

$$r_3(t) = A(q)y(t) - B(q)u(t)$$

This is the *computational form*, showing how the residual is computed. The *internal form* shows how the residual is affected by a fault and is

$$r_3(t) = A(q)f(t)$$

The parity equation (4) is just one of several linearly independent parity equations that can be derived from the system description (3). By using different parity equations for different residuals they can be made insensitive to different faults. Depending of what time history a parity equation contain, it can use the principle of direct redundancy or temporal redundancy but is always a non recursive filter.

3. EARLIER WORKS ON MODEL BASED DIAGNOSIS FOR SI-ENGINES

In the late 80's, the first results on trying to solve automotive diagnosis problems by means of model based diagnosis, were reported. The reason for this increased interest was probably the legislative regulations. This section reviews some results on model based solutions to automotive engine diagnosis. The most reports about this subject have been published by the groups around Rizzoni and Gertler. Other reports, dealing with model based diagnosis applied to automotive engines and not treated here, are [19][20][21][22][23].

PAOLELLA AND CHO 1990 [24], they develop a diagnosis scheme based on non-linear extended Kalman filters. However they first tried a linear version that did not provide satisfactory performance. They apply the non-linear scheme to an automotive powertrain to diagnose faults in engine speed, transmission, and wheel speed sensors, which are also the inputs to the diagnosis system. The scheme is validated in simulations.

In 1991 [25], they test their approach in a real vehicle and they are able to diagnose 100% faults, i.e. the case when sensor outputs are set to zero.

RIZZONI ET AL. Partly supported by Ford, 1989-1991 Rizzoni [26][27] uses an extended version of the detection filter derived from a fourth order linear state space engine model. The diagnosis system measures throttle angle, manifold pressure and engine speed to produce two residuals. The scheme is shown to be able to diagnose 10% faults in throttle angle sensor and manifold pressure sensor on a real Ford 3.0 liter engine. The use of a linear model restricts the operating range of the model, and data is shown for 56-60 kPa and 1050-1130 rpm. It seems like the residuals are sensitive to engine transients.

In 1993 [28] he uses a non-linear model to generate five residuals based on parity equations. Throttle angle, engine speed, manifold pressure and injected fuel is measured to diagnose faults on a real 4 cylinder 1.3 liter engine. The diagnosed components are throttle angle sensor, engine speed sensor, manifold pressure sensor and fuel injectors. The load is decoupled. Plots of the residuals are shown for the case of 10% faults of the diagnosed components.

1994 [18][29][30] they start using a non-linear discrete NARMAX model which is a linear combination of second order polynomials, i.e. it is linear in the parameters. Inputs to the diagnosis system are demanded throttle angle, injected fuel and measured air mass flow and engine speed. Forward and inverse models are used to generate four corresponding parity equation residuals. The load is decoupled. The scheme is tested on a real Ford 3.0 liter engine over a standardized test schedule and is reported to be able to detect 10% faults in the air mass flow sensor, 20% faults in the engine speed sensor, 15% faults in the throttle actuator and 40% faults in the fuel injector.

GERTLER ET AL. Gertler and his group has been involved in project with GM. During 1991-1993 [17][31] a simulation study was done. After concluding that a linear model is not sufficient because of its limited operating range, they use a hybrid model with linear core. Five residuals in the form of parity equations are used to diagnose faults in throttle angle sensor, EGR-valve, fuel injectors, manifold pressure sensor, engine speed sensor and lambda sensor. These are also the inputs to the diagnosis system. The residual structure is able to distinguish all faults except fuel injector and lambda sensor faults. A potential problem is that they rely on that the lambda sensor characteristic is known.

1994-1995 [9][32] they try this scheme on a real 3.1 liter V6 engine in a production vehicle. Both off-line and on-line versions are developed and they stress the importance of simple algorithms because of limited on-board computing power. Instead of using five linear parity equations, they now use six non-linear parity equations as residuals. The same faults as before are diagnosed and the new residual structure provides isolation between all faults. To increase robustness, the residuals are low-pass filtered and threshold crossings are counted. They report that they are able to diagnose faults of 10% size.

4. THE AIR INTAKE SYSTEM MODEL

The SI-engine is a non-linear plant and it has been indicated in a pre-study that diagnosis based on a linear model is not sufficient for diagnosing an engine. This has also been concluded by other authors [17][18]. This motivates the choice of a non-linear model in this work.

For the purpose of diagnosis, a simple and accurate model is desirable. In the air system application there is no need for extremely fast fault detection, therefore a so called *mean value model* [33] is chosen. This means that no within cycle variations are covered by the model. The engine is a 2.3 liter 4 cylinder SAAB production engine mounted in a test bench together with a Schenk AC dynamometer. The measured variables are the same as the ones used for engine control. A schematic picture of the whole engine is shown in Figure 2. The engine has electronic throttle control (drive-by-wire), which is basically a DC-servo controlled by a PID controller. The part that is considered to be the air intake system is everything to the left of the dashed line. Also the engine speed must be taken into account because it affects the amount of air that is drawn into the engine.

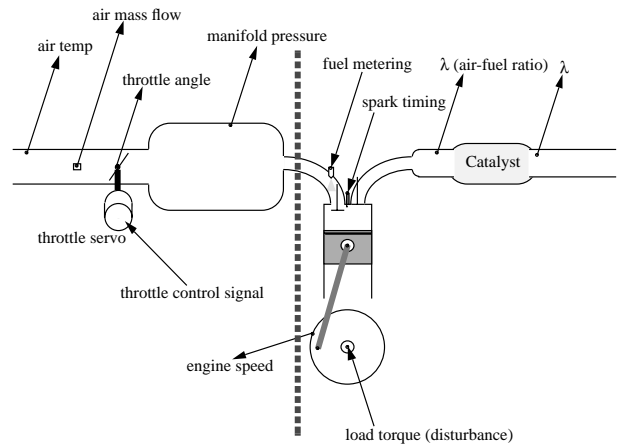


Fig. 2. The basic SI-engine.

The model of the air intake system is continuous and consists of two parts, the throttle model and the air dynamics. The throttle dynamics is modeled as a second order linear system in which the states are the throttle angular speed and the throttle angle. The air dynamics is derived from the ideal gas law and has one state which is the manifold pressure. In Figure 3, the model of air intake system together with the throttle controller is shown. The process inputs are the throttle control signal u , and the engine speed n . The outputs are throttle angle sensor α_s , mass air flow sensor $\dot{m}_{air,s}$, and manifold pressure sensor $p_{man,s}$. The faults are modeled as additive faults. The equations describing the fault free model can be written as

$$\dot{\omega} = a\omega + b(u(t) - \tau_{air}) \quad (5)$$

$$\dot{\alpha} = \omega \quad (6)$$

$$\tau_{air} = h(p_{man}, \dot{m}_{air}, \alpha) \quad (7)$$

$$\dot{p}_{man} = \frac{RT_{man}}{V_{man}}(\dot{m}_{air} - \dot{m}_{ac}) \quad (8)$$

$$\dot{m}_{air} = f(p_{man}, \alpha) \quad (9)$$

$$\dot{m}_{ac} = g(p_{man}, n) \quad (10)$$

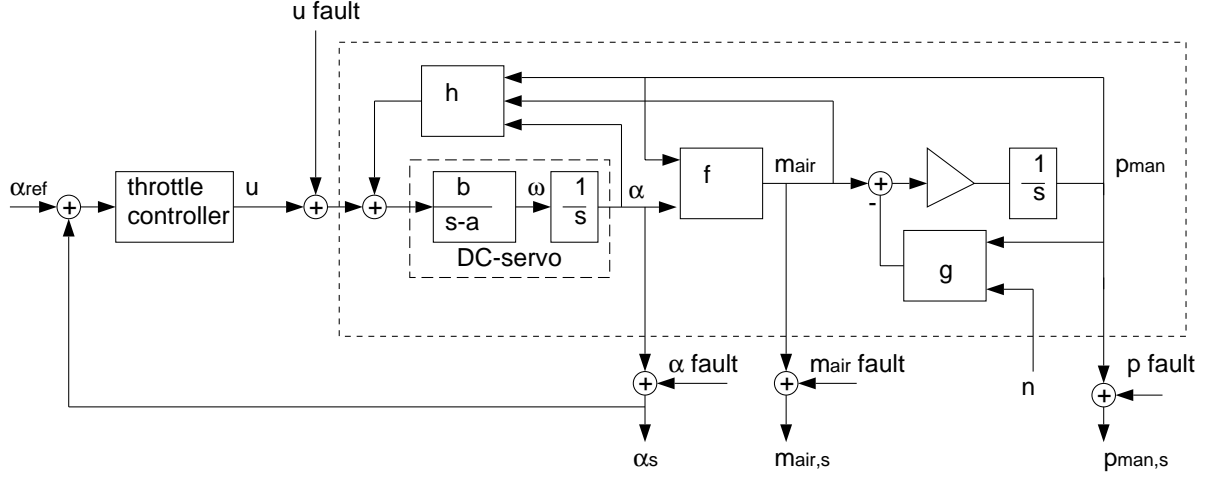


Fig. 3: The model of the air intake system and the throttle controller.

The variables and its units are summarized and ex-

u [V]	the output from the throttle controller
p_{man} [kPa]	manifold pressure
R [J/(g K)]	the gas constant
T [K]	manifold air temperature which is assumed to be equal to the ambient temperature
V [m ³]	manifold volume
\dot{m}_{air} [kg/s]	air mass flow into the manifold and is equal to the air flow past the air mass flow meter
\dot{m}_{ac} [kg/s]	air mass flow out from the manifold
f	static function describing the flow past the throttle
g	static function describing the flow into the cylinders
α [deg]	throttle angle
τ_{air}	normalized torque on the throttle plate generated by the air flow past the throttle
h	static function describing the torque on the throttle plate generated by the air flow past the throttle
n [rpm]	engine speed

Table 1 Symbols and units.

plained in Table 1. The model consists of a physical part, the equations (5), (6) and (8), and a black box part, the functions (7), (9) and (10). Even if variations in ambient pressure and temperature do affect the system, they are here assumed to be constant.

5. MODEL IDENTIFICATION

To find the static functions f , g and h , a steady state experiment was performed. 12 equally spaced engine speeds (1500 to 3000 rpm) and 8 equally spaced manifold pressures (35 to 60 kPa) was used to build a map of u , p_{man} , \dot{m}_{air} , α and n measurements in 96 different steady state operating points. To represent the static functions, there is a choice between interpolating in the map or fit for example polynomials to the map. Here interpolation was chosen for g and polynomials for f and h . The choice of terms to be included in the polynomials was guided by studying the correlation coefficient and the final choice was based on a validation against another data set of 96 operating points. The function g is closely related to volumetric efficiency. The resulting polynomials are

$$\begin{aligned}
 f(p_{man}, \alpha) &= \eta_0 + \eta_1 \alpha + \eta_2 \alpha^2 + \eta_3 \alpha^3 + \\
 &\quad + \eta_4 \alpha p_{man} + \eta_5 p_{man} \\
 h(p_{man}, \dot{m}_{air}, \alpha) &= \nu_0 + \sqrt{p_{man}} + \nu_1 p_{man} + \nu_2 p_{man}^2 + \\
 &\quad + \nu_3 p_{man} \sqrt{\alpha} + \nu_4 p_{man}^2 \alpha + \\
 &\quad + \nu_5 p_{man} \alpha + \nu_6 p_{man} \sqrt{\dot{m}_{air}}
 \end{aligned}$$

For the air dynamics, the only remaining constant to identify is V_{man} . To identify this constant, a dynamic test was performed. A test-cycle consisting of 4 throttle steps in different operating points was constructed and the already identified functions, f and g , were used to find the value of V_{man} that gave the best fit between measured pressure and estimated pressure. A comparison between measured and simulated pressure for the final air dynamic model is shown in Figure 4.

The parameters, a and b , used in the throttle model, were identified by applying a pseudorandom binary signal to the throttle input and measuring the throttle angle. This was done when the engine was not running so the term τ_{air} in (5) was zero. The throttle model is not

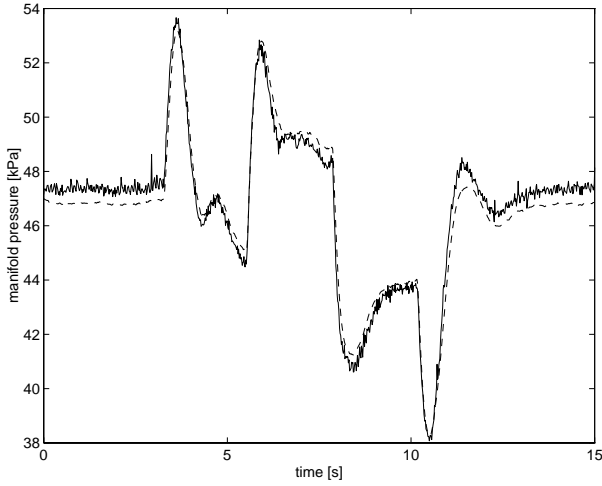


Fig. 4. Measured (solid line) and simulated (dashed line) pressure.

as easily validated as the air dynamics, because it is a marginally stable system. However, closed loop experiments showed that this model is less accurate than the air dynamics model. This is also reflected in diagnosis performance as can be seen in Section 7.

6. THE DIAGNOSIS SYSTEM

For the design of the diagnosis system a number of approaches, listed in Section 2, exists. For non-linear systems, the number is reduced to just a few. Direct redundancy and non-linear diagnostic observers, including the *dedicated observer* (explained below), are used in this work because they are naturally derived from the non-linear model and constrain the structure of the model little compared to other approaches. For example, the static functions do not need analytical expressions, they can be represented by maps, which is the case for the function g in this work. In the popular parity equation approach, the residual generator is constrained to be a non recursive filter, which can be unnecessary complicated to derive from the recursive model.

The inputs to the diagnosis system are $\dot{m}_{air,s}$, u , α_s , $p_{man,s}$, and n . The diagnosed components are the throttle actuator, throttle angle sensor, air mass flow meter, and manifold pressure sensor. This means that one actuator fault and three sensor faults are considered. It is assumed that only one fault can occur at the same time. The engine speed sensor is not diagnosed because it is diagnosed sufficiently well by existing diagnosing techniques. This is because the speed is measured by counting pulses from a magnetic sensor on a toothed wheel. A magnetic sensor failure is detected from the devastating effects on the pulse train. However the residual structure (see Table 2) of the diagnosis system has the property that also the engine speed sensor could be in-

cluded among the diagnosed components.

RESIDUAL GENERATION By using for example static relationships, i.e. direct redundancy, and diagnostic observers estimating measurable signals, it is possible to construct a large number of residuals. However many of them are based on the same part of the model that makes them redundant. It is thus desirable to find a small, non redundant set of residuals with good isolation properties, without losing robustness. Starting with 18 such residuals only six were selected to be used in the residual generator. The computational form of these six residuals are described next.

The first residual is a direct redundancy residual:

$$r_1 = \dot{m}_{air,s} - f(p_{man,s}, \alpha_s)$$

This residual relies on a static relationship in the model and checks the consistency of Equation (9).

All other residuals are derived using temporal redundancy. The second residual checks the consistency of Equations (8) and (10):

$$r_2 = \dot{m}_{air,s} - g(p_{man,s}, n) - \frac{V_{man}}{RT_{man}} \hat{p}_{man}$$

It is assumed that the derivative of p_{man} can be estimated with sufficient accuracy. If the derivative is computed as a difference, r_2 is a parity equation.

Next are two dedicated observer residuals. The point of dedicated observer residuals is to make the residuals sensitive to only one sensor fault by measuring only one output (α_s is not considered to be an output in this case). This can be seen in r_3 , which measures only p_{man} , and r_4 , which measures only \dot{m}_{air} . The residuals are formed as

$$\hat{p}_{man,3} = \frac{RT_{man}}{V_{man}} (f(\hat{p}_{man,3}, \alpha_s) - g(\hat{p}_{man,3}, n) + K_1(p_{man,s} - \hat{p}_{man,3}))$$

$$r_3 = p_{man,s} - \hat{p}_{man,3}$$

$$\hat{p}_{man,4} = \frac{RT_{man}}{V_{man}} (f(\hat{p}_{man,4}, \alpha_s) - g(\hat{p}_{man,4}, n) + K_2(\dot{m}_{air,s} - \hat{m}_{air,4}))$$

$$\hat{m}_{air,4} = f(\hat{p}_{man,4}, \alpha_s)$$

$$r_4 = \dot{m}_{air,s} - \hat{m}_{air,4}$$

In r_3 , p_{man} is estimated by means of a non-linear diagnostic observer. The pressure p_{man} is measured and the estimation error fed back into the observer. The residual r_3 equals the estimation error. The fourth residual, r_4 , is constructed similarly but the estimation error of \dot{m}_{air} , that equals r_4 , is fed back into the observer.

The fifth residual, r_5 , is also based on an observer but in contrast to r_3 and r_4 , the estimation error does not equal r_5 . This means that to compute r_5 , both p_{man}

and \dot{m}_{air} are measured so r_5 is not a dedicated observer residual and therefore becomes sensitive to both manifold pressure sensor and air mass flow sensor faults. The residual is formed as

$$\begin{aligned}\dot{\hat{p}}_{man,5} &= \frac{RT_{man}}{V_{man}} (f(\hat{p}_{man,5}, \alpha_s) - g(\hat{p}_{man,5}, n) + \\ &\quad K(\dot{m}_{air,s} - \hat{m}_{air,5})) \\ \hat{m}_{air,5} &= f(\hat{p}_{man,5}, \alpha_s) \\ r_5 &= p_{man,s} - \hat{p}_{man,5}\end{aligned}$$

The remaining residual is observer based as well, and its purpose strictly for diagnosis of the throttle actuator:

$$\begin{aligned}\dot{\hat{\omega}} &= a\hat{\omega} + b(u(t) - h(p_{man,s}, \dot{m}_{air,s}, \alpha_s)) + k_1(\alpha_s - \hat{\alpha}) \\ \dot{\hat{\alpha}} &= \hat{\omega} + k_2(\alpha_s - \hat{\alpha}) \\ r_6 &= \alpha_s - \hat{\alpha}\end{aligned}$$

The throttle subsystem has only one output and therefore, it is not relevant to discuss if the residual is a dedicated observer or not.

The choice of the observer gains k_1 , k_2 and also K_1 and K_2 were done by pole placement. The observer poles are functioning like a low-pass filter so the pole placement is a compromise between fast fault response and sensitivity to disturbances and noise.

The six residuals form a set of structured residuals. Different residuals are sensitive to different faults. This can be seen by studying the equations of the residuals and is summarized in Table 2. Some residuals should be sensitive to specific faults due to the model equations, but it turned out in the experiments that the effect of a fault on some residuals was not significant enough to guarantee a robust diagnosis system. These cases are marked with an X (don't care) in Table 2. In one case, marked X_1 , this has a straightforward physical explanation, the air flow past the throttle is not dependent on manifold pressure for supersonic air speeds [34]. This is the case for manifold pressures below 50 kPa. It can be seen that this residual structure is strongly isolating as long as only α_s , $\dot{m}_{air,s}$, and $p_{man,s}$ -faults are considered. If also u -faults are considered, it is not strongly isolating. The throttle model has a larger model uncertainty, compared to the other parts of the model, so that if any of the other residuals were fed by u , then it would degrade the performance of these residuals. That is the reason for the many zeros in the u column. In practise this can have the effect that a small α_s , $\dot{m}_{air,s}$ or $p_{man,s}$ -fault is interpreted as a u -fault.

RESIDUAL EVALUATION The six residuals should respond due to Table 2 if a fault occurs and take the value zero in the fault free case. The purpose of the residual evaluation is to check this and generate a fault decision. Taking disturbances into account this is not trivial. The X:s in Table 2 are important. For a specific

	u	α_s	$\dot{m}_{air,s}$	$p_{man,s}$	n
r_1	0	1	1	X_1	0
r_2	0	0	1	1	1
r_3	0	1	0	1	1
r_4	0	1	1	0	1
r_5	0	1	1	X	1
r_6	1	X	X	X	0

Table 2 The residual structure of the diagnosis system. Each column represents a faulty signal.

fault, residuals marked with an X have an uncertain response. This means that the fault decision must not depend on these residuals.

The residual evaluation scheme chosen in this work is shown in Figure 5. First, the residuals are low-pass filtered with 0.8 Hz cut-off frequency and then normalized with their standard deviations in the fault free case. The last step is fuzzy thresholding [10]. The fuzzy thresholding is chosen before regular thresholding to make the output from the diagnosis system contain more information. With fuzzy thresholding the output is not just a fault or not a fault, but a relative degree of how probable it is that a fault has occurred.

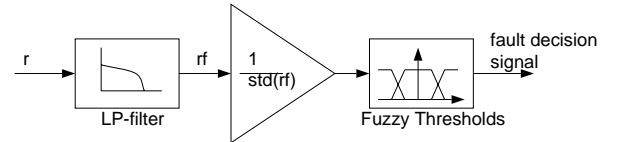


Fig. 5. Residual evaluation.

Three fuzzy sets are used and they are shown in Figure 6. The way they are chosen, only the constant γ has to be determined. The criterion used to determine γ , is that in the fault free case, there should be no false alarm. Also in the case of a fault, all residuals that should respond due to Table 2, must reach the fuzzy *high* value and residuals that should not respond must be fuzzy *low*. A value of 4 proved experimentally to be good choice. Table 2 is transferred to fuzzy rules and an example of this is the $p_{man,s}$ -column that corresponds to the fuzzy rule

IF r_2 is *high* AND r_3 is *high* AND r_4 is *low*
THEN *pressure sensor fault*

The fault decision is then computed by executing all the rules, one for each fault. The connective AND is implemented as multiplication. Neither fuzzy implication nor defuzzification is needed because the result of each fuzzy rule is a crisp value, not a membership function. Further, note that no aggregation between the rules is needed because each rule is a fuzzy system on its own and therefore executed independently of the other fuzzy

rules. The fault decision output of the diagnosis system is then four fault decision signals, one for each fault, representing how probable it is that the particular fault has occurred.

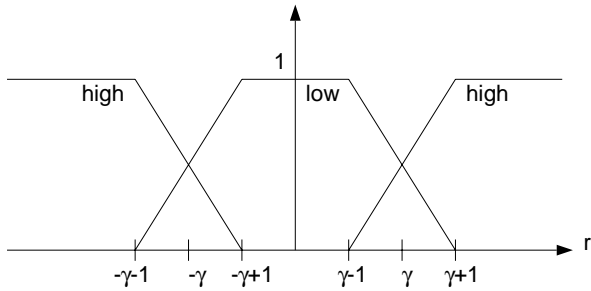


Fig. 6. The fuzzy sets used in the fuzzy thresholding.

IMPLEMENTATION The diagnosis system was implemented in Matlab Simulink. However all measurements were made on the real engine and sampled at 240 Hz with 40 Hz first order anti-alias filters. The sampling frequency 240 Hz is probably unnecessary high compared to what is needed to satisfy OBDII. The diagnosis was then performed off-line.

An important issue is how much computing power that is needed to compute the residuals. By studying the equations of the residuals and assuming that the static functions are implemented by maps, it can be concluded that there is about one multiplication, some table look-ups, and additions in each residual. This should not be a problem in an on-board implementation in production vehicles.

7. VALIDATION

To validate the diagnosis system a short test-cycle was constructed. It was the ambition that the test-cycle should represent realistic driving and cover a large part of the range of the model. The test-cycle lasts for 60 seconds and is shown in Figure 7.

All faults are simulated by adding a pulse to the corresponding signal. The pulse starts after 10 seconds and lasts for 40 seconds. It has an amplitude of about 10% of the signal mean for all cases except for the throttle actuator fault where 20% is used. The actuator fault is larger because of the comparably larger model uncertainty in the throttle model. The throttle actuator fault is simulated by adding a pulse to the output of the throttle controller. Due to the closed loop throttle control, the throttle angle returns to its reference value shortly after the start of the pulse. Although the operation of the throttle is only marginally affected by the fault, the diagnosis is not degraded as can be seen in Figure 9. The sensor faults are simulated by adding a constant to

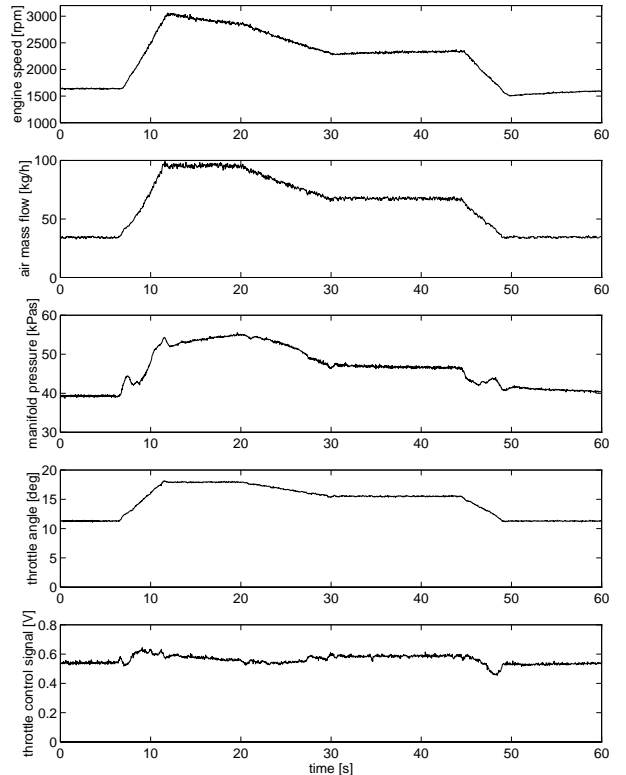


Fig. 7. The 60 seconds test-cycle.

the measured value. How the faults are simulated is also marked in Figure 3.

In Figure 8 to 12, the responses (the absolute values) of the six residuals are shown for no fault, u -fault, α_s -fault, $\dot{m}_{air,s}$ -fault, and $p_{man,s}$ -fault respectively. The residuals have been filtered and normalized so it is really the input to the fuzzy thresholding that is shown. Only the range from 0 to 8 is shown in the plots. Therefore residuals values higher than 8 is not visible. It can be seen that the residuals are noisy and non zero even in the case of no fault. The reason for this is the model error, which is the dominating factor that limits the performance of the diagnosis system.

In each plot, the two horizontal solid lines represent the value of $\gamma \pm 1$ in the membership functions (see Figure 6). The dashed line is the crossing between the membership functions. Also shown in the figures (in the 7:th plot) is the fault decision signal from the diagnosis system. Only the fault decision signal corresponding to the simulated fault is visible, because the other fault decision signals are constantly zero. The dashed line in this plot is the 0.5 level. This can be used as a criterion when to trigger the fault alarm. If this criterion is used, it can be seen that for these 10% and 20% faults the response of the diagnosis system is distinct.

All faults are detected and correctly isolated. Also there are no false alarms. Even if in several cases, one or two residuals incorrectly indicates false, there is no

erroneous diagnosis because the residual structures dictated by Table 2 do not coincide. In this way, a kind of robustness against false alarms is achieved by means of the residual structure. In some cases a residual that should react, is not very high. This can be seen in for example Figure 9 and 11 where a residual reaches the fuzzy threshold area, which results in a weaker fault decision signal.

The level of robustness achieved encourage to go on with experiments involving several engines in real vehicles to study vehicle to vehicle variations, environmental changes and aging. A result regarding this issue is from Gertler [9] that reports that atmospheric variables, as ambient temperature, pressure and humidity, affects the residuals at a maximum by 10% to 15%. Further expansions of the work include how to make the diagnosis system more robust by using adaptation and find ways to utilize known properties of the model uncertainty.

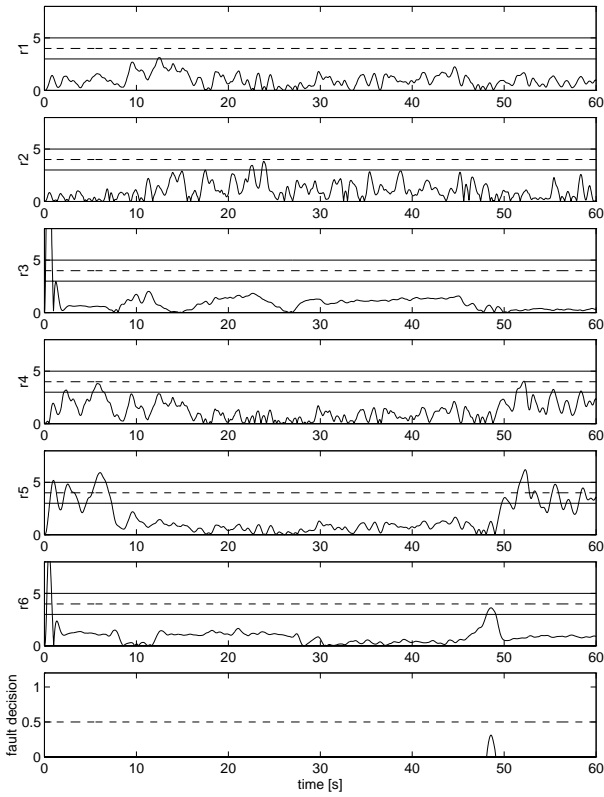


Fig. 8. The residuals for no fault.

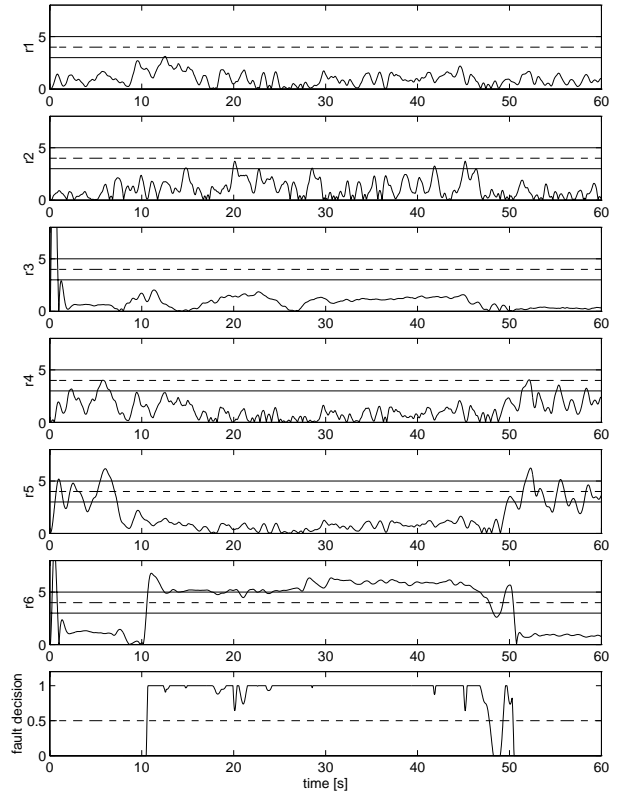


Fig. 9. The residuals for 20% throttle actuator fault.

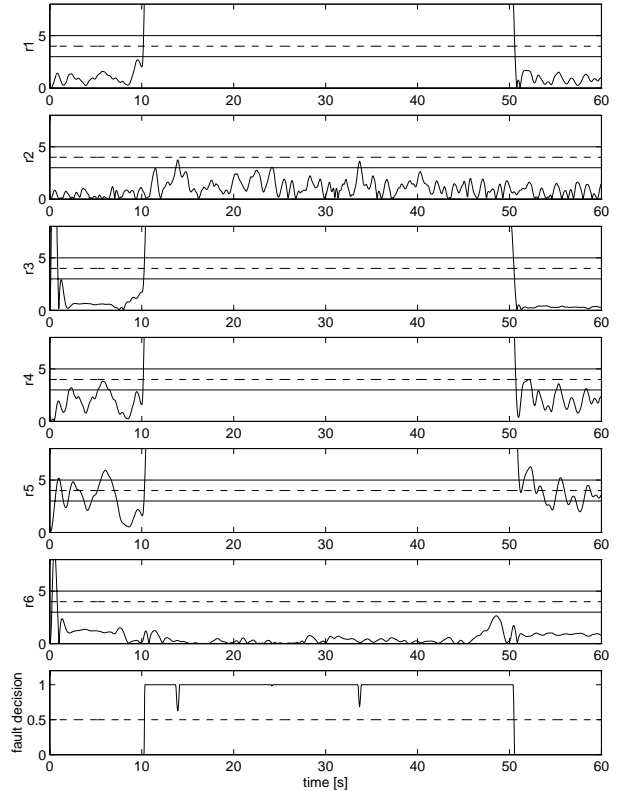


Fig. 10. The residuals for 10% throttle sensor fault.

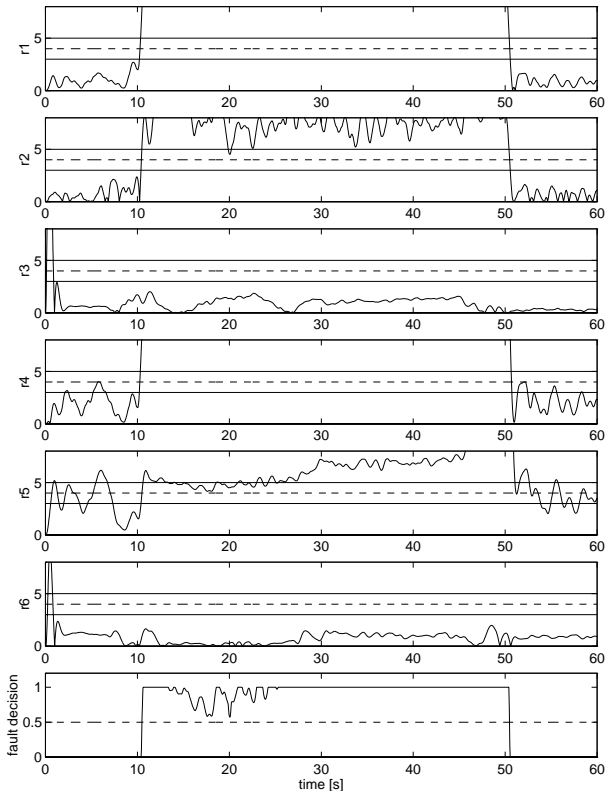


Fig. 11. The residuals for 10% air mass flow sensor fault.

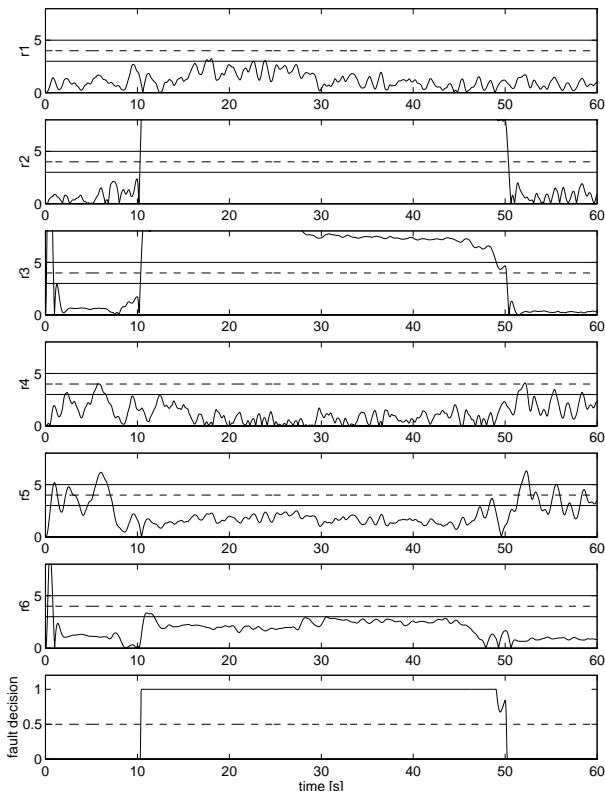


Fig. 12. The residuals for 10% manifold pressure sensor fault.

8. CONCLUSIONS

To increase the performance of the engine diagnosis system, a model based approach has been suggested for the air intake system. Important issues are modeling concepts, residual generation and evaluation, overall performance, and limiting factors. It is argued that model based diagnosis is appropriate for the air intake system which is not necessarily the case for the other parts of the SI-engine.

Much effort has been put into building a semi-physical model of the system. It is believed that a better diagnosis system is obtained by using physical insight of the process.

Instead of restricting the residual generator design to only one method, a combination of different residual generation methods have been used. This has the advantage of fewer model constraints, which make it possible to use less restrictive model building, which in turn lead to possibly better diagnosis performance.

It is noted that the response of some residuals to specific faults is uncertain and the reason is explained. A residual evaluation strategy, that makes the fault decision not dependent on these uncertain residuals, is introduced. In the residual structure this is marked X (don't care).

All experiments were performed on a single production engine in a test cell. The validation shows that the diagnosis system performs well.

9. ACKNOWLEDGMENTS

Erik Frisk and Andrej Perkovic are gratefully acknowledged for their help with experiments and insightful discussions. This research is supported by NUTEK (Swedish National Board for Industrial and Technical Development).

REFERENCES

- [1] California's OBD-II regulation (section 1968.1, title 13, california code of regulations), resolution 93-40, july 9. pages 220.7 – 220.12(h), 1993.
- [2] A. Unger and K. Smith. The OBDII system in the volvo 850 turbo. *SAE Paper*, (932665), 1993.
- [3] Ronald Jurgen. *Automotive Electronics Handbook*. McGraw-Hill, 1994.
- [4] R. Patton, P. Frank, and R. Clark, editors. *Fault diagnosis in Dynamic systems*. Systems and Control Engineering. Prentice Hall, 1989.
- [5] J. Gertler. Analytical redundancy methods in fault detection and isolation; survey and synthesis. IFAC

- Fault Detection, Supervision and Safety for Technical Processes, pages 9–21, Baden-Baden, Germany, 1991.
- [6] P.M. Frank. Advances in observer-based fault diagnosis. Proc. TOOLDIAG'93, pages 817–836, Toulouse, France, 1993. CERT.
- [7] R.J. Patton. Robust model-based fault diagnosis: the state of the art. IFAC Fault Detection, Supervision and Safety for Technical Processes, pages 1–24, Espoo, Finland, 1994.
- [8] J. Auzins, H. Johansson, and J. Nytomt. Ion-gap sense in misfire detection, knock and engine control. *SAE-Technical Paper Series*, (950004), 1995.
- [9] J. Gertler and M. Costin. Model-based diagnosis of automotive engines. IFAC Fault Detection, Supervision and Safety for Technical Processes, pages 393–402, Espoo, Finland, 1994.
- [10] P.M. Frank. Application of fuzzy logic to process supervision and fault diagnosis. Fault Detection, Supervision and Safety for Technical Processes, pages 507–514, Espoo, Finland, 1994. IFAC.
- [11] E.Y. Chow and A.S. Willsky. Analytical redundancy and the design of robust failure detection systems. *IEEE Trans. on Automatic Control*, 29(7):603–614, 1984.
- [12] R.N. Clark. The dedicated observer approach to instrument fault detection. Proc. of the 15th CDC, pages 237–241, 1979.
- [13] J.E. White and J.L. Speyer. Detection filter design: Spectral theory and algorithms. *IEEE Trans. Automatic Control*, AC-32(7):593–603, 1987.
- [14] P.M. Frank and J. Wünnenberg. *Robust fault diagnosis using unknown input observer schemes*, chapter 3. In Patton et al. [4], 1989.
- [15] N. Viswanadham, J.H. Taylor, and E.C. Luce. A frequency-domain approach to failure detection and isolation with application to GE-21 turbine engine control systems. *Control - Theory and advanced technology*, 3(1):45–72, March 1987.
- [16] W.C. Merrill, J.C. DeLaat, and W.M. Bruton. Advanced detection, isolation, and accommodation of sensor failures - real-time evaluation. *AIAA J. of Guidance, Control, and Dynamics*, 11(6):517–526, 1988.
- [17] J.J. Gertler, M. Costin, X. Fang, R. Hira, Z. Kowalczyk, and Q. Luo. Model-based on-board fault detection and diagnosis for automotive engines. IFAC Fault Detection, Supervision and Safety for Technical Processes, pages 503–508, Baden-Baden, Germany, 1991.
- [18] V. Krishnaswami, G.C. Luh, and G. Rizzoni. Fault detection in IC engines using nonlinear parity equations. Proceedings of the American Control Conference, pages 2001–2005, Baltimore, Maryland, 1994.
- [19] L.C. Shen and P.L. Hsu. Design of the robust unknown input observer for automotive engines. pages 129–134, San Francisco, USA, 1996. IFAC.
- [20] P.S. Min and W.B. Ribbens. A vector space solution to incipient sensor failure detection with applications to automotive environments. *IEEE trans. on Vehicular Technology*, 38(3):148–158, 1989.
- [21] G.F. Mauer. On-line performance diagnostics for internal combustion engines. *Int. Conf. on Ind. Electronics, Control, Instrumentation and Automation, San Diego*, 3:1460–1465, 1992.
- [22] R.N. Riggins and G. Rizzoni. The distinction between a special class of multiplicative events and additive events: Theory and application to automotive failure diagnosis. American Control Conf., pages 2906–2911, San Diego, California, 1990.
- [23] P.L. Hsu, K.L. Lin, and L.C. Shen. Diagnosis of multiple sensor and actuator failures in automotive engines. *IEEE trans. on Vehicular Technology*, 44(4):779–789, 1995.
- [24] D. Cho and P. Paoletta. Model-based failure detection and isolation of automotive powertrain systems. *Proc. of ACC*, pages 2898–2905, 1990.
- [25] P. Paoletta and D. Cho. A robust failure detection and isolation method for automotive power train sensors. IFAC Fault Detection, Supervision and Safety for Technical Processes, pages 509–515, Baden-Baden, Germany, 1991.
- [26] G. Rizzoni and R. Hampo. Real time detection filters for onboard diagnosis of incipient failures. *SAE Paper*, (890763), 1989.
- [27] G. Rizzoni and P.S. Min. Detection of sensor failures in automotive engines. *IEEE Trans. on Vehicular Technology*, 40(2):487–500, 1991.
- [28] G. Rizzoni, P.M. Azzoni, and G. Minelli. On-board diagnosis of emission control system malfunctions in electronically controlled spark ignition engines. *Proc. of the American Control Conference*, pages 1790–1795, 1993.
- [29] V. Krishnaswami, G. Chun-Luh, and G. Rizzoni. Diagnosis of exhaust emissions control systems during the E.P.A. tailpipe inspection program. IFAC Fault Detection, Supervision and Safety for Technical Processes, pages 381–386, Espoo, Finland, 1994.

- [30] G.C. Luh and G. Rizzoni. Identification of a non-linear mimo ic engine model during I/M240 driving cycle for on-board diagnosis. Proceedings of the American Control Conference, pages 1581–1584, Baltimore, Maryland, 1994.
- [31] J. Gertler, M. Costin, X. Fang, R. Hira, Z. Kowalalczuk, and Q. Luo. Model-based on-board fault detection and diagnosis for automotive engines. *Control Engineering Practice*, 1(1):3–17, 1993.
- [32] J. Gertler, M. Costin, X. Fang, R. Hira, Z. Kowalalczuk, M. Kunwer, and R. Monajemy. Model based diagnosis for automotive engines - algoritm development and testing on a production vehicle. *IEEE Trans. on Control Systems Technology*, 3(1):61–69, 1995.
- [33] E. Hendricks. Mean value modelling of spark ignition engines. *SAE-Technical Paper Series*, (900616), 1990.
- [34] John B. Heywood. *Internal Combustion Engine Fundamentals*. McGraw-Hill series in mechanical engineering. McGraw-Hill, 1992.