

Modular Fault Diagnosis System for Engine Test Bed Measurements

Michael Wohlthan and Gerhard Pirker
 LEC GmbH

Andreas Wimmer
 Graz University of Technology

ABSTRACT

To achieve high power output and good efficiency and to comply with increasingly stricter emission standards, modern combustion engines require a more complex engine design, which results in a higher number of control parameters. As the measurement effort and the number of sensors for engine development at the test bed continue to increase, it is becoming nearly impossible for the test bed engineer to manually check measurement data quality. As a result, automated methods for analysis and plausibility checks of measurement data are necessary in order to find faults as soon as they occur and to obtain test results of the highest possible quality. This paper presents a methodology for automated fault diagnosis on engine test beds. The methodology allows reliable detection of measurement faults as well as the identification of the root cause of faults. This methodology is based on a modular concept that combines several modules and methods consisting of physical principles as well as statistical or data-based models. The theoretical principles and functionality of the different modules and the combinational logic between the modules are discussed in detail. A standardized setup of the modules facilitates the combination of module results, thereby improving the expandability and flexibility of the system. Finally, a thorough evaluation of the fault diagnosis system using various parameters is provided and the performance of the individual modules are compared and discussed. In conclusion, the best diagnosis results can be obtained with a combination of all the modules presented in the paper.

CITATION: Wohlthan, M., Pirker, G., and Wimmer, A., "Modular Fault Diagnosis System for Engine Test Bed Measurements," *SAE Int. J. Commer. Veh.* 10(1):2017, doi:10.4271/2017-01-0386.

INTRODUCTION

Given the higher number of control values and regulating values, the growing complexity of modern internal combustion engines requires a higher number of sensors for development work on engine test beds. This increasingly complicates manual monitoring of the quality of measurement data. For reasons of cost, it is also imperative to reduce test bed development time as much as possible. Thus it is attempted to best maximize the significance of the measurements in all tests. Measurement faults frequently do not become apparent until the postprocessing stage and give rise to a useless test result or the need for an expensive repetition of the test. Faulty measurement results are thus to be avoided if possible. For these reasons, the need is growing for a fully automatic measurement data monitoring system. Faults should be detected as soon as possible and the fault diagnosis system should also be capable of giving test bed technicians immediate information about where or in which measurement channel the cause of the fault is located so that the right measures to eliminate faults can be taken directly.

The topic of fault diagnosis in measurement systems has already been widely discussed in the scientific literature. Isermann [1] defines the most important terms related to fault diagnosis. For example, the term fault is defined as such: an unpermitted deviation of at least one

characteristic property or parameter of the system from the acceptable/usual/standard condition. The fundamentals of fault diagnosis can be found in [2]. The special case of fault diagnosis on engine test beds has also been dealt with extensively, cf. [3,4,5,6]. LEC and Graz University of Technology have also published papers on this topic [7,8] that are concerned with the detection and correction of measurement faults using Gaussian correction in order to improve the accuracy of engine cycle analysis. One application of this correction is presented in [9]. In [10], on the other hand, a method for fault diagnosis on engine test beds whose basic principle is to combine physical and statistical methods, is presented.

This paper presents a fault diagnosis procedure for steady state engine test bed measurements that expands the method presented in [10]. The most important requirements for this procedure can be summarized as follows:

- High general validity: Fault diagnosis should make do with a small number of system parameters in order to enable its quick use with only little configuration effort.
- Easy expandability: The general architecture of the procedure should be organized so that only little effort is required to integrate further methods into the fault diagnosis system.

- High scope of diagnosis: Fault diagnosis should be capable of monitoring a maximum number of measurement channels
- High recall: Faults should be detected reliably and as soon as possible
- Few false alarms: Fault diagnosis should only set off an alarm when a fault has actually occurred. This requirement stands in direct conflict with the previous one for high recall. The ratio of recall to the number of false alarms can be set by the level of threshold values.

The procedure presented here has been developed to meet these requirements.

GENERAL APPROACH

Measurement data is examined that consists of n measured values ($x_1, \dots, x_p, \dots, x_n$). This measurement data is checked using a procedure for fault diagnosis shown in Figure 1. Additional input data is provided by the relative uncertainties ($u_1, \dots, u_p, \dots, u_n$) of the measurement channels. As shown below, measurement uncertainties are required to determine threshold values. The result of fault diagnosis is a classification of the measurement data as either faulty or fault-free. In addition, the error measures ($f_1, \dots, f_p, \dots, f_n$) are determined with fault diagnosis. The error measure values lie in a range of zero to one. The higher the error measure f_p , the more likely the corresponding measured value x_i is the cause of the fault.

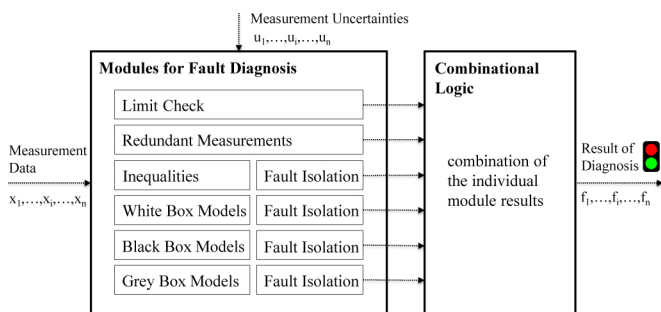


Figure 1. Structure of the fault diagnosis procedure

As visible in Figure 1, the procedure is based on a modular setup; modularity appears to be a promising layout that has been previously used for similar implementations, cf. [11]. The fault diagnosis procedure consists of different modules, each of which checks the quality of the measurement data. In the combinational logic step, a final result of diagnosis is obtained from the individual results. This structure makes it possible to easily combine a large number of different methods. The modules that are used are of varying complexity. They range from very simple modules such as limit check or the check of simple inequalities to complex model-based modules (white box, black box and gray box models). The modules are described in detail later in this paper.

All modules have the same basic setup. Fault diagnosis can be divided into two consecutive substeps [10]:

- **Fault detection:** Establishment of whether a fault has occurred

- **Fault isolation:** Determination of the faulty measurement channel

As can be seen in Figure 2, this division also determines the general setup of a fault diagnosis module. A module generally consists of a fault detection step and a fault isolation step. The main task of fault detection is to check fault conditions. If a fault condition is met, the module has detected a fault, whose cause must still be determined in the fault isolation step. In addition, n module-specific error measures ($f_1^{mod}, \dots, f_i^{mod}, \dots, f_n^{mod}$) are generated from fault isolation. An error measure f_i^{mod} is assigned to each measured value x_i . The module-specific error measures also lie in a range of zero to one. Here too it is true that the higher the error measure f_i^{mod} , the more probable it is that the measured value x_i is the cause of the fault. In the event that no fault condition is met, all error measures in the module equal zero.

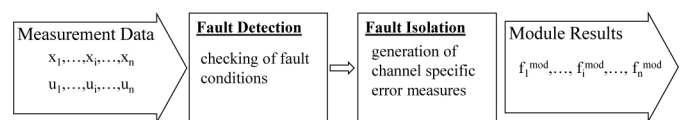


Figure 2. Setup of a fault diagnosis module

As seen in Figure 1, not every module requires a fault isolation step. The reason for this is also explained in more detail below in the description of the modules. The uniform design of the modules with module-specific error measures as output makes it easier to link results in the subsequent combinational logic step and thus improve the expandability of the procedure. Additional modules can be easily added without any change to the combinational logic.

DESCRIPTION OF THE MODULES

The following section describes the individual fault diagnosis modules in detail. A basic distinction is made between non-model-based modules and model-based modules. Fault detection and fault isolation are each presented. Finally, it is explained how the combinational logic is used to link up the results from the individual modules.

Non-Model-Based Modules

There are three very simple methods for fault diagnosis that are only capable of recognizing major faults. For comprehensive fault diagnosis, however, these checks are very practical and well justified.

Limit Check

The main task of this module is to recognize serious faults. The measured value x_i of a channel is compared with an upper limit $Lim_{u,i}$ and a lower limit $Lim_{l,i}$. The fault condition is:

$$Lim_{l,i} > x_i > Lim_{u,i} \quad (1)$$

The user must set the limits for each channel that is to be checked. Since the faulty measurement channel is isolated as soon as the limit is exceeded, this module does not require an extra fault isolation step, cf. Figure 1. In the event that the limit is exceeded, the error measure of a measurement channel is $f_i = 1$, otherwise it is $f_i = 0$.

Redundant Measurements

This module can only be used if redundant measured values exist. In exceptional cases, important measured values are measured several times on engine test beds. In this case, it is appropriate to use this information for automatic fault diagnosis. The fault condition that is checked in this case is:

$$|x_i - x_{i,red}| > thr_i \quad (2)$$

The absolute deviation between the measured value x_i and the redundant measured value $x_{i,red}$ is compared with a threshold value thr_i . This threshold value is determined from the default measurement uncertainties of the measuring instruments:

$$thr_i = k_{red} \sqrt{u_i^2 + u_{i,red}^2} \quad (3)$$

whereby u_i indicates the measurement uncertainty of the measured value x_i and $u_{i,red}$ the measurement uncertainty of the redundant measured value $x_{i,red}$. The diagnostic sensitivity can be set with k_{red} . The quadratic addition follows from Gaussian error propagation. Separate fault isolation is not required for this module as well. Nevertheless, unambiguous fault isolation is not possible if the fault condition is met (Eq. 2) since both the measured value x_i and the redundant measured value $x_{i,red}$ are possible causes of the fault. Thus both error measures yield $f_i = 0.5$ and $f_{i,red} = 0.5$.

Inequalities

The exact physical relationship between two measured values is usually unknown, yet they can often be expressed in an inequality form. A concrete example of this is the arrangement of several temperature measuring points in the exhaust gas system. Since heat from the exhaust is dissipated into the environment downstream, it is known that the temperature decreases. Thus both of these temperatures can be combined with an inequality. The exact temperature difference can only be determined from complex physical modeling. For any two measured values x_a and x_b , the fault condition of an inequality is:

$$x_a < x_b \quad (4)$$

In this case, the point of fault isolation is simply to ascertain whether a measured value is found in only one violated inequality or in several. If the former is the case, the error measures are $f_a = 0.5$ and $f_b = 0.5$. If a measured value, for example x_a , is found in several inequalities that are not fulfilled, it can be isolated as the cause of the fault with high probability ($f_a = 1, f_b = 0$).

Model-Based Modules

Three more complex modules will now be presented that depend on model-based methods. A distinction is made between white box, black box and gray box models depending on the empirical share. The next

three sections concentrate on the fault detection step alone. Since the process of fault isolation is similar for all model-based modules, the fault isolation step is explained later in a separate section.

White Box Models

White box models are those whose physical relationships are completely known. This means the model structure as well as all required model parameters are given. This module completely dispenses with empirical data. Physical relationships of this type can be obtained at engine test beds in the form of balance equations for mass and energy.

Carbon balance is a tried and true relationship that can be used to check the plausibility of measurement results. The carbon discharged in the constituents of the exhaust gas should be equal to the carbon introduced by the fuel.

$$0 = C_{out} - C_{in} = \left(\frac{M_C}{M_{CO}} \mu_{CO} + \frac{M_C}{M_{CO2}} \mu_{CO2} + \mu_{C,fuel} \mu_{HC} \right) (\dot{m}_{fuel} + \dot{m}_{air}) - (\mu_{C,f} \dot{m}_{fuel}) \quad (5)$$

A balance equation for atomic oxygen can be similarly formulated:

$$0 = O_{out} - O_{in} = \left(2 \frac{M_O}{M_{CO2}} \mu_{CO2} + \frac{M_O}{M_{H2O}} \mu_{H2O} + \mu_{O2} + \frac{M_O}{M_{CO}} \mu_{CO} + 2 \frac{M_O}{M_{NO2}} \mu_{NO2} + \frac{M_O}{M_{NO}} \mu_{NO} \right) (\dot{m}_{fuel} + \dot{m}_{air}) - \mu_{O,fuel} \dot{m}_{fuel} - \dot{m}_{air} \left(\mu_{O2,air} + \frac{M_O}{M_{H2O}} \mu_{H2O,air} \right) \quad (6)$$

In the equations above, $\dot{m}$ represents mass flow, M molar mass and μ mass fraction. If the water content of the exhaust is not measured it can be calculated based on the laws of thermodynamics. A detailed description of the procedure can be found in [12].

Other relationships can be obtained from energy balances. If the system boundaries are set for the complete engine, the following is obtained:

$$0 = E_{in} - E_{out} = P_e + \dot{H}_{out} + \dot{Q}_{unb} + \dot{Q}_{oil} + \dot{Q}_w + \dot{Q}_{env} - (\dot{m}_{fuel} LCV + \dot{H}_{in}) \quad (7)$$

whereby P_e represents the effective power output, $\dot{H}_{in}$ and $\dot{H}_{out}$ the inflowing or discharge enthalpy, $\dot{Q}_w$ and $\dot{Q}_{oil}$ the heat that dissipates via the coolant and oil circuit and $\dot{Q}_{env}$ the heat that dissipates into the environment via radiation and convection. Since the last term is normally not measured, the energy balance of the complete engine has higher uncertainties like the mass balances shown above.

In classical pressure curve analysis, however, the system boundary is only set to the combustion chamber. The rate of heat release is calculated from the measured pressure curve dQ_h . It should be equal



to the measured fuel energy $m_{fuel} \cdot LCV$ minus the unburned energy remaining in the exhaust Q_{unb} . The energy balance of the combustion chamber is:

$$0 = \int dQ_b - (m_{fuel}LCV - Q_{unb}) \quad (8)$$

Since in practice every measurement is fraught with measurement uncertainties and the models and balances represent a simplification of reality, the amount quantity added will never correspond exactly to the amount discharged even in fault-free operation. Thus there will be a deviation between those two values. The relative deviation is referred to as a residual and can be determined for balance equations as follows:

$$res_j = \frac{out_j - in_j}{in_j} \quad (9)$$

The index j stands for an arbitrary model. In addition to the above mentioned balance equations, the excess air ratios can be compared by using a lambda probe in the exhaust. The value calculated from air and fuel mass should correspond to the measured value obtained with the lambda probe. The residual of this relationship is as follows:

$$res = \frac{\frac{\dot{m}_{air}}{L_{st} \cdot \dot{m}_{fuel}} - \lambda}{\lambda} \quad (10)$$

As an alternative, excess air ratio could also be calculated from the exhaust components as Brettschneider [13] published. The residuals are now defined for the different models; they represent fault symptoms and must be compared with threshold values in the form of a fault conditions:

$$|res_j| > thr_j \quad (11)$$

The threshold value can be estimated using the following equation:

$$thr_j = k_{wb} \sum_{i=1}^n \left(\frac{\partial res_j}{\partial x_i} \right) u_{rel,i} + u_{model,j} \quad (12)$$

whereby the first term represents the share of the threshold value that is caused by measurement uncertainties. This can be estimated with Gaussian error propagation using the relative measurement uncertainties $u_{rel,i}$. The second term $u_{model,j}$ represents the share caused by model faults themselves. While mass balances have relatively minor model faults, higher values are used with energy balances. The role of factor k_{wb} is to set the sensitivity of fault diagnosis.

Black Box Models

In contrast to white box models, neither the model structure nor model parameters of black box models are known. These models are data based. The relationship between model input and output values

is purely mathematical and created based on measurement data. Continuous model training is required; this means that after each measurement, the models are updated based on the measurement value history.

The procedure presented here uses principle component regression [14] (PCR) to generate the models. This consists of principle component analysis (PCA) followed by multiple linear regression (MLR). In principle component analysis, a linear transformation of variables is performed so that the relevant information can be transferred with the smallest number of new variables. The advantages of PCR over pure MLR include the minimizing of the degrees of freedom, which reduces the tendency for overfitting effects to occur.

In general, all available measured values and manipulated values are used for regression. This results in models of the following type:

$$\hat{x}_i = f_i(x_1, \dots, x_k, \dots, x_n, z_1, \dots, z_l, \dots, z_o) \quad k \in \{1, n\} \setminus \{i\} \quad (13)$$

where $\hat{x}_i$ is the predicted value, x are the measured values, n the number of measured values, z the regulating values and o the number of regulating values. Before the models are used for fault diagnosis, it is necessary to check the quality of the models. This is done with the coefficient of determination R^2 . Only models with a sufficiently high coefficient of determination are allowed for fault diagnosis. The value $R^2 = 0.95$ has proved to be suitable as a threshold.

The relative fault between predicted value and measured value is referred to as a residual and represents the fault symptom:

$$res_i = \frac{\hat{x}_i - x_i}{x_i} \quad (14)$$

As above for white box models, the residual is compared to a threshold value. The threshold value thr_i can be determined as follows:

$$thr_i = k_{bb} \cdot RRMSE = k_{bb} \sqrt{\frac{1}{n-p} \sum_{l=1}^n \left(\frac{(\hat{x}_{i,l} - x_{i,l})^2}{x_{i,l}} \right)} \quad (15)$$

whereby the relative root mean squared error (RRMSE) is an estimation of the standard deviation of the model, n the number of measurement points upon which the model is based and p the number of parameters to be estimated. The diagnostic sensitivity can be adjusted with the factor k_{bb} . If the residual exceeds a threshold value at several measurement points in a row, it can be assumed that there is a fault. The number of times the threshold must be exceeded before an alarm is set off is referred to as the alarm delay ad . This alarm delay is necessary for two reasons. Because of model faults, the threshold can be exceeded at individual measurement points even if a physical fault is not present. In addition, it is possible that when there is a change in operating point, the models are not yet valid for the current state of operation, for example the first measurement is