MODEL-BASED FAULT DETECTION AND DIAGNOSIS OF SELECTIVE CATALYTIC REDUCTION SYSTEMS FOR DIESEL ENGINES

BY

RUI CHEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Agricultural and Biological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Xinlei Wang, Chair
Professor Alan C. Hansen
Professor Chia-Fon Lee
Assistant Professor Alejandro Dominguez-Garcia

ABSTRACT

The continuously stringent emission regulations call for the adaptation of the Selective Catalytic Reduction (SCR) system by many diesel engine manufacturers. In order to show the EPA the latest emission compliance, the NO_x sensors are required to be installed upstream and downstream of the SCR. As a result, the NO_x reduction efficiency is also required to be monitored by the On-Board Diagnostics (OBD) regulations. Specifically, a diagnostic algorithm is required to detect and isolate the SCR system faults that may cause emission violations.

In this research, two model-based fault detection and isolation algorithms were developed to detect and isolate the dosing fault and the outlet NOx sensor fault for the SCR system. The dosing fault was treated as an actuator additive fault, while the outlet NOx sensor drift and/or offset fault was treated as a sensor additive fault.

First, a 0-D SCR nonlinear dynamic model was developed to facilitate the model-based approaches. A parity equation residual generator was designed based on the linearized SCR model and the fault transfer function matrix. Then, a sliding mode observer based residual generator was designed directly based on the nonlinear model. The two diagnostic algorithms were then implemented in the Matlab/Simulink environment for validation. A high fidelity nonlinear 1-D SCR model was used to generate system outputs and to simulate the plant. The simulation results showed that the two model-based fault diagnosis approaches were capable of detecting and isolating the outlet NOx sensor and dosing faults in real time.

ACKNOWLEDGEMENTS

This research was guided by my Ph.D. advisor, Dr. Xinlei Wang. Without his guidance I could never have finished my Ph.D. study. Dr. Wang is very knowledgeable in the area of diesel engine emission control and helped me get started at the beginning of the project, and his continuous financial support during my Ph.D. years made this research possible. I was given a lot of freedom to choose the research topic and to manage the progress. I also learned a lot from him about how to work on industrial projects and to communicate effectively.

I would like to also thank my committee members: Dr. Alan C. Hansen from the Agricultural and Biological Engineering Department, Dr. Chia-Fon Lee from the Mechanical Engineering Department and Dr. Alejandro Dominguez-Garcia from the Electrical Engineering Department. I appreciate the time and effort they put into our discussions and their valuable suggestions on my research.

It is also necessary to extend my appreciation to staff and co-workers from industry. I want to specifically thank Dr. Xinqun Gui, John Deere Power Systems, for giving me the opportunity to work on a diesel engine particulate filter project during the summer of 2010. I also want to thank James Lenz from the John Deere Technology Innovation Center in Champaign, Illinois for his effort in bringing about this SCR diagnostic project.

This work would not have been possible without the support of my family. Among them, I want to thank my mother, Shumin Liu, who comforted and encouraged me during the most difficult times. Her wisdom guided me throughout my Ph.D. years. I also would like to thank my girlfriend, Yuanyuan, who later became my wife, for her understanding and kindness during my dissertation study.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1
- **1.1 Background** 1
- **1.2 Motivation** 3
- **1.3 Objectives** 6

CHAPTER 2: LITERATURE REVIEW 7
- **2.1 Fault Detection Methodologies** 7
- **2.2 Fault Diagnosis Methodologies** 18
- **2.3 Selective Catalytic Reduction System** 23
- **2.4 SCR System Modeling and Control** 24

CHAPTER 3: APPROACHES 28
- **3.1 SCR System Modeling** 28
- **3.2 Residual Generators Design** 29
- **3.3 Simulation Validation** 30

CHAPTER 4: SCR SYSTEM MODELING 32
- **4.1 Diesel Engine Aftertreatment System** 32
- **4.2 SCR System Model Development** 35

CHAPTER 5: RESIDUAL GENERATOR DESIGN 43
- **5.1 Parity Equation Residual Generation** 43
- **5.2 Observer Based Residual Generation** 49

CHAPTER 6: SIMULATION VALIDATION 56
- **6.1 Simulation Environment** 56
- **6.2 Validation Results** 60

CHAPTER 7: CONCLUSIONS 68

REFERENCES 71

APPENDIX A: NOMENCLATURE 78

APPENDIX B: ACRONYMS 79

CHAPTER 1

INTRODUCTION

1.1 Background

Compared with gasoline engines, the diesel engine inherently has a much higher thermal efficiency, longer life cycle, and uses safer fuel. Due to these advantages, it has been favored in many applications, including stationary power generation, heavy duty vehicles and marine applications. Although a diesel engine generates minimum carbon monoxide, Nitrogen Oxides (NO_x) and Particulate Matter (PM) in the exhaust emission are major concerns (Khair and Majewski, 2006). NO_x is a serious environmental cause for smog formation. Nitrogen dioxide (NO_2) is especially toxic when inhaled and can cause lung edema or even death. PM is almost totally inspirable due to its small size, and increases the risk of human cancer and heart diseases.

One of the major sources of diesel emissions is off-road vehicles equipped with a diesel engine. To regulate this part of diesel emissions, the Environmental Protection Agency (EPA), which represents the United States government, has enforced Tier 1 ~ Tier 4 diesel emission regulations on non-road diesel engine manufacturers. The emission limits of these regulations have become gradually more stringent from 1996 through 2008, as shown in Fig 1.1 (Gui et al., 2010). The Tier 1 ~ Tier 3 emission regulations were met by manufacturers mostly through in-cylinder combustion improvements, Exhaust Gas Recirculation (EGR), turbo-charger, and intercooler, with limited use of exhaust aftertreatment technologies. As a result, diesel emissions have improved dramatically.

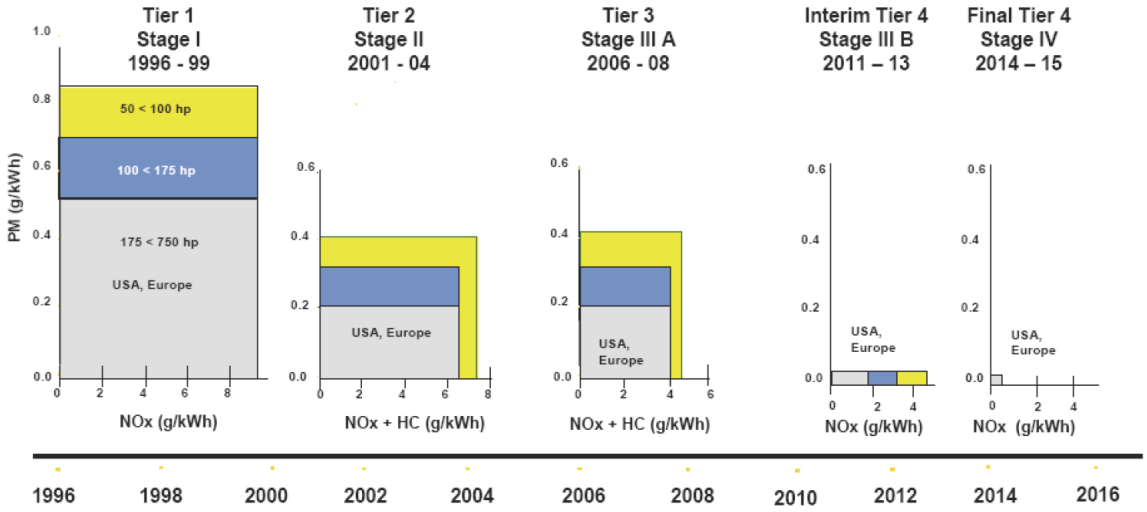

Fig 1.1 Non-road emission regulations (Gui et al., 2010)

Recently, the EPA introduced the Tier 4 emission standard for non-road vehicles equipped with a diesel engine. The regulation is being phased in from 2011 to 2015, during which it is divided into two phases. The first one is called Interim Tier 4 (IT4), which requires further reduction of PM in the diesel exhaust. The second phase is called Final Tier 4 (T4), which requires further reduction of NO_x by 80% for engines with rated power of 130kW and above.

While in-cylinder combustion refinements helped manufacturers achieve Tier 2 and Tier 3 standards, the IT4 and T4 regulations require new exhaust aftertreatment systems to further reduce the PM and NO_x. For NO_x reduction, the Selective Catalytic Reduction (SCR) system has been a proven solution in stationary internal combustion engine applications for its high conversion efficiency. The SCR works by mixing the exhaust gases with ammonia in the form of urea (or Diesel Emissions Fluid, DEF) and passing the mixture through a catalyst. The result is that the majority of NO_x in the exhaust will be converted to harmless elemental nitrogen and

water vapor. The SCR technology has been selected for highway applications by most of the diesel engine manufacturers to meet EPA regulations.

1.2 Motivation

The SCR system is a complex system and contains sensors and actuators that are liable to drift, with aging and other faults. The SCR system uses a consumable reagent (urea), which is an added cost to the vehicle operation. This may motivate the vehicle operator to tamper with the SCR system by diluting or re-routing the urea line to reduce the cost (Nebergall et al., 2005). Eventually, these faults will lead to a decrease in NO_x reduction efficiency and will, therefore, exceed the T4 emission standard. The EPA realized that, and has already enforced On-Board Diagnostics (OBD) regulations for 2010 and later heavy-duty diesel engines used for highway applications. The goal is to monitor the SCR conversion efficiency and NO_x sensor performance, as well as the actual injection of the SCR reductant. When low conversion efficiency is detected, the Malfunction Indicator Light (MIL) on the instrument panel will illuminate and a diagnostic trouble code is recorded. This OBD regulation will be extended to non-road Tier 4 engine applications in the near future. This will be the first time that an OBD regulation is enforced in the non-road diesel engine sector. Therefore, currently all major diesel engine manufacturers are actively seeking solutions in order to meet the regulations.

The SCR system is a newer aftertreatment system added to the mobile diesel engine (Khair, 2006). Unlike stationary applications, the load and speed of mobile engines vary dramatically in an unpredictable way. Accordingly, the exhaust flow rate, temperature and emission gas components will also vary dramatically. This characteristic makes the fault detection task using traditional single threshold monitoring very difficult. Unfortunately, since the SCR system has

recently been introduced to non-road applications, there has not been much effort in the development of monitoring methods for SCR systems (Mohammadpour et al., 2011; Chen and Wang, 2014).

To monitor the NO_x conversion efficiency, one straightforward way is to install two NO_x sensors, before and after the SCR, and to compute the NO_x conversion efficiency onboard. From a diagnostic system point of view, once a low NO_x conversion rate and/or an abnormal sensor reading is detected, a decision is expected to be made to indicate the location and the nature of the fault. However, this fault isolation task can be difficult using traditional single threshold monitoring of NO_x sensor readings. For example, without knowing the state of the plant (SCR catalyst) and that of the actuator (doser), it cannot be determined if an abnormal outlet NO_x sensor reading is caused by the sensor itself or the doser. Installing more sensors will increase the cost and system complexity and sometimes will lead to more reliability issues. Therefore, a more dedicated diagnostic method that can describe the interaction between SCR components is needed.

In order to comply with the final Tier 4 NO_x emission requirements and future OBD requirements, NO_x sensors have been installed at both the inlet and outlet of the SCR system. The outlet NO_x sensor reading is critical for determining regulation compliance. Unfortunately, the outlet NO_x sensor reading is very sensitive to small sensor drift and/or offset due to lower NO_x concentration at the SCR outlet. This kind of sensor drift and/or offset can be treated as a sensor additive fault f_1. Another common failure of the SCR system is due to dosing fault. This

fault can be caused by various reasons from blockage on the injector nozzle to human tampering of DEF quality. All of these dosing faults can be treated as a urea dosing quantity additive fault f_2.

According to the preliminary research of the fault tree analysis, the symptom of outlet NO_x sensor reading exceeding the limit could be the result of three major faults within the SCR system. They are dosing fault, SCR catalyst fault and outlet NO_x sensor fault. Each major fault may be caused by a subset of faults, as illustrated in Fig 1.2.

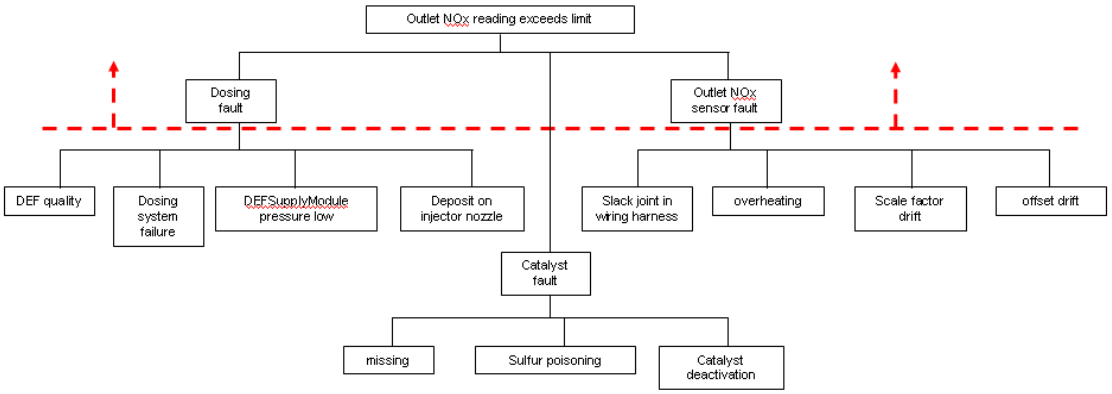

Fig 1.2 Fault tree analysis

This research is intended to detect and isolate the SCR dosing fault and the outlet NO_x sensor fault. The SCR catalyst is assumed to be fault free, while only a dosing fault or the outlet NO_x sensor fault may be present.

The scope of fault detection and isolation will remain in the SCR system and will assume all other engine and aftertreatment components are fault free. The task of fault isolation will not go below the dotted line as shown in Fig 1.2 (no further isolation within f_1 and f_2).

1.3 Objectives

In this research, other parts of the engine and aftertreatment system except the SCR system are assumed to be fault free. The SCR system can only contain a dosing fault and/or the outlet NO_x sensor fault.

The objectives in this research are:

1. Develop a SCR catalyst model for the target system.
2. Design a parity equation residual generator.
3. Design an observer based residual generator.
4. Implement the two residual generators in a Matlab/Simulink environment.
5. Validate and evaluate the detectability and the isolability of the two residual generators in a Matlab/Simulink environment.

CHAPTER 2

LITERATURE REVIEW

Extensive research has been done in the field of fault detection and diagnosis (Isermann, 2006). Among those research efforts, model-based fault detection and diagnosis methods have shown many advantages in solving engineering system diagnostic problems (Gertler, 1998). Unfortunately, most of the model-based fault detection and identification (FDI) results are only applicable to linear systems (Patton, et al., 2000). A summary of model based FDI methodologies for nonlinear systems is reviewed at the end of section 2.1.

2.1 Fault Detection Methodologies

A classification of the fault detection methodologies is presented in Fig 2.1 (Isermann, 2006). In this section, the typical methodologies in each of the categories will be reviewed.

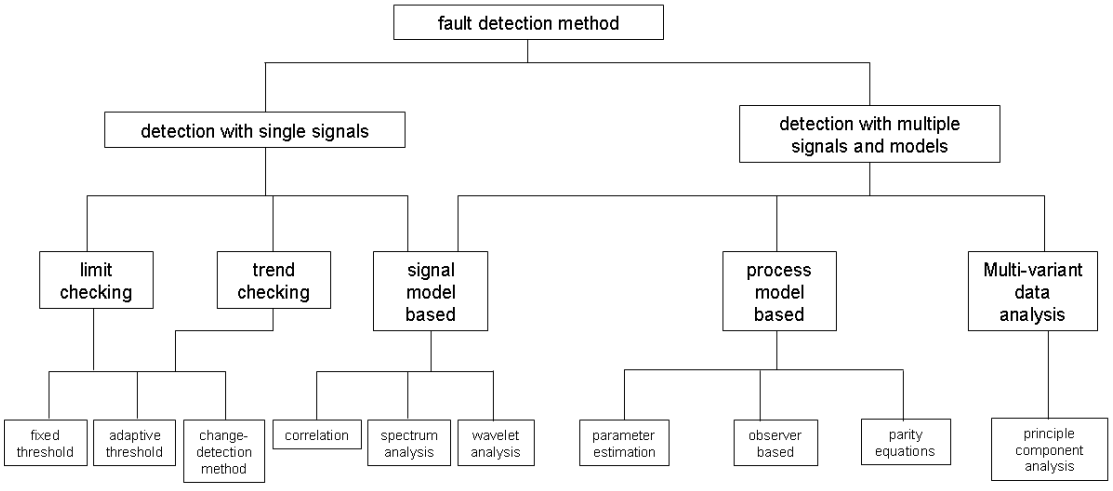

Fig 2.1 Classification of fault detection methods (Isermann, 2011)

2.1.1 Limit checking:

The simple limit checking approach is widely used in practice. It can also be extended to checking of selected variables, which is checking the derivatives of the selected variables. Though simple and straightforward, the limit checking approach has two serious drawbacks (Isermann, 2006):

1) Because the measured variables may vary widely while still in a normal operation condition, the fixed thresholds have to be set very conservatively to avoid false alarms.

2) One component fault can propagate too many measured variables, and thus exceed many thresholds in a sequence determined by the system dynamics. This property makes the fault isolation extremely difficult.

An improvement of a fixed threshold is the adaptive threshold (Hofling, 1996). When implementing the model based FDI methods, generated residuals could deviate from zero even when no fault has occurred due to model uncertainties. These deviations usually depend on amplitude and frequencies of the input excitation. The adaptive threshold approach applies to a variable gain and a high pass filter (HPF) to the input excitation in order to enlarge the threshold, and then a low pass filter (LPF) to smooth the threshold. This approach usually is not implemented alone because it does not include the internal dynamics of the process.

Another revision of the fixed threshold checking is the change detection method (Himmelblau, 1970). In Fig 2.2, normal distributions are assumed for the observed variable Y for the nominal state (index 0) and the changed (faulty) state (index 1).

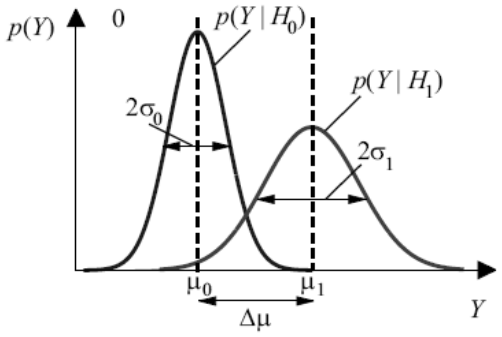

Fig 2.2 Change detection (Isermann, 2006)

The main idea of this approach is to perform statistical tests (t and F tests) for the observed variable to see if the mean and/or the variance of the variable has/have changed. The mean and variance of the observed variable can be obtained by LPFs. Then, given the level of significance, the decision of whether the observed variable has changed from a nominal value could be drawn by checking the range of acceptance. The change detection method can be further improved by using fuzzy thresholds, because there is seldom a definite border between both states. Due to the wide operating conditions of the aftertreatment system, the implementation of the change detection method has to rely on extensive field test data for all faulty and normal states.

2.1.2 Signal model based fault detection:

Many measured engineering signals are harmonic in nature, such as machine vibration measurement, engine misfire (Ribbens, 1990) and cylinder knocking (Samimy, 1996), as well as bearing faults. These faults will cause changes to some of the properties of the signals. The signal model based fault detection is suited for these kinds of applications.

The first task of the signal model based fault detection is to develop or assume a special model for the signal. Then, suitable features such as amplitudes, phases, spectrum frequencies, and

correlation functions of the signals are generated. Finally, comparison between the observed features and normal behavior will provide the analytical symptoms for further diagnosis.

Specific signal model based fault detection methods that have been studied include bandpass filtering, FFT/spectrum analysis for periodic signals (Janik, 1991), wavelet analysis for non-stationary periodic signals (Willimowski, 2000), and correlation analysis for stochastic signals (Isermann, 2010).

Generally, signal based fault detection methods are applied in vibration related faults or processes such as miss fire and knocking, which generate periodic signals. Recently, researchers from General Motors Company (Sun, 2012) demonstrated detecting urea injection faults in the SCR system using the signal model based methods. However, the NO_x sensor measurements in the SCR system are not periodic in nature. Therefore, studying the signal model features, including amplitude, phase, spectrum frequency, and correlation function, of the measured signals may not help in detecting and isolating this kind of fault in the SCR.

2.1.3 Principle component analysis (PCA)

In a large scale process, such as chemical plants, the available measurements are highly correlated but only a small number of faults produce unusual patterns. In such cases, the model based fault detection will then require too much effort to be justified. Then a data driven multivariate analysis could be applied. Specifically, the principle component analysis has attracted some attention from FDI researchers (Dunia, 1998; Gertler, 1997). The main idea of the principle component analysis is to reduce the dimensionality of a data set considering a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming the measured data to a new set of variables, the principle components, which are uncorrelated.

2.1.4 Process model based fault detection

The process model based fault detection approach originated from a chemical process control (Mah, 1976; Himmelblau, 1978) and aerospace research (Willsky, 1976; Lou, 1986). Since then, different approaches for fault detection using mathematical process models have been developed (Isermann, 2005). They have also been widely adopted in engine applications (McDowell, 2007). The task consists of detecting faults in the process, actuators and sensors by using the dependencies between different measurable signals. These dependencies are expressed by mathematical process models.

Essentially, the model based fault detection is the generation of residual, which is the deviation variable from its normal value. As for the quality of residual generation, Gertler (1998) reviewed the isolability, sensibility, robustness, and stability.

There are three categories of model based fault detection methods. They are parity equation, observer based and parameter estimation methods. As pointed out earlier, most of the model-based fault detection and identification (FDI) results are only applicable to linear systems (Patton, et al., 2000). At the end of this section, a brief review of fault detection for nonlinear systems is included.

Parity equation

This is a straightforward method. The idea is to use a model and run it in parallel to the process, thereby generating an output error (residual). If the model agrees with the process, the residual generated will only be affected by faults (Gertler, 1998). Gertler (1990) developed parity equation methods for input-output model structure. For state-space models, there is a so called Chow-Willsky scheme (Chow, 1984). Patton and Chen examined the relationship between the parity equation and the observer based approaches (Patton and Chen, 1991a and 1991b).

For linear systems, if the parity equations are formulated for more than one input and one output, it will be possible to generate structured residuals such that each residual only responds to a subset of faults (Ben-Haim, 1980; Chow, 1984; Gertler, 1985). The structured residuals help to better isolate the different faults, and give the model based method an advantage compared to the traditional limit checking method. Definitions of strong isolation and canonical structure, and concepts of non-attainable structures and non-isolable faults first appeared in (Gertler, 1985, 1990). Rank conditions for structured residual design were studied in a paper by Luo (1990).

Observer based

Observer based approaches reconstruct the outputs of the system from the input vector and sensor measurements by using observers (Luenberger 1963, 1966). Kalman filters, an optimal observer technique, can be used to optimize the outputs if sensor and process noise are considered. The error signals between the actual outputs and the estimated outputs are then used as residuals for fault detection and isolation.

Clark first published several papers (1975, 1978a, 1978b) using observer based schemes for instrument failure detection. In Clark's original work (1978a), the single observer fault detection scheme, as shown in Fig 2.3, was included. This scheme uses one observer driven by the plant input and a sensor output. This requires that the system be observable from the chosen sensor alone, which limits the applicability of this method. Furthermore, since there is a feedback inherent in the observer design, model uncertainty makes all error terms non-zero and necessitates non-zero thresholds. The choice of the threshold size is dependent on the size of the fault to be detected. So there is usually a tradeoff between the number of false alarms and the size of the fault that must be detected.

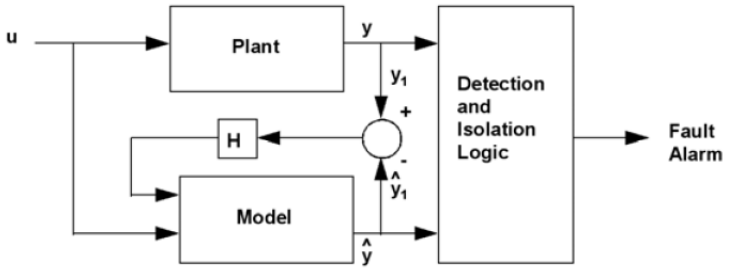

Fig 2.3 Single observer fault detection scheme

Clark (1978b) published the dedicated observer fault detection scheme, as shown in Fig 2.4. Each of the sensors has a reduced order observer dedicated to it. The i-th observer's inputs are the plant input vector and the i-th sensor value. The observer's output is an estimate of the states of the plant that are used in the decision logic to form estimates of the other sensor outputs. Therefore, two separate faults can be detected simultaneously in this scheme, which was not possible in the single observer scheme.

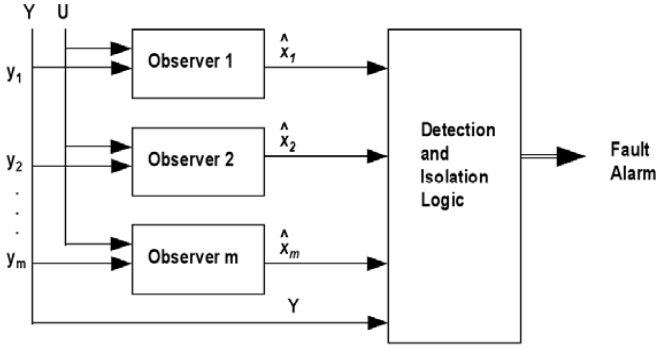

Fig 2.4 Dedicated observer fault detection scheme

Although the single observer scheme and the dedicated observer scheme provide a small degree of robustness to small modeling uncertainties, there is no freedom in their design that may be

used to increase this robustness. Therefore, it is said that neither of these schemes has robustness to unknown inputs (Wuennenberg1990; Frank, 1990).

The generalized observer scheme (Frank 1987) uses a bank of m observers similar to the dedicated observer scheme with m equal to the number of sensors as shown in Fig 2.5. Each observer uses the plant input, plus the outputs from all but the i-th sensor to estimate the value of the i-th sensor output. By using additional inputs, as compared to the dedicated observer scheme, there is additional design freedom to make the observer sensitive to all but the i-th fault and insensitive to up to m-2 unknown inputs. The observers are designed in the same way as the unknown input observers. Therefore, it has better robustness to unknown inputs. However, one disadvantage of this scheme is that the generalized observer scheme is limited to single fault detection. Furthermore, for the SCR system equipped with two sensors, this method loses its insensitivity to an unknown input.

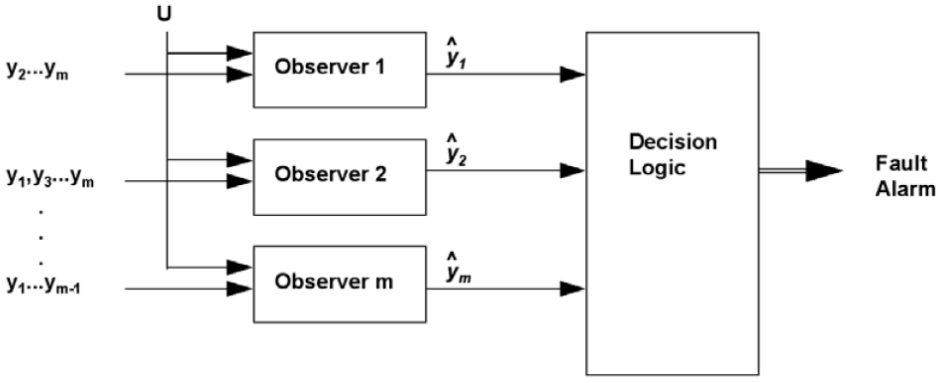

Fig 2.5 Generalized observer fault detection scheme

The unknown input observer (UIO) is one of the most popular techniques to be implemented in fault detection and isolation due to its effectiveness in sensor and actuator fault isolation

(Witczak, 2007). As an example, the scheme of UIO based actuator fault detection is shown in Fig 2.6.

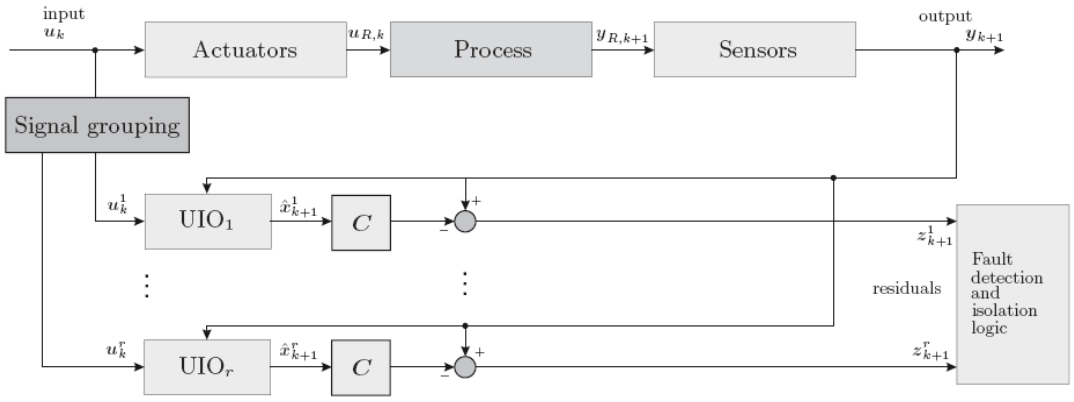

Figure 2.6 UIO based actuator fault detection scheme (Witczak, 2007)

In this UIO fault detection scheme, r UIOs need to be designed first. Each UIO will generate a residual that is driven by all inputs but one. So each residual will then respond to all actuator faults except the one that is missing. Therefore, the actuator faults can be strongly isolated. The output sensor fault isolation scheme is similar to that of the actuator.

Parameter estimation

The Parameter estimation method is best suited for parametric faults. It usually needs a dynamic process input excitation, and is especially suitable for the detection of multiple faults. Since the process parameters usually depend on physically defined properties, this fault detection method allows a deeper insight and makes the later diagnosis easier (Gertler, 1995). In most applications, process model parameters are not known exactly or not known at all. Still, given the basic model structure, parameter estimation can be performed by using input and output signals (Isermann, 1984). In this paper (Isermann, 1984), a generalized structure of FDI based on process models

and un-measurable quantities was given. The estimation could be both direct (fast) or iterative (robust). A link between parity equation and parameter estimation was shown in Delmaire et al. (1994). The relationship was later investigated in depth in Gertler (1995). Relationships between observer based diagnostics and parameter estimation were investigated in Magni (1995) and Alcota Garcia et al. (1996).

Nonlinear system fault detection

It should be noted, that most of the results of the available model based fault detection methods are limited to linear systems. To overcome this limitation, many approaches have been proposed. One approach is to linearize the model at an operating point or the current estimated point (Patton et al., 2000). This approach works well only if the linearized model does not have a large mismatch with the nonlinear model.

Another approach is to use a suitable change of coordinates (Witczak, 2007) to bring the original system into a linear (or pseudo-linear) one. For the nonlinear system:

$$\begin{aligned} x_{k+1} &= g(x_k, u_k) \\ y_{k+1} &= h(x_{k+1}) \end{aligned} \qquad (2 - 1)$$

Find the coordinate change (at least locally defined) of the form:

$$\begin{aligned} s &= \phi(x) \\ \bar{y} &= \varphi(y) \end{aligned} \qquad (2 - 2)$$

Such that in the new coordinates, the system is described by:

$$s_{k+1} = A(u_k)s_k + \Psi(y_k, u_k)$$
$$\bar{y}_{k+1} = Cs_{k+1} \qquad (2 - 3)$$

where $\Psi(\cdot)$ is a nonlinear (in general) function. There is no doubt that the observer design problem is significantly simplified by this coordinate change. The main drawback of such an approach is related to strong design conditions that limit its application to a small class of nonlinear systems.

From a mathematical point of view, the precise fault detection and isolation of a nonlinear dynamic system directly is a formidable one (Chen and Patton, 1999). Therefore, research has to restrict the class of nonlinear systems in the study of FDI problems. Among them, nonlinear systems with bilinear dynamics have been studied extensively. Important studies in this class could be found in Yu (1994, 1996), Bennett (1996), Yang (1997), and Mechmeche (1997). The main idea is to treat nonlinear terms as disturbances and decouple their effects from the residual using an unknown input observer.

Krishnaswami (1995 and 1997) and Guernez (1997) extended the parity equation methods to nonlinear cases. The generalized parity equation residual generation procedure may be described as the calculation of the faulty outputs and inputs from some subset of the measurements and then checking the resulting solutions for consistency, as shown in the Fig 2.7. It shows the schematic of the residual generator for a particular combination of faults. The term “non-faulty” means the set of variables that are assumed to be without faults for that particular residual generator and the term “faulty” is the converse. However, this extension is limited because of the strict existence condition of the inverse model of the original nonlinear model.

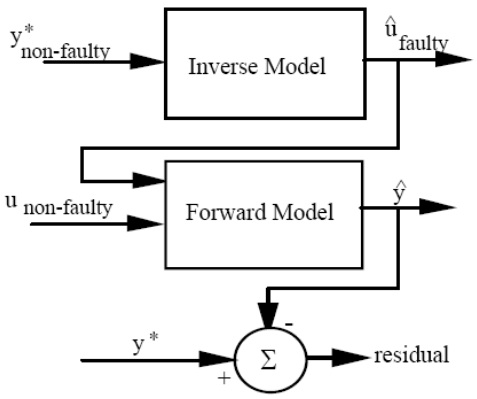

Fig 2.7 Generalized parity equation residual generation

Due to its robustness, the sliding mode observer has been applied to nonlinear FDI problems (Edwards, 1997; Zhang, 2009). Unlike the linear systems, where the parity equation method can be derived from the observer based method, there is no direct link between the two (Chen and Patton, 1999).

2.2 Fault Diagnosis Methodologies

Once a residual has been generated, which indicates there is a faulty condition, a fault diagnosis method needs to be applied to evaluate the residual in order to isolate and identify the exact fault. A classification of the fault diagnosis methodologies (Isermann, 2006) is presented in Fig 2.8. In this section, three typical classification methods and two inference methods will be reviewed, followed by a short summary.

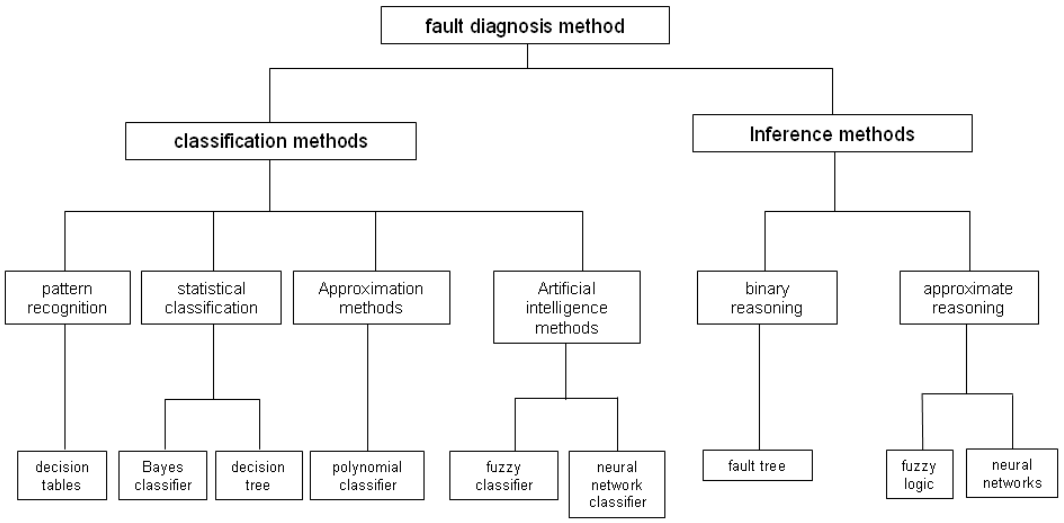

Fig 2.8 Classification of fault diagnosis methods (Isermann, 2006)

2.2.1 Bayes classifier

This approach is based on a reasonable assumption of the statistical distribution of the residuals, usually normal probability density functions. The procedure (Isermann, 2006) is first to determine the maximum likelihood estimations for its parameters. It can be shown that the minimum of wrong decision is achieved if a maximum of $p(F_j \mid s)$ is selected. And this posterior probability can be obtained by the Bayes Law:

$$f_j(s) = p(F_j \mid s) = \frac{p(s \mid F_j)P(F_j)}{p(s)} \qquad (2 - 4)$$

Therefore, the diagnosis decision comes down to, given the residual s, finding the index j that maximize the posterior probability $p(F_j \mid s)$ expressed by the above equation.

The class specific densities $p(s \mid F_j)$ can be estimated from labeled reference data using those data points belonging to the fault F_j. See Fig 2.9 for an example; the statistical distribution of single symptom "s" belongs to two different faults F1 and F2.

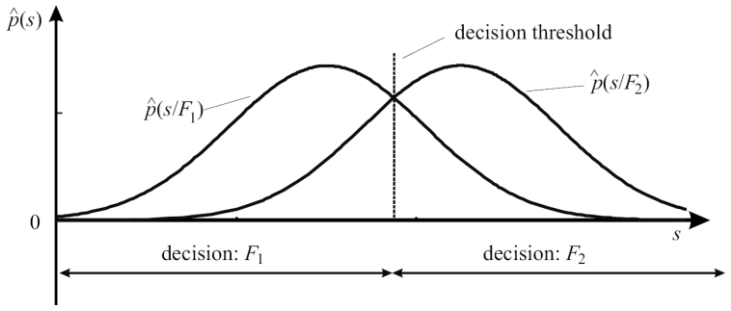

Fig 2.9 Example of two class specific densities (Isermann, 2006)

Since only the maximum of f_j is of interest, the denominator p(s) is not important (serves as a normalizing factor) because it does not depend on F_j. The prior probabilities $P(F_j)$, however, are important. It can be estimated by experimental data or can be assumed by equal probability.

In spite of its naive design and apparently over-simplified assumptions, the Bayes classifier would work well in many complex real-world situations (Zhang, 2004).

2.2.2 Decision tree

This approach basically relies on a series of questions that have to be answered and depending on the answer the next question narrows the classes until the exact fault is determined (Yuan, 1995). The collection of all questions forms the complete decision tree. However, the building of this tree can be tedious.

2.2.3 Polynomial classifier

The polynomial classifier (Isermann, 2006) employs polynomial approximation for the posterior probability of the fault instead of normal distribution assumed for the Bayes classification scheme. The polynomials are

$$p(s \mid F_j) = f_j = a_{j,0} + a_{j,1}s_1 + a_{j,2}s_2 + \cdots + a_{j,n+1}s_1s_2 + \cdots \qquad (2\text{-}5)$$

The coefficients a_j could be determined using reference data points and a least square approach.

Fig 2.10 shows a typical decision boundary on a two-residual plane for a two-fault problem. The boundary is given by the line of equal polynomial values

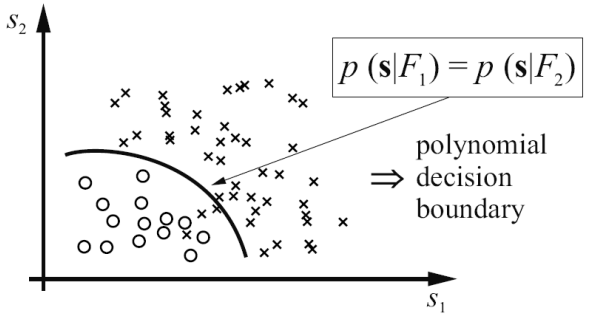

Fig 2.10 Decision boundary of a polynomial classifier (Isermann, 2006)

Generally, the biggest problem of polynomial classifier is to choose the appropriate order of the polynomial and select the correct terms. A complete polynomial would be unnecessarily large in most cases.

2.2.4 Fault tree

Fault tree analysis is the most widely used fault diagnosis method in the industry. It can demonstrate the causal-effect relationship in a physical system (Isermann, 2006), as shown in Fig 2.11 (a).

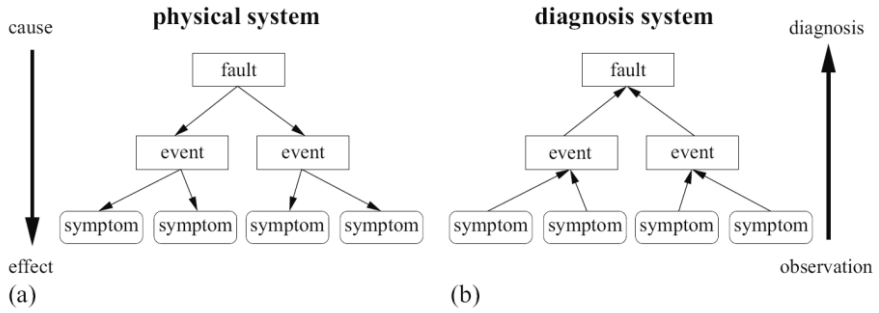

Fig 2.11 Fault tree (Isermann, 2006)

However, most of the underlying physical laws are not known in analytical form, or too complicated for calculation. Therefore, the fault diagnosis system needs to proceed in the reversed way. It has to be concluded from the observed symptoms to the faults, see Fig 2.11(b). Usually, the fault-symptom tree will be summarized to form a fault-symptom table for easy implementation. By analyzing the signs of the residuals, the isolation of different faults can be achieved. It is a simple and transparent diagnosis system.

2.2.5 Fuzzy logic

For rule-based fault diagnosis of continuous processes with continuous variable symptoms, methods of approximate reasoning may be more appropriate than a binary decision (Isermann, 2006). One common approach is fuzzy logic reasoning for fault diagnosis.

2.2.6 Summary

During the fault diagnosis process, two different sources of information can be used: Expert knowledge and measured data from fault experiments (Isermann, 2006). Each of the above mentioned methods mainly makes use of one of the two. This leads to highly transparent systems that are tedious to design or non-transparent data-based classifiers. It is the designer's duty to recognize which kind is more suitable for the application.

2.3 Selective Catalytic Reduction System

The NO_x conversion efficiency of a SCR system is usually above 90% after it reaches normal operating conditions (Hsieh, 2010). The urea-SCR does not require additional fuel for NO_x reduction in regeneration such as a lean NO_x trap (LNT). Therefore, the SCR system has become one of the most promising NO_x reduction aftertreatment devices that have been proven effective.

The SCR injects urea into the upstream of a catalyst chamber as a reductant to convert the NO_x in the exhaust to nitrogen and water vapor, as shown in Fig 2.12. Before entering the SCR system, the exhaust gas usually passes through a filter called the Diesel Particulate Filter (DPF) to trap the PM. Due to the hazardous nature of the pure anhydrous ammonia, the DEF (diesel emission fluid), which is a 32.5% aqueous urea solution, is widely used in the industry for this application. Several chemical reactions will occur inside the SCR catalyst, and is discussed in Chapter 4. Finally, the NO_x in the exhaust will be reduced to elemental nitrogen and water vapor.

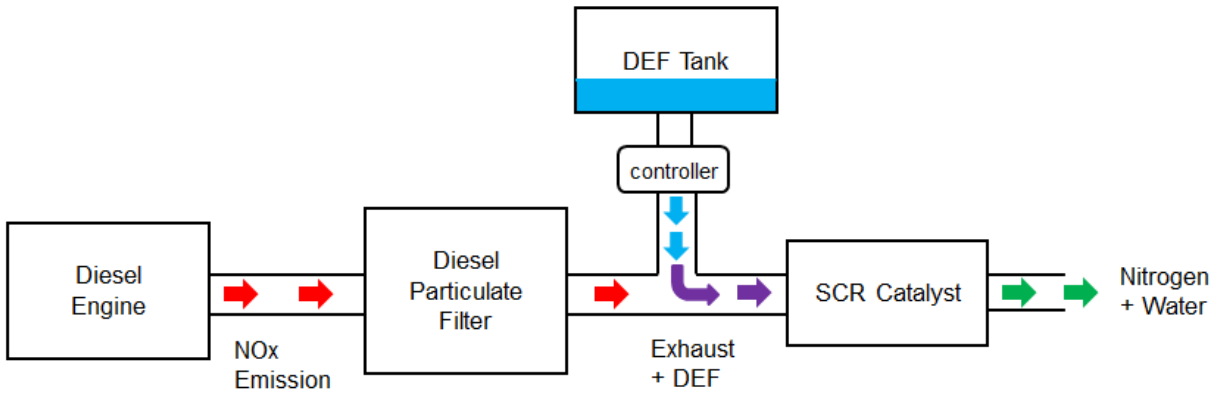

Fig 2.12 SCR in a diesel aftertreatment system

In order to achieve high NO_x conversion efficiency, a sufficient reductant needs to be delivered. On the other hand, excessive dosing of DEF will cause ammonia slip, which is also considered a pollutant, such as EuroVI (Johnson, 2010). These challenges require more sophisticated control methods for the SCR, and will be discussed in the following section.

2.4 SCR System Modeling and Control

SCR models can be classified into research and design models for component design and analysis, distributed models and embedded models for control implementation (Chi, 2005; Devarakonda, 2008a). Among the distributed and embedded models, various control oriented models have been developed recently, which also could be used to facilitate model-based fault diagnosis.

Some important control oriented modeling work includes: Schär, 2003 and 2004; Upadhyay, 2006; Devarakonda, 2008a and 2008b; Guzzella, 2010; Shost, 2008; Hsieh, 2010 and 2011; and Willems, 2011. Most of them are three or four state models, with the exception of Schär's (2004).

The latter assumes gaseous phase contents are faster than ammonia storage dynamics, thereby reducing the model to one state.

All of these control oriented models are “0-D” ODE models, except for Willems’, based on different assumptions on different reactions. A comparison of their assumptions is shown in Table 2.1.

Table 2.1 Overview of SCR control oriented models

affiliation	Standard reaction	Fast reaction	Slow reaction	NH3 oxidation to N2	NH3 oxidation to NO	NO oxidation to NO2	reference
ETH Zurich	●	●		●			Schär, 2003 and 2004; Guzzella, 2010
FORD	●				●		Upadhyay, 2006
MTU & International	● ●	●	●	● ●			Devarakonda, 2008a and 2008b
TNO	●	●	●	●	●		Willems, 2011
OSU	●	●	●	●		●	Hsieh, 2010 and 2011

The SCR system control system has been widely studied due to the increasing demand of performance from the industry (Skaf et al., 2014). Both model-based feed forward control (Krijnsen, 2000; Song, 2002; and Schär, 2004) and feedback control (Upadhyay, 2006; Shost,

2008; Devarakonda, 2008b; Hsieh, 2010; and Stadlbauer, 2014) have been implemented to ensure the NO_x emission does not exceed the EPA's standard. It is believed that closed-loop control is necessary to ensure proper diagnosis of all failure modes (Nebergall, 2005).

Besides improving NO_x reduction efficiency, diesel engine manufacturers also seek to reduce tailpipe ammonia slip. Upadhyay (2006) and Devarakonda (2008b) introduced state feedback for closed loop control of ammonia, without the installation of an additional ammonia sensor. Their state observer is based on the outlet NO_x sensor measurement. Recently, researchers have realized the importance of this additional control objective. An outlet ammonia sensor has been implemented in closed loop control by Shost (2008), Wang (2008), Hsieh (2010), and Willems (2011).

In addition to the tailpipe ammonia sensor and the tailpipe NO_x sensor, Herman (2009) took a further step in using an ammonia sensor at the mid-brick position of the catalyst for a faster and more sensitive closed loop control. The author concluded that the ammonia sensor has potential for both ammonia slip detection as well as catalyst NO_x conversion efficiency degradation for OBD.

Hsieh (2011) designed an observer based on both the tailpipe NO_x sensor and ammonia sensor to estimate mid-SCR ammonia concentration to achieve high NO_x reduction efficiency and low tailpipe ammonia slip. The author also indicates that this mid-SCR ammonia concentration observer and outlet ammonia sensor can be used for ammonia distribution control as well as mid catalyst fault diagnosis purposes.

CHAPTER 3

APPROACHES

As has been stated in the introduction, the diesel engine manufacturing industry is expecting a new OBD regulation for non-road vehicles in the near future, after such an OBD regulation has been enforced in highway applications since 2010. Among many emission monitors in the new OBD regulation, the NO_x conversion efficiency monitoring has become the most challenging one. This is partly because of the new NO_x reduction system, which is the SCR system in most applications, has just been introduced to the non-road mobile application to meet the recently enforced Final Tier 4 emission regulation.

Diesel engine manufacturers are actively seeking monitoring approaches that can detect and isolate faults that lead to reduction in NO_x conversion efficiency, while maintaining the cost of the entire system and complexity at reasonable levels.

After completing the survey of fault detection and diagnosis methodology, a decision was made to tackle the problem using a model based fault detection and isolation (FDI) method.

This research focuses on detecting and isolating the SCR dosing fault and outlet NO_x sensor fault. The main tasks include SCR modeling, residual generators design and simulation validation.

3.1 SCR System Modeling

In order to facilitate the model based FDI method, an SCR system model needs to be developed.

Preliminary research gathered the setup information of the aftertreatment system. Available signals for the SCR diagnosis system include:

1. DPF outlet NO.
2. DPF outlet NO_2.
3. Urea dosing rate command (potential fault).
4. Exhaust gas flow rate.
5. SCR inlet NO_x sensor reading.
6. SCR outlet ammonia sensor reading.
7. AOC outlet NO_x sensor reading (potential fault).
8. SCR inlet and outlet temperature.
9. SCR inlet and outlet pressure.

The SCR dosing control scheme and operating conditions also were investigated. The result reveals potential simplification of the SCR system dynamics.

A control-oriented SCR model is proposed to be utilized for this diagnosis purpose (Hsieh, 2010). After the survey of previous embedded SCR modeling work, a "0-D" lumped parameter nonlinear SCR model was developed according to the application.

3.2 Residual Generators Design

Two model based fault detection and isolation methods for FDI were developed for the SCR input dosing fault and the output NO_x sensor fault. Both faults are modeled as an additive fault.

The first approach uses parity equations to decouple the two faults. In order to apply, the SCR model needs to be linearized first. The residual generator is then designed based on the fault transfer function matrix.

The second approach was an observer based FDI scheme. It generated two equivalent controls using two sliding mode observers. Each equivalent control was sensitive to only one fault and not to the other. So it is possible to achieve fault detection and isolation by checking their magnitudes.

3.3 Simulation Validation

The two residual generator algorithms were implemented in a Matlab/Simulink environment for validation. The SCR model parameters were obtained from publications. A high fidelity 1-D SCR model (Willems and Cloudt, 2011) was implemented to generate system outputs and, therefore, simulate the plant. The same 1-D SCR model was also used in the SCR dosing control scheme on the current ECU. Two faults were added into the simulation. The simulation scheme is shown in Fig 3.1.

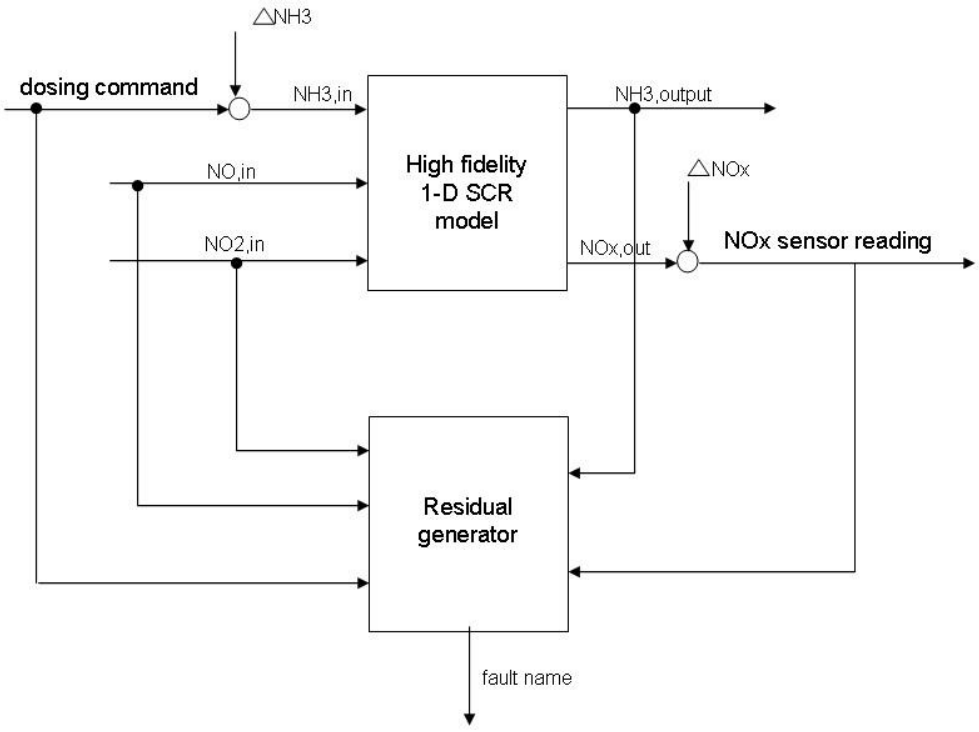

Fig 3.1 FDI algorithm simulation validation

The responses of two residual generators to the same set of faults were compared. A comparison of their applicability in different operating conditions and the capability of isolating different faults were investigated.

CHAPTER 4

SCR SYSTEM MODELING

In this chapter, the targeted diesel engine aftertreatment system was first analyzed to understand the operating conditions for the system and the requirements for the SCR model. Then, a 0-D nonlinear SCR model was developed based on the system mass flow.

4.1 Diesel Engine Aftertreatment System

The targeted Tier 4 diesel engine aftertreatment system consists of a Diesel Oxidation Catalyst (DOC), Diesel Particulate Filter (DPF), urea decomposition tube, SCR catalyst, and Ammonia Oxidation Catalyst (AOC). The Diesel Emission Fluid (DEF) as a urea solution is injected through a nozzle to the exhaust gas upstream of the SCR catalyst. When the exhaust temperature is high enough, the urea droplets will go through decomposition and hydrolysis reactions to form gaseous NH_3. Therefore, this part of the exhaust pipe is also called the decomposition tube, where the urea droplets are mixed with exhaust gas and decomposed.

Two NO_x smart sensors were installed after the DPF and AOC, respectively. The smart NO_x sensor measures NO_x concentration as well as oxygen concentration. Unfortunately, the NO_x sensor also is sensitive to ammonia (Hsieh, 2010). An ammonia sensor was installed between the SCR catalyst and the AOC for diagnostic purpose. The outlet NO_x sensor is not used for closed loop dosing control purposes in this SCR system.

In order to eliminate undesirable ammonia slip released to the environment, the AOC was installed after the SCR. The exhaust flow is fed into the AOC before the outlet NO_x sensor. It is

assumed that all ammonia slip from the SCR is eliminated by the AOC, thus the AOC eliminates the cross sensitivity of the outlet NO_x sensor to ammonia. The system block diagram, including mass flow, available measurements and estimated signals, is shown in Fig 4.1.

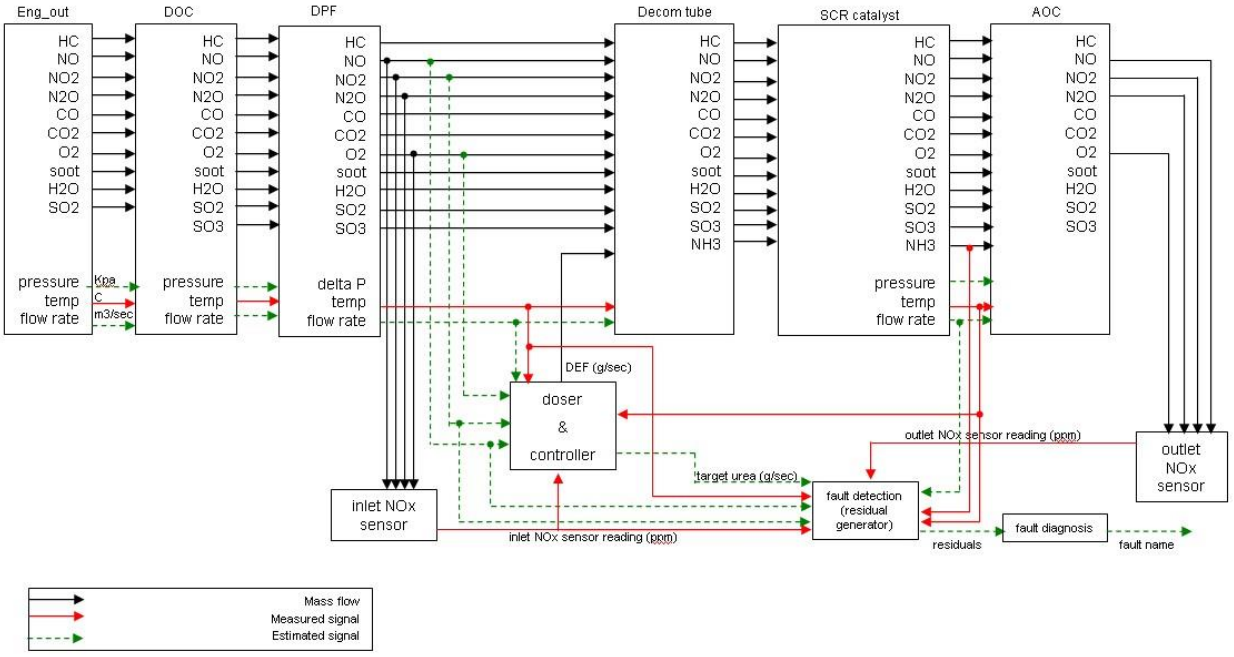

Fig 4.1 Aftertreatment system block diagram

The DPF is installed to trap particulate matter (PM) in the exhaust flow. In order to improve fuel economy and to extend the life of the DPF catalyst, more DPF passive regeneration is encouraged. To achieve this, DOC is added at the upstream of the DPF to convert NO in the exhaust flow into NO_2. Though the total NO_x concentration is not changed, this conversion causes the NO/NO_2 ratio to vary from the engine output NO/NO_2 ratio during different DPF regeneration patterns and soot loading state. This varying NO/NO_2 ratio makes ammonia

dosing control more difficult because the stoichiometric ratios of NO and NO_2 reduction by ammonia in the SCR are different. To overcome this problem, a model based feed forward control system was implemented for the SCR dosing control, as shown in Fig 4.2.

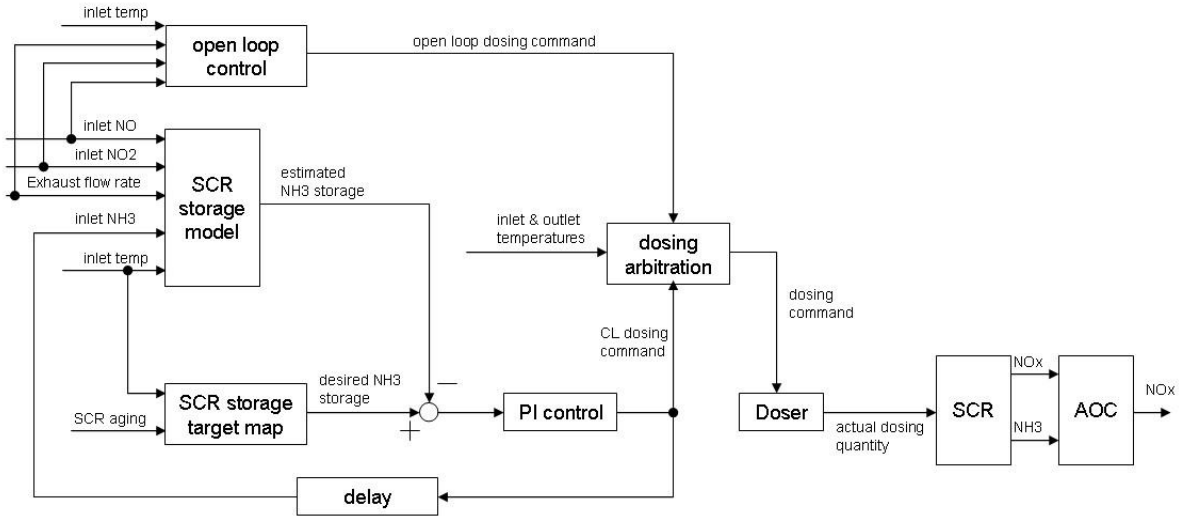

Fig 4.2 SCR dosing control diagram

When the DPF is not in active regeneration, closed loop dosing control is adapted. The closed loop control law first calculates the target storage ratio from the inlet temperature and SCR aging factors. And this target is compared with the storage estimation from an embedded SCR storage model. This model takes the NO and NO_2 as inputs from the existing DPF soot model in the ECU, as well as SCR inlet temperature and exhaust flow rate. According to the fault-free assumption made in Chapter 1.3, the estimation of NO/NO_2 ratio obtained from the DPF soot model is assumed to be accurate. Then the ammonia storage error is fed to a PI controller to generate the DEF dosing command to the doser. This dosing command is then delayed and fed

back to the SCR storage model for ammonia storage estimation. However, the outlet NO_x sensor and ammonia sensor readings are not used for control.

The SCR system operates at a quasi-stationary state condition when the exhaust temperature is between $200^{\circ}C$ and $450^{\circ}C$. During this time, the closed loop dosing control is enabled. The dosing control switches to open loop when the DPF is in the active regeneration mode. The exhaust temperature is typically above $450^{\circ}C$ during DPF active regeneration. This high temperature during active regeneration decreases the SCR ammonia storage capability. However, because the active regeneration of the DPF increases fuel consumption and reduces DPF catalyst life, the passive regeneration of the DPF is the preferred way to clean up soot. The active regeneration of DPF is limited to a minimum possible frequency and duration. Therefore, the SCR is assumed to be operated at a quasi-stationary state condition in this research.

4.2 SCR System Model Development

In this section, a 0-D lumped parameter SCR model was developed. A lumped parameter SCR model was determined to be a good representative of the physical system and is sufficient for model based control purposes (Devarakonda, 2008a). Therefore, it has the potential to facilitate the model based fault diagnosis.

In this model, the SCR catalyst is assumed to be a continuously stirred tank reactor (CSTR). This assumption is valid when the catalyst is not large (Hsieh, 2011). The exhaust flow is assumed to be a homogeneous, incompressible flow of ideal gas. Due to high exhaust flow velocity, axial diffusion and dispersion are neglected. The chemical kinetics in the catalyst is reaction controlled.

The objective of the SCR model is to predict the gaseous NO, NO_2 and NH_3 concentration at the SCR outlet by modeling the reaction kinetics. It takes NO, NO_2 and NH_3 molar flow rates as inputs.

In this model, urea injected to the exhaust flow is assumed to be completely decomposed and hydrolyzed before entering the SCR catalyst, as cited by Piazzesi (2006). Therefore, the gaseous molar input NH_3 to the SCR catalyst is two times the injected urea molar flow, as determined by the decomposition and hydrolysis reactions:

$$\begin{aligned} H_4N_2CO &\rightarrow NH_3 + HNCO \\ HNCO + H_2O &\rightarrow NH_3 + CO_2 \end{aligned} \qquad (4 - 1)$$

As the exhaust gas flows through the SCR catalyst wall supported on a honeycomb substrate, gaseous NH_3 can be absorbed at the surface and reacts with gaseous NO and NO_2. The catalytic reaction product is desorbed and subsequently transported to the bulk flow, and then exits. To capture the NO_x reduction, even during the period of no urea injection, concentrations of both gaseous and surface phase of the participating species were tracked in this model.

The absorption and desorption reactions of ammonia on the catalyst surface are:

$$\begin{aligned} NH_3 + S &\rightarrow NH_3^* \\ NH_3^* &\rightarrow NH_3 + S \end{aligned} \qquad (4 - 2)$$

where S stands for active site available on the catalyst surface, and NH_3^* is the absorbed NH_3 site in the catalyst.

In order to better describe this surface phase of NH_3, the ammonia coverage ratio θ is introduced. It is defined as the ratio of number of stored sites filled with NH_3 to the total number of storage sites in the catalyst.

There are three main selective reductions in the catalyst that occur to reduce NO_x. They are standard, fast and slow reactions respectively:

$$\begin{aligned} &4NH_3 + 4NO + O_2 \rightarrow 4N_2 + 6H_2O \\ &2NH_3 + NO + NO_2 \rightarrow 2N_2 + 3H_2O \\ &8NH_3 + 6NO_2 \rightarrow 7N_2 + 12H_2O \end{aligned} \qquad (4 - 3)$$

The slow reaction is considered here because of the varying NO / NO_2 ratio discussed previously.

In the SCR catalyst, there are also non-selective reactions with oxygen, which is abundant in the diesel exhaust. These reactions consume ammonia and are undesirable:

$$\begin{aligned} &4NH_3 + 3O_2 \rightarrow 2N_2 + 6H_2O \\ &4NH_3 + 5O_2 \rightarrow 4NO + 6H_2O \\ &2NH_3 + 2O_2 \rightarrow N_2O + 3H_2O \end{aligned} \qquad (4 - 4)$$

The first reaction consumes NH_3 unnecessarily, while the second one generates a secondary emission. They are all unwanted reactions, but are not negligible during higher temperature situations. The third oxidation generates secondary emission N_2O, and is not considered in this model.

Any side reactions involving the formation of N_2O and NH_4NO_3 are neglected. The oxygen concentration is in a range where variations do not affect the reaction rates. A sketch of the mass flow within the SCR cell is shown in Fig 4.3.

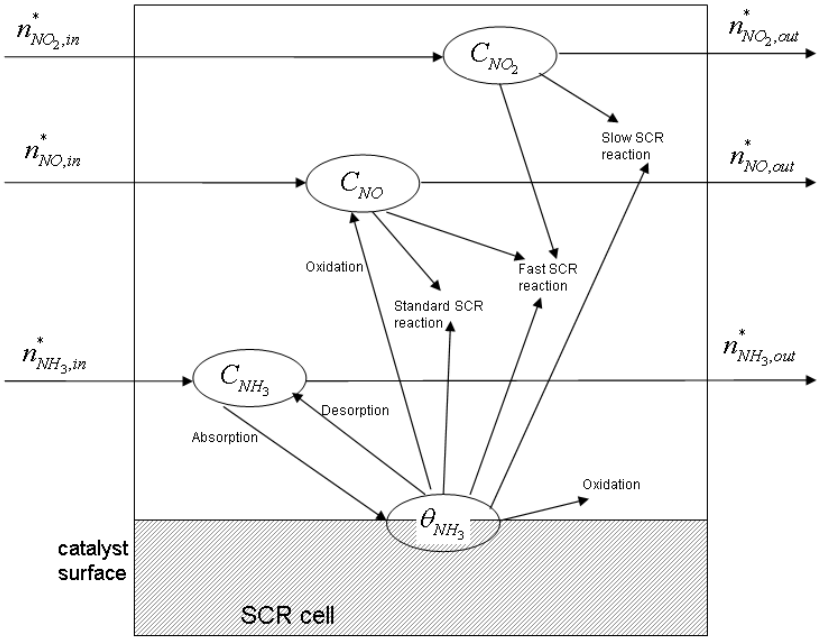

Fig 4.3 Mass flow and reactions within SCR cell

The reaction rates of the above discussed absorption, desorption, the standard/fast/slow SCR reductions and two oxidation reactions, (4 - 2) ~ (4 - 4), can be expressed using the Eley-Rideal mechanism:

$$
\begin{aligned}
r_{Ads} &= a_1 C_{NH_3}(1-\theta) \\
r_{Des} &= a_2 \theta \\
r_{st\_SCR} &= a_3 C_{NO} \theta \\
r_{fa\_SCR} &= a_4 C_{NO} C_{NO_2} \theta \\
r_{sl\_SCR} &= a_5 C_{NO_2} \theta \\
r_{Oxn2} &= a_6 \theta \\
r_{Oxno} &= a_7 \theta
\end{aligned}
\qquad (4-5)
$$

The reaction rate constants $a_1 \sim a_7$ are functions of temperature, and are defined using the Arrhenius equation:

$$
\begin{aligned}
a_1 &= k_{Ads} e^{(\frac{-E_{Ads}}{RT})}, \\
a_2 &= k_{Des} e^{(\frac{-E_{Des}}{RT})}, \\
a_3 &= k_{st} e^{(\frac{-E_{st}}{RT})}, \\
a_4 &= k_{fa} e^{(\frac{-E_{fa}}{RT})}, \\
a_5 &= k_{sl} e^{(\frac{-E_{sl}}{RT})}, \\
a_6 &= k_{Oxn2} e^{(\frac{-E_{Oxn2}}{RT})}, \\
a_7 &= k_{Oxno} e^{(\frac{-E_{Oxno}}{RT})}
\end{aligned}
\qquad (4-6)
$$

Where k_x is the reaction rate constant, E_x is the activation energy, and R is the universal gas constant.

According to Schär et al. (2004), energy released by the chemical reactions in the SCR catalyst is small and can be ignored. Therefore, the temperature model is decoupled from the mass flow part of the SCR model.

Since the targeted SCR convertor is double insulated, the radiation energy released to the atmosphere is also ignored. The SCR experiment data showed that the difference between the SCR inlet and outlet temperature sensor readings was always within $10^{\circ}C$ when the DPF upstream was not in active regeneration, which justified this assumption. Due to the limited volume of the catalyst convertor, the thermal inertia of the gas contained in the catalyst was also ignored. Therefore, in this model, the exhaust gas temperature in the SCR catalyst was assumed to be uniform within the SCR catalyst. It was averaged from the measured SCR inlet and outlet temperature:

$$T = \frac{T_{in} + T_{out}}{2} \qquad (4 - 7)$$

According to Fig 4.3, the dynamics of the ammonia coverage ratio θ, gaseous NO, NO_2 and NH_3 concentration, can be determined:

$$
\begin{aligned}
\dot{\theta} &= r_{Ads} - r_{Des} - r_{st\_SCR} - 2r_{fa\_SCR} - \frac{4}{3} r_{sl\_SCR} - r_{Oxn2} - r_{Oxno} \\
&= a_1 C_{NH_3}(1-\theta) - a_2\theta - a_3 C_{NO}\theta - 2a_4 C_{NO} C_{NO_2}\theta - \frac{4}{3} a_5 C_{NO_2}\theta - a_6\theta - a_7\theta \\
\dot{C}_{NH_3} &= a_8 n_{NH_3,in} - a_8 Q\, C_{NH_3} - a_1 C_{NH_3}(1-\theta) + a_2\theta \\
\dot{C}_{NO} &= a_8 n_{NO,in} - a_8 Q\, C_{NO} - a_3 C_{NO}\theta - a_4 C_{NO} C_{NO_2}\theta + a_7\theta \\
\dot{C}_{NO_2} &= a_8 n_{NO_2,in} - a_8 Q\, C_{NO_2} - a_4 C_{NO} C_{NO_2}\theta - a_5 C_{NO_2}\theta
\end{aligned}
\qquad (4 - 8)
$$

where n is the molar flow rate of each species. Constant $a_8 = \frac{1}{V_c}$, where V_c is the volume of the catalyst convertor. The exhaust volumetric flow rate Q is treated as a parameter, according to the quasi-steady state operating condition assumption.

According to the assumption made about the urea decomposition, the input ammonia molar flow rate $n_{NH_3,in}$ can be calculated directly from DEF dosing rate. The input NO and NO_2 molar flow rates are available from a DPF model running on the current engine ECU.

In summary, this is a four states, lumped parameter nonlinear model of the SCR catalyst:

$$
\begin{aligned}
\dot{x}_1 &= -2a_4 x_3 x_4 x_1 - (a_1 x_2 + a_3 x_3 + \frac{4}{3} a_5 x_4) x_1 - (a_2 + a_6 + a_7) x_1 + a_1 x_2 \\
\dot{x}_2 &= a_1 x_1 x_2 + a_2 x_1 - (a_1 + a_8 Q) x_2 + a_8 u_1 \\
\dot{x}_3 &= -a_4 x_1 x_4 x_3 - a_3 x_1 x_3 - a_8 Q x_3 + a_7 x_1 + a_8 u_2 \\
\dot{x}_4 &= -a_4 x_1 x_3 x_4 - a_5 x_1 x_4 - a_8 Q x_4 + a_8 u_3
\end{aligned}
\quad , \qquad (4 - 9)
$$

Where

$$\vec{x} = \begin{bmatrix} \theta \\ C_{NH_3} \\ C_{NO} \\ C_{NO_2} \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} n_{NH_3,in} \\ n_{NO,in} \\ n_{NO_2,in} \end{bmatrix}$$

Compared with the models listed in Table 2.1, this model contained all the reactions that were considered essential in the targeted diesel aftertreatment system. Meanwhile, this model is still an ODE model, and therefore, can be easily implemented in an automotive control unit for diagnosis purpose.

CHAPTER 5

RESIDUAL GENERATOR DESIGN

In this chapter, two types of residual generators were designed independently. The first one was a parity equation based residual generator based on a linearized SCR system model, while the second one was an observer based residual generator based on a nonlinear SCR system model.

5.1 Parity Equation Residual Generation

5.1.1 SCR system linearization

In order to facilitate the parity equation fault detection method for this nonlinear system, the first step was to linearize the nonlinear model (4-9). When the SCR system is operating in a quasi-steady state, a linearization can be made around the operating point "op". The linearized system takes the form of the following linear state space representation:

$$\begin{aligned} \dot{x} &= Ax + Bu,\ x \in \Re^4,\ u \in \Re^3 \\ y &= Cx,\ y \in \Re^2 \end{aligned} \tag{5 - 1}$$

where

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & 0 & 0 \\ A_{31} & 0 & A_{33} & 0 \\ A_{41} & 0 & 0 & A_{44} \end{bmatrix},\ B = a_8 \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

and

$$A_{11}=\left[-2a_4x_3x_4-a_1x_2-a_3x_3-\frac{4}{3}a_5x_4-(a_2+a_6+a_7)\right]_{op}$$
$$A_{12}=\left[a_1(1-x_1)\right]_{op}$$
$$A_{13}=\left[-a_3x_1-2a_4x_1x_4\right]_{op}$$
$$A_{14}=\left[-\frac{4}{3}a_5x_1-2a_4x_1x_3\right]_{op}$$
$$A_{21}=\left[a_1x_2+a_2\right]_{op}$$
$$A_{22}=\left[a_1x_1-a_1-a_8Q\right]_{op}$$
$$A_{31}=\left[-a_4x_3x_4-a_3x_3+a_7\right]_{op}$$
$$A_{33}=\left[-a_4x_1x_4-a_3x_1-a_8Q\right]_{op}$$
$$A_{41}=\left[-a_4x_3x_4-a_5x_4\right]_{op}$$
$$A_{44}=\left[-a_4x_1x_3-a_5x_1-a_8Q\right]_{op}$$

The observability matrix of the linearized system (5–1) is:

$$O=\begin{bmatrix}C\\CA\end{bmatrix}=\begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 1\\ A_{21} & A_{22} & 0 & 0\\ A_{31}+A_{41} & 0 & A_{33} & A_{44}\end{bmatrix} \qquad (5-2)$$

The observability matrix (5-2) will lose its full rank (n=4) under certain conditions, which are listed in Table 5.1:

Table 5.1 Loss of rank conditions and interpretations

	condition	physical interpretation
1	$A_{21}=0$	$x_1=0$, the catalyst is empty
2	$A_{33}=A_{44}$	$x_1=0$, the catalyst is empty Or $x_3-x_4=\frac{a_3-a_5}{a_4}$

The first conclusion from the above analysis is that, the linearized system (5-1) is not observable when the ammonia storage in the catalyst is empty, which rarely happens in the real world operation.

According to the system model calibration data, for the same temperature, the reaction rate constant of fast reduction is considerably faster than that of the standard and slow reductions, while the standard and slow reduction rates are close to each other. This indicates that when NO and NO_2 concentrations are close to each other, the system states may not be observable based on the system outputs.

Except for the above two operating conditions, it is possible to estimate the current system operating points in real-time by constructing a state observer. Therefore, the linearized model can be implemented in a broader range of operating conditions.

5.1.2 Parity equation based residual generator design

The parity equation based fault detection scheme is shown in Fig 5.1. ΔNH_3 and ΔNO_x are input and output fault respectively. The linearized model takes the dosing command as well as NO and NO_2 inputs to generate NH_3 and NO_x model predictions. These predictions are then fed into a residual generation module along with NH_3 and NO_x sensor measurements to detect and isolate the fault.

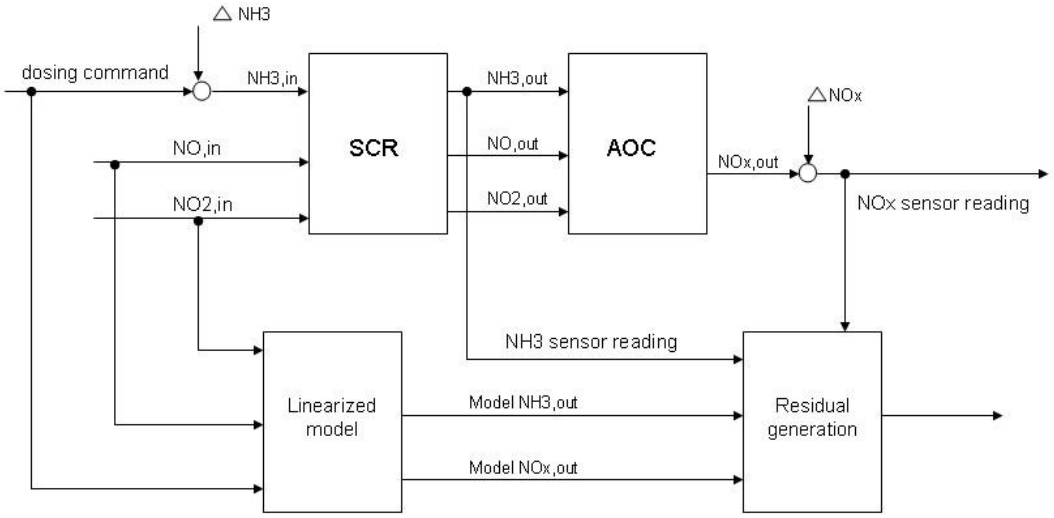

Fig 5.1 Parity equation based FDI scheme

With the presence of input fault (dosing fault p_1) and output fault (NO_x sensor fault p_2), the linearized SCR model (5-1) becomes:

$$\begin{aligned} \dot{x} &= Ax + Bu + E_f p \\ y &= Cx + F_f p \end{aligned} \qquad (5 - 3)$$

where $p = (p_1, p_2)^T$ is the fault vector.

The model (5-3) can also be expressed in terms of transfer functions in the frequency domain. The transfer function matrix can be obtained by:

$$M(s) = C\left[sI - A\right]^{-1} B = \begin{bmatrix} m_{11}(s) & m_{12}(s) & m_{13}(s) \\ m_{21}(s) & m_{22}(s) & m_{23}(s) \end{bmatrix} \tag{5 - 4}$$

With the presence of the two faults, the system output becomes:

$$\begin{aligned} Y(s) &= \begin{bmatrix} m_{11}(s) & m_{12}(s) & m_{13}(s) \\ m_{21}(s) & m_{22}(s) & m_{23}(s) \end{bmatrix} \begin{bmatrix} u_1(s) + p_1(s) \\ u_2(s) \\ u_3(s) \end{bmatrix} + \begin{bmatrix} 0 \\ p_2(s) \end{bmatrix} \\ &= \begin{bmatrix} m_{11}(s) & m_{12}(s) & m_{13}(s) \\ m_{21}(s) & m_{22}(s) & m_{23}(s) \end{bmatrix} \begin{bmatrix} u_1(s) \\ u_2(s) \\ u_3(s) \end{bmatrix} + \begin{bmatrix} m_{11}(s) & 0 \\ m_{21}(s) & 1 \end{bmatrix} \begin{bmatrix} p_1(s) \\ p_2(s) \end{bmatrix} \\ &= M(s)U(s) + S_f(s)p(s) \end{aligned} \tag{5 - 5}$$

where $S_f(s)$ is the fault transfer function matrix, and $p(s)$ is the fault vector in the frequency domain.

Therefore, the parity equation residuals are given by:

$$\begin{aligned} R(s) &= \begin{bmatrix} r_1(s) \\ r_2(s) \end{bmatrix} = W(s)[Y(s) - M(s)U(s)] = W(s)S_f(s)p(s) \\ &= \begin{bmatrix} w_{11}(s) & w_{12}(s) \\ w_{21}(s) & w_{22}(s) \end{bmatrix} \begin{bmatrix} m_{11}(s) & 0 \\ m_{21}(s) & 1 \end{bmatrix} \begin{bmatrix} p_1(s) \\ p_2(s) \end{bmatrix} \end{aligned} \tag{5 - 6}$$

where $Y(s)$ is the measurement, $M(s)U(s)$ is the linearized model prediction, and $W(s)$ is a design matrix.

The objective is to design the w_{ij} (i, j=1, 2) such that each residual responds to only one fault and not the other, thereby decoupling the two faults. Specifically, in order to achieve strong isolation of the two faults $W(s)$ needs to be designed as:

$$\begin{cases} w_{21}(s)m_{11}(s) + w_{22}(s)m_{21}(s) = 0 \\ w_{12}(s) = 0 \\ w_{11}(s)m_{11}(s) \neq 0 \\ w_{22}(s) \neq 0 \end{cases} \tag{5 - 7}$$

Further tuning of $W(s)$ may be needed to suppress the residual response to the change of volumetric flow rate Q, which is treated as a varying parameter in the SCR model.

If the fault signature is defined as:

$$s_1 = \begin{cases} 1 & \text{if } |r_1| > th_1 \\ 0 & \text{if } |r_1| \leq th_1 \end{cases} \qquad s_2 = \begin{cases} 1 & \text{if } |r_2| > th_2 \\ 0 & \text{if } |r_2| \leq th_2 \end{cases} \tag{5 - 8}$$

Then the fault diagnosis problem can be reduced to the following fault-symptom table:

Table 5.2 Fault-symptom table for parity equation FDI scheme

	S_1	S_2
No fault	0	0
Dosing faults	1	0
NO_x sensor fault	0	1
Both fault	1	1

5.2 Observer Based Residual Generation

The second FDI approach was based on an unknown input observer (UIO) and dedicated observer scheme. It utilizes two sliding mode observers to generate two equivalent controls for the dosing input, as shown in Fig 5.2.

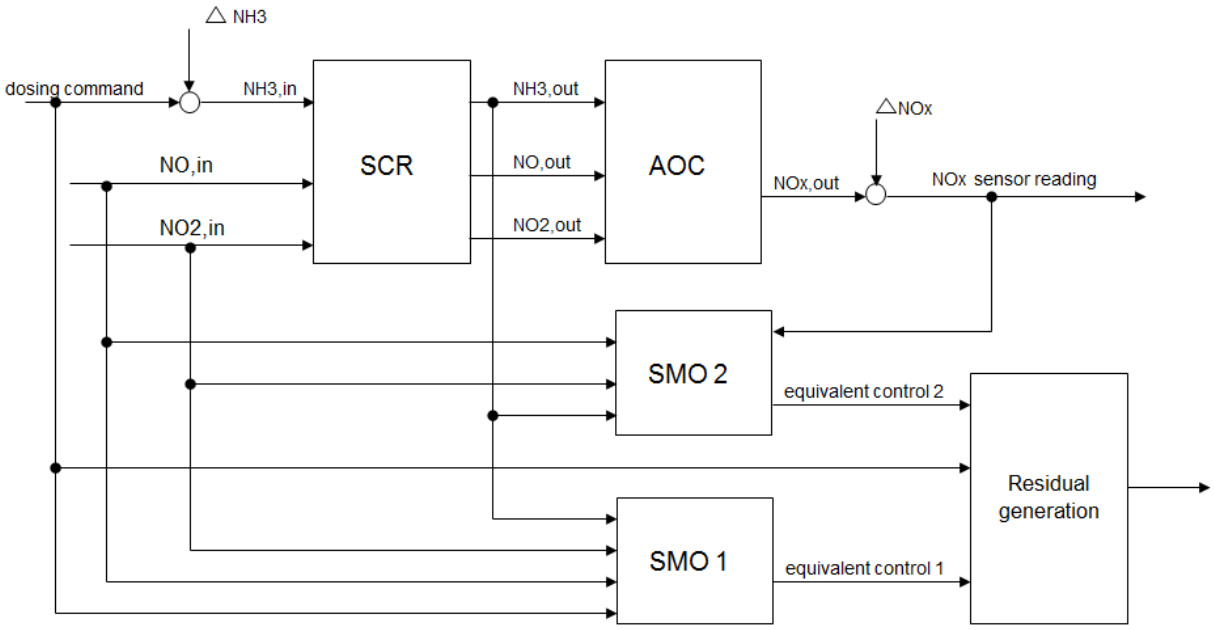

Fig 5.2 Observer based FDI scheme

The first sliding mode observer took only the fault information of dosing input and was robust to the output NO_x sensor fault. The second sliding mode observer took the fault information of both the dosing input and the output NO_x sensor. Therefore, it is possible to separate the two faults by analyzing the magnitudes of the two equivalent controls.

In control theory, sliding mode control is a type of variable structure control (VSC) where the dynamics of a nonlinear system are altered via application of a high-frequency switching control. This is a state feedback control scheme where the feedback is not a continuous function of time. The gains in each feedback path switch between two values according to the switching rule. The purpose of the switching control law is to drive the state trajectory of the nonlinear plant onto a specified surface in the state space called sliding surface, as shown in Fig 5.3. Because of its

robustness and capability of nonlinear system applications, the sliding mode control has been widely adapted (Utkin et al., 1999) since its development in the 1960s.

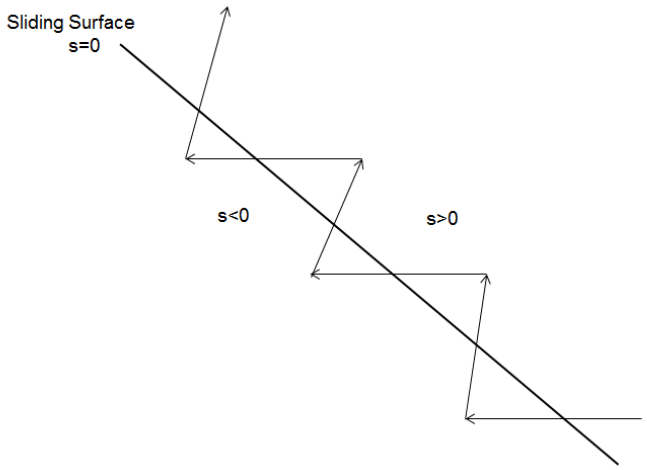

Fig 5.3 Controlled state on the sliding surface

The sliding mode approach also can be used to design asymptotic observers for estimating system state using the measured components. The sliding mode observer has robustness with respect to nonlinearity and bounded model uncertainty. This is a great advantage when dealing with a nonlinear system with an unknown input. Unlike the parity equation method, the sliding mode observer based method can be applied directly in nonlinear systems, which may provide more potential for more accurate fault detection and isolation.

5.2.1 Design of Sliding Mode Observer 1 (SMO 1)

According to the nonlinear SCR system state space equation (4-9), the ammonia concentration C_{NH_3} dynamics is:

$$\dot{x}_2 = a_1 x_1 x_2 + a_2 x_1 - (a_1 + a_8 Q) x_2 + a_8 u_1 \quad (5 - 9)$$

The SMO 1 only takes ammonia output sensor readings as feedback. With the dosing input fault, the dynamics (5-9) becomes:

$$\dot{x}_2 = a_1 x_1 x_2 + a_2 x_1 - (a_1 + a_8 Q) x_2 + a_8 (u_1 + \Delta u) \quad (5 - 10)$$

Consider the sliding mode observer:

$$\dot{\hat{x}}_2 = a_1 x_1 \hat{x}_2 + a_2 x_1 - (a_1 + a_8 Q) \hat{x}_2 + a_8 u_1 + \eta_1 \quad (5 - 11)$$

where η_1 is the equivalent control, and is defined by:

$$\eta_1 = K_1 sign(e_2) = K_1 sign(x_2 - \hat{x}_2) \quad (5 - 12)$$

Comparing the system (5-10) and the observer (5-11), the error dynamics is:

$$\dot{e}_2 = \dot{x}_2 - \dot{\hat{x}}_2 = \left[(x_1 - 1) a_1 - a_8 Q \right] e_2 + a_8 \Delta u - \eta_1 \quad (5 - 13)$$

Because the ammonia coverage ratio $x_1 \leq 1$, when the gain K_1 is large enough, the error will converge to $e_2 = 0$ after transient. The equivalent control η_1 can be obtained from (5-13) as:

$$\eta_1 = a_8 \Delta u \quad (5 - 14)$$

It can be seen from (5-14), the equivalent control of SMO 1 η_1 is proportional to the input dosing fault Δu.

5.2.2 Design of Sliding Mode Observer 2 (SMO 2)

The sliding mode observer 2 takes input from the outlet NO_x sensor output as well as the dosing command to estimate the system state variable, as was shown in Fig 5.2.

First, the original nonlinear system model (4-9) needs to be revisited. It was suggested that the time constants of the storage of mass in the gas phase are about two orders of magnitude smaller than that of the NH_3 storage on the catalyst surface (Schär et al., 2004). Therefore, the dynamic elements that determine the NH_3, NO and NO_2 concentrations in the nonlinear model (4-8) can be replaced by static elements:

$$\begin{aligned}
&\dot{\theta} = a_1 C_{NH_3}(1-\theta) - a_2\theta - a_3 C_{NO}\theta - 2a_4 C_{NO} C_{NO_2}\theta - \frac{4}{3}a_5 C_{NO_2}\theta - a_6\theta - a_7\theta \\
&a_1 C_{NH_3}(1-\theta) - a_2\theta = a_8 n_{NH_3,in} - a_8 Q\ C_{NH_3} = a_8(n_{NH_3,in} - n_{NH_3,out}) \\
&a_3 C_{NO}\theta + a_4 C_{NO} C_{NO_2}\theta - a_7\theta = a_8 n_{NO,in} - a_8 Q\ C_{NO} = a_8(n_{NO,in} - n_{NO,out}) \\
&a_4 C_{NO} C_{NO_2}\theta + a_5 C_{NO_2}\theta = a_8 n_{NO_2,in} - a_8 Q\ C_{NO_2} = a_8(n_{NO_2,in} - n_{NO_2,out})
\end{aligned} \qquad (5 - 15)$$

According to the calibration data, the reaction rate constant of a slow SCR reaction a_5 is at least one order of magnitude smaller than those of other reactions. Therefore, the slow SCR reaction terms can be dropped. Then, substitute the right hand side of the ammonia coverage state equation in (5-15) with the rest of three:

$$\dot{\theta} = a_8 \left[n_{NH_3,in} - n_{NO_x,in} - n_{NH_3,out} + n_{NO_x,out} \right] - (a_6 + 2a_7)\theta \qquad (5 - 16)$$

Therefore, the nonlinear system model (4-9) is reduced to:

$$\begin{aligned} \dot{x}_1 &= a_8(u_1 - u_2) - (y_1 - y_2) - (a_6 + 2a_7)x_1 \\ x_2 &= \frac{a_2 x_1 + a_8 u_1}{a_1(1 - x_1) + a_8 Q} \\ x_3 &= \frac{a_7 x_1 + a_8 u_2}{a_4 x_1 x_4 + a_3 x_1 + a_8 Q} \\ x_4 &= \frac{a_8 u_3}{a_4 x_1 x_3 + a_8 Q} \end{aligned} \qquad (5 - 17)$$

With dosing input fault and the outlet NO_x sensor fault present, the dynamics in (5-17) becomes:

$$\dot{x}_1 = a_8(u_1 + \Delta u - u_2) - (y_1 - y_2 - \Delta y) - (a_6 + 2a_7)x_1 \qquad (5 - 18)$$

Similar to the design of SMO 1, consider the observer:

$$\dot{\hat{x}}_1 = a_8(u_1 - u_2) - (y_1 - y_2) - (a_6 + 2a_7)\hat{x}_1 + \eta_2 \qquad (5 - 19)$$

where η_2 is the equivalent control defined by:

$$\eta_2 = K_2 sign(e_1) = K_2 sign(x_1 - \hat{x}_1) \qquad (5 - 20)$$

Comparing the system (5-18) and the observer (5-19), the error dynamics is:

$$\dot{e}_1 = \dot{x}_1 - \dot{\hat{x}}_1 = -(a_6 + 2a_7)e_1 + a_8\Delta u + \Delta y - \eta_2 \qquad (5 - 21)$$

Similarly, because $-(a_6 + 2a_7) < 0$, when the gain K_2 is large enough, the error will converge to $e_1 = 0$ after transient. The equivalent control η_2 can be obtained from (5-21) as:

$$\eta_2 = a_8\Delta u + \Delta y \qquad (5 - 22)$$

It can be seen from (5-22) the equivalent control of SMO 2 η_2 has two terms. The first one is proportional to the input dosing fault Δu, while the second one has the size of the outlet NO_x sensor fault. To decouple these two types of faults, consider the following transformation:

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \bar{T}\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \Delta u \\ \Delta y \end{bmatrix} \qquad (5 - 23)$$

where $\bar{T} = \begin{bmatrix} \frac{1}{a_8} & 0 \\ -1 & 1 \end{bmatrix}$. Then, each residual will only react to one fault and not the other, and therefore, achieve the fault isolation.

CHAPTER 6

SIMULATION VALIDATION

In this chapter, the residual generators and the SCR model developed in the previous chapters were implemented in simulation for validation. The simulation was run in different operating conditions and with different faults presented. The simulation results show that both residual generators have the capability to detect and isolate both faults, while both have their own limitations.

6.1 Simulation Environment

As shown in Fig 6.1, the simulation validation of the proposed residual generators was implemented in the Matlab/Simulink environment.

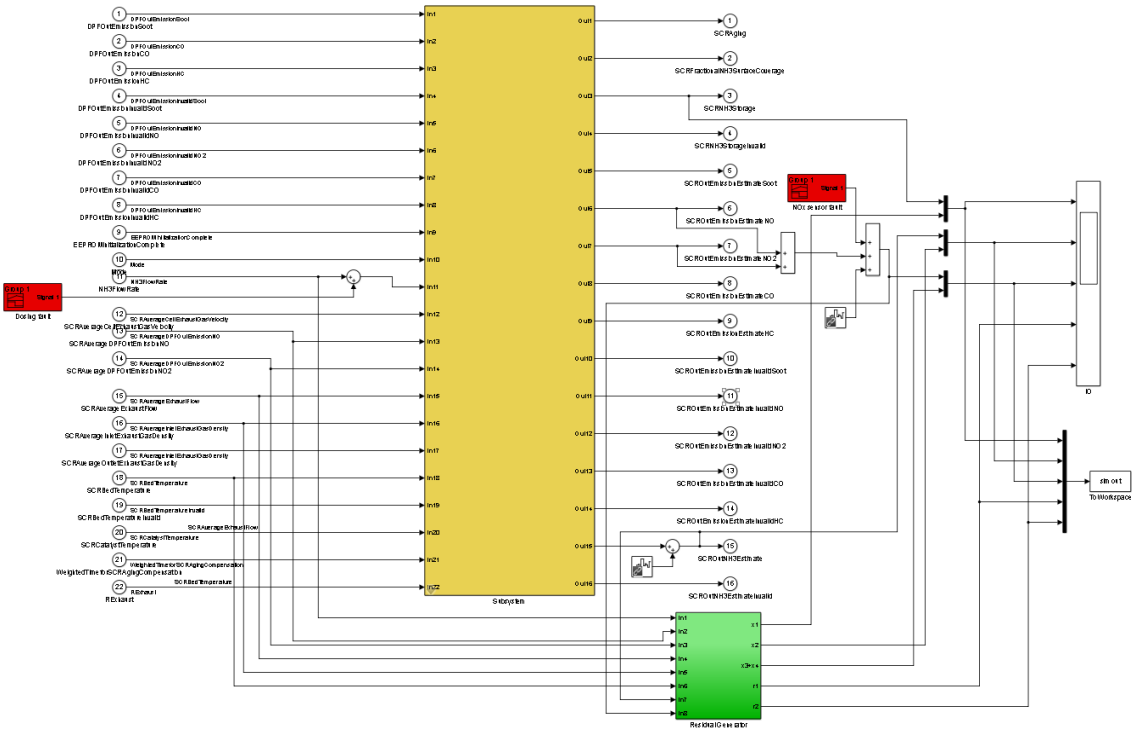

Fig 6.1 Residual generator simulation validation

In order to simulate the plant response, a 1-D SCR model (Willems and Cloudt, 2011) was implemented in the same environment. The SCR catalyst was divided into 20 idealized longitudinal segments. All variables were assumed to be homogenous within each segment, and were calculated using the same 0-D SCR catalyst model developed in Chapter 4.

The two faults, dosing fault and outlet NO_x sensor fault, were added to the plant input and output respectively. The proposed residual generator takes the same plant inputs and outputs except for the faulty dosing rate, as well as the faulty outlet NO_x sensor reading. Two white noise sources were added to the outputs to simulate measurement noise.

In order to validate the sliding mode observer based residual generator, the simplified nonlinear SCR model (5-17) was implemented in the SIMULINK environment, as shown in Fig 6.2. The reaction rates were calculated individually. The temperature and volumetric flow rate were taken from the drive cycle. The molar flow rates were then calculated using the Ideal Gas Law.

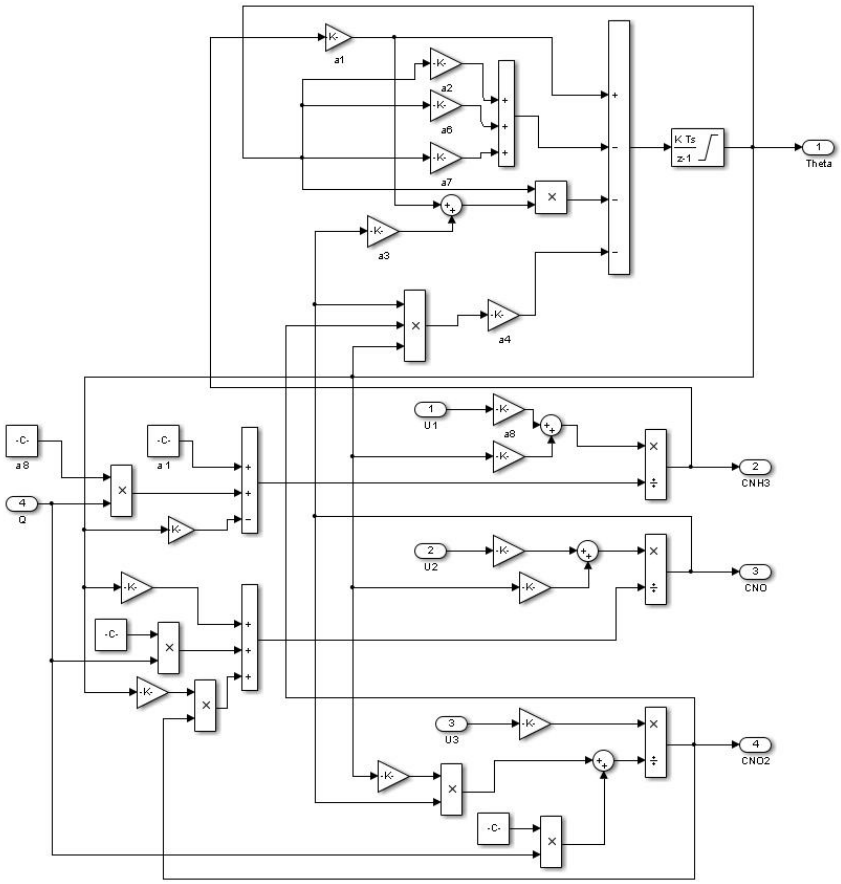

Fig 6.2 Nonlinear SCR model implemented in SIMULINK

The SCR model parameters used for the simulation validation were derived from the regression analysis based on the bench testing data for the targeted SCR system. They are listed in the Table 6.1.

Table 6.1 SCR model parameter

Name	Symbol	Value	Unit
SCR catalyst diameter		0.325	m
SCR catalyst length		0.38	m
SCR catalyst void fraction		0.6889	N/A
SCR NH_3 adsorption capacity		130.216	$mol \cdot m^{-3}$
Universal gas constant	R	8.314	$J \cdot mol^{-1} \cdot K^{-1}$
Absorption reaction rate constant	k_{Ads}	3.8723	$m^3 \cdot mol^{-1} \cdot s^{-1}$
Desorption reaction rate constant	k_{Des}	0.0080344	s^{-1}
Fast reaction rate constant	k_{fa}	11110	$m^6 \cdot mol^{-2} \cdot s^{-1}$
Slow reaction rate constant	k_{sl}	19.5685	$m^3 \cdot mol^{-1} \cdot s^{-1}$
Standard reaction rate constant	k_{st}	11.3417	$m^3 \cdot mol^{-1} \cdot s^{-1}$
Oxidation to N_2 reaction rate constant	k_{Oxn2}	0.000004311	s^{-1}
Oxidation to NO reaction rate constant	k_{Oxno}	0.00000011	s^{-1}
Absorption activation energy	E_{Ads}	35.03	$J \cdot mol^{-1}$
Desorption activation energy	E_{Des}	15255	$J \cdot mol^{-1}$
Fast activation energy	E_{fa}	10020.97	$J \cdot mol^{-1}$
Slow activation energy	E_{sl}	59182.3	$J \cdot mol^{-1}$

Table 6.1 (cont.)

Name	Symbol	Value	Unit
Standard activation energy	E_{st}	60584	$J \cdot mol^{-1}$
Oxidation to N_2 activation energy	E_{Oxn2}	225590	$J \cdot mol^{-1}$
Oxidation to NO activation energy	E_{Oxno}	199355	$J \cdot mol^{-1}$

6.2 Validation Results

In this section, the two residual generators were used to detect and isolate the same input and output faults. In all cases, the SCR was initially operating in a steady state condition. The dosing fault was added as a step decrease in dosing quantity by 1/3, and then gradually returned to the set point dosing quantity after the outputs achieved steady state. The outlet NO_x sensor fault was achieved by a step increase in the outlet NO_x sensor reading by 200 ppm, and then gradually returned to the true reading.

6.2.1 Parity equation residual generator validation

The simulation result of under-dosing fault detection using parity equation based approach is shown in Fig 6.3. After reaching the steady state of the SCR, the dosing quantity (in NH3 molar flow rate) was reduced by 1/3. As a result, the outlet NO_x sensor reading increased after a time lag. This was because the ammonia storage reaction has a large time constant, and the SCR catalyst ammonia storage was full at the beginning of the fault. The first residual generated by the parity equation approach responded to the change right after the outlet NO_x sensor response,

while the second residual remained unchanged. The result shows that the PE based residual generator is capable of detecting and isolating the dosing fault.

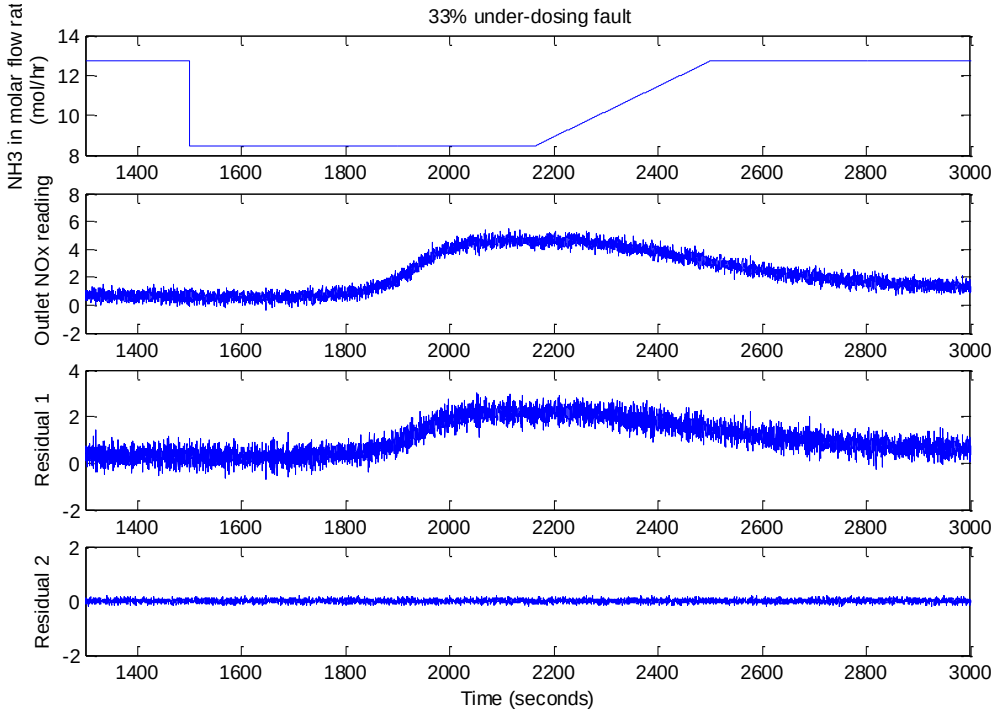

Fig 6.3 PE based residuals in dosing fault

The simulation result of the outlet NO_x sensor using a parity equation based approach is shown in Fig 6.4. The outlet NO_x sensor reading was manually increased by 200 ppm at 1500 s. The rise of the outlet NO_x sensor reading immediately triggered the second residual, while the first residual remained unchanged. The result shows that the PE based residual generator is capable of detecting and isolating the outlet NO_x sensor fault.

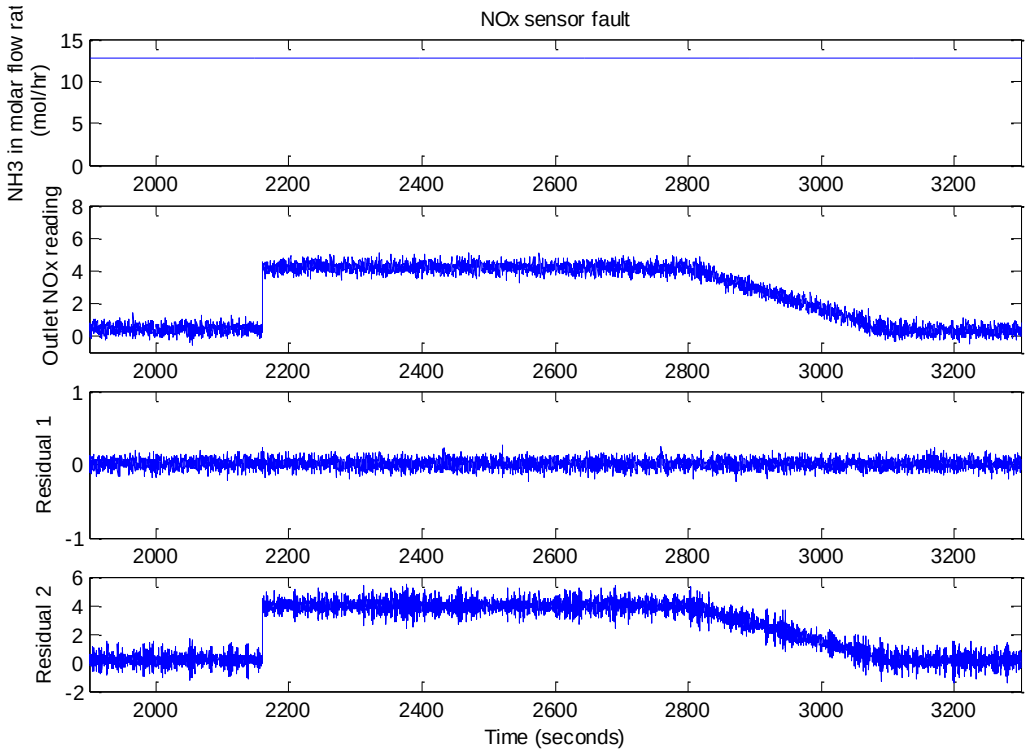

Fig 6.4 PE based residuals with outlet NO_x sensor fault

Assume that κ_i is the threshold for residual r_i. The response of the same residual to a fault $p_j(t)$, with no other fault or nuisance input present, will be

$$r_i(t \mid p_j) = w_{i\cdot}^{'}(\phi) s_{F\cdot j}(\phi) p_j(t) \qquad (6 - 1)$$

where $w_{i\cdot}^{'}(\phi)$ is the i-th row of design matrix $W(\phi)$ and $s_{F\cdot j}(\phi)$ is the j-th column of the fault transfer function $S_F(\phi)$.

Then the residual sensitivity may be characterized by the ratio of the nominal-fault (p_j^0) residual response to the threshold, assuming steady state gain ($\phi = 1$) of the fault-to-residual transfer function. The ratio is

$$\xi_{ij} = \frac{p_j^0 \left[w_{i\cdot}^{'}(\phi) s_{F\cdot j}(\phi) \right]_{\phi=1}}{\kappa_i} \qquad (6 - 2)$$

The thresholds are set such that the ratios in (6-2) are slightly above one for all faults (Gertler, 1998).

However, the parity equation method has its draw backs. When the SCR actual operating point deviates further from the linearized point of the SCR model, the coefficient matrices in (5-1) of the linearized model may change significantly. As a result, the performance of the residual generator will be degraded due to modeling error. This limits the application of the parity equation based residual generator to smaller faults detection and diagnosis. As shown in Fig 6.5, when the dosing fault reaches more than 85% and the SCR catalyst is near empty, the outputs of the linearized model deviates far from the 1-D SCR model outputs.

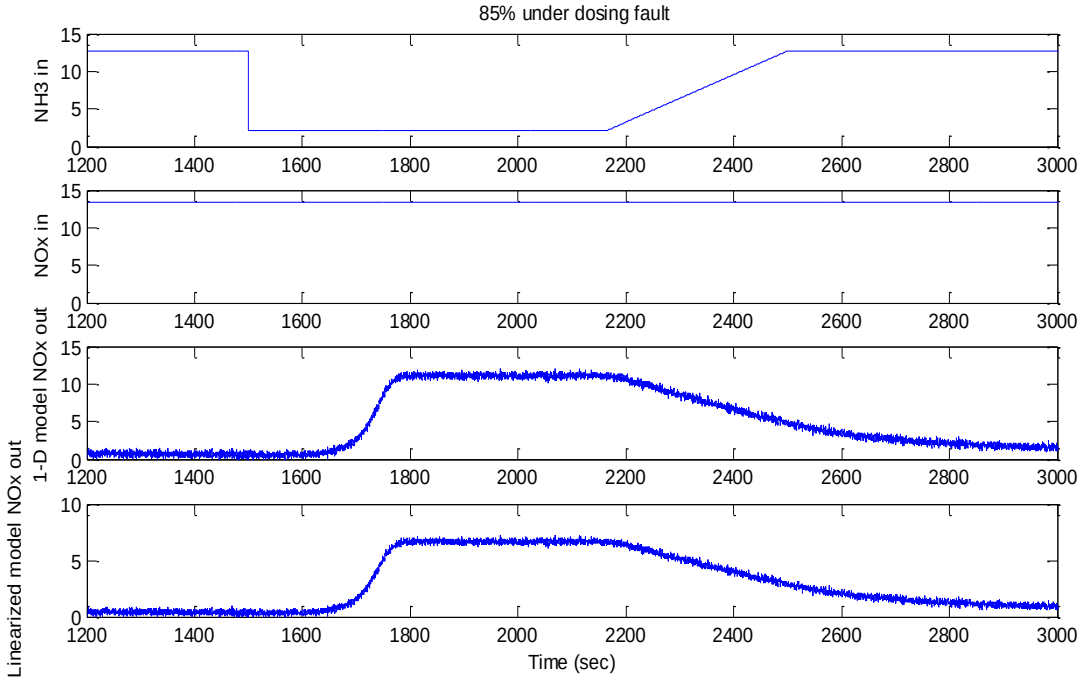

Fig 6.5 Modeling error occurred in large fault with PE method

More sophisticated online state estimation and arbitration methods need to be employed to determine the SCR operating point and linearized point in real time. As stated in section 5.1.1, the online estimation is feasible because of the observability of the linearized model.

6.2.2 Observer based residual generator validation

Similarly, the same faults and operating conditions were used to validate the observer based residual generator. The results are shown in Fig 6.6 and Fig 6.7.

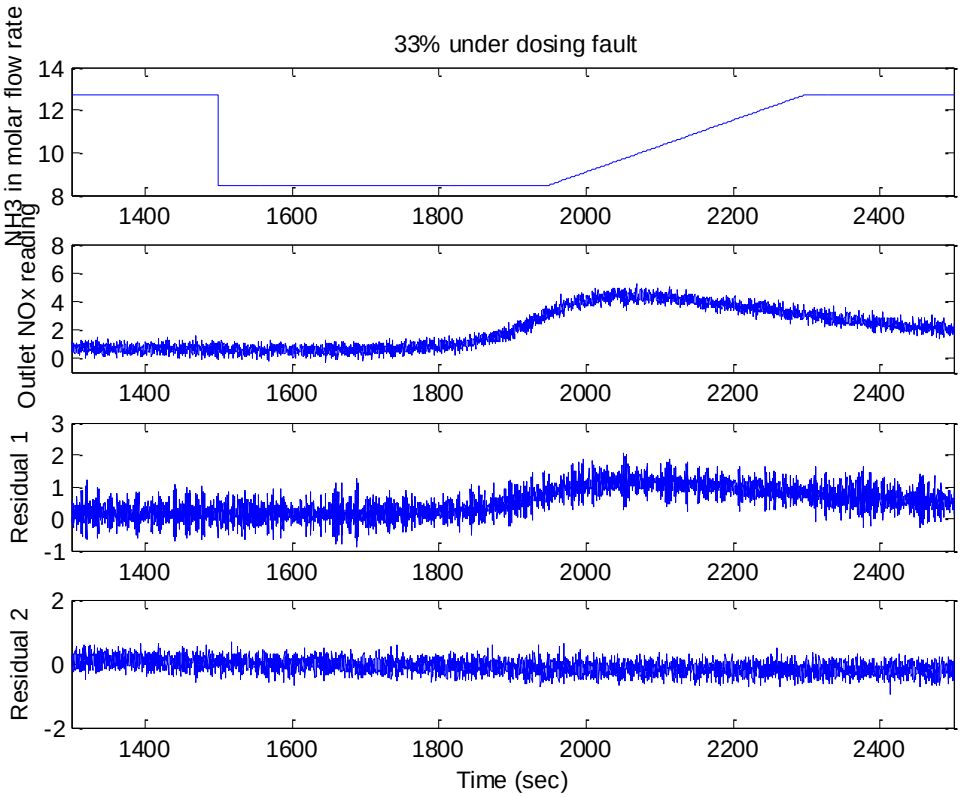

Fig 6.6 Observer based residuals with dosing fault

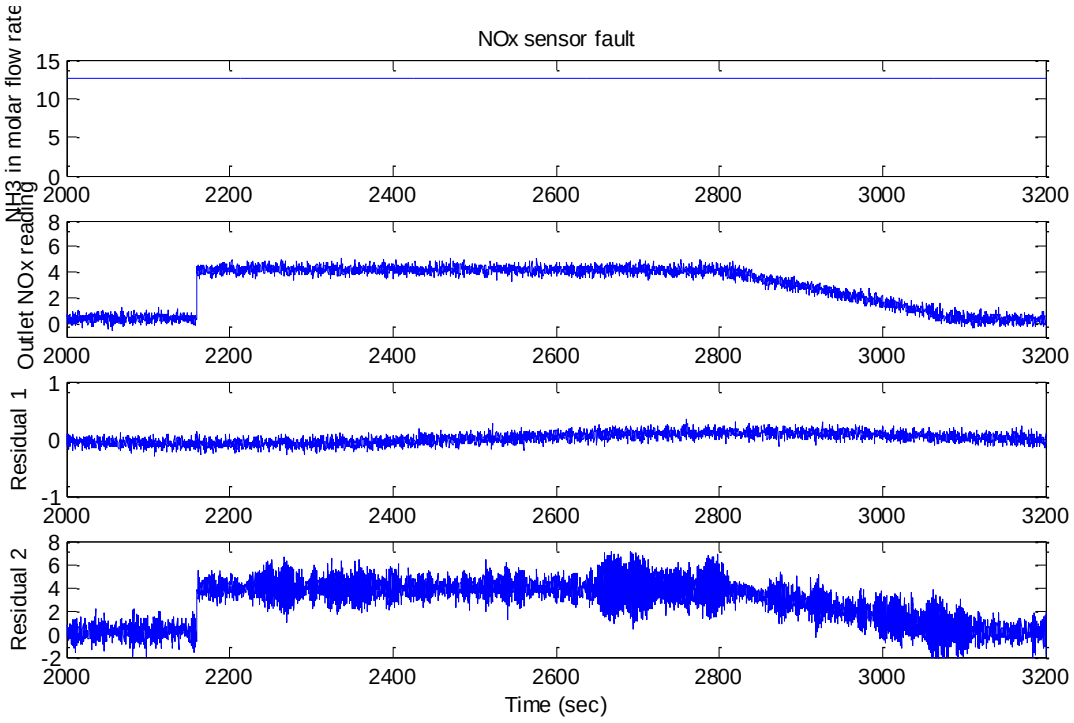

Fig 6.7 Observer based residuals with outlet NO_x sensor fault

As shown in Fig 6.6 and Fig 6.7, the observer based residuals are capable of detecting and isolating the same under dosing fault and outlet NO_x sensor fault. However, it can be seen from Fig 6.7 that the second observer based residual had noticeable noise and chattering. This is due to the switching nature of the sliding mode control.

When implementing the sliding mode observers, it was important to carefully tune the switching gains of the observers. Otherwise, the observer based residuals could contain more chattering or even become unstable, as shown in Fig 6.8.

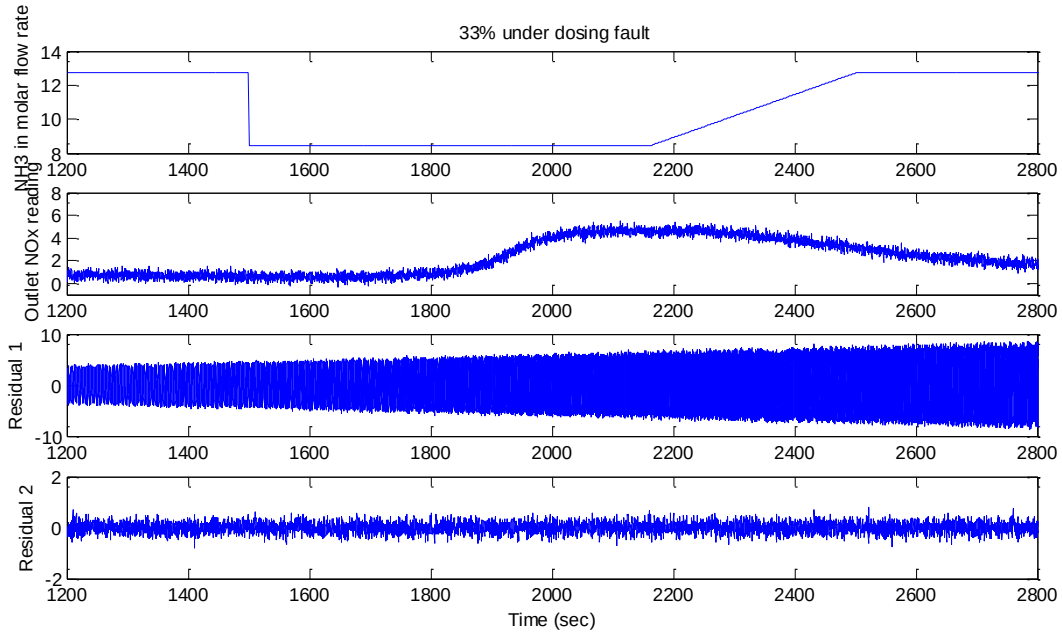

Fig 6.8 Unstable observer based residual with under dosing fault

Compared to the parity equation approach, the residuals using the sliding mode observer approach were showing more chattering when the two approaches had comparable transient responses. However, when the SCR operating condition such as the flow rate deviated further from the linearized operating point used in the parity equation based method, the coefficient matrices in (5-1) changed. This will cause significant modeling error for the parity equation based approach, and therefore, hinders the capability of the approach. On the other hand, the sliding mode observer based approach was a time-variant approach in nature. The sliding mode observers were constructed directly from the nonlinear system, and therefore, possessed more robustness towards model nonlinearity and errors.

CHAPTER 7

CONCLUSIONS

In order to meet the final Tier 4 emission regulation, a Selective Catalytic Reduction (SCR) system has been adopted by many diesel engine manufacturers. Ammonia solution is dosed upstream of the SCR catalyst. The ammonia will be decomposed and react with the NO_x emissions inside the SCR catalyst to generate nitrogen and water vapor. The reduction efficiency of the SCR system will be monitored, as required by the On Board Diagnostic regulations.

The NO_x emission compliance is monitored by using the outlet NO_x sensor. When the outlet NO_x sensor reading exceeds the limit, a diagnostic program must be able to detect and locate the fault. According to the preliminary research, this can be attributed to three major causes. This research is dedicated to developing a diagnostic method to detect and isolate the two most common faults, the dosing fault and the outlet NO_x sensor fault.

After extensive survey and review of the available fault detection and diagnosis methodologies, as well as the current SCR modeling and control research, a decision has been made to utilize model-based fault detection and isolation methods to tackle this problem. Specifically, two model based fault detection and isolation methods, including parity equation and observer based methods, have been developed for the SCR input dosing fault and the output NO_x sensor fault.

In order to facilitate the model based fault detection methods, first a 0-D nonlinear SCR dynamic model was developed. The model was then linearized for the parity equation method. By analyzing the transfer function matrix of the faults, a parity equation based residual generator

was designed so that each residual will react to one fault but not to the other. The observer based method was developed directly from a simplified nonlinear SCR model. Two sliding mode observers were developed to generate equivalent controls based on a different set of inputs. The two equivalent controls were then used to formulate two residuals to isolate the two faults.

Simulation validation of the two diagnostic algorithms was performed using a high fidelity 1-D SCR model. The results show that the proposed model-based fault diagnosis methods have the capability of detecting and isolating the targeted faults. The parity equation method has better stability over the other, while the observer based method can be applied to more general situations.

The main conclusions of this research can be summarized as:

1. The developed control oriented SCR model is shown to work for the proposed model based diagnosis methods.
2. The proposed parity equation method and the linearized model are capable of detecting and isolating the dosing fault and the outlet NO_x sensor fault in steady state conditions, but have limitations when the operating point deviates further from the linearized point.
3. The proposed observer based method has good robustness toward the system nonlinearity and different operating conditions, but presents noticeable chattering in the residuals.

The main innovations and contributions of this research are:

1. Applying model based diagnostic methodologies on the SCR system for the first time, and

2. Design of the ammonia coverage sliding mode observer directly based on the nonlinear SCR model.

Further research can be concentrated in the following directions:

1. Take other failure modes into consideration, such as exhaust line leakage, catalyst aging and failure.
2. Optimize the sensitivity of the proposed residual generators towards smaller faults.
3. Improve the robustness of the proposed residual generators toward certain disturbances, such as drastic change of engine operating conditions.

REFERENCES

Alcota Garcia, E. and P. M. Frank, 1996: On the relationship between observer based and parameter identification based approaches to fault detection. Proc. IFAC 13th world congress, page 25-30.

Ben-Haim, Y. 1980: An algorithm for failure location in a complex network. Nuclear Science and Engineering, Vol. 75, page 191-199.

Bennett, S. M., R. J. Patton, S. Daley and D. A. Newton, 1996: Model based intermittent fault tolerance in an induction motor drive, Proc. of IMACS, Lille, France, page 678-683

Chen, Jie and Ron J. Patton, 1999: Robust model-based fault diagnosis for dynamic systems. Kluwer Academic Publishers, Norwell, MA.

Chi, J. and H. DaCosta, 2005: Modeling and control of urea-SCR aftertreatment system. SAE paper number 2005-01-0966.

Chen, R. and Wang, X., 2014: Model-Based Fault Diagnosis of Selective Catalytic Reduction Systems for Diesel Engines, *SAE Int. J. Passeng. Cars – Electron. Electr. Syst.* 7(2)

Chow, E.Y. and A.S. Willsky, 1984: Analytical redundancy and the design of robust failure detection systems. IEEE Trans. On Automatic Control, Vol AC-29, page 603-614.

Clark, R. N. 1975: Detecting Instrument Malfunctions in Control Systems, IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-11, No. 4, page 465-473.

Clark, R. N. 1978a: A Simplified instrument failure detection scheme, IEEE Transactions on Aerospace and Electronic Systems, AES-14, page 558-563.

Clark, R. N. 1978b: Instrumant fault detection, IEEE Transactions on Aerospace and Electronic Systems, AES-14, 456-465.

Delmaire, G., J. P. Cassar and M. Staroswiecki, 1994: Comparisom of identification and parity space approaches for failure detection in single-input single-output systems. Proc. 33rd IEEE conference on decision and control, page 2279-2285.

Devadas, M., O. Krocher, M. Elsener, A. Wokaun, N. Soger, M. Pfeifer, Y. Demel, and L. Mussmann., 2006: Influence of NO2 on the selective catalytic reduction of NO with NH3 over FeZSM5. Applied Catalysis B : Environmental, 67: page 187 – 196.

Devarakonda, M., G. Parker, J. H. Johnson, V. Strots and S. Santhanam, 2008a: Adequacy of reduced order models for model-based control in a urea-SCR aftertreatment system. SAE paper number 2008-01-0617.

Devarakonda, M., G. Parker, J. H. Johnson, V. Strots and S. Santhanam, 2008b: Model-based estimation and control system development in a urea-SCR aftertreatment system. SAE paper number 2008-01-1324.

Ding, S. X., 2008: Model-based fault diagnosis techniques: design schemes, algorithms, and tools. Springer-Verlag, Berlin, Germany.

Dunia, R. and S. Qin, 1998: Joint diagnosis of process and sensor faults using PCA, Control Engineering Practice, 6:457-469.

Edwards, C., S. K. Spurgeon, R. J. Patton and P. Klotzek, 1997: Sliding mode observers for fault detection, Proc. of IFAC Sympo. on Fault Detection, Supervision and Safety for Technical Process: SAFEPROCESS'97, Univ. of Hull, UK, page 875-880.

Frank P. M. 1987: Fault detection in dynamic systems via state estimation -a survey, in Tzafestas, S., M. Singh and G. Schmidt (Eds), System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches, Reidel, Dordrecht, Vol. 1, 35-98.

Frank, P. M. 1990: Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy-a survey and some new results, Automatica, Vol. 26, No.3, 459-474.

Gertler, J. 1985: Fault detection and isolation in complex technical systems – a new model error approach. Proc. Conference on Information Sciences and Systems, John Hopkins University, page 68-73.

Gertler, J. and D. Singer, 1990: A new structural framework for parity equation based failure detection and isolation. Automatica, 26(2): 381-388.

Gertler, J and K. Yin, 1995: Diagnosing parametric faults - from identification to parity relations. Proc. American control conference, page 1615-1620.

Gertler, J. and T. McAvoy, 1997: Principal component analysis and parity relations – a strong duality. In Proc. 3rd IFAC Symposium, Vol 2: 837.

Gertler, J. 1998: Fault detection and diagnosis in engineering sytems. Marcel Deckker, Basel, Switzerland.

Guernez, C., J. P. CASSAR and M. Staroswiecki, 1997: Extension of parity space to nonlinear polynomianl dynamic system, Proc. of IFAC Sympo. on Fault Detection, Supervision and Safety for Technical Process: SAFEPROCESS'97, Univ. of Hull, UK, page 857-862.

Gui, X., D. Dou, R. Winsor, 2010: Non-road diesel engine emissions and technology options for meeting them. ASABE Distinguished Lecture #34, page 1-24. Agricultural Equipment Technology Conference, Orlando, Florida, USA.

Guzzella, L. and C. Onder, 2010: Introduction to modeling and control of internal combustion engine systems. 2nd. Ed., Springer-Verlag, Berlin.

Herman, A., M. Wu, D. Cabush and M. Shost, 2009: Model based control of SCR dosing and OBD strategies with feedback from NH3 sensors. SAE paper number: 2009-01-0911.

Himmelblau, D.M. 1970: Process analysis by statistical methods. John Wiley & Sons, New York.

Himmelblau, D.M. 1978: Fault detection and diagnosis in chemical and petrochemical processes. Chemical Engineering Monograph Vol8, Elsevier Science Ltd.

Hofling, T. and R. Isermann, 1996: Fault detection based on adaptive parity equations and single parameter tracking. Control Engineering Practice – CEP, 4(10):1361-1369.

Hsieh, M.-F., 2010: Control of diesel engine urea selective catalytic reduction systems. Dissertation, Ohio State University, Columbus.

Hsieh, M.-F. and J. Wang, 2011: Sliding-mode observer for urea-selective catalytic reduction (SCR) mid-catalyst ammonia concentration estimation, Int. J. of Automotive Technology, Vol 12, Issue 3, page 321-329

Isermann, R. 1984: Process fault detection on modeling and estimation methods – a survey, Automatica, 20(4): 387-404.

Isermann, R. 2005: Model-based fault-detection and diagnosis - Status and applications. Annual Reviews in Control

Isermann, R. 2006: Fault-diagnosis systems: an introduction from fault detection to fault tolerance, Springer-Verlag, Berlin.

Isermann, R. 2010: Identification of dynamic systems: An introduction with applications. Springer-Verlag, Berlin.

Isermann, R. 2011: Fault-diagnosis applications – Model-based condition monitoring: actuators, drives, machinery, plants, sensors and fault-tolerant systems. Springer-Verlag, Berlin.

Janik, W and Fuchs. 1991: Process- and signal-model based fault detection of the grinding process. IFAC Symposium on Fault detection Supervision and Safety for Technical Process (SAFEPROCESS), Vol 2, pages 299-304, Baden-Baden, Germany.

Johnson, T. V., 2010: Review of diesel emission and control. SAE 2010 World Congress, SAE paper 2010-01-0301.

Khair, M. and A. Majewski. 2006: Diesel emissions and their control, SAE International

Krijnsen, M., 2000: Advanced control of NOx diesel emissions. Ph.D. dissertation, Delft Univ. Technol., Delft, Netherlands.

Krishnaswami, V., G. C. Luh and G. Rizzoni, 1995: Nonlinear parity equation based residual generation for diagnosis of automotive engine faults. Control Engineering Practice 3(10): page 1385-1392.

Krishnaswami, V. and G. Rizzoni, 1997: Robust residual generation for nonlinear system fault detection and isolation, Proc. of IFAC Sympo. on Fault Detection, Supervision and Safety for Technical Process: SAFEPROCESS'97, Univ. of Hull, UK, page 163-168.

Krocher, O., M. Devadas, M. Elsener, A. Wokaun, N. Soger, M. Pfeifer, Y. Demel, and L. Mussmann. 2006: Investigation of the selective catalytic reduction of NO by NH3 on Fe-ZSM5 monolith catalysts. Applied Catalysis B : Environmental, 66: page 208 – 216.

Lou, X.C., Willsky, A.S. and Verghese, G.C. 1986 Optimally robust redundancy relations for failure detection in uncertain systems. Automatica.

Luenberger, D. G. 1963: Observing the State of a Linear System, IEEE Transactions on Military Electronics, MIL-8, 75-80.

Luenberger, D. G. 1966: Observers for Multivariable Systems, IEEE Transactions onAutomatic Control, AC-11, 190-197.

Luo, Q., 1990: Parity equation approach to failure detection and isolation in dynamic systems. Master thesis, George Mason University, Fairfax, VA.

Magni, J. F., 1995: On continuous-time parameter identification using observers. IEEE Trans. on automatic control, Vol. AC-40, page1789-1792.

Mah, R.S., Stanley, G.M. and Downing, D.M. 1976: Reconciliation and rectification of process flow and inventory data. Ind. Eng. Chemistry, Process Design, Vol. 15.

Mohammadpour, J., M. Franchek and K. Grigoriadis, 2011: A survey on diagnostics methods for automotive engines. Proceeding of American Control Conference (ACC), page 985-990.

Mechmeche, C. and S. Nowakowski, 1997: Residual generator synthesis for bilinear systems with unknown inputs, Proc. of IFAC Sympo. on Fault Detection, Supervision and Safety for Technical Process: SAFEPROCESS'97, Univ. of Hull, UK, page 765-770.

McDowell, N., McCullough, G., Wang, X., Kruger, U. et al. 2007: Fault diagnostics for internal combustion engines - current and future techniques, SAE 2007-01-1603

Nebergall, J., E. Hagen, and J. Owen, 2005: Selective catalytic reduction On-Board Diagnosis: past and future challenges, SAE 2005-01-3603.

Patton, R. J. and J. Chen, 1991a: A review of parity space approaches to fault diagnosis, Preprints of IFAC/IMACS sympo, page 239-255. Invited survey paper.

Patton, R. J. and J. Chen, 1991b: A re-examination of the relationship between parity space and observer based approaches in fault diagnosis, European J. of diagnosis and safety in automation) 1(2): page 183-200.

Patton, R. J. et al. 2000: Issues of fault diagnosis for dynamic systems, Springer-Verlag, London.

Piazzesi, G., M. Devadas, O. Krocher, M. Elsener and A. Wokaun, 2006: Isocyanic acid hydrolysis over Fe-ZSM5 in urea-SCR. Catalyst Communications, 7(8): page 600-603.

Ribbens, W. and Rizzoni, G. 1990: Onboard diagnosis of engine misfires. SAE paper number 901768.

Samimy, B and G. Rizzoni, 1996: Mechanical signature analysis using time-frequency signal processing: application to internal combustion engine knock detection. Proceeding of IEEE, 84(9): page1330-1340.

Schär, C. M., 2003: Control of a Selective Catalytic Reduction Process, Dissertation. ETH No. 15221, Swiss Federal Institute of Technology ETH, Zurich.

Schär, C. M., C. H. Onder, H. P. Geering, and M. Elsener, 2004: Control-oriented model of an SCR catalyst convertor system. SAE paper number 2004-01-0153.

Shost, M., J. Noetzel, M. Wu, T. Sugiarto, T. Bordewyk, G. Fulks and G. B. Fisher, 2008: Monitoring, feedback and control of urea SCR dosing systems for NOx reduction -- utilizing an embedded model and ammonia sensing. SAE paper number 2008-01-1325

Skaf, Z., Aliyev, T., Shead, L., and Steffen, T., 2014: The State of the Art in Selective Catalytic Reduction Control, SAE Technical Paper 2014-01-1533

Stadlbauer, S., Waschl, H., and del Re, L., 2014: NO/NO2 Ratio based NH3 Control of a SCR, SAE Technical Paper 2014-01-1565

Sun, Y., Y. Wang, C. Chang and S. Levijoki, 2012: Detection of urea injection system faults for SCR systems. SAE paper number 2012-01-0431.

Song, Q. and G. Zhu, 2002: Model-based closed-loop control of urea SCR exhaust aftertreatment system for diesel engine. SAE paper number 2002-01-0287

Upadhyay, D. and M. Van Nieuwstadt, 2006: Model based analysis and control design of a urea-SCR deNOx aftertreatment system. ASME J. of Dynamic system, measurement and control, Vol 128: page 737-741

Utkin, V., 1978: Sliding modes and their application in electromechanical systems. Mir Publishers, Russia.

Utkin, V., Guldner, J., Shi, J., 1999: Sliding mode control in electromechanical systems, Taylor

& Francis, London.

Wang, D., S. Yao, M. Shost, J. Yoo, et al, 2008: Ammonia sensor for close-loop SCR control. SAE paper number 2008-01-0919.

Willems, F. and R. Cloudt, 2011: Experimental Demonstration of a new model-based SCR control strategy for cleaner heavy-duty diesel engines, IEEE transactions on control system techonology, Vol. 19, page 1305-1313.

Willimowski, M. and Isermann, R. 2000: A time domain based diagnostic system for misfire detection in spark-ignition engines by exhaust-gas analysis. SAE 2000-01-0366.

Willsky, A.S. and Jones, H.L. 1976: A generalized likelihood ratio approach to detection and estimation of jumps in linear systems. IEEE Trans. On Automatic Control, AC-21.

Witczak, M., 2007: Modelling and estimation strategies for fault diagnosis of non-linear systems, Springer-Verlag, Berlin.

Wuennenberg, J. 1990: Observer-based fault detection in dynamic systems, Ph.D. dissertation, Univ. Duisburg, Duisburg, Germany.

Yang, H. L. and M. Saif, 1997: State observation, failure detection and isolation (FDI) in bilinear systems, Int. J. Contr. 67(6): 901-920

Yu, D. L., D. N. Shields and J. L. Mahtani, 1994: A nonlinear fault detection method for a hydraulic system, Proc. of IEE int. Conf.: Control'94, Peregrinus Press, Conf. Pub. No. 389, Warwick, UK, page 1318-1322.

Yu, D. L. and D. N. Shields, 1996: A bilinear fault detection observer, Automatica 32(11): 1597-1602.

Yuan, Y. and M. J. Shaw, 1995: Introduction of fuzzy decision trees, fuzzy sets and systems 69 (1995), page 125–139.

Zhang, H., 2004: The optimality Naïve Bayes. Proceedings of the 17th International FLAIRS conference (FLAIRS2004).

Zhang, X. and P. Pisu, 2009: Model-based fault diagnosis for a vehicle chassis system, Proc. of American Control Conference, page 1116-1221.

APPENDIX A: NOMENCLATURE

Symbol	Description	Unit
θ	Ammonia coverage ratio	
a_i	Reaction rate constant of ith reaction, (i=1~7)	
C_x	Gas concentration of the “x” species	$mol \cdot m^{-3}$
E_x	Activation energy in the reaction rate constant of the “x” reaction	$J \cdot mol^{-1}$
k_x	Pre-exponential factor in the reaction rate constant of the “x” reaction	
n_x	Molar flow rate of the “x” species	$mol \cdot s^{-1}$
Q	Volumetric flow rate of exhaust gas	$m^3 \cdot s^{-1}$
R	Universal gas constant	8.314 $J \cdot mol^{-1} \cdot K^{-1}$
r_x	Reaction rate of the “x” reaction	$mol \cdot s^{-1} m^{-3}$
T	SCR average temperature	K
V_c	SCR catalyst volume	m^3

APPENDIX B: ACRONYMS

Acronym	Full Name
AOC	Ammonia Oxidation Catalyst
DEF	Diesel Emission Fluid
DOC	Diesel Oxidation Catalyst
DPF	Diesel Particulate Filter
EPA	Environmental Protection Agency
FDI	Fault Detection and Isolation
MIL	Malfunction Indication Light
NOx	Nitrogen Oxides
OBD	On-Board Diagnostics
ODE	Ordinary Differential Equation
PE	Parity Equation
PM	Particulate Matter
SCR	Selective Catalytic Reduction
SMO	Sliding Mode Observer
UIO	Unknown Input Observer