ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

MODEL BASED DIAGNOSIS OF THE OXYGEN SENSORS

M.Sc. THESIS

Kübra EKİNCİ

Department of Mechanical Engineering

System Dynamics and Control Programme

MAY 2018

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

MODEL BASED DIAGNOSIS OF THE OXYGEN SENSORS

M.Sc. THESIS

Kübra EKINCI
(503121609)

Department of Mechanical Engineering

Ssytem Dynamics and Control Programme

Thesis Advisor: Prof. Dr. Şeniz ERTUĞRUL

MAY 2018

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

OKSİJEN ALGILAYICILARIN MODEL TABANLI ARIZA TEŞHİSİ

YÜKSEK LİSANS TEZİ

Kübra EKİNCİ
(503121609)

Makine Mühendisliği Anabilim Dalı

Sistem Dinamiği ve Kontrol Programı

Tez Danışmanı: Prof. Dr. Şeniz ERTUĞRUL

MAYIS 2018

Kübra Ekinci, a M.Sc. student of ITU Graduate School of Science Engineering and Technology student ID 503121609, successfully defended the thesis entitled "MODEL BASED DIAGNOSIS OF THE OXYGEN SENSORS", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Şeniz ERTUĞRUL** ....
Istanbul Technical University

Jury Members : **Assoc. Prof. Özgen Akalın** ....
Istanbul Technical University

Assoc. Prof. Alp Tekin Ergenç ....
Yildiz Technical University

Date of Submission : 04 May 2018
Date of Defense : 04 June 2018

To my family,

FOREWORD

I would like to show my gratitude to my thesis advisor Prof. Dr. Şeniz Ertuğrul for her guidance, advice, generously spent time throughout this study. Her support, not only scientific contributions but also in personal motivation has a great impact in my life. Also, special thanks to Uğur Yavaş and Amedeo Bianchimano due to contributions of AVL Research and Engineering Turkey during the development process of this study.

My sincere thanks go to Martin Lamprecht for his patience with reviewing this work, his valuable feedback with helpful discussions, and being always with me for the endless support.

Last but not least, I would like to thank my dear family members Nurten Aksakal and Volkan Ekinci whom I always feel the sincere support, care, and love.

May 2018 Kübra EKİNCİ

TABLE OF CONTENTS

Page

ABBREVIATIONS

AFR	: Air-Fuel-Ratio
AIC	: Akaike-Information-Criterion
ARX	: Auto-Regressive-Exogeneous-Model
BR	: Bayesian-Regularization-Backpropagation
CARB	: California-Air-Regulation-Board
CNS	: Central-Nervous-System
CO	: Carbon-Monoxide
CO_2	: Carbon-Dioxide
ECU	: Electronic-Control-Unit
EOBD	: European-On-Board-Diagnosis
EPA	: Environmantal-Protection-Agency
GDX	: Adaptive-Learning-Rate-Backpropagation
GPF	: Gasoline-Particulate-Filter
HC	: Hydrocarbons
HEGO	: Heated-Exhaust-Gas-Oxygen-Sensor
ICE	: Internal-Combustion-Engine
LM	: Levenberg-Marquardt-Backpropagation
MAF	: Mass-Air-Flow
MDL	: Minimum-Description-Length
MISO	: Multi-Input-Single-Output
MSE	: Mean-Squared-Error
NARX	: Nonlinear-Autoagressive-Exogenous-Model
NN	: Neural-Network
NO_x	: Nitrogen-Oxide
OBD	: On-Board-Diagnosis
PCA	: Principal-Component-Analysis
RNN	: Recurrent-Neural-Network
SCG	: Scaled-Conjugate-Gradient-Backpropagation
TWC	: Three-Way-Catalytic-converter
UEGO	: Universal-Exhaust-Gas-Oxygen-Sensor
YSZ	: Ytrria-stabilised-zirconia

SYMBOLS

$(A/F)_{stoch}$	: Stochiometric air-to-fuel ratio
μ	: Mean value
μ_r	: Mean value of residuals
$e(k)$	: System zero-mean white noise disturbance based on k
e_t	: White noise time disturbance
$G(z^{-1}, \theta)$	: Determistic transfer function
$H(z^{-1}, \theta)$	: Stochastic transfer function
J_μ	: Mean threshold
J_σ	: Standard deviation threshold
k_a	: Model order a
k_b	: Model order b
k_c	: Model order c
k_d	: Model order d
k_f	: Model order f :
m_a	: Mass of air
$m_{comb, air}$	: Air-mass in combustion chamber
m_f	: Mass of fuel
$\dot{m}_{fuel}$	: Fuel-massflow
$\dot{m}_{O2}$	: Oxygen-massflow
$m_{standardized}$	: Standardized mass
n_a	: Number of poles
n_b	: Number of zeros
n_k	: Input delay
p	: Pressure
p_0	: Norm pressure
p_{comb}	: Combustion chamber pressure
p_{man}	: Intake manifold pressure
p_{res}	: Internal resudial gas pressure
R	: Gas constant
T	: Temperature
t	: Time
T_0	: Norm temperature
T_c	: Time constant
T_{comb}	: Combustion chamber gas temperature
T_d	: Time delay
$u(k)$	: System input
V	: Volume
V_h	: Cylinder displacement volume
x_i	: Input value
x_{mean}	: Mean value of input x
$y(k)$	: System output
λ	: Lambda

$\boldsymbol{\lambda^{I}}$: Model output
$\boldsymbol{\sigma}$: Standard deviation
$\boldsymbol{\sigma_r}$: Standard deviation of residuals

LIST OF TABLES

Page

LIST OF FIGURES

Page

MODEL BASED DIAGNOSIS OF THE OXYGEN SENSORS

SUMMARY

Targets such as complying with legislations on emissions regulated for the protection of the environment and human health, increasing fuel consumption performance, and being competitive in the market, have consistently encouraged the experts to contemplate on air-fuel ratio control for decades. Therefore, many advanced systems, components, and methodologies have been developed and used in the industry. Oxygen sensor is one of these systems.

Oxygen sensors in the exhaust layout of the vehicles are used as feedback elements for precise control of lambda calculation in engine cylinders to be sure that the three-way catalyst works in optimum range and to optimize the power output and fuel consumption of the engine. Both OBDII and EOBD regulations oblige the monitoring of oxygen sensors and warning the driver in case of detection of a faulty sensor.

The primary purpose of the thesis is to develop a methodology for precise and accurate monitoring and diagnosis of the oxygen sensor to meet the legislations and the performance targets while the required calibration effort is reduced regarding time, cost and human resources. The thesis consists of six different chapters. Chapter 1 explains the role of air-fuel ratio in emission control in detail. The aftertreatment layout is presented for a better understanding. Besides, emission gases and their damages are explained. Finally, different methodologies in the literature about the model-based diagnosis applied to various applications are analyzed.

Chapter 2 introduces the oxygen sensors in detail. There are generally two types of oxygen sensors used in vehicles. The first type is a binary sensor and gives information whether the mixture is rich or lean. The type of the oxygen sensor studied in this thesis is the second type called as Universal Exhaust Gas Oxygen Sensor. Therefore, the chapter explains the working principle of Bosch LSU4.9 sensor and importance of the

sensor in closed loop control of the air-fuel ratio. Besides, the factors causing oxygen sensor to fail, the defect types and the conclusion of having a malfunctioning sensor in the aftertreatment layout are described in the section.

The benefits of modeling the dynamics systems are indicated in Chapter 3. Besides, the advantages of the system identification for both linear and nonlinear systems are detailed for some specific methodologies such as impulse response, state space and ARX models for linear system identification and NARX models for nonlinear system identification. System identification uses a set of experimental data to correlate predefined inputs and the system output. System identification is a helpful approach especially when there is not enough information about the dynamic characteristic of the system, or mathematical equations to represent the physic behind are not known.

Chapter 4, firstly, describes the location of the UEGO sensor in the current exhaust layout and the characteristic of experimental data. Then, input selection for system identification methods described in the previous section is realized by using principal component analysis, and fundamentals of combustion and ideal gas law. Using linear system identification follows the input selection step. In this step, three different linear identification model is created by using different combinations of preselected inputs. Since no satisfactory result could be reached, the data is split small portions to analyze the existence of local linearity. Afterward, due to lack of improvement in the performance nonlinear system identification approach is decided to utilize. Nonlinear-autoregressive-exogenous model is designed with several different configurations of the number of hidden layers, the number of the neurons, delays, learning algorithms, etc. to obtain the best suitable model representing the UEGO sensor. In the end, a NARX model with two hidden layers and eight neurons in each hidden layers is decided to use in residual generation phase of sensor diagnosis.

After a successful model has been achieved to a sufficient degree, the residual generation and residual evaluation steps are continued in Chapter 5. Residuals are produced by comparing the sensor output to model output in two WLTC data. The first data set is collected in a WLTC with a good sensor on the vehicle whereas a faulty sensor is located on the vehicle during the second WLTC. After that, these two residual sets are analyzed statistically to decide on if the sensor is defective or not. It is decided to use the decision-making approach by using two different threshold values based on the standard deviation and mean values calculated for each residual data set.

In future work, some other residual evaluation methods can be tried to see the effectiveness of different approaches. In addition, it is recommended that validations be performed under various environmental conditions.

OKSİJEN ALGILAYICILARIN MODEL TABANLI ARIZA TEŞHİSİ

ÖZET

Çevre ve insan sağlığının korunmasına yönelik düzenlenmiş emisyon mevzuatlarına uyma, yakıt tüketiminin iyileştirilmesi, araç performansının artırılması ve piyasada rekabet edebilme gibi hedefler, uzmanları hava-yakıt oranı kontrolü üzerine yoğun çalışmalar gerçekleştirmeye yöneltmiştir. Bunun sonucu olarak otomotiv endüstrisinde hava-yakıt oranı kontrolü üzerine pek çok sistem geliştirilmekte ve kullanılmaktadır. Oksijen algılayıcı bu sistemlerden bir tanesidir.

Oksijen algılayıcı, araçların egzoz hattında yer alır ve egzoz gazında bulunan oksijen miktarını ölçerek silindir içerisindeki hava yakıt oranı hesaplamaları için geri besleme elemanı olarak kullanılır. Böylece, üç yollu katalitik konvertörün, emisyon gazlarının indirgenebilmesi için optimum aralıkta çalışması sağlanırken, motor gücü ve yakıt tüketimi performansı optimize edilir. OBDII ve EOBD yönetmeliklerinin her ikisi de oksijen algılayıcıların çalışma süresince gözlemlenmesini ve algılayıcıda arıza teşhis edilmesi halinde sürücünün uyarılmasının zorunlu olduğunu birdirmektedir.

Bu çalışmanın temel amacı, oksijen algılayıcının emisyon yönetmeliklerine uygun olarak performans hedeflerini karşılayabilmek için hassas ve doğru bir şekilde izlenmesini ve arıza durumunda algılayıcıya teşhis konulmasını sağlarken, zaman, maliyet ve insan gücü açısından kalibrasyon için gerekli iş gücünü azaltacak bir yöntem geliştirmektir. Tez altı farklı bölüm içermektedir. Birinci bölümde, emisyon kontrolünde hava yakıt oranının rolü detaylı olarak açıklanmaktadır. Ayrıca, sistemin anlaşılabilirliğini kolaylaştırmak için genel elemanları ile bir egzoz hattı gösterilirken; emisyon gazları ile emisyon gazlarının zararlarına yer verilmektedir. Son olarak ise, literatürde yer alan model tabanlı teşhis uygulamalarına yönelik çeşitli çalışmalar incelenmektedir.

İkinci bölüm oksijen algılayıcıları ayrıntılı olarak tanıtmaktadır. Araçlarda genellikle iki tip oksijen algılayıcı kullanılmaktadır. Bunlardan ilki ikili sistemde çalışmaktadır, ve karışımın zengin ya da fakir karışım olduğunu belirtmektedir. Bu tezde incelenen oksijen algılayıcının tipi ise ikinci tip olan Üniversal Egzoz Gazı Oksijen Algılayıcısıdır. Bu algılayıcı tipine örnek olarak Bosch LSU4.9 algılayıcının çalışma prensibi ve algılayıcının kapalı çevrim hava-yakıt oranı kontrolündeki önemi açıklanmaktadır. Ek olarak, oksijen algılayıcının arızalanmasına neden olabilecek etkenler, hata tipleri, ve egzoz hattında arızalı bir algılayıcı kullanılmasının neden olabileceği sonuçlar anlatılmaktadır.

Üçüncü bölümde dinamik sistemlerin modellenmesinin önemine değinilmektedir. Doğrusal ve doğrusal olmayan sistemler için sistem tanılama yöntemlerinin kullanılmasının sağladığı avantajlar belirli metodlar üzerinden anlatılmaktadır. Doğrusal sistemler için dürtü yanıtı, durum uzay ve ARX modelleri açıklanırken; doğrusal olmayan sistemler için NARX modeli detaylandırılmaktadır. Sistem tanılama, bir sistemin önceden tanımlanmış girdileri ile sistemin çıktı ya da çıktılarını ilişkilendirmek için bir dizi deneysel veri kullanılması esasına dayanmaktadır. Sistem tanılama özellikle sistemin dinamik karakteristiği hakkında yeterli bilgiye sahip olunmadığı ya da sistemi fiziksel olarak temsil eden matematiksel denklemlerin bilinmediği durumlarda tercih edilen faydalı bir yaklaşımdır.

Mevcut projede kullanılan UEGO algılayıcısının egzoz hattındaki konumundan dördüncü bölümde bahsedilmektedir. Ayrıca, bu bölümde şasi dinamometresinde araç üzerinde toplanmış verilerin özellikleri belirtilmektedir. Üçüncü bölümde anlatılan sistem tanılama yöntemlerinde kullanılmak üzere girdi seçimi analizi de bu bölümde yapılmaktadır. Girdi seçimi temel bileşen analizi, yanma dinamikleri ve ideal gaz kanunu kullanılarak gerçekleştirilmektedir. Girdi seçiminin ardından doğrusal sistem tanılama yöntemi kullanılmaktadır. Bu adımda, önceden belirlenen olası girdi sinyallerinin farklı kombinasyonları kullanılarak üç farklı doğrusal tanılama modeli oluşturulmaktadır. Farklı çalışma noktalarında toplanmış veri seti kullanılarak farklı girdi kombinasyonu ve modeller ile başarılı bir sonuca ulaşılamamıştır. Bu nedenle, veri seti yerel doğrusallık durumunu incelemek üzere farklı motor hızları için daha küçük parçalara ayrılmıştır. Bu adımda da tatmin edici bir sonuç elde edilemediğinden bir sonraki adımda doğrusal olmayan sistem tanılama yaklaşımının kullanılması kararlaştırılmıştır. Doğrusal olmayan sistem tanılama yaklaşımı olarak bir yapay sinir ağı yöntemi olan "nonlinear-autoregressive-exogenous model" kullanılmaktadır. UEGO algılayıcısını en iyi çıktı performansını elde edebilecek şekilde modelleyebilmek amaçlanmıştır. Bu amaçla, çeşitli sayıda saklı katman ve nöronlar için girdilere farklı gecikmeler ve farklı öğrenme yöntemleri uygulanarak bir çok konfigürasyon oluşturulmuştur. Bu konfigürasyonların performansları karşılaştırılarak en iyi modele karar verilmiştir. İki saklı katmanlı, her saklı katmanda sekiz adet nöron bulunduran, sistem gecikmesinin beş olarak belirlendiği NARX modelinin, algılayıcı arıza teşhisinin artık üretme aşamasında kulanılmasına karar verilmiştir.

Beşinci bölümde, NARX modeli ile yeterli seviyede başarılı bir model elde edilmesinin ardından artık üretimi ve artık değerlendirme aşamaları ile devam edilmektedir. Artık ya da kalan değerler, WLTC sırasında toplanan model ve algılayıcı çıktılarının karşılaştırılması ile elde edilmektedir. Çalışma için iki farklı WLTC verisi toplanmıştır. İlk WLTC sırasında araç üzerinde normal çalışan bir oksijen algılayıcı ile veri toplanırken ikinci WLTC'de arızalı bir oksijen algılayıcı kullanılarak veri toplanmaktadır. Her iki test çevrimindeki model ve algılayıcı çıktılarının karşılaştırılması ile elde edilen iki farklı artık veri seti, algılayıcı arıza teşhisi hakkında karar verebilmek amacıyla istatistiksel olarak analiz edilmektedir. Her bir artık veri seti için hesaplanan standart sapma ve ortalama değerlere bağlı iki farklı eşik değer kullanılarak karar verme yaklaşımının kullanılmasına karar verilmiştir.

Gelecekteki çalışmalarda, artık değerlendirme aşamasında farklı yaklaşımlar denenerek bu yaklaşımların verimliliklerinin ve uygulanabilirliklerinin gözlemlenebileceği düşünülmektedir. Ayrıca, çeşitli çevresel koşullar altında validasyon testlerinin yapılması tavsiye edilmektedir.

1. INTRODUCTION

It is considered that it is possible to achieve the most effective trade-off between engine power, fuel economy, and emissions when the air-to-fuel ratio is equal to the theoretically required amount for complete combustion (stoichiometric) [1]. Lambda is the ratio used to specify how far the actual air-to-fuel mixture deviates from that stoichiometric ratio. If the engine is not running in that idealized point, some unfavorable effects occur. In lean condition, which means the air-fuel mixture containing a relatively low proportion of fuel, the emissions increase, particularly NOx, there will be heat increase as well as the high possibility of engine knock coupled with a slight drop in engine power. On the other hand usually fuel economy increases. If the engine is running in the rich condition, which means the air-fuel mixture containing an excessive proportion of fuel, there will be emissions and fuel consumption increase, a slight decrease in the heat with lower risk for engine knocking. Engine power will be slightly higher in this case.

The air-to-fuel ratio is defined as

$$(A/F) = \frac{m_a}{m_f} \tag{1.1}$$

Where m_a and m_f are the mass of air and fuel entering the engine respectively.

Lambda is defined as:

$$\lambda = \frac{(A/F)}{(A/F)_{stoch}} \tag{1.2}$$

Where $(A/F)_{stoch}$ is the stochiometric air-to-fuel ratio, which is obtained when complete combustion (theoretically) occurs.

To be able to specify the lambda value and to make calculations about air-to-fuel ratio, or AFR in short, a measurement system or sensor within the system is required.

The oxygen sensor (or lambda sensor) is an electronic device that is located in the exhaust system, close to the engine as shown in Figure 1.1. Its purpose is to monitor the concentration of residual oxygen within the exhaust gases produced by the engine.

In internal combustion engines the sensor is used in order to estimate, and dynamically tune the air-to-fuel ratio. In this way, the three-way catalytic converters can work optimally. Additionally, a second sensor behind the catalytic converter can also determine whether a catalyst is performing correctly or not [1].

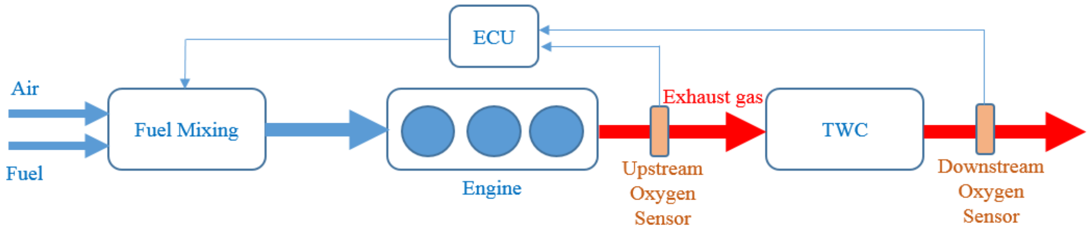

Figure 1.1: Engine and after treatment layout

To reach optimal combustion several inputs, depending on the operation point, need to be considered. Such as engine coolant temperature, throttle position, air mass volume and engine speed (rpm). To further refine combustion, additional values like changes in altitude, humidity, ambient temperature and fuel quality can be used. The primary Oxygen sensor is located upstream of the catalytic converter and delivers the actual AFR to allow the ECU to provide the necessary air-to-fuel ratio for optimal combustion.

Using a three-way catalytic converter (TWC) is the most widespread technique utilized by vehicle manufacturers to decrease exhaust emissions especially in gasoline engines. The TWC uses precious metals to convert harmful emissions into non-toxic end products. It contains ceramic or metallic monolithic structures which surface areas and walls are coated with platinum, palladium, and rhodium. In a TWC uncombusted fuel residues as hydrocarbons (HC). This causes a depression of the central nervous system (CNS). The hydrocarbons are oxidized with residual oxygen to produce carbon dioxide and water. The nitrogen oxides (NOx), which cause irritations in the respiratory system of humans and animals, are converted to ubiquitous nitrogen, and toxic carbon monoxide (CO) preventing delivery of oxygen to bodily tissues [2].

If the exhaust gases are within specified tolerances on air/fuel ratio and the catalyst operates at a specific temperature range, the TWC can perform the conversion of exhaust gases efficiently. Figure 1.2 represents the general trend of conversion efficiencies in relations to the air-fuel ratio [3]. Utilizing an exhaust gas oxygen sensor installed upstream of the catalyst, the required air/fuel ratio to operate the system is provided precisely. By receiving the feedback signal from upstream of the catalyst, the control unit can adjust the fuel quantity to keep the whole system in the tight range of air/fuel ratio. Therefore, having an oxygen sensor that works correctly in the exhaust system is substantial to be able to meet emission targets defined by authorities.

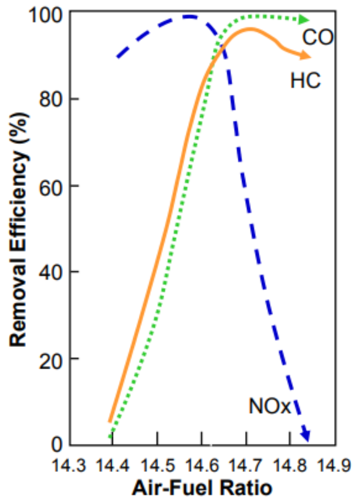

Figure 1.2: Conversion efficiency of TWC vs Air-Fuel Ratio [3]

In addition to not passing the emission testing, due to an undetected failure or damage of the catalytic converter and wrong AFR, a malfunctioning oxygen sensor may cause increased fuel consumption (by up to 15 % on average, considerably more for city driving), failure or poor driving characteristic.

Detecting when there is a sensor which is not working correctly in the system prevents these above mentioned unfavorable consequences. By means of warning the driver and indicating the necessity of going to car care service. In vehicles, detection of malfunctioning component and taking the further actions are carried out by an onboard diagnostic system (OBD), the vehicle's self-diagnostic and reporting capability.

The legislations about exhaust emission regulations, represented by institutions like California Air Regulation Board (CARB), Environmental Protection Agency (EPA) are the primary motivations for the function and methodology developers for OBD. Stringent emission regulations require monitoring the upstream exhaust gas oxygen sensor (UEGO, Universal Exhaust Gas Oxygen Sensor) for any possible malfunction causing the exhaust emissions to exceed certain thresholds.

Due to increased environmental consciousness, and the need to be competitive in the global market to meet engine performance targets and driver expectations, the complexity of components and control algorithms have been increasing drastically. In the automotive industry, experts are used to rely on extensive experimental calibration to fine-tune performance. However, they are moving towards different approaches nowadays. Data based or model-based methods to control and calibrate control parameters to minimize costs, time and effort and systematically incorporate advances in powertrain technology are increasingly common.

1.1 Purpose of the Thesis

In compliance with both the European and US legislations, motor vehicles having On Board Diagnostics are required to monitor and detect the slow response of Oxygen Sensor. Detection of slow sensor response is one of the fundamental parts of On-Board Diagnostics to ensure that the oxygen sensors in an automotive emission control system continue to operate properly with age. After possessing this complex functionality in control structure, it usually requires the use of sophisticated measurement methods and the post-processing of the measured signals. The effort to make the sensor diagnostics functional is time-consuming and often needs engineering skills at a certain level.

The aim of this thesis is to examine and explore dynamics of Universal Exhaust Gas Oxygen Sensors and the feasibility and effectiveness of different modeling and diagnostic approaches to provide simplicity to control and calibrate the system even if the designer does not have enough experience.

This study is organized as follows: Chapter 1 gives a detailed description of lambda, oxygen sensors, and OBD, and also provides reasons of the necessity of oxygen sensors in-vehicle systems and motivation behind the monitoring functions and the need of new diagnostic approaches. Different approaches to monitoring various components and systems in automotive environments are detailed. The aim of the thesis and the literature survey are mentioned. Chapter 2 presents a comprehensive description of oxygen sensors and environmental legislations. System Identification methods for linear and nonlinear systems are briefly mentioned in Chapter 3. Chapter 4 provides the steps followed to get a sufficient system model, while chapter 5 reports the proposed solution to detect slow response failure of the oxygen sensor.

Finally, chapter 6 outlines the conclusions derived from this research, and therefore the original contributions of this work.

1.2 Literature Survey

Continuous developments to improve the performance of the three-way catalyst provides a field of study to researchers since the eighties. This is due to the significant effect of precise air-to-fuel ratio calculation on emission control. The UEGO sensor located in the exhaust system substituted for the heated exhaust oxygen sensors has a great contribution to the AFR control algorithm [4]. Therefore, the diagnosis of the UEGO sensor takes a considerable place in the literature. Fault detection and diagnosis of sensors as well as actuators and comprehensive components of automotive systems have been very popular. It is proved useful to prevent misleading the control system due to the misrepresented measurements. Different approaches developed or proposed for fault detection and diagnosis in literature are discussed in this chapter. The current methodologies can be branched two main subcategories called model-free and model-based fault diagnosis. Applications and limitations of these different approaches are also detailed. Figure 1.3 summarizes the model-free and model-based diagnostic methods.

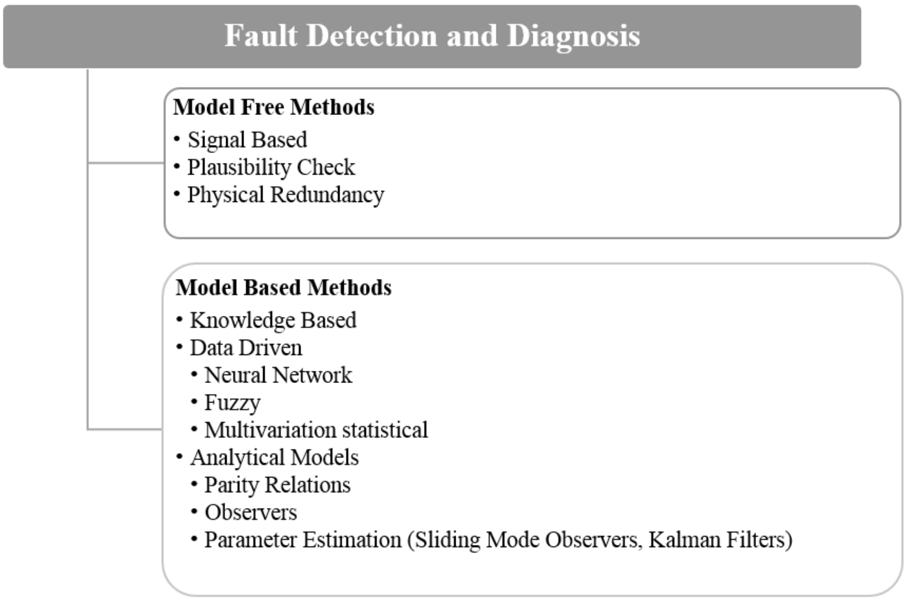

Figure 1.3: Fault Detection and Diagnosis Methods

Model-free methods rely on system characteristics and heuristics approaches to decide if the system is running faulty or as expected. A signal based diagnosis approach analyzes the feature of the affected system by using signal processing techniques or filter the final output signal to decide on the faulty or healthy system [5,6,7]. The signal based fault diagnosis methodology usually does not take the internal dynamics of the system during decision making into consideration. It needs a database for fault scenarios. A diagnosis method similar to the signal based methods is the plausibility check, since it uses the sensor output signal against the physical laws to analyze the conditions for fault detection. Analogically, the plausibility check is not interested in dynamic relations between system variables [8]. The major trends are mostly far from using a physical redundancy approach. Using several sensors to measure the quantity of same physical feature needs more space and increases cost and complexity. Nevertheless, it can be efficient regarding accurate detection and isolation of the faults [9].

Model-based methods mostly take advantage of mathematical models which use inputs and outputs of the real system. Once obtaining the model, outputs of the model and the actual system are compared and a decision is made based on the difference between two signals. One of the popular model-based methods is the knowledge-based model. Even though model-based techniques are preferred in case of imprecise measurement data and environmental conditions, the main hassle is the necessity of prior knowledge and experience with deep understanding of the system. In the study of Nybarg and Nielse, direct redundancy and nonlinear diagnostic observers are used in the diagnosis of air intake system. It utilizes mean value modeling of the system including throttle model and air dynamics. The throttle is modeled as second order linear system, and air dynamics is derived from ideal gas law, whereas physical knowledge about the air intake system and mathematical equations come into prominence. After obtaining the residuals using mean value model, residual evaluation is realized based on fuzzy thresholding [10]. The analytical models are another common way used in literature for fault detection and diagnosis. Analytical models utilize dynamical models of the system or components based on physical laws. These dynamical models are used to generate residuals to decide if the system is faulty or not. Analytical models can be examined by separating several different groups. These groups mainly include parity relations, observers, and parameter estimation methods, such as sliding mode observers or Kalman Filter. In the structured parity equation approach, inputs such as sensor bias, actuator bias and disturbances coming from simulation data, are used to get residuals in the system on the purpose of online detection of sensors and actuator faults in automobile engine [11]. Using parity equation is a cost-effective way with a trace of performance loss compared to more complex methods. Even despite the difficulties on working with the physical systems, due to the issues researchers do not face in the simulation environments. The exhaust oxygen sensor, which is one of the main components subjected to study, has the main difficulty on identification due to its nonlinear characteristic. A suitable method for nonlinear systems is using sliding mode observers, as it is used to detect misfiring faults [12]. However, a sufficient engine model is essential to be able to get satisfying results with sliding mode observers. Structured hypothesis testing is another methodology being used in literature. In the application of air path diagnosis in a diesel engine, the study shows validated results on a vehicle in real-world driving conditions. It is based on structured hypothesis testing, which observers used to calculate error parameters [13]. Another structured hypothesis test based diagnosis is exemplified in order to demonstrate that many different faults can be detected by the same diagnosis system modeled and diagnosed by using different methods [14]. To achieve this goal, several behavioral modes are generated. Mainly for cases such as no fault as well as different faults that might be observed in the air intake like leakage of boost and disconnection of the manifold pressure sensor. Since separate hypothesis tests are needed for different behavioral modes, the burden of excess test numbers and computation is the main drawback. Statistical methods to detect a failure of the UEGO sensor provides a good way to separate different sensor faults under different operation. However, this approach requires modeling the distribution function of the difference between consecutive sensor measurements and its parameters. It also needs to model a probability density function [15]. This non-intrusive statistical based methodology involves two different steps based on the operation. In the first step, parameters which belong to the distribution function in case of asymmetric operation are calculated. Considering these parameters a decision is made whether a fault exists. To be able to specify the fault type, a system identification process is applied for the case of symmetric operation in the second step. According to central peak and shape of the distribution functions, the type of the operation can be defined. Principal component analysis (PCA) provides another approach for fault detection and diagnosis of sensors utilizing Q-statistic and/or squared prediction error estimations [9]. In the study two PCA models were build for diagnosis of air handling unit sensors. The main idea is to minimize the impact of the system nonlinearity and improve the robustness, while the Q contribution plot is used to isolate the fault. In such diagnostic structures using only a single PCA model is not enough to be able to detect all sensor failures. Thus, it is important to use extensive PCA models in parallel. Other than methods mentioned above, estimating fuel film dynamics in the intake port of an SI engine, by using adaptive estimator, gives a good performance in predicting AFR by exploiting extended Kalman filter [16]. However, this method needs to be extended for a wide range of throttle operation. Fuzzy-based pattern recognition method for real-time detection of abnormal injection pressure patterns is another approach revealed in the literature [17].

In order to overcome the experimental and computational workloads, and decrease the complexity of physical equations, applying data-driven methodologies can be another solution. Especially when the topic comes to the nonlinear systems, neural network based fault diagnosis can be efficient and robust as well. An example of the recurrent neural network model is proposed to predict the air-to-fuel ratio to use in closed loop fuel calculation, and diagnostic application in a port fuel injection spark ignited engine [18]. In this study, it is observed that, by means of RNN based methods, robust system models for a broad range of operating conditions can be obtained with no remarkable delay and with high accuracy. Consequently, the behavior of the system dynamics which are close to the real system can be achieved with no complicated calculations and a high number of experimental data. Therefore, improving such data-driven model-based approaches to detect sensor failure has become primary motivation for this study. During the next chapters, it is focused to enhance a suitable diagnostic methodology which does not require in-depth analysis of various data to be able to calibrate system or control parameters. Therefore, firstly applicability of different modeling approaches are investigated and then residual generation and decision making are analyzed based on the likeliest model. Once the best model is decided, ideally it requires less time to collect data on the vehicle or engine during calibration. In this way, a calibration engineer can tune the necessary parameters for fault detection and diagnosis easily. Consequently, there is less effort to understand the complex physical equations of the system, with no high burden of statistical examinations.

2. OXYGEN SENSORS

2.1 Overview Of Oxygen Sensors

Oxygen sensors located in the exhaust layout are commonly used as sensing devices in ICE systems. The primary purpose is to deduce the AFR via sampling the exhaust gas from the engine. Thus, the feedback of the sensor is used to calculate the amount of injected fuel by the closed-loop controller to provide a sufficient operating condition to the TWC and allow it to convert the residual exhaust gases efficiently. Besides that, the fuel consumption and engine power output are optimized as well. The sensor is usually positioned after the turbocharger and before the TWC in turbocharged engines. The high-temperature exhaust gases are leaving the cylinder during the exhaust stroke under high pressure, then travel through the exhaust manifold before they come in contact with the oxygen sensor placed ahead of the TWC. Another reason to have an oxygen sensor in the exhaust layout is monitoring the performance of the TWC by placing another sensor after the catalyst. The output signals of one sensor positioned upstream, and one sensor downstream of the catalyst are compared to detect any degradation in catalyst efficiency [19].

There are several different types of oxygen sensors used in literature for the abovementioned purposes [20]. One of the most common types of sensors is the switched type or binary oxygen sensor, also known as Heated Exhaust Gas Oxygen (HEGO) sensor. This sensor signal shows whether the exhaust gas is the output of rich or lean combustion. Another common oxygen sensor type is wideband oxygen sensor, also called Universal Exhaust Gas Oxygen (UEGO) sensor. The UEGO sensor uses relatively new technology compared to HEGO sensor. Differently, from the HEGO sensor, it provides exact information of how lean or rich the combustion is. These two different sorts of sensing devices can be used interchangeably in fuel calculations, as well as in the same system for TWC diagnostics. Within the scope of this thesis, UEGO sensors used for regulating the fuel supply is focused.

In this study, the considered sensor design is the Bosch LSU4.9 whose physical and electrical structure is as shown in Figure 2.1 [21]. The reference and pump cells, the measurement cavity, and the diffusion barrier are the crucial parts of the sensor. Within a steel shell, the sensing element interacting with the exhaust gas of UEGO is composed of yttria-stabilized zirconia (YSZ) to make use of its feature as a good conductor of oxygen ions. Down the line, the ceramic sensing element of YSZ is connected to platinum electrodes as well as wire leads. After flowing through the holes in the steel shell the exhaust gases, consisting of oxygen molecules, contacts the sensing element. The ambient air is forced to flow through the gaps between the connecting cables while being heated to enable the ions to produce voltage. Due to the difference in concentration of oxygen molecules in the exhaust gas and the ambient air an ion exchange takes place. This exchange happens from the higher to the lower concentration between the exhaust gas and reference gas. The actual measuring occurs in the Nernst cell. As a result of the movement of oxygen ions from one platinum layer to the other, a potential difference is generated in the Nernst cell, which enables a current to flow. Generated voltage signals are fed to the engine control unit, or ECU for short. The ECU compares the voltage signal with the pre-stored standard data to decide whether the mixture is rich or lean.

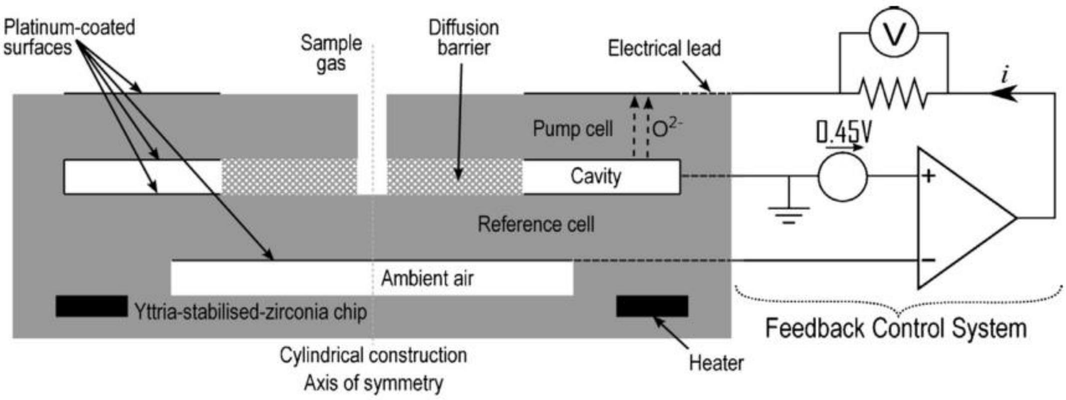

Figure 2.1: Electrical structure of Bosch LSU4.9 UEGO Sensor [21]

The standard way to model the UEGO sensor dynamics is considering a linear response around stoichiometry and approximated by a first order transfer functions with a delay [22]. The ECU has a cascaded AFR controller algorithm including inner and outer control loops. The outer loop provides the set-point value of AFR for the inner loop. The inner loop controller keeps the AFR at the set-point by using feedback from UEGO sensor [23]. In this feedback loop, HEGO sensor is also used as a corrector to enhance the robustness of UEGO sensor. The feedback component decreases the steady-state error to minimize the differences between the desired AFR and the actual AFR even though it is slower than feedforward component which might not always be accurate since it may be controlled independently from the actual AFR [24]. The inner loop controller is shown in Figure 2.2, and the normalized AFR and the overall open loop dynamics are stated in equations 2.1 and 2.2 respectively whereas T_c stands for the time constant and T_d stands for the time delay [15].

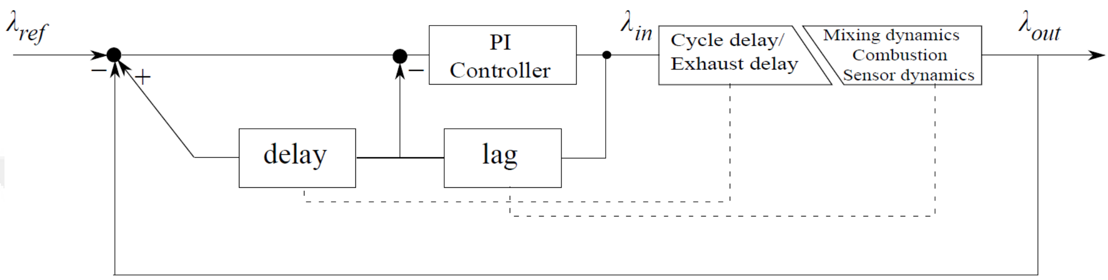

Figure 2.2: Inner loop air-fuel ratio controller

$$\lambda = \frac{1}{K_{stoch}} * \frac{\dot{m}_{O_2}}{\dot{m}_{fuel}} \tag{2.1}$$

$$G(s) = \frac{\lambda_{out}(s)}{\lambda_{in}(s)} \approx \frac{e^{-T_d s}}{T_c s+1} \tag{2.2}$$

The time delay is the difference between the inclusion of fuel-air mixture into the cylinder and when it is observed at the UEGO sensor. The first reason for this delay is the physical plant delay due to the combustion dynamics while the second reason is the delay of exhaust gas transportation [24]. In further steps of this study, understanding of these controller loops and the time delay is essential to comprehend the dynamics of the system.

2.2 Oxygen Sensor Failures

The functioning of the oxygen sensor may be affected by high thermal overload and poor fuel quality causing poisoning. Besides other environmental influences, residues in the exhaust gas like chemicals, soot, and oil as well as vibration on the exhaust after-treatment line can result in severe damage. Eventually, the performance of the oxygen sensors is diminished over the years. Like every other component on a vehicle, they are subjected to a certain amount of aging and wear. Therefore, it is essential to monitor the performance of oxygen sensors to detect any failure. Using the sensor diagnostics, in case of failure, the ECU can take further action and can inform the driver. This prevents some side effects like increased fuel consumption with resulting higher CO2 emissions, possible engine and/or catalytic converter damage, poor driving characteristics and the possibility of failure to pass the emissions testing.

2.3 Environmental Legislation

On-Board Diagnostic, or OBD in short, means that the diagnostic is done self-reliant respectively stand-alone by the engine control unit. The OBD is divided into two subgroups as OBD I and OBD II. While OBD I includes mainly electrical faults, OBD II stands for emission-related defects. The main drivers of OBD are the US which are mainly the California Air Resources Board (CARB) and US Environmental Protection Agency (EPA) and Europe certification authorities. The most stringent OBD requirements, especially OBD II regulations, are featured by the CARB since 1996. OBD regulations for Europe called EOBD adjust its requirements mostly based on US standards. If any failure that may occur in one of the components of the vehicle affects the emission performance of the vehicle, both of the emission regulation for US and EU will require the diagnosis of this component.

Since the oxygen sensors have a crucial role in AFR control for the TWC to work with the maximum efficiency by providing feedback from exhaust line, a well established diagnostic function must monitor the sensor activity. Thus, the control unit may ensure a proper mixture for the optimal operating conditions of TWC. That allows the TWC to convert the harmful emissions efficiently at the end.

Regarding diagnosis of the dynamic behavior of oxygen sensors, OBD legislations require the detection of two fault cases shown in Figure 2.3. The transient-time faults consist in a reduction of the oxygen sensor signal gradient concerning the gradient of the real lambda value, while the response-time defects include in a retarded reaction of the oxygen sensor signal to lambda changes. Both kinds of faults have to be detected regardless of the direction of the lambda changes. Said that either rich-to-lean or lean-to-rich, as well as changes in both directions, are covered.

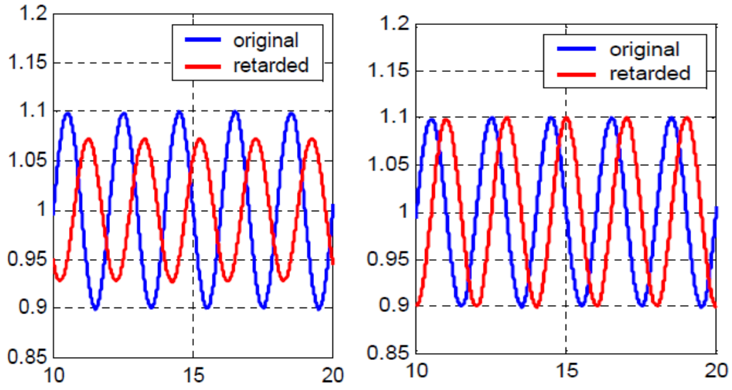

Figure 2.3: Transient Time Fault, Response Time Fault

In this study, European legislation is taken into consideration especially while deciding on faulty component and error thresholds.

3. MODELING OF DYNAMICAL SYSTEMS

Modeling of dynamical systems provides convenient solutions in many different working areas, since the models allow the internal dynamics of the systems to be simulated and analyzed by mathematical representations, without physically building the system. Also, it is possible to run models way more quickly. Thus they save time and cost.

Dynamical system modeling can be performed by using physical laws and equations or by system identification. System identification is a principal to obtain the system model by use of the inputs and outputs of the system. In the present study, system identification is examined to model the dynamical system of the oxygen sensor due to the inconvenience of applying physical principals based equations.

3.1 System Identification

System identification is a process to find the mathematical functions correlating to the inputs and outputs of a dynamical system [25]. This system might also include disturbances. A system identification characterizes the behavior of the system and internal dynamics such as delays, speed, oscillations, and so on. The model based on experimental data should be suitable and compact with respect to the certain purpose it is to be used for.

If the mathematical model is derived from physical principles, it is referred to white box modeling approach. On the other hand, it is called black-box modeling approach, if the relationship between input and output is estimated from experimental data only. As a result of mixing the white box modeling principle and empirical estimation of the model parameters is called as grey-box modeling.

A typical procedure of system identification includes input and output selection first. Then an experiment should be designed to collect required input and output data. After getting enough amount of experimental data, the model structure is selected from a predefined set of possible solutions. As a next step, the relevant parameters of the model are estimated. Finally, the model is validated with a set of data which is different than the first data used in modeling [25].

Difference or differential equations can describe the connections between the system variables of the mathematical models. These models can be separated as time continuous or time discrete, linear or nonlinear, deterministic or stochastic, etc. These features determine the type of the equations. The steps to follow are different from the systems having various characteristics. Therefore, before starting the model construction, the specific characteristics of the system should be determined. For example, the system should be classified as continuous or intermittent time, dynamic or static system that changes with time or does not change with time. The methods applied in the identification depend on the assumptions that are made on the model structure [25]. To differentiate the system identification methods clearly, they can be separated as nonparametric and parametric identification. To estimate the first parameters of the already defined model, a parametric identification is often preferred over the less accurate nonparametric model estimation [25]. Still its useful to apply the nonparametric model estimation to gain more detailed pieces of information about the system to help to define parameters like time delay, model order or if preconditioning is required. Even though there are several different identification techniques determined under parametric and nonparametric identification methods in the literature, only the methods discussed in the scope of this study are highlighted in the 3.2 and 3.3. In the following two sections, linear and nonlinear system modeling methods are detailed.

3.2 Linear System Identification

A linear model complies with the principles of superposition and homogeneity. This can be represented by the equations 3.1 and 3.2, respectively [25].

$$y_1 = f(u_1)$$

$$y_2 = f(u_2)$$

$$f(u_1 + u_2) = f(u_1) + f(u_2) = y_1 + y_2 \tag{3.1}$$

$$f(a_1 u_1) = a_1 f(u_1) = a_1 y_1 \tag{3.2}$$

The system inputs are named with u1 and u2 while y1 and y2 are the system outputs. A discrete linear polynomial model can be described by the following equation 3.3:

$$y(k) = z^{-n} * G(z^{-1}, \theta) * u(k) + H(z^{-1}, \theta) * e(k) \tag{3.3}$$

With u(k) as input and y(k) as the output of the system. The disturbance e(k) of the system is usually zero-mean. The transfer function of the deterministic part of the system is formed with white noise G(z–1, θ) while the stochastic part of the system is shown in the transfer function H(z–1, θ).

The deterministic transfer function specifies the relationship between the output and the input signal. The stochastic transfer function determines the random disturbances which affect the output signalfunction. In many cases, the deterministic part is referred to system dynamics, respectively the stochastic components of a system to stochastic dynamics. The backward shift operator term z–1 is defined by the equations 3.4, 3.5, 3.6.

$$z^{-1}x(k) = x(k-1) \tag{3.4}$$

$$z^{-2}x(k) = x(k-2) \tag{3.5}$$

$$\dots$$

$$z^{-n}x(k) = x(k-n) \tag{3.6}$$

z–n defines the number of delay samples between the input and the output. G(z–1, θ) and H(z–1, θ) are rational polynomials as defined by the equations 3.7 and 3.8.

$$G(z^{-1}, \theta) = \frac{B(z,\theta)}{A(z,\theta) * F(z,\theta)} \tag{3.7}$$

$$H(z^{-1}, \theta) = \frac{C(z,\theta)}{A(z,\theta) * D(z,\theta)} \tag{3.8}$$

The vector θ is the set of model parameters. Equations in the following sections of this manual do not display θ to make the equations easier to read. The equations from 3.9 to 3.13 define A(z), B(z), C(z), D(z), and F(z).

$$A_{(z)} = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_{k_a} z^{-k_a} \tag{3.9}$$

$$B_{(z)} = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_{k_{b-1}} z^{-(k_b-1)} \tag{3.10}$$

$$C_{(z)} = 1 + c_1 z^{-1} + c_2 z^{-2} + \dots + c_{k_c} z^{-k_c} \tag{3.11}$$

$$D_{(z)} = 1 + d_1 z^{-1} + d_2 z^{-2} + \ldots + d_{k_d} z^{-k_d} \quad (3.12)$$

$$F_{(z)} = 1 + f_1 z^{-1} + f_2 z^{-2} + \ldots + f_{k_f} z^{-k_f} \quad (3.13)$$

Equation 3.14 describes a general-linear polynomial model where ka, kb, kc, kd, and kf are the model orders.

$$A_{(z)} y_{(k)} = \frac{z^{-n} B_{(z)}}{F_{(z)}} * u_{(k)} + \frac{C_{(z)}}{D_{(z)}} * e_{(k)} = \frac{B_{(z)}}{F_{(z)}} * u_{(k-n)} + \frac{C_{(z)}}{D_{(z)}} * e_k \quad (3.15)$$

Particular attention is given to the impulse response model which a nonparametric identification method, state space model and finally auto regressive exogeneous (ARX) model within the scope of linear system identification.

3.2.1 Impulse response model

An impulse in a discrete system specifies a physical impulse using unit-amplitudes for the first sample period and zero-amplitudes all other times. To compute the output y(k) of the system, the equation 3.16 can be used if the input signal u(k) and impulse response h(n) is known.

$$y_{(k)} = \sum_{k=-\infty}^{\infty} u_{(k-n)} * h_{(n)} + e_{(k)} \quad (3.16)$$

where e(k) is the disturbance of the system.

3.2.2 State space model

Difference or differential equations with auxiliary state vector are used within the discrete state-space model. This model is described with the equations 3.17 and 3.18.

$$x_{(k+1)} = Ax_{(k)} + Bu_{(k)} + Ke_{(k)} \quad (3.17)$$

$$y_{(k)} = Cx_{(k)} + Du_{(k)} + e_{(k)} \quad (3.18)$$

where

- x: the state vector
- k: the model sampling time multiplied by the discrete time step, where the discrete time step equals 0, 1, 2, ...
- t: the time for the continuous model
- A: the system matrix that describes the dynamics of the states of the system

- B: the input matrix that relates the inputs to the states
- C: the output matrix that relates the outputs to the states
- D: the transmission matrix that relates the inputs to the outputs
- K: the Kalman gain.

Physical characteristics of a system are reflected by the state-space transfer matrices (A, B, C, D). For describing multivariable systems the state-space model is known for its convenience and therefore preferably used, especially in modern control applications, as polynomial models focusing on multivariable systems.

3.2.3 ARX model

For reducing the general-linear-polynomial model to an ARX model, C(z), D(z), and F(z) need to be equal 1.

$$A_{(z)}y_{(k)} = z^{-n}B_{(z)}u_{(k)} + e_{(k)} = B_{(z)}u_{(k-n)} + e_{(k)} \tag{3.20}$$

$$\mathrm{y(t)} + a_{(1)}\mathrm{y(t-1)} + \cdots + a_{(na)}\mathrm{y}(\mathrm{t} - n_{\mathrm{a}}) = b_1\mathrm{u}(\mathrm{t} - n_{\mathrm{k}}) + \cdots + b_{nb}\mathrm{u}(\mathrm{t} - n_{\mathrm{k}} - n_b + 1) + \mathrm{e(t)} \tag{3.21}$$

For a discrete-time SISO system, y(t) is the output at time t whereas u(t) is the input at time t, na represents the number of poles, nb represents the number of zeros plus 1, nk shows the input delay. Input delay means the number of samples before the input affects the system output. e(t) is the white-noise disturbance.

It is possible to bias the estimation of the ARX model while using the coupling between deterministic and stochastic dynamics, if the disturbance e(k) of a system is not white noise. Especially if the signal-to-noise ratio is low, a way to minimize the estimation error can be to set the model higher than the actual model order. However, some dynamic characteristics, like the overall stability of the model, can be affected by an increased model order. The method to identify the ARX model is a particular case of the prediction error method called the least-squares approach.

3.3 Nonlinear System Identification

The principles of superposition or homogeneity do not comply with nonlinear models, unlike the linear models. Real-world systems such as relays, switches and rate limiters are showing nonlinear effects like saturation, dead-zone, friction, backlash, and quantization. Despite the fact that most real-world systems are nonlinear, the use of a linear model, to simplify the analysis procedure or even the design, can simulate most real-world systems. For defining an identification method for a nonlinear system with effective function approximators, neural networks are known to be reliable if there is a lack of physical insight about the system or mathematical difficulties with physical modeling. Neural networks just require the behavior of the system to be modeled, with no need for any physical modeling.

3.3.1 Neural network modeling

Over time many necessities of automatic control like to work with more and more complicated systems and satisfy stricter design criteria have occurred. At this point, the neural network modeling provides solutions to the previously mentioned concerns and gives the opportunity to fulfill the requirements with less and less a priori knowledge of the plant dynamics. The excellent performance for the approximation of nonlinear functions is the most critical capability of neural networks because of their ability to learn.

Currently, neural networks based on multilayer feedforward, back propagation learning or more efficient variations of this algorithm are used mostly for system identification in the industry [27].

The basic units of neural networks are called neurons. A number of those highly interconnected, identical or similar, simple processing units (neurons) are the structure of neural networks. Those neurons are arranged in ordered topology and are doing the local processing. To acquire the knowledge from their environment, they also have a learning algorithm using examples as well as a recall algorithm to be able to use the learned knowledge later on.

Artificial neurons can be visualized as shown in Figure 3.1.

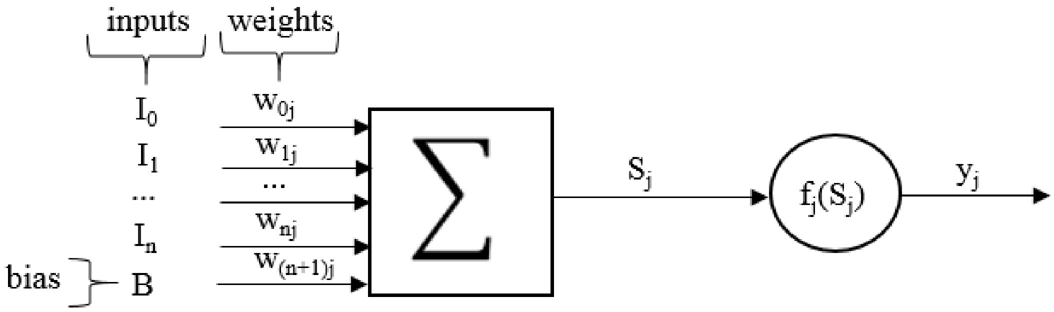

Figure 3.1: Internal structure of an artificial neuron

A specific weight representing the strength of connections between units is added to each connection and each input. The corresponding connections weights are multiplied with all the inputs to compute the state when the neuron is activated as shown in equation 3.21. Additionally, an extra and separate weighted input, the bias, is included and contains a constant value of one.

$$s_j = \sum_i w_{ij} * u_i \tag{3.21}$$

Where ui represents all the inputs including bias. After the computing of the state, the neuron needs to pass through its activation function. Activation function also normalizes the results. Equation 3.22 gives the output.

$$y_i = f_j(S_j) \tag{3.22}$$

The activation functions can be summarized as:

Linear Function

$$f_{(u)} = au + b \tag{3.23}$$

Sigmoid Function

$$f_{(u)} = \frac{1}{(1+e^{-u})} \tag{3.24}$$

Hyperbolic Tangent Function

$$f_{(u)} = \frac{2}{(1+e^{-2u})-1} \tag{3.25}$$

Rectified Linear Unit (ReLU) Function

$$f_{(u)} = max_{(0,u)} \tag{3.26}$$

The components of an artificial neural network are the input and output layers as well as hidden layers in between as in Figure 3.2. Starting with the first layer (as input layer) which receives its inputs from the environment and provides its outputs as input for the next layer, until the final layer is reached. The end of the process gives the output through the output layer.

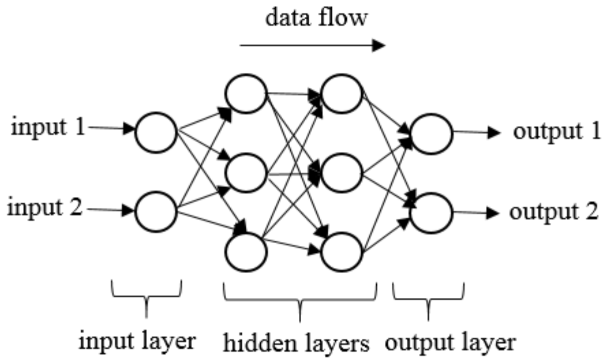

Figure 3.2: Multi-layer artificial neural network

The algorithm to make the learning possible by training, is called backpropagation. To produce an output, an input is given first. The teaching part contains a comparison of the actual output to what should have been ideal or target output for the specific input. To produce a more accurate output layer, the network can adjust the weights by going backward, starting with the output layer until the input layer is reached. When providing the same input to the network, it will give an output much closer to the ideal one used to train to the network. This process needs to be repeated until the calculated error between the network output and the typical output is at an acceptable level. Figure 3.3 shows a typical learning structure.

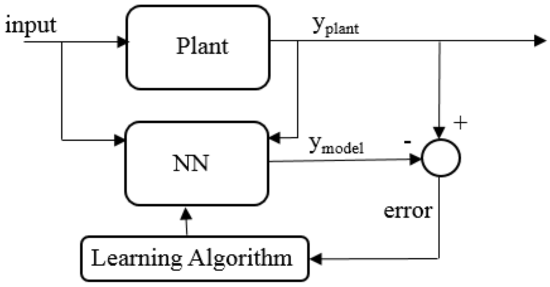

Figure 3.3: Learning structure

There are different types of learning in artificial neural networks. Supervised learning is the learning method where both inputs and outputs are provided to the neural network whereas in unsupervised learning only inputs are provided.

There are seven steps within the workflow of the general neural network design process:

- data collection
- network creation
- network configuration
- Initial values for weights and biases
- network training
- network validation or post-training analysis
- network usage

3.3.2 NARX

The NARX or nonlinear-autoregressive-exogenous model is one of the feedforward neural networks. The NARX network is used to model time series as a special form of the linear ARX [28]. This model relates both, the past and current values of the inputs as well as the past values of the time series, to the present output of a time series. Figure 3.4 represents the generalize NARX network [28].

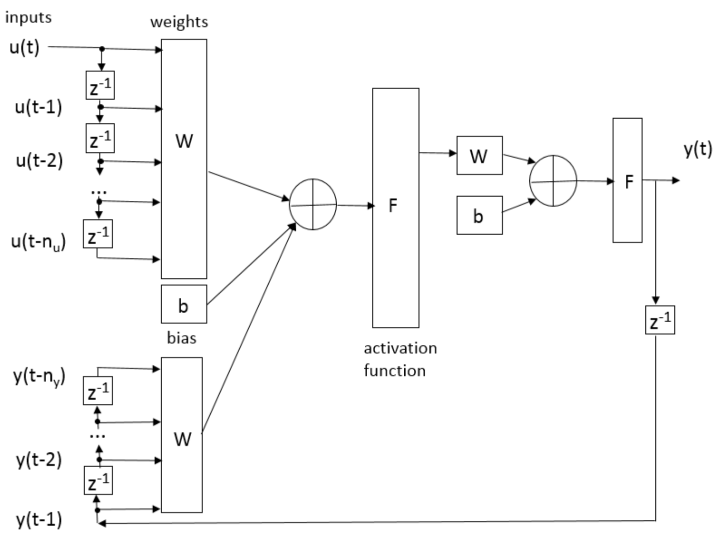

Figure 3.4: NARX network

Equation 3.27 shows the formulation of NARX network.

$$y_{(t)} = f\left(y_{(t-1)}, y_{(t-2)}, \dots\dots, y_{(t-n_y)}, u_{(t-1)}, u_{(t-2)}, \dots\dots, y_{(t-n_u)}\right) \quad (3.27)$$

3.4 Proposed Method

Designation the most suitable model for the system is the essential problem of the system identification. First, the input and output data should be observed, and then a convenient model should be searched among a set of candidate model structures. To be able to select the most suitable model, some criteria should be set. Therefore, the abovementioned linear system identification techniques are first tested, and then neural network based nonlinear system identification technique is applied to obtain a sufficient model of the oxygen sensor.

4. MODELING OF OXYGEN SENSOR

4.1 Experimental Design

The project studied in this thesis is subject to the EOBD regulations. For this reason, the error cases of lag in sensor response during the transition for both directions, rich-to-lean and lean-to-rich have been taken into consideration.

In this application, the system identification is handled for a UEGO sensor on a light duty gasoline vehicle with different approaches. As shown in Figure 4.1, the UEGO sensor is located after the turbocharger and before the TWC in the exhaust line. After the TWC, there is a Gasoline Particulate Filter (GPF) followed by a HEGO sensor. The UEGO sensor measures the oxygen concentration in the exhaust gas resulting from the combustion of the air and fuel mixture.

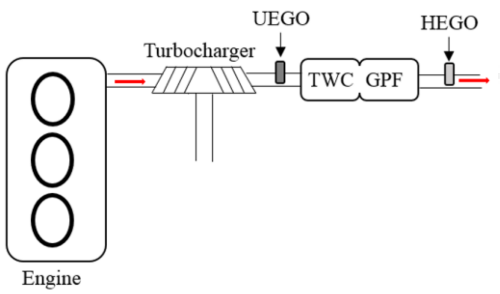

Figure 4.1: Illustration of the experimental setup

The ECU calculates the amount of air and fuel involved in the combustion based on the torque calculation algorithm and other parameters affecting the combustion dynamics. A wide variety of data is collected on the vehicle before system identification, so that the model to be constructed can work properly for different combustion conditions. Instead of testing on the road, data is collected in chassis dynamometer under different steady state and transient driving maneuvers. Figure 4.2, Figure 4.3 and Figure 4.4 show the vehicle speed, engine speed and torque traces of the data used in this study.

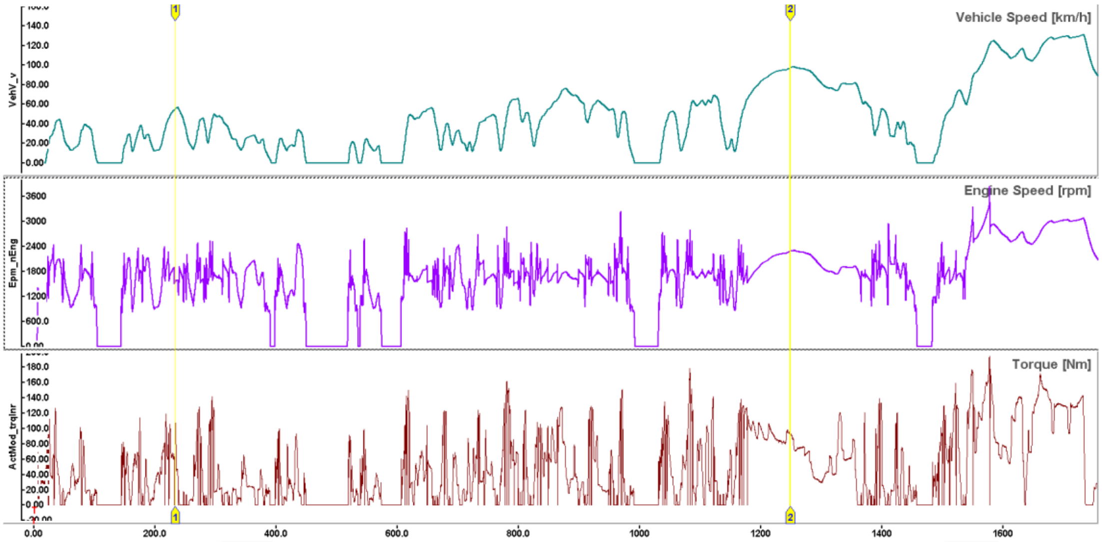

Figure 4.2: Worldwide Harmonized Light Vehicles Test Cycle (WLTC)

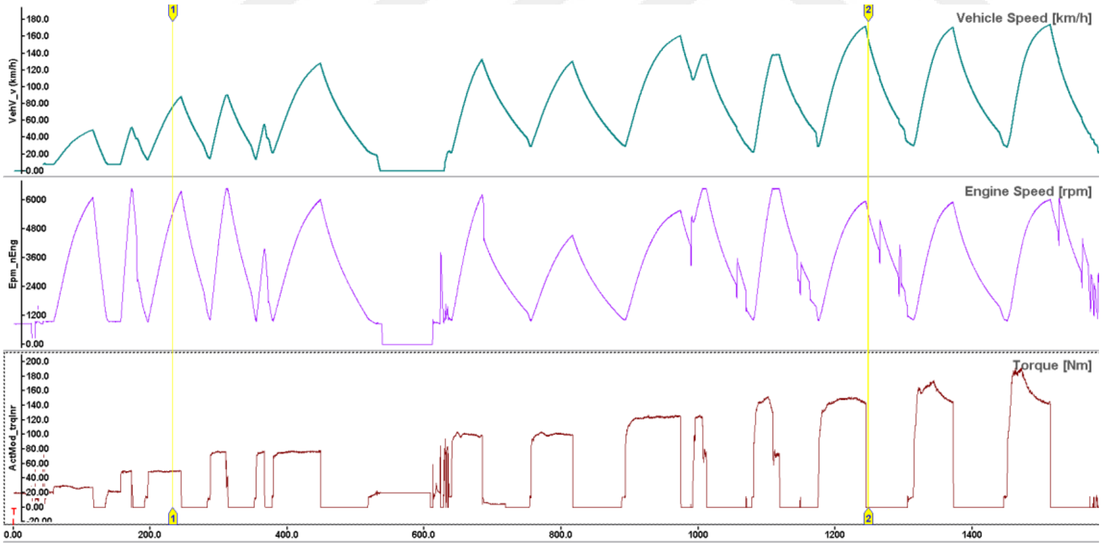

Figure 4.3: Engine Speed – Torque Sweep Measurement

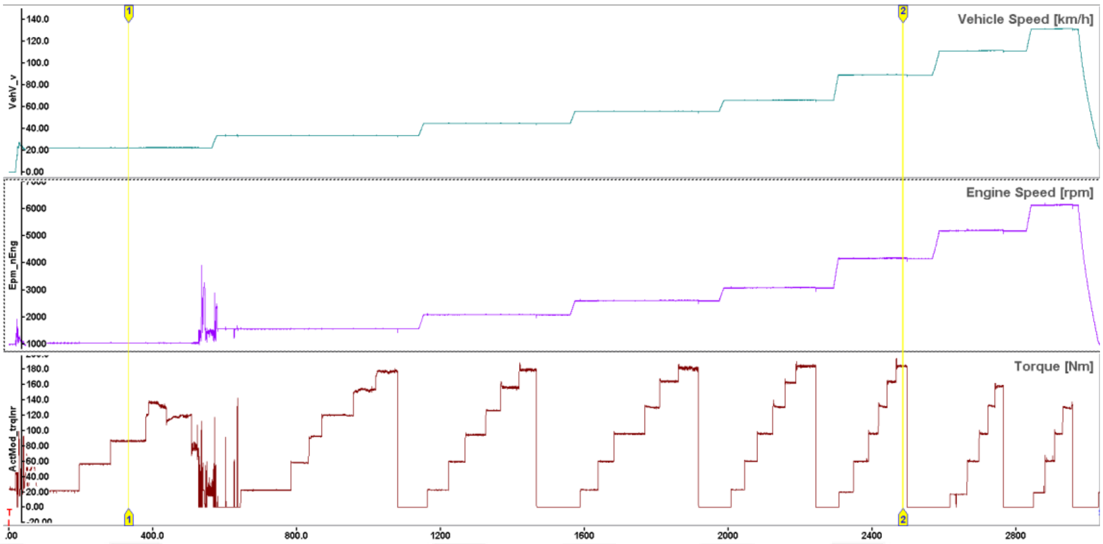

Figure 4.4: Steady State Mapping Measurement

Calibration software ATI is used to collect data on the vehicle. MATLAB is used to process the data for system identification and model-based diagnostic.

4.2 Input Selection for System Identification

One of the most important parameters to get a sufficient system identification is choosing the correct inputs. Otherwise, it is possible to face obtaining irrelevant and unsatisfactory results. Firstly, the physics of the system is considered, and signals that might influence the output are determined. In the current application of UEGO sensor, the system is a Multi Input Single Output (MISO) system. The output is the predicted UEGO sensor signal. Regarding inputs, first, some general parameters such as vehicle speed, engine speed, and torque are selected. Then, air and fuel mass are added to the input lists as main contributors of the combustion, to get more accurate and general outputs. By taking the physics of the combustion into consideration, a reflection of the environmental effects and disturbances to the system is possible. The ideal gas equation is used to add the parameters affecting the amount of air in the cylinder to the input list. Equation (4.1) and equation (4.2) show the general gas equation and standardized air mass calculation respectively.

$$p * V = m * R * T \tag{4.1}$$

$$m_{standardized} = \frac{p_0 . V_h}{R. T_0}, \quad p0=1013 \text{ hPa}, \quad T0=273 \text{ K} \tag{4.2}$$

In equation (4.1), p and V stand for pressure and volume respectively whereas m stands for mass, R is gas constant, and T is temperature. Vh in Equation (4.2) is the cylinder displacement volume. When this ideal gas equation is referred to the combustion chamber, the air mass in the combustion chamber, mcomb,air can be shown by equation (4.3).

$$m_{comb,air} = \frac{p_{comb} . V_h}{R. T_{comb}} = \frac{(p_{man} - p_{res}) . V_h}{R. T_{comb}} \tag{4.3}$$

Where p_{comb} is partial air pressure in the combustion chamber, p_{res} is partial pressure of internal residual gas and p_{man} is intake manifold pressure. Also, T_{comb} stands for gas temperature in the combustion chamber.

As a result, it is seen that Mass Air Flow (MAF) sensor signal, manifold temperature, manifold pressure, and throttle angle might influence the system identification process regarding air mass calculation. Finally, rail pressure is also taken into consideration since it affects the fuel quantity injected into the combustion chamber and the lambda controller output due to its influence on closed loop controller as well. After deciding on possible input parameters, Principal Component Analysis (PCA) method is applied to specify the final inputs to use in system identification process. PCA is one of the common techniques in input selection due to its positive impact on improving the robustness of the identification [30]. PCA helps to find which variables are most strongly correlated with each component. During the calculations, PCA assumes that the input data set is a linear combination of the variables. According to PCA, the components with larger variance correspond to remarkable dynamics whereas the lower variation corresponds to noise. In this thesis, the PCA Toolbox for MATLAB is used for the analysis of the input data [30].

Inputs for PCA analysis are vehicle speed, engine speed, torque, air mass, fuel mass, intake manifold temperature and pressure, throttle valve position, MAF output, rail pressure and lambda controller output whereas the only output is UEGO sensor signal.

It is seen in Figure 4.5 that PC1, PC2 and PC3 cover approximately 94% of the total variance.

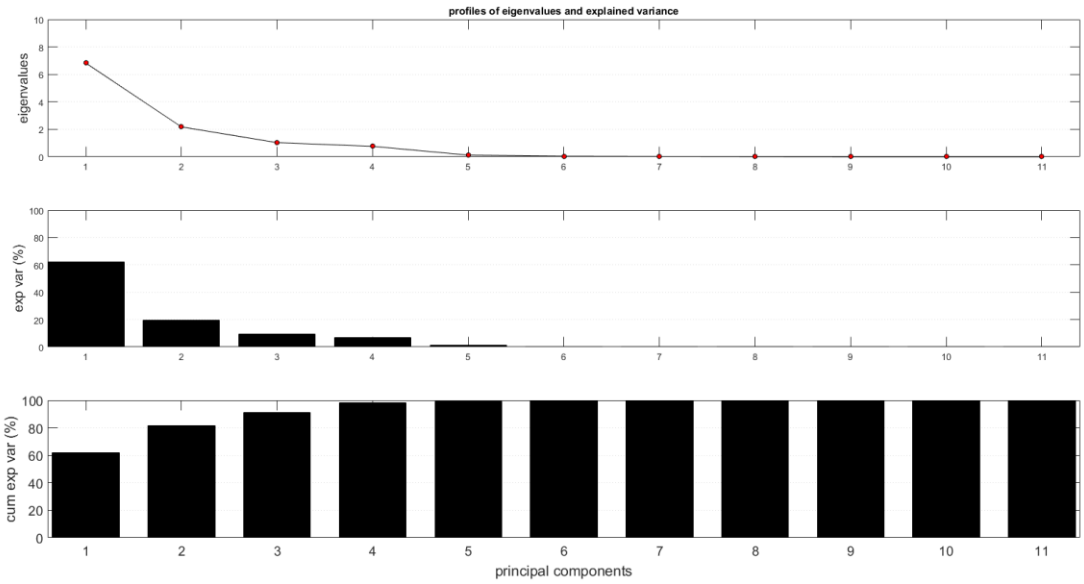

Figure 4.5: Variations of the components

Table 4.1 represents the correlations between abovementioned input parameters. In the Table 4.1, the numbers having large absolute magnitudes are the most strongly correlated parameters with the relevant component.

Table 4.1: Eigenvector analysis of the correlation matrix.

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Torque	0.38	0.06	-0.04	0.12	-0.10	-0.02	0.10	-0.40	-0.53	0.61	0.00
Engine Speed	0.01	-0.68	-0.03	-0.08	0.06	-0.01	0.17	-0.07	-0.01	0.00	0.71
Air	0.38	0.09	-0.03	0.10	-0.08	-0.01	0.29	-0.29	-0.29	-0.76	0.00
Fuel	0.38	0.10	-0.02	0.11	-0.09	0.01	0.35	-0.27	0.78	0.16	0.00
MAF	0.36	-0.18	-0.02	0.07	-0.02	-0.26	-0.82	-0.22	0.16	-0.11	0.00
Rail Pressure	0.34	-0.15	-0.14	0.36	-0.46	0.41	-0.06	0.57	-0.04	0.02	0.00
Intake Manifold Temperature	0.28	0.04	0.29	-0.70	0.03	0.57	-0.12	-0.05	0.00	0.00	0.00
Intake Manifold Pressure	0.35	0.04	0.19	-0.38	-0.14	-0.65	0.18	0.47	-0.06	0.05	0.00
Throttle	0.36	0.04	-0.11	0.18	0.86	0.09	0.03	0.29	-0.03	0.04	0.00
Lambda Controller	-0.02	-0.08	0.92	0.38	0.04	0.02	0.01	0.00	-0.01	0.00	0.00
Vehicle Speed	0.01	-0.68	-0.03	-0.08	0.06	-0.01	0.17	-0.07	-0.01	0.00	-0.71

Loading plot in Figure 4.6, can be used to define which variables have the most significant effect on each component as well.

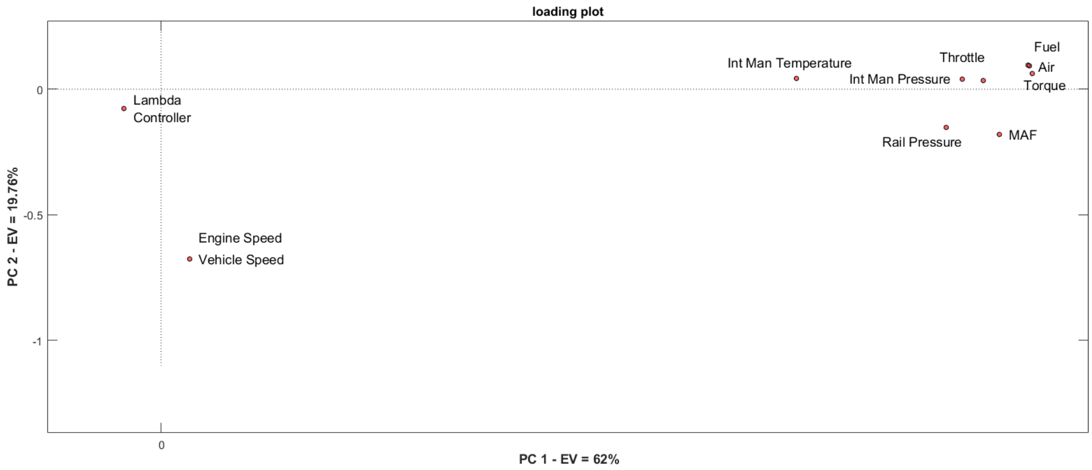

Figure 4.6: Loading plot of PC1 vs. PC2

The range for loading values is from -1 to 1. The loading values near to -1 and 1 shows the variable strongly affects the component. On the other hands, loading close to 0 means that the variable has a weak impact on the component. In this loading plot, air, fuel, torque, MAF, Throttle, Rail pressure, intake manifold pressure, and temperature have large positive loadings on component 1, so this component mainly identifies the UEGO sensor output. Engine speed and vehicle speed have large negative loadings on component 2. By considering the PCA results and combustion dynamics together, initial input variables are defined as air, fuel, MAF, rail pressure, intake manifold pressure, intake manifold temperature and lambda controller output.

4.3 System Identification with Linear System Approach

The system is a procedure to construct the mathematical model of a system. A model refers to a series of mathematical equations between inputs and outputs of a system. Usually, such model structure involves unknown parameters. The relevant model is often a dynamic system. When the system is a dynamic system, the magnitude of the outputs consists in both the instantaneous values of the inputs and the past behavior of the system. And system identification makes use of the system's premeasured inputs and outputs.

System identification procedure includes steps that should be followed one by one. Firstly, the inputs and outputs of the system are measured either in time or frequency domain. Then a candidate model structure is chosen. An estimation method which is suitable for the preselected model structure is applied to calculate the unknown parameters. Finally, the obtained model and the parameters are evaluated to see the suitability with the application. In case of getting unsatisfying results, another model structure might be selected. This modeling approach also used in this thesis is called black-box modeling. System Identification Toolbox in Matlab is used for system identification steps.

As mentioned in the previous section, input selection is an essential step before starting the system identification procedure, but it is not enough. These preselected inputs should process and prepared for the identification. As a first step, data can be plotted and checked with bare eyes to see if there are any outliers in the data. Besides, the data needs to be filtered in case of noise in the system. In this study, the Fourier Transform Analysis is used to check the existence of noise. The Fourier Transform analyzes the signal in the frequency domain. In Figure 4.7, the amplitude vs. frequency plot shows which frequencies are dominant. A specific frequency that involves noise is determined in this plot, and a suitable filter can be constructed to filter the noise in the system out. The data is filtered for the frequency range 2.4-2.5 Hz based on the analysis result.

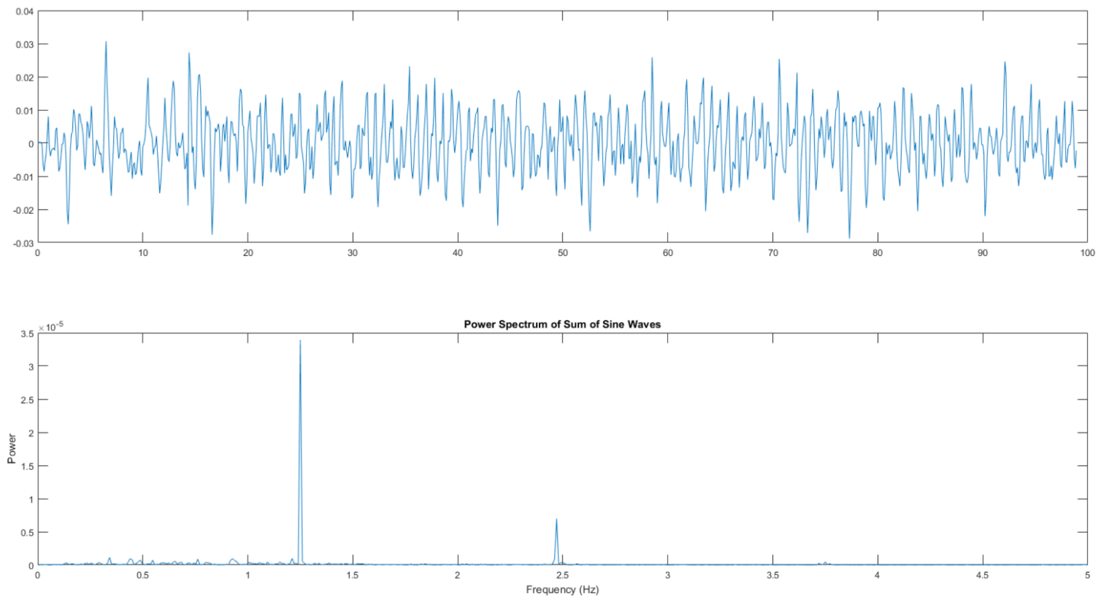

Figure 4.7: Fourier Transform Analysis Results

After removing the outliers, and filtering the data, the scale of the different variables should be considered. Since the input parameters define different physical features like pressure, mass flow, temperature and so on, there are big differences between the magnitudes of the signal. Therefore, both input and output data series should be normalized to a common scale. Otherwise, the differences in a variable range might affect the modeling performance negatively. At this stage, the z-score normalization method shown in equation 4.4 is used for each input and output parameters.

$$\boldsymbol{z_i = \frac{x_i - x_{mean}}{s}} \tag{4.4}$$

Where z_i is the normalized value of input x_i , x_{mean} is the mean value of the relevant signal and s is the standard deviation. Another preprocessing operation is detrending the data. Detrending means removing means, offsets or linear trends. The detrending provides more accurate linear models because of eliminating the random differences between the input and output levels. After all these preprocessing steps, data is split as identification and validation dataset.

In black-box modeling, the process is usually based on a trial-and-error method by estimating the model parameters for different model structures and finding the best model reflecting the input-output relationship. The best way is the starting with simple structures and then moving to more complex models. At this step, model order and model delay can be estimated as well. In this study, various input combinations are also tried next to different model structures. As the starting point, Quick Start function of the System Identification Toolbox is used. The Toolbox uses minimization criterion to estimate model parameters by minimizing the deviation between the model response and the measured output. By using Quick Start, State-space model using n4sid, fourth order ARX model and step response over a period, the impulse algorithms are estimated. In Table 4.2, the performance values of these models for various inputs are listed. 100 indicates a perfect fit, and 0 is a poor fit.

Table 4.2: Linear System Identification Results from 1000 RPM to 3100 RPM.

Model Input (s):	Air Fuel MAF Rail Pressure Intake Manifold Pressure Intake Manifold Temperature Lambda Controller	Air Fuel	Air Fuel MAF	Air Fuel Lambda Controller	Air Fuel MAF Lambda Controller	Air Fuel Rail Pressure	Air Fuel Rail Pressure Lambda Controller	Air Fuel Rail Pressure Intake Manifold Pressure Intake Manifold Temperature	Air Fuel MAF Intake Manifold Pressure Intake Manifold Temperature
Model Output:	Oxygen Sensor Output								
Impulse Response Model:	imp: 32	imp: 23	imp: 21	imp: 35	imp: 34	imp: 22	imp: 34	imp: 18	imp: 16
Discrete time – identified State – Space Model:	n4s6: 29	n4s6: 15	n4s8: 15	n4s8: 31	n4s6: 31	n4s8: 11	n4s5: 30	n4s5: 12	n4s6: 13
ARX Model:	arxqs: 21	arxqs: 9	arxqs: 9	arxqs: 20	arxqs: 20	arxqs: 10	arxqs: 20	arxqs: 10	arxqs: 9

As seen in Table 4.2 and Figure 4.8 model performance values with the max value of 40% are not sufficient for system identification. Therefore, data is split in smaller portions based on engine speed ranges and processed separately.

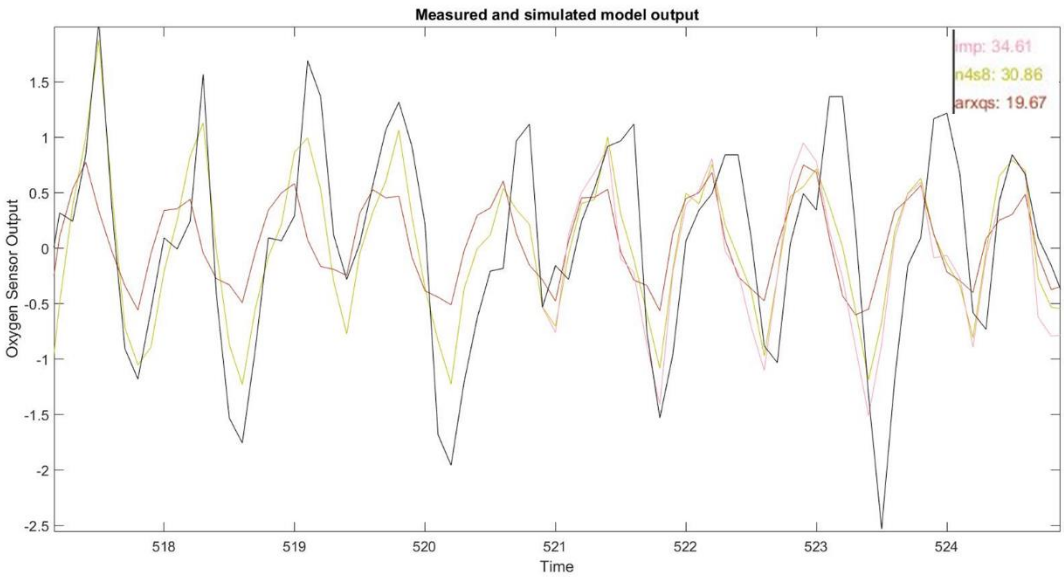

Figure 4.8: Best-Fit Analysis of the data range from 1000 RPM to 3100 RPM.

The measured and simulated model output signals for the input configuration, giving the best results shown for different engine speed ranges in Table 4.3 - Table 4.7. Even though the results are improved for different engine speed ranges, there are only slight differences with max 50% best-fit value.

Table 4.3: Linear System Identification Results from 1000 RPM.

Model Input (s):	Air Fuel MAF Rail Pressure Intake Manifold Pressure Intake Manifold Temperature Lambda Controller	Air Fuel	Air Fuel MAF	Air Fuel Lambda Controller	Air Fuel MAF Lambda Controller	Air Fuel Rail Pressure	Air Fuel Rail Pressure Lambda Controller	Air Fuel Rail Pressure Intake Manifold Pressure Intake Manifold Temperature	Air Fuel MAF Intake Manifold Pressure Intake Manifold Temperature
Model Output:	Oxygen Sensor Output								
Impulse Response Model:	imp: 8	imp: 27	imp: 31	imp: 22	imp: 26	imp: 20	imp: 26	imp: 14	imp: 5
Discrete time – identified State – Space Model:	n4s5: 29	n4s7: 31	n4s6: 30	n4s6: 32	n4s6: 29	n4s7: 35	n4s5: 32	n4s5: 32	n4s6: 16
ARX Model:	arxqs: 28	arxqs: 25	arxqs: 26	arxqs: 29	arxqs: 19	arxqs: 20	arxqs: 13	arxqs: 32	arxqs: 25

Table 4.4: Linear System Identification Results from 1500 RPM.

Model Input (s):	Air Fuel MAF Rail Pressure Intake Manifold Pressure Intake Manifold Temperature Lambda Controller	Air Fuel	Air Fuel MAF	Air Fuel Lambda Controller	Air Fuel MAF Lambda Controller	Air Fuel Rail Pressure	Air Fuel Rail Pressure Lambda Controller	Air Fuel Rail Pressure Intake Manifold Pressure Intake Manifold Temperature	Air Fuel MAF Intake Manifold Pressure Intake Manifold Temperature
Model Output:	Oxygen Sensor Output								
Impulse Response Model:	imp: 27	imp: 25	imp: 20	imp: 31	imp: 30	imp: 27	imp: 35	imp: 21	imp: 11
Discrete time – identified State – Space Model:	n4s5: 25	n4s7: 11	n4s5: -4	n4s8: 30	n4s8: 28	n4s5: 10.5	n4s 8:31	n4s5: 13	n4s5: 8
ARX Model:	arxqs: 27	arxqs: 10	arxqs: 1	arxqs: 25	arxqs: 26	arxqs: 20	arxqs: 27	arxqs: 12	arxqs: 11

Table 4.5: Linear System Identification Results from 2000 RPM.

Model Input (s):	Air Fuel MAF Rail Pressure Intake Manifold Pressure Intake Manifold Temperature Lambda Controller	Air Fuel	Air Fuel MAF	Air Fuel Lambda Controller	Air Fuel MAF Lambda Controller	Air Fuel Rail Pressure	Air Fuel Rail Pressure Lambda Controller	Air Fuel Rail Pressure Intake Manifold Pressure Intake Manifold Temperature	Air Fuel MAF Intake Manifold Pressure Intake Manifold Temperature
Model Output:	Oxygen Sensor Output								
Impulse Response Model:	imp: 49	imp: 31	imp: 31	imp: 52	imp: 52	imp: 30	imp: 52	imp: 23	imp: 21
Discrete time – identified State – Space Model:	n4s7: 43	n4s6: 29	n4s5: 27	n4s8: 52	n4s7: 47	n4s6: 27	n4s8: 51	n4s5: 26	n4s6: 24
ARX Model:	arxqs: 35	arxqs: 8	arxqs: 8	arxqs: 39	arxqs: 22	arxqs: 10	arxqs: 40	arxqs: 1	arxqs: 15

Table 4.6: Linear System Identification Results from 2600 RPM.

Model Input (s):	Air Fuel MAF Rail Pressure Intake Manifold Pressure Intake Manifold Temperature Lambda Controller	Air Fuel	Air Fuel MAF	Air Fuel Lambda Controller	Air Fuel MAF Lambda Controller	Air Fuel Rail Pressure	Air Fuel Rail Pressure Lambda Controller	Air Fuel Rail Pressure Intake Manifold Pressure Intake Manifold Temperature	Air Fuel MAF Intake Manifold Pressure Intake Manifold Temperature
Model Output:	Oxygen Sensor Output								
Impulse Response Model:	imp: 16	imp: 29	imp: 27	imp: 35	imp: 33	imp: 28	imp: 35	imp: 1.5	imp: -7
Discrete time – identified State – Space Model:	n4s7: 31	n4s7: 23	n4s6: 24	n4s6: 29	n4s6: 30	n4s8: 25	n4s7: 30	n4s8: 18	n4s5: 20
ARX Model:	arxqs: 33	arxqs: 25	arxqs: 25	arxqs: 26	arxqs: 26	arxqs: 25	arxqs: 30	arxqs: 26	arxqs: 22.5

Table 4.7: Linear System Identification Results from 3100 RPM.

Model Input (s):	Air Fuel MAF Rail Pressure Intake Manifold Pressure Intake Manifold Temperature Lambda Controller	Air Fuel	Air Fuel MAF	Air Fuel Lambda Controller	Air Fuel MAF Lambda Controller	Air Fuel Rail Pressure	Air Fuel Rail Pressure Lambda Controller	Air Fuel Rail Pressure Intake Manifold Pressure Intake Manifold Temperature	Air Fuel MAF Intake Manifold Pressure Intake Manifold Temperature
Model Output:	Oxygen Sensor Output								
Impulse Response Model:	imp: 24	imp: 24	imp: 15	imp: 37	imp: 31	imp: 6	imp: 36	imp: -2	imp: -12
Discrete time – identified State – Space Model:	n4s8: 44	n4s9: 25	n4s6: 30	n4s8: 39	n4s6: 40	n4s6: 28	n4s7: 41	n4s6: 10	n4s7: 16
ARX Model:	arxqs: 26	arxqs: 25	arxqs: 20	arxqs: 32	arxqs: 32	arxqs: 28	arxqs: 40	arxqs: 32	arxqs: 28

The best-fit analyses of the model output for different engine speed ranges are also shown in Figure 4.9 – 4.13.

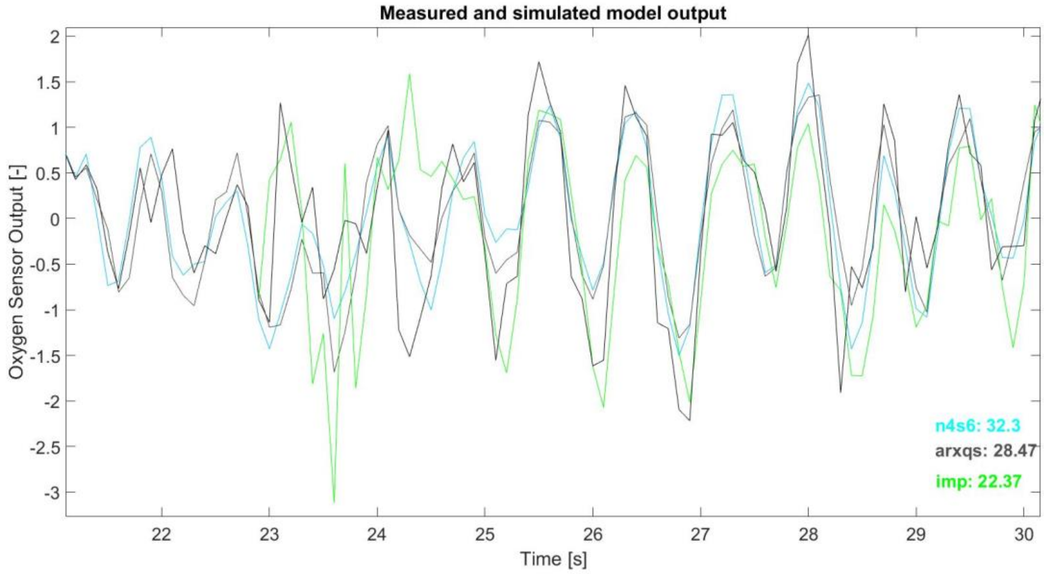

Figure 4.9: Best-Fit Analysis of the modeled output for 1000 RPM

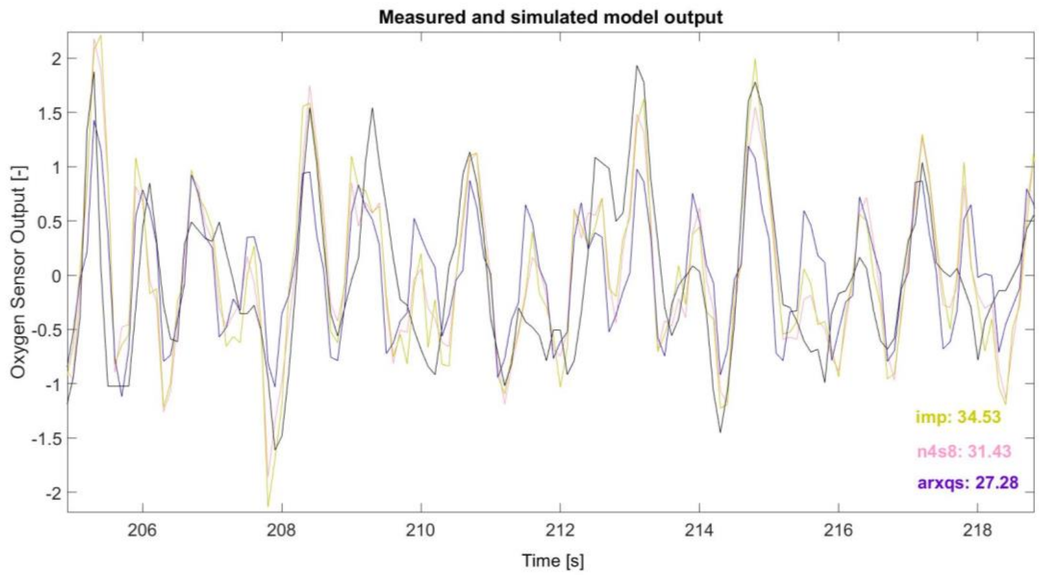

Figure 4.10: Best-Fit Analysis of the modeled output for 1500 RPM

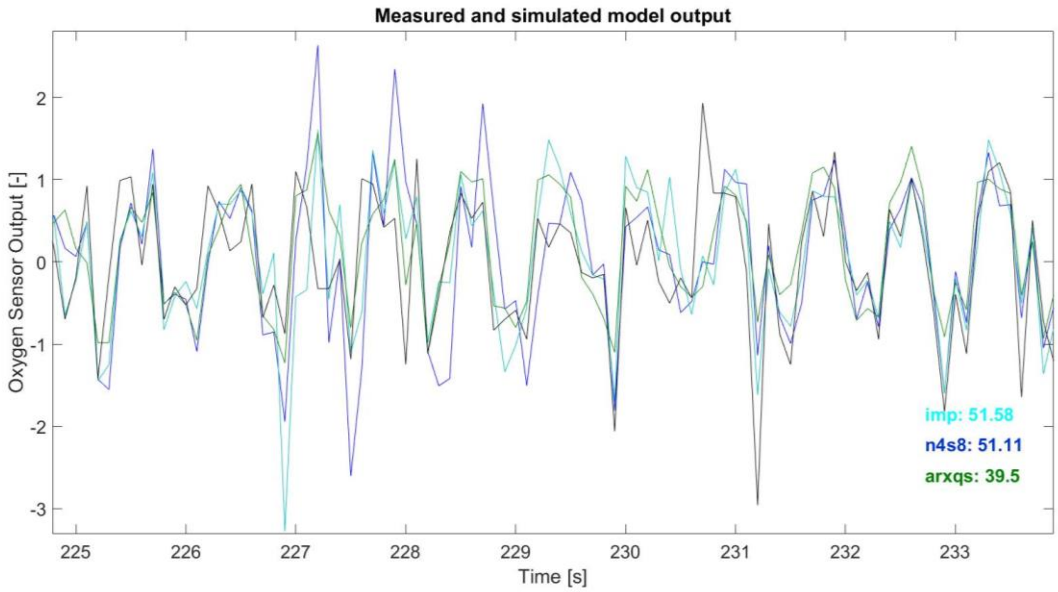

Figure 4.11: Best-Fit Analysis of the modeled output for 2000 RPM

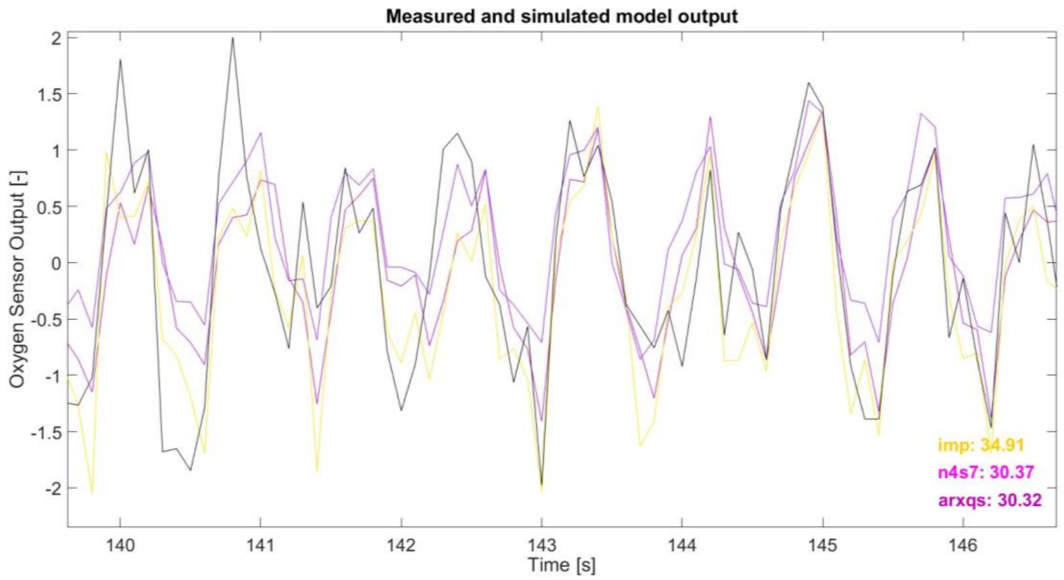

Figure 4.12: Best-Fit Analysis of the modeled output for 2600 RPM

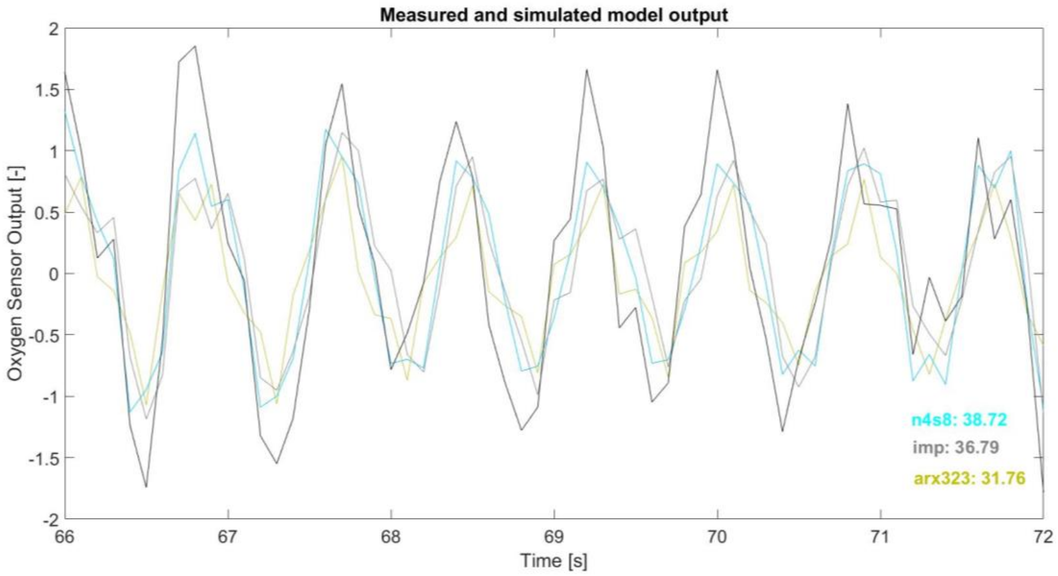

Figure 4.13: Best-Fit Analysis of the modeled output for 3100 RPM

There might be several reasons why the results are not sufficient so far. The ARX or state-space model may be unstable. These two models may ignore the effect of the feedback signal from the output to the input. The disturbances may have a significant impact on the system, and so they may need to be modeled as well. Higher order models might be required. The nonlinearities in the system might not be modeled with linear system identification.

At this point, firstly, the system delay is calculated, and the calculated system delay is compared with the raw data. Then, a few different model orders are tried for ARX model.

Figure 4.14 shows the graphic to select the best-fit ARX model. In Figure 4.14, the x-axis is the total number of poles, na, nb and zeros. The y-axis is unexplained output variance in percentage. The unexplained output variance means the portion of the output is not explained by the model. Three bars highlighted on the plot in green, blue, and red show a type of different best-fit criterion. The red bar is the best-fit minimizing the sum of the squares of the difference between the validation data output and the model output. The green bar indicates the best-fit minimizing Rissanen MDL criterion, whereas the blue bar is for the best-fit minimizing Akaike AIC criterion. Among all these three bars, the red bar has the overall best-fit. Besides, system delay, nk is calculated as three here.

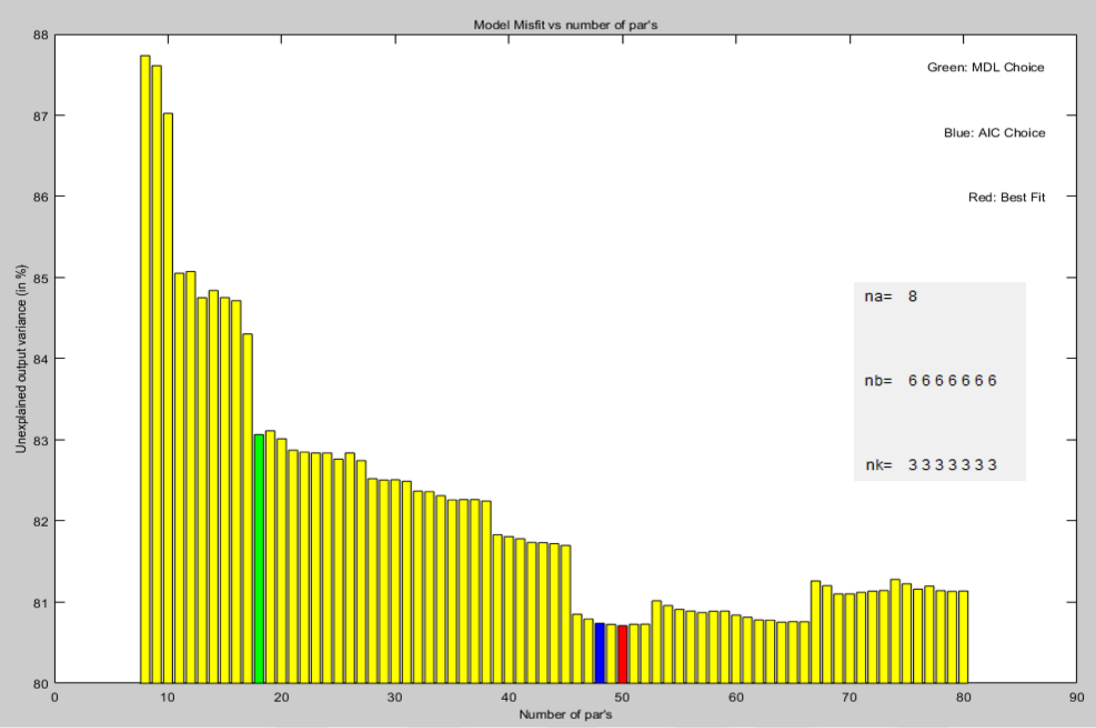

Figure 4.14: ARX model structure selection

In Figure 4.15, the sensor output and model outputs of various ARX models with different system orders are shown.

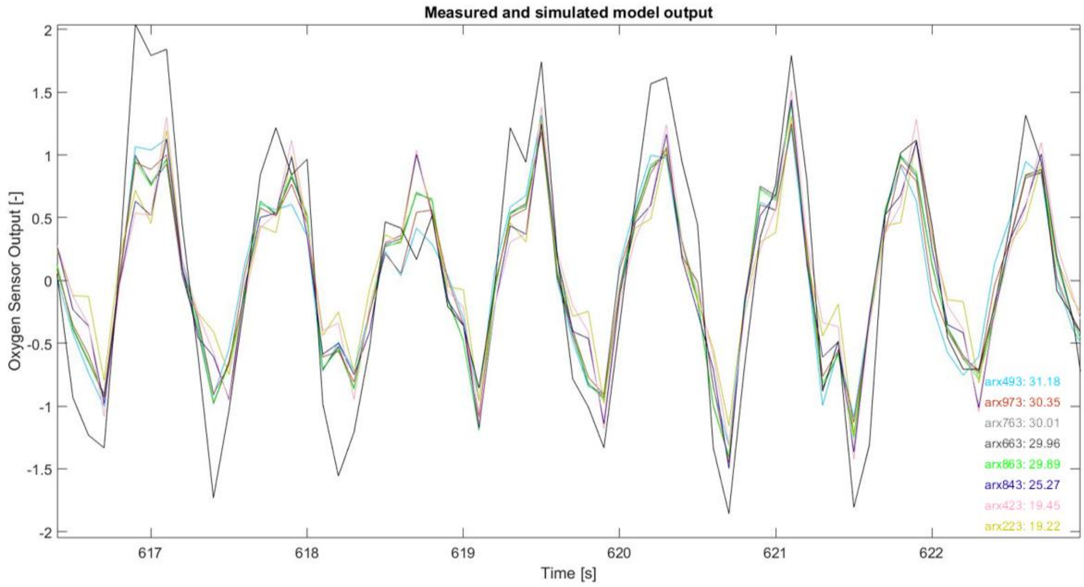

Figure 4.15: Measured and simulated ARX model outputs for different system orders

Since the results are far away from any acceptable values to improve with linear system identification method, nonlinear system identification methods are decided to be used.

4.4 System Identification with Neural Network

In this part of the thesis, nonlinear system identification is used for modeling since the linear model approaches being discussed in the previous section cannot capture the system dynamics. As a nonlinear system identification method, Neural Network approach is used.

One data set is used to train the recurrent neural network model, and one data set is used for validating the model. The PCA calculations and the results being obtained based on different inputs configuration in the previous sections are considered to select the inputs of the recurrent neural network model. As a result, system input parameters are selected as air, fuel, MAF sensor signal, rail pressure and lambda controller output. The output of the system is the oxygen sensor signal. Before moving through the training process, data normalization is an essential step. Better results and significantly fast calculations can be provided because of normalization of input data [31]. Min-max normalization method is used to preprocess the input data by using equation 4.5.

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{4.5}$$

Figure 4.16 shows the normalized inputs and output variables of the neural network.

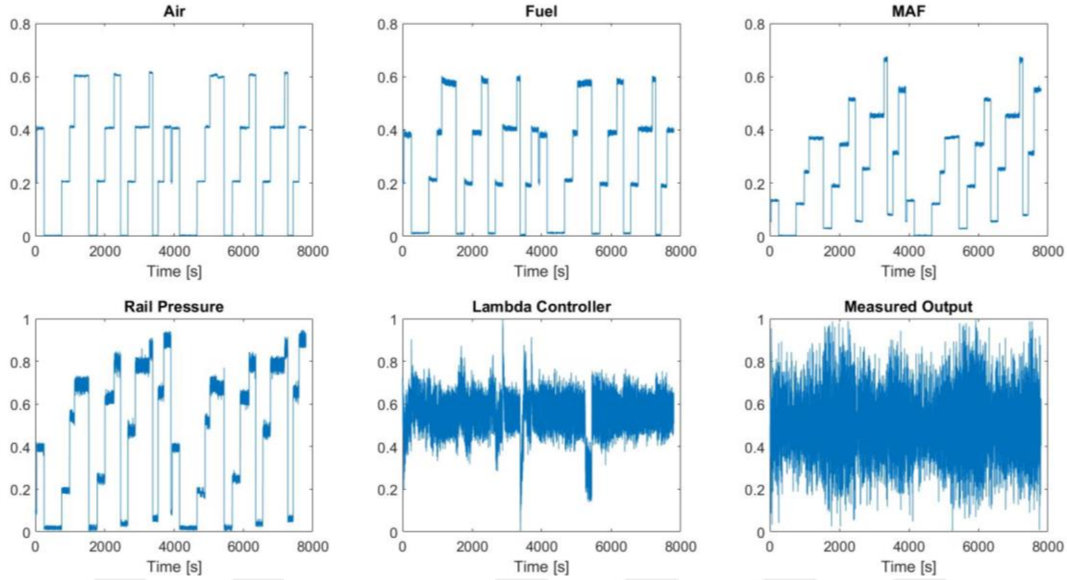

Figure 4.16: Normalized inputs and output variables of the neural network

Open loop structure where the delayed target data is used as an additional input is selected. In the open loop network structure, the output is generated based on the common time extent of the input signals and target output data. Other parameters of neural network structure to have a successful model are number of hidden layers, number of neurons and the training methods.

The selection of a proper number of hidden layers and an adequate amount of neurons is a significant problem to have a good model. In case of less hidden layers and neurons than necessary, considering the complexity of the system, the model might fail to meet the target output on a large scale of the data. This case is called as underfitting. On the other hand, in case of having too many hidden layers and neurons might cause overfitting. Overfitting causes to have output values that are so tightly fit a limited set of input data. Therefore, such model might have issues to fit another additional data set other than training data. Trial-and-error method to define the numbers of the hidden layers and neurons is a well-accepted way in the literature. The trial-and-error approach starts with training and testing the neural network by using a small number of hidden layers and neurons. Then according to the results, the number is increased. These attempts continue until getting sufficient model outputs [32].

Training algorithms are another crucial factor to enhance the model performance. Different algorithms might provide performances in a wide range due to their varied characteristics. Even though there are many different training algorithms available, only four of them are used to compare the performances in this study. These are Levenberg-Marquardt Backpropagation (LM) algorithm, Gradient Descent with Momentum and Adaptive Learning Rate Backpropagation (GDX) algorithm, Bayesian Regularization Backpropagation (BR) algorithm, and Scaled Conjugate Gradient Backpropagation (SCG) algorithm. The LM is a second-order numerical optimization technique. The advantage of LM is to combine features of Gauss-Newton and steepest descent algorithms. The LM has better convergence features than the conventional backpropagation algorithm with its very efficient performance when the network has maximum few hundredweights. The performance of GDX training algorithm depends on the learning rate. Besides, there is a momentum coefficient input since the GDX combines the advantage of adaptive learning rate and momentum coefficient. The BR training algorithm considers the goodness of fit next to network architecture. Therefore, the BR usually has the successful performance to overcome the overfitting problems in the neural network structure. Firstly, the BR minimizes the combination of squared errors and weights. After that, the BR specifies the appropriate combination. Thus, the network generalizes better. The conjugate gradient algorithms are mostly expensive regarding computations because they need that the network output to all training inputs is computed several times for each search. However, the SCG training algorithm avoids this time-consuming line search and reduces the number of computations performed in each iteration [33]. The entire dataset is divided into three different parts. The first 70% of the dataset is used to train the network while the next 15% is used for validation and last 15% is used for testing. Table 4.8 shows the results of correlation coefficients and performances for a wide variety of neural network structures with different number of hidden layers, number of nodes, training methods and delay values. After this trial-and-error method, the best NN model with high convergence performance is selected.

Since the system delay is calculated as three and validated by using raw data in previous sections, the delay value is started with three in neural network structure as well. Then the increased delay values are applied to the system to improve the performance output. In Table 4.8 correlation coefficients refers to the relationship between input dataset and output.

Table 4.8: NN Model performances of different configurations.

#	Correlation Coefficients			MSE	Input Delay	Output	# of HL	# of Neurons	LR	TM
	Training	Val.	Test							
1	0.735	0.678	0.690	0.013	3	y(k)	1	10	0.2	lm
2	0.760	0.695	0.653	0.013	3	y(k)	1	15	0.2	lm
3	0.752	0.705	0.706	0.012	3	y(k)	1	20	0.2	lm
4	0.784	0.696	0.673	0.012	3	y(k)	2	10-10	0.2	lm
5	0.748	0.708	0.690	0.012	3	y(k)	2	15-15	0.2	lm
6	0.666	0.664	0.651	0.015	3	y(k)	1	10	0.2	gdx
8	0.679	0.655	0.631	0.015	3	y(k)	1	20	0.2	gdx
9	0.679	0.654	0.635	0.015	3	y(k)	2	10-10	0.2	gdx
10	0.668	0.664	0.652	0.015	3	y(k)	2	15-15	0.2	gdx
11	0.705	0.754	0.639	0.014	3	y(k)	1	10	0.2	br
12	0.756	0.770	0.649	0.012	3	y(k)	1	20	0.2	br
13	0.889	0.891	0.876	0.006	4	y(k)	1	20	0.2	lm
14	0.932	0.865	0.797	0.005	5	y(k)	1	20	0.2	lm
15	0.908	0.908	0.886	0.005	5	y(k)	1	20	0.2	scg
16	0.919	0.896	0.861	0.005	5	y(k)	1	30	0.2	lm
17	0.910	0.908	0.869	0.005	5	y(k)	1	30	0.2	scg
18	0.927	0.916	0.872	0.004	5	y(k)	2	15-15	0.2	lm
19	0.966	0.904	0.745	0.005	5	y(k)	2	20-20	0.2	lm
20	0.946	0.896	0.787	0.005	5	y(k)	2	15-15	0.4	lm
21	0.908	0.905	0.881	0.005	5	y(k)	2	15-15	0.1	lm
22	0.927	0.914	0.874	0.004	5	y(k)	2	20-20	0.2	br
23	0.925	0.914	0.884	0.004	5	y(k)	2	20-20	0.2	scg
24	0.952	0.914	0.880	0.003	5	y(k)	2	30-30	0.2	scg
25	0.804	0.841	0.787	0.009	5	y(k)	2	20-20	0.2	gdx
26	0.970	0.940	0.884	0.002	5	y(k)	2	13-13	0.2	lm
27	0.963	0.937	0.905	0.002	5	y(k)	2	12-12	0.2	lm
28	0.961	0.952	0.905	0.002	5	y(k)	2	8-8	0.2	lm
29	0.965	0.945	0.903	0.002	5	y(k)	2	10-10	0.2	lm
30	0.939	0.911	0.823	0.004	8	y(k)	1	20	0.2	lm

The value can change from 0 to 1. While 1 indicates that two variables are perfectly correlated a 0 means poor correlation. While choosing the best-fit model, a trade-off between overfitting to training data and underfit to test and validation data must be considered. If the accuracy of the training data set is increasing, but the accuracy of validation dataset stays same or decreases, then that means the current neural network structure overfits. In addition to correlation coefficients, mean squared error (MSE) calculation is used to measure the performance of the model structure. The MSE takes the errors between the target value and the model output and squares them to remove any negative signs. Then the MSE calculates the average of a set of errors. The smaller MSE value indicates having a better fit. According to the results of the models 1, 6 and 11, the LM gives the best performance compared to GDX and BR for this application. On the other hand, if models 14-17 are analyzed, it is seen that there are no significant differences in the performance of LM and SCG. In this case, LM is recommended due to the speed of data processing. When models 18, 20, 21 is analyzed, it is seen that increase learning rate improves the correlation coefficients for training dataset, but also decreases the performance of test and validation sets. Therefore, the learning rate is selected as 0.2. Also, it is observed that increasing the delay value from three to five enhances the performance significantly as past values of the system have a substantial impact on the results. Using two hidden layers instead of only one hidden layer gives better results for different configurations of the number of neurons, delays, learning rates or training algorithms. While deciding on the number of neurons, the optimization between overfitting and improved overall results are taken into consideration. As a result, the Model 28 is selected to use for model-based diagnostics calculations. Figure 4.17 represents the model structure. Figure 4.18– Figure 4.20 show the performance outputs of the Model 28.

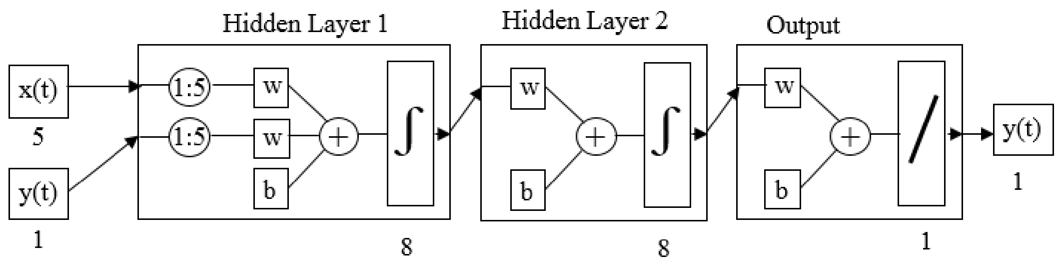

Figure 4.17: The chosen NARX model structure

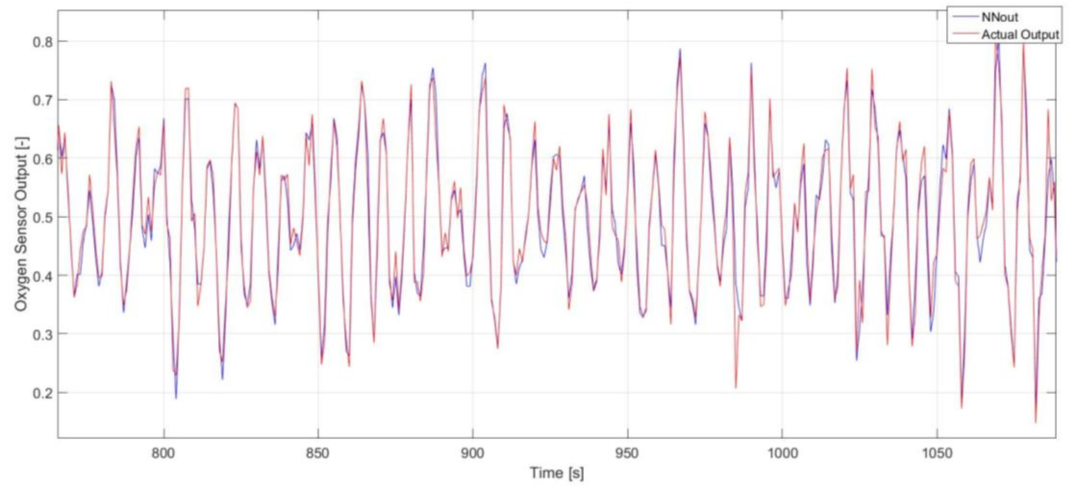

Figure 4.18: Comparison of the model output and measured signal

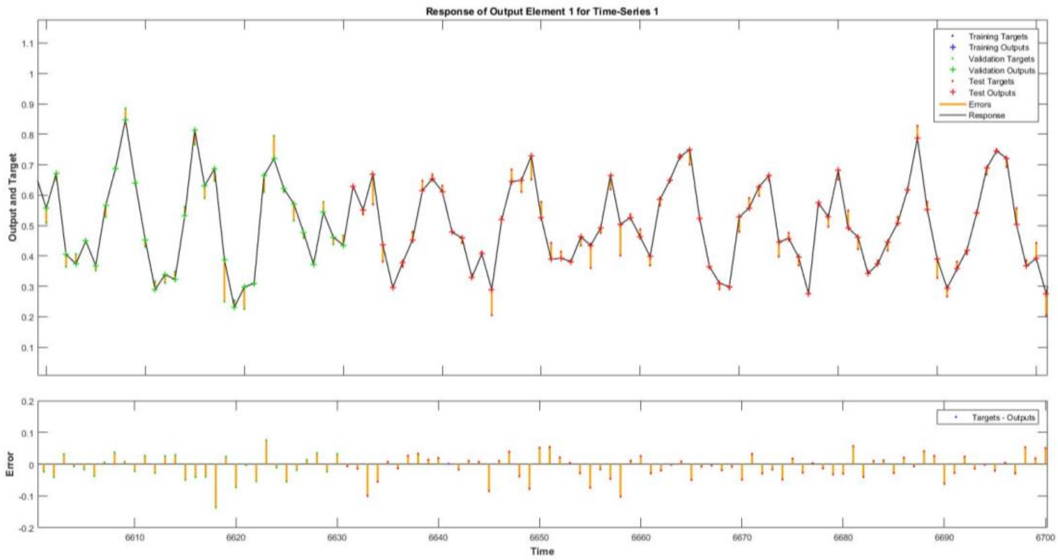

Figure 4.19: Comparison of the model output and measured signal for test and validation data sets

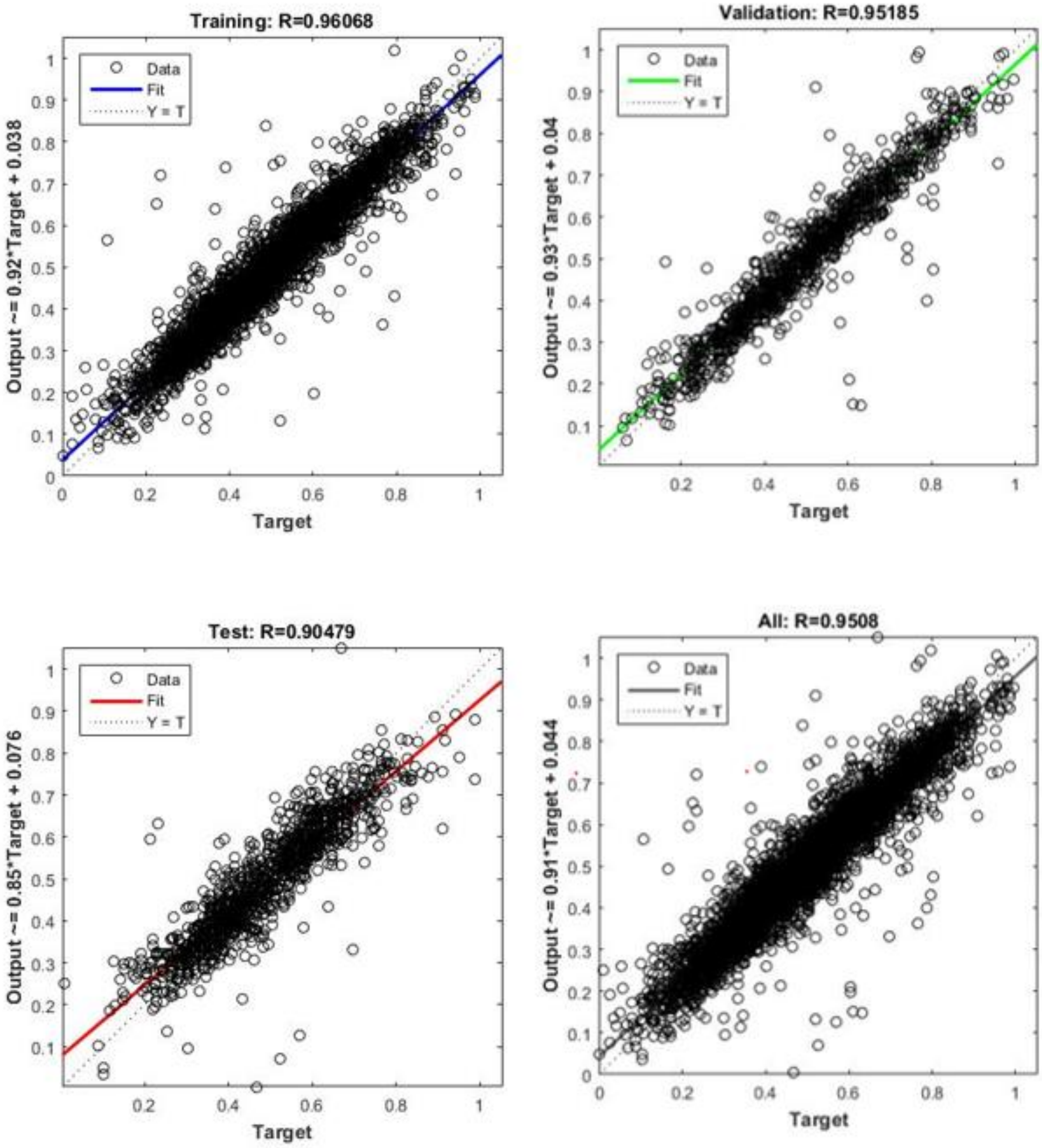

Figure 4.20: Correlation coefficient

5. MONITORING OF OXYGEN SENSOR

Chapter 5 firstly describes how to use artificial neural network model of the oxygen sensor to generate residuals after getting an accurate-enough model. As a second step, Chapter 5 illustrates how the residuals can be evaluated to develop a sufficient fault diagnosis approach for the oxygen sensor. Figure 5.1 represents the overall process of residual generation and evaluation.

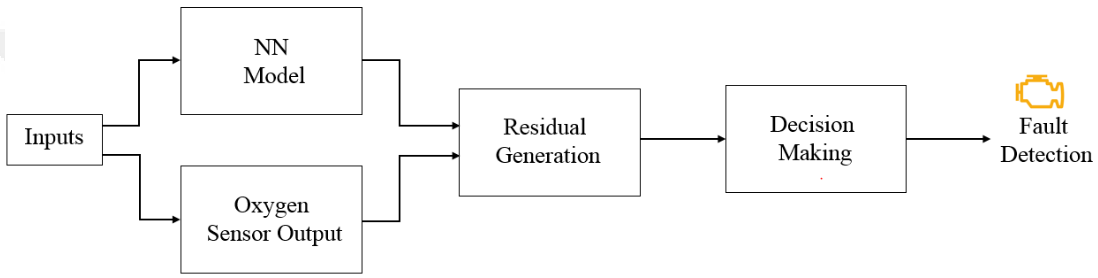

Figure 5.1: Block diagram of residual generation and evaluation process

5.1 Residual Generator

Residuals indicate the amount of deviation of the estimated signal from the actual system output. Most of the diagnostic approaches in literature are based on the residual evaluation. For this purpose residuals in the system, at different working ranges and conditions, need to be produced. The neural network-based model of the oxygen sensor is the fundamental component in residual generator step. The model structure is composed as an open loop in the training phase in Chapter 4. To be able to use the model with other datasets of a functional and damaged sensor, the model is converted to a closed loop system. In the closed-loop system a delayed output connection is used directly, instead of the delayed target input. Figure 5.2 shows the closed loop model structure.

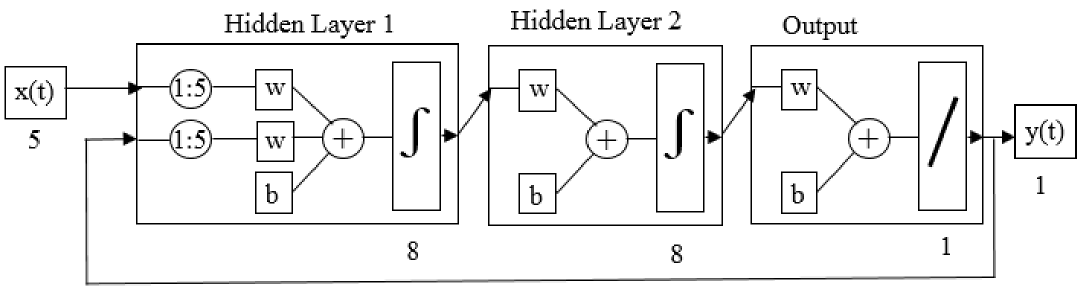

Figure 5.2: Closed-loop model structure of the oxygen sensor

By using closed-loop model and new datasets, the residual generator is designed as shown in Figure 5.3.

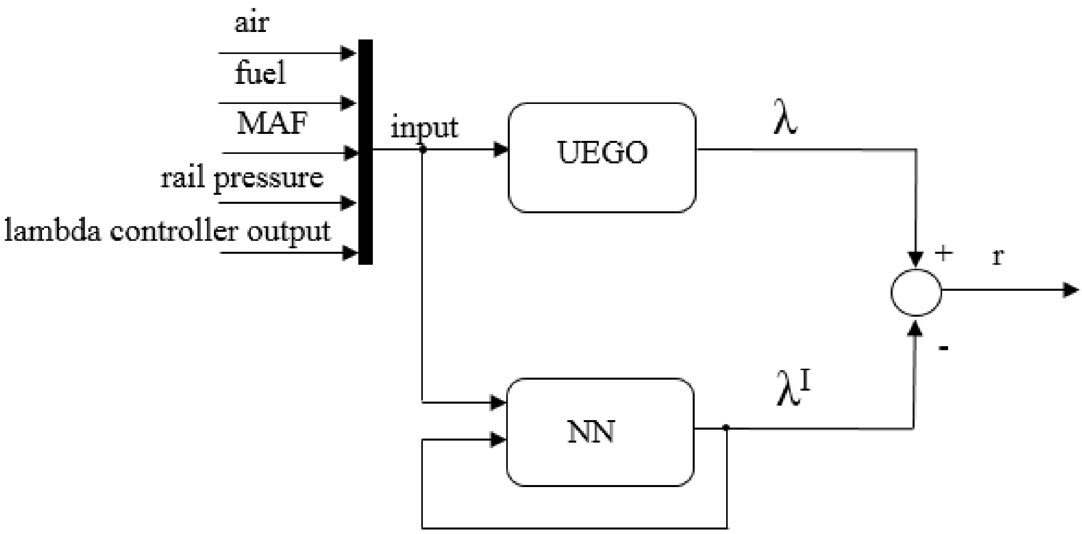

Figure 5.3: Residual generator

λ stands for the actual signal of oxygen sensor on the vehicle whereas λ^I represent model output for sampling time t. The residuals according to the residual generator is calculated as equation 5.1.

$$r(t)= \lambda(t) - \lambda^I(t) \tag{5.1}$$

The magnitude of residuals must be large enough and last long enough. So, the fault can be detected robustly. Residual generator collects the output signals of the actual component and model for a specific range of time or event and calculates the error between these two signals. Afterward, the residuals are analyzed for decision making.

The main neural network is modeled by using steady-state data in different operating points of engine speed and torque. During residual generation, this steady-state data-based model is used with two different sets of transient emission cycle data. The first data-set is collected during WLTC with a good oxygen sensor. The second data-set is gathered again in a WLTC, but by using a faulty sensor this time. Emission limits are considered to decide the degree of malfunctioning of the faulty sensor.

The normalized values of vehicle speed, engine speed, and oxygen sensor outputs are in Figure 5.4 and Figure 5.5 for good and faulty sensors respectively.

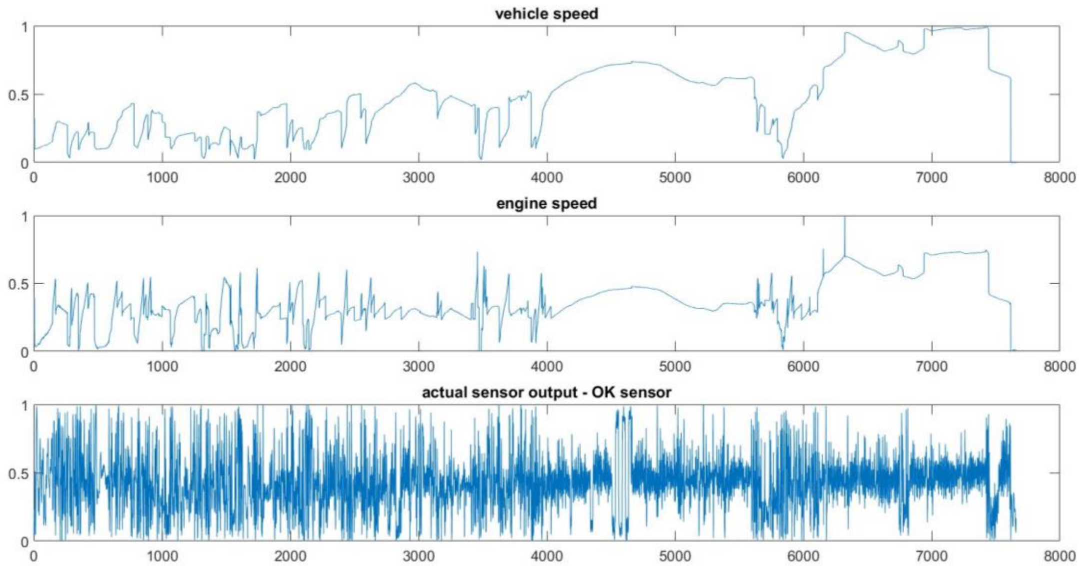

Figure 5.4: Normalized data of WLTC with the good sensor

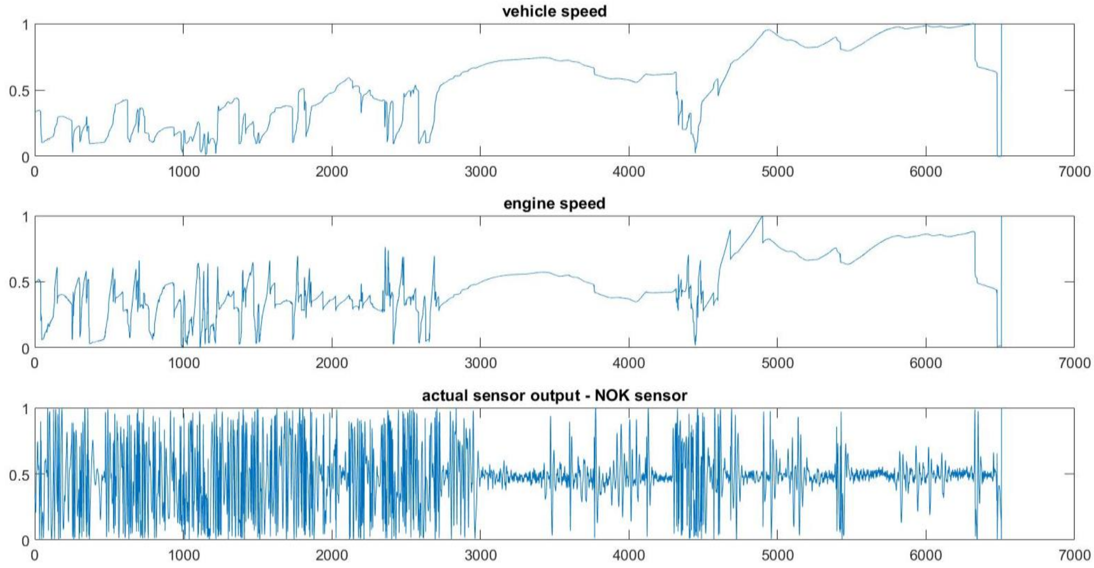

Figure 5.5: Normalized data of WLTC with the faulty sensor

5.2 Residual Evaluation

Residual evaluation is necessary to decide if the system is faulty or functioning in an acceptable range. The residual evaluation process involves assessment of residual for a predetermined range of data by comparing to a threshold value. The threshold can be a single value, a varying signal, or a variety depending on the application. In the ideal case, the residual is expected as zero for a component or system with no fault. However, in real-practices, this is almost impossible. Unknown disturbances, dynamics that cannot be modeled, modeling errors and the degree of the precision of input measurements are the factors causing non-zero residuals for adequately working components. Even though there is no residual because of the functioning component, the residual might differ from zero due to these factors. Therefore, a robust residual evaluation algorithm is essential to prevent any misdetection. Besides, the residual evaluation should not identify any malfunctioning component as a component working correctly.

Several ways are studied in the literature for robust residual assessment. Statistical data processing, adaptive thresholds, fuzzy clustering, pattern recognition data correlation are some of the standard residual evaluation methods [35], [36], [37], [38]. The residual evaluation algorithm directly affects the performance of fault detection.

Therefore, the selected function must be suitable and applicable for the relevant component.

In this study, fault detection with limit checking is focused. A statistical method based on the estimation of mean and variance of the normal distribution is investigated**.**

5.2.1 Statistical Approach

The main idea is based on the comparison of the probability distribution of the residuals for functional and malfunctioning components. For this purpose, an appropriate area is chosen on WLTC, and the comparison is realized for that specific operating region. Figure 5.6 shows the normal distribution plots of residuals for functional and malfunctioning components. The differences concerning mean and standard deviation are seen clearly.

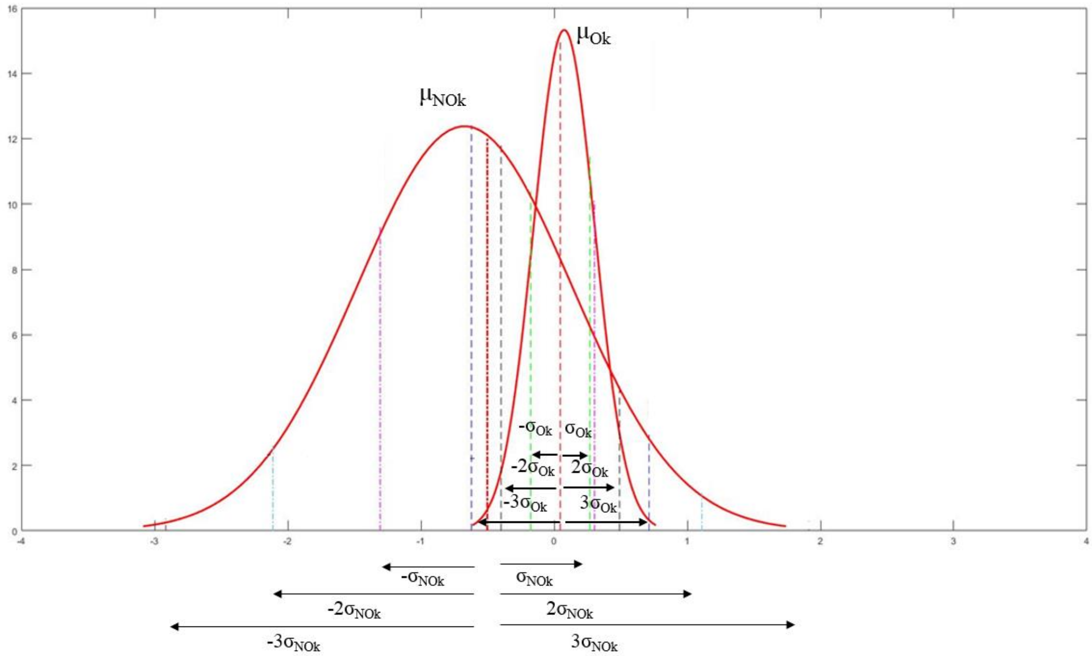

Figure 5.6: Probability distributions of residuals of OK and not-OK oxygen sensors

The mean is attained by the addition of values of each sample for a specific range of data and then dividing the total value by the number of the samples, as shown in the equation 5.2. Any change in mean value or average of the residuals is related to the accuracy.

$$\boldsymbol{\mu} = \frac{\sum_{i=0}^{N} \mathbf{xi}}{\boldsymbol{N}} \;\; , \;\; \text{N=1,2,3,…} \tag{5.2}$$

The standard deviation shows how the samples spread around the mean value. The smaller value of the standard deviation of the residuals reflects a good precision. Equation 5.3 indicates the estimation of the standard deviation.

$$\boldsymbol{\sigma} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} ((\boldsymbol{x_i} - \boldsymbol{\mu})^{\mathbf{2}}) \tag{5.3}$$

In this study, the calculation of mean and standard deviation is based on taking 50 samples after monitoring function is enabled into account. After estimating the characteristic features such as mean and standard deviation values of the normal distributions, a quantitative threshold must be set to decide about the component.

In order to get a more robust evaluation of the residuals, an approach where two different thresholds are defined is built.

$$\boldsymbol{J_\mu} \leq \; |\boldsymbol{\mu_r}| \;\wedge\;\; \boldsymbol{J_\sigma} \leq \boldsymbol{\sigma_r} \;\Rightarrow\; \text{Fault detection} \tag{5.4}$$

J_μ and J_σ stand for threshold values for mean and standard deviation respectively, whereas μ_r is the mean value of the residuals in a specified data range and σ_r is the standard deviation of the residuals. Equation 5.4 and the Figure 5.7 represent the residual evaluation approach.

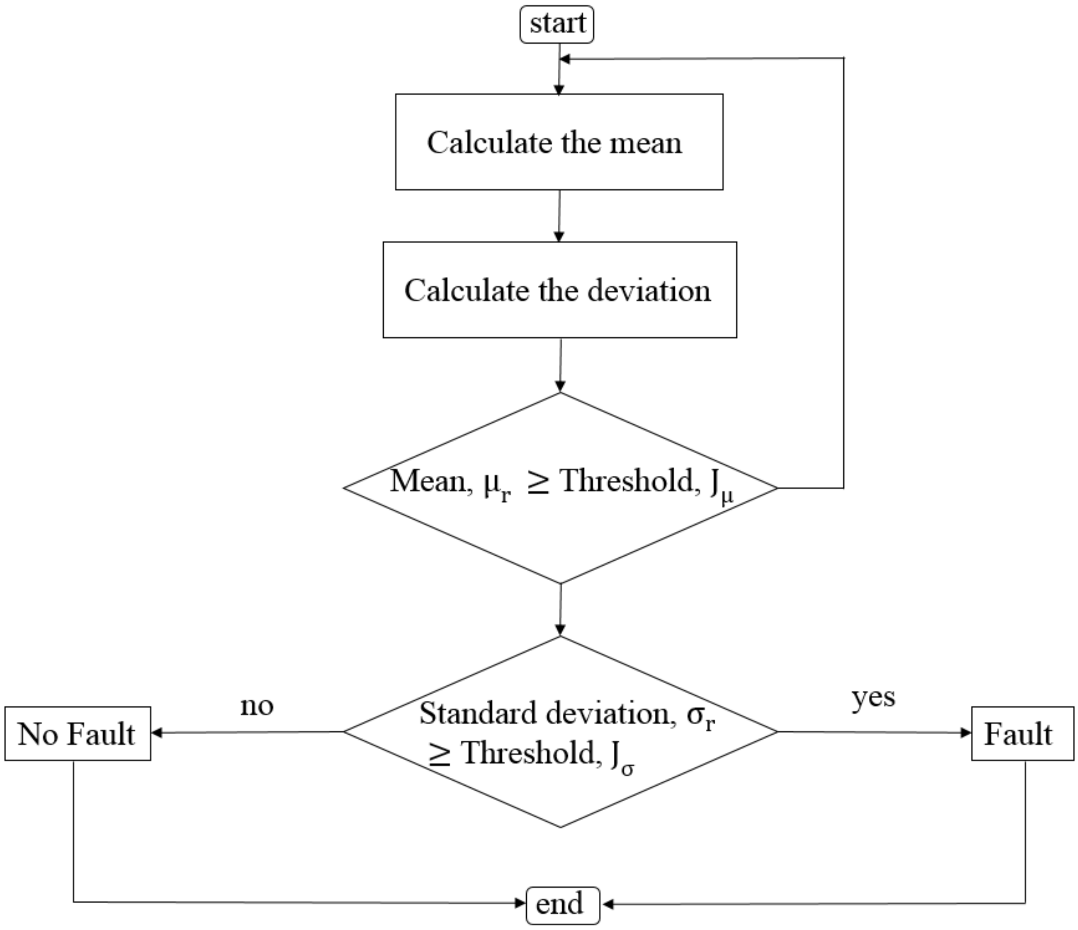

Figure 5.7: Decision-making process

The decision-making process is based on two criteria. The calculated mean value of the residuals is compared to a predefined threshold J_μ, and the standard deviation of the probability distribution is compared to another predefined threshold value J_σ. Based on the comparison results the final decision is given to the oxygen sensor.

6. CONCLUSION AND RECOMMENDATIONS

Oxygen sensors on aftertreatment systems of motor vehicles play an essential role in emission control and fuel economy. Due to the impact on emissions, the authorities like CARB, EPA or EU apply strict rules regarding on-board diagnostics of the oxygen sensors. Therefore, the monitoring of the sensor and detection the slow response due to a malfunctioning sensor is essential as well. Damaging of the sensor might occur due to aging, environmental influences, temperature fluctuations of the exhaust gas or residues in the exhaust gas such as soot and oil.

In literature, there are many different methods for fault detection and diagnosis on various areas including automotive industry and related components. However, most of the proposed techniques are questionable regarding applying to the real implementations in life, even though they recommend sophisticated and successful approaches in the lab environment with virtual data. The main reasons usually are due to physical limitations in dynamic systems during measurements and data collections, the required excessive calibration effort for the parameters involved in the proposed solutions or the complexity of the proposed method. In the meantime, the diagnostics method for oxygen sensor used by the manufacturer in the current project show good performance regarding fault detection, but in a limited operating area. Also, the present method requires extreme calibration effort coming with excessive costs due to the usage of chassis dynamometer and vehicle, besides it is time-consuming. Furthermore, the calibration engineer must have a certain level of experience to be able to process the measurements and interpret the collected data in a right way. Considering the time constraint, a qualified engineer has a significant impact on meeting calibration delivery targets before the start of production by using the current diagnostic method.

This study has presented an accurate-enough and successful design of the diagnostic system for monitoring the oxygen sensor with less effort to collect data and calibrate the monitoring function in less time. The entire steps of design, including the modeling of the sensor, residual generation, and decision making has been discussed. First of all, the system identification methods are investigated to model the oxygen sensor with the black-box approach. The black-box approach provides the system to be modeled with no previous knowledge about the physical background of the sensor or deep understanding of the system dynamics. For this purpose, a wide range of steady-state data is collected on chassis dynamometer for only once. The PCA approach is applied to see the correlation between different possible input signals and sensor output. A group of inputs is specified as initial input set. Then, an overview of varying system identification approaches is given with the theory behind of each method. Initially, different linear system identification methods with differing configurations of input are analyzed to see the applicability to the relevant system. At this point, it is understood that the results of PCA give an idea about what the correct inputs of the model should be. However, it does not have to be exactly same in real application. In addition, the linear system identification does not give sufficient results regarding modeling of the oxygen sensor, even with different input selections. Splitting the whole data into smaller pieces, based on engine speed range, does not improve the outcomes either. Therefore, nonlinear system identification approach is the next focus as a further step. NARX neural network model is used as a nonlinear system identification method. Different neural network structures including a variety of number of hidden layers, number of neurons, learning rates, training algorithms, input delay are investigated for fault detection. The best performance is obtained by using two hidden layers with eight neurons, while the system delay is five and the training algorithm is chosen as Levenberg-Marquardt.

Designing of the monitoring function follows the obtaining a successful model. Residual generation and evaluations are the successor steps. One WLTC ran with well-functioning sensor and data collected whereas the next data is collected using a faulty sensor in the same emission test cycle. Two different residual sets are generated for these two data-sets by using the neural network model. In the residual evaluation process, the statistical approach based on mean and standard deviation calculations is taken advantage of decision making between faulty and healthy oxygen sensors. The design of neural network models for healthy and malfunctioning sensors is proved to be applicable for fault detection in such nonlinear system.

In further studies, the environmental conditions like temperature in hot and cold climates, different altitudes, or piece to piece variations can be considered to build a more robust model while collecting data at the beginning. Also, more data can be gathered with sensors having a different level of degradations. Regarding residual evaluation method, different approaches known in literature such as fuzzy clustering can be investigated to compare the performances.

REFERENCES

[1] **Robert Bosch GmbH** (2014). Bosch Automotive Handbook - 9th Edition

[2] **Harris, J.A., and Collings, N.** (2015). The transient start-up response of a universal exhaust gas oxygen sensor to investigate the Nernst equation in platinum/zirconia cells. Sensors and Actuators, B: Chemical 221, 81–87.

[3] **Mooney, J.J.** (2007).The 3-Way Catalytic Converter: a) Invention and Introduction into Commerce - Impacts and Results, b) Barriers Negotiated California Air Resources Board Chairman Invitational Seminar Series

[4] **R. Heck and R. Farrauto** (2001) Automobile Exhaust Catalysts, Applied Catalysis A: General, volume 221, pp. 443-457.

[5] **McDowell, N., McCullough, G., Wang, X., Kruger, U., and Irwin, G.W.** (2007). Fault diagnostics for internal combustion engines - Current and future techniques. SAE Technical Papers 2007.

[6] **Al Ahmar, E., Choqueuse, V., Benbouzid, M.E.H., Amirat, Y., El Assad, J., Karam, R., and Farah, S.** (2010). Advanced signal processing techniques for fault detection and diagnosis in a wind turbine induction generator drive train. 2010 IEEE Energy Conversion Congress and Exposition 3576–3581.

[7] **Abbas, M., Ferri, A.A., Orchard, M.E., and Vachtsevanos, G.J.** (2007). An Intelligent Diagnostic/Prognostic Framework for Automotive Electrical Systems. In Intelligent Vehicles Symposium, 2007 IEEE, pp. 352–357.

[8] **Versmold, H., and Saeger, M.** (2006). Plausibility Checking of Sensor Signals for Vehicle Dynamics Control Systems. … Symposium on Advanced Vehicle Control

[9] **Wang, S., and Xiao, F.** (2004). AHU Sensor Fault Diagnosis Using Principal Component Analysis Method. Energy and Buildings 36, 147–160.

[10] **Nybarg, M., and Nielsen, L.** (1997). Model-Based Diagnosis for the Air Intake System of the SI Engine. SAE Transactions 106, 9–20.

[11] **Gertler, J.J., Costin, M., Fang, X., Hira, R., Kowalczuk, Z., and Luo, Q.** (1993). Model-based on-board fault detection and diagnosis for automotive engines. Control Engineering Practice 1, 3–17.

[12] **Wang, Y., and Chu, F.** (2005). Real-time misfire detection via sliding mode observer. Mechanical Systems and Signal Processing 19, 900–912.

[13] **Nyberg, M., and Stutte, T.** (2004). Model based diagnosis of the air path of an automotive diesel engine. Control Engineering Practice 12, 513–525.

[14] **Nyberg, M.** (2002). Model-based diagnosis of an automotive engine using several types of fault models. IEEE Transactions on Control Systems Technology 10, 679–689.

[15] **Jammoussi, H., Makki, I., Filev, D., Franchek, M.** (2013). Statistics Based Detection and Isolation of UEGO Sensor Faults. Proceedings of the ASME 2013 Dynamic Systems and Control Conference Palo Alto, California, USA.

[16] **Arsiea, I., Pianesea, C., Rizzoa, G., Cioffib, V.** (2002). An adaptive estimator of fuel film dynamics in the intake port of a spark ignition engine. ELASIS, 80038 Pomigliano D'Arco, NA, Italy.

[17] **He, Y., Feng, L.** (2004). Diesel fuel injection system faults diagnosis based on fuzzy injection pressure pattern recognition. 5th World Congress on Intelligent Control and Automation, 1654–1657.

[18] **Arsie, I., Sorrentino, M., and Pianese, C.** (2007). A neural network air-fuel ratio estimator for control and diagnostics in spark-ignited engines. In IFAC Proceedings Volumes (IFAC-PapersOnline), pp. 227–234.

[19] **Moriya, K., and Sako, T.** (2001). Oxygen sensor monitoring a deterioration of a three-way catalyst in natural gas fueled engines. Sensors and Actuators, B: Chemical 73, 142–151.

[20] **Lee, J.H.** (2003). Review on zirconia air-ful ratio sensors for automotive applications. Journal of Materials Science 38, 4247–4257.

[21] **Harris, J.A., and Collings, N.** (2015). The transient start-up response of a universal exhaust gas oxygen sensor to investigate the Nernst equation in platinum/zirconia cells. Sensors and Actuators, B: Chemical 221, 81–87.

[22] **H. Germann, S. Tagliaferri, H. P. Geering.** (1996). Differences in Pre- and Post-Converter Lambda Sensor Characteristics. SAE International Congress, SAE-paper 960335.

[23] **Okazaki, S., Kato, N., Kako, J., and Ohata, A.** (2009). Development of a New Model Based Air-Fuel Ratio Control System. 2, 335–343.

[24] **Yildiz, Y., Annaswamy, A., Yanakiev, D., and Kolmanovsky, I.V.** (2008). Adaptive Air Fuel Ratio Control for Internal Combustion Engines. System 2058–2063.

[25] **Ljung L.** (1999) System identification. theory for the user., 2nd edn. Prentice Hall, New Jersey

[26] **Labview,** (2006): System identification toolkit user manuel, National Instruments Corporation instruments, 371001C-01

[27] **Purwar, S., Kar, I.N., and Jha, A.N.** (2007). Nonlinear system identification using neural networks. IETE Journal of Research 53, 35–42.

[28] **Beale, M. H., Hagan, M. H., and Demuth, H. B.** (2010): Neural network toolbox: User's guide, The MathWorks, Inc., Natick, MA

[29] **Perreault, E.J., Westwick, D.T., Pohlmeyer, E.A., Solla, S.A., and Miller, L.E.** (2005). Optimal input selection for MISO systems identification:

Applications to BMIs. In 2nd International IEEE EMBS Conference on Neural Engineering, pp. 167–170.

[30] **Ballabio, D.** (2015). A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure. Chemometrics and Intelligent Laboratory Systems 149, 1–9.

[31] **Sola, J., and Sevilla, J.** (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Transactions on Nuclear Science 44, 1464–1468.

[32] **Panchal, F.S., and Panchal, M.** (2014). Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network. International Journal of Computer Science and Mobile Computing 311, 455–464.

[33] **Gopalakrishnan, K.** (2010). Effect of training algorithms on neural networks aided pavement diagnosis. International Journal of Engineering, Science and Technology 2, 83–92.

[34] **Schoppe, O., Harper, N.S., Willmore, B.D.B., King, A.J., and Schnupp, J.W.H.** (2016). Measuring the Performance of Neural Models. Frontiers in Computational Neuroscience 10.

[35] **Frank, P.M., Ding, X.** (1997). Survey of robust residual generation and evaluation methods in observer-based fault detection systems, J. Proc. Cont. Vol. 7, No. 6, 403-424.

[36] **Isermann, R.** (2006). Fault-diagnosis systems, an introduction from fault detectionto fault tolerance, Springer.

[37] **Svärd, C., Nyberg, M., Frisk, E., and Krysander, M.** (2014). Data-driven and adaptive statistical residual evaluation for fault detection with an automotive application. Mechanical Systems and Signal Processing 45, 170–192.

[38] **Patan, K.** (2008). Artificial neural networks for the modelling and fault diagnosis of technical processes, Springer.

CURRICULUM VITAE

Name Surname	**:** Kübra Ekinci
Place and Date of Birth	**:** Istanbul, 1989
E-Mail	**:** ekinci.kubra1@gmail.com

EDUCATION	**:**
• **B.Sc.**	**:** 2012, Hacettepe University, Engineering Faculty, Mechanical Engineering, Automotive Engineering Program

OTHER PUBLICATIONS:

- Şenakın H., Kal İ., Ekinci K. (2017) 'Genetic Algorithm Usage in Threshold Calibration', 2nd International Congress of the Mediterranean Science and Engineering (IMSEC'17), Adana.
- Ekinci K., Karabıyık E., (2017) 'Relay Based Auto Tuner for Calibration of SCR Pump Controller Parameters in Diesel After Treatment Systems', International Journal of Automotive Engineering and Technologies, pp. 26 – 32.