

Institutionen för systemteknik

Department of Electrical Engineering

Examensarbete

Black-Box Modeling of the Air Mass-Flow Through the Compressor in A Scania Diesel Engine

Examensarbete utfört i Fordonssystem
vid Tekniska högskolan i Linköping
av

Oskar Törnqvist

LiTH-ISY-EX--09/4175--SE

Linköping 2009



Linköpings universitet
TEKNISKA HÖGSKOLAN

Black-Box Modeling of the Air Mass-Flow Through the Compressor in A Scania Diesel Engine

Examensarbete utfört i Fordonssystem
vid Tekniska högskolan i Linköping
av

Oskar Törnqvist

LiTH-ISY-EX--09/4175--SE

Handledare: **Erik Höckerdal**
ISY, Linköpings universitet
Scania CV AB

Examinator: **Associate Professor Erik Frisk**
ISY, Linköpings universitet

Linköping, 27 November, 2009

	Avdelning, Institution Division, Department Division of Vehicular Systems Department of Electrical Engineering Linköpings universitet SE-581 83 Linköping, Sweden		Datum Date 2009-11-27
	Språk Language <input type="checkbox"/> Svenska/Swedish <input checked="" type="checkbox"/> Engelska/English <input type="checkbox"/> _____	Rapporttyp Report category <input type="checkbox"/> Licentiatavhandling <input checked="" type="checkbox"/> Examensarbete <input type="checkbox"/> C-uppsats <input type="checkbox"/> D-uppsats <input type="checkbox"/> Övrig rapport <input type="checkbox"/> _____	ISBN _____ ISRN LiTH-ISY-EX--09/4175--SE Serietitel och serienummer ISSN Title of series, numbering _____
URL för elektronisk version http://www.vehicular.isy.liu.se/ http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-52125			
Titel Title Svartboxmodellering av luftmassflödet förbi kompressorn i en Scania dieselmotor Black-Box Modeling of the Air Mass-Flow Through the Compressor in A Scania Diesel Engine			
Författare Oskar Törnqvist Author			
Sammanfattning Abstract <p>Stricter emission legislation for heavy trucks in combination with the customers demand on low fuel consumption has resulted in intensive technical development of engines and their control systems. To control all these new solutions it is desirable to have reliable models for important control variables. One of them is the air mass-flow, which is important when controlling the amount of recirculated exhaust gases in the EGR system and to make sure that the air to fuel ratio is correct in the cylinders.</p> <p>The purpose with this thesis was to use system identification theory to develop a model for the air mass-flow through the compressor. First linear black-box models were developed without any knowledge of the physics behind. The collected data was preprocessed to work in the modeling procedure and then models with one or more inputs where built according to the ARX model structure.</p> <p>To further improve the models performance, non-linear regressors was developed from physical relations for the air mass-flow and used to form grey-box models of the air mass-flow.</p> <p>In conclusion, the performance was evaluated through comparing the estimated air mass-flow from the best model with the estimate that an extended Kalman filter together with a physical model produced.</p>			
Nyckelord Keywords Air Mass-Flow, System Identification, Black-Box, Grey-Box, Linear Regressors, Non-linear Regressors, Diesel Engine			

Abstract

Stricter emission legislation for heavy trucks in combination with the customers demand on low fuel consumption has resulted in intensive technical development of engines and their control systems. To control all these new solutions it is desirable to have reliable models for important control variables. One of them is the air mass-flow, which is important when controlling the amount of recirculated exhaust gases in the EGR system and to make sure that the air to fuel ratio is correct in the cylinders.

The purpose with this thesis was to use system identification theory to develop a model for the air mass-flow through the compressor. First linear black-box models were developed without any knowledge of the physics behind. The collected data was preprocessed to work in the modeling procedure and then models with one or more inputs where built according to the ARX model structure.

To further improve the models performance, non-linear regressors was developed from physical relations for the air mass-flow and used to form grey-box models of the air mass-flow.

In conclusion, the performance was evaluated through comparing the estimated air mass-flow from the best model with the estimate that an extended Kalman filter together with a physical model produced.

Sammanfattning

Hårdare utsläppskrav för tunga lastbilar i kombination med kundernas efterfrågan på låg bränsleförbrukning har resulterat i en intensiv utveckling av motorer och deras kontrollsystem. För att kunna styra alla dessa nya lösningar är det nödvändigt att ha tillförlitliga modeller över viktiga kontrollvariabler. En av dessa är luftmassflödet som är viktig när man ska kontrollera den mängd avgaser som återcirkuleras i EGR-systemet och för att se till att kvoten mellan luft och bränsle är korrekt i motorns cylindrar.

Syftet med det här examensarbetet var att använda systemidentifiering för att ta fram en modell över luftmassflödet förbi kompressorn. Först togs linjära svartboxmodeller fram utan att ta med någon kunskap om den bakomliggande fysiken. Insamlade data förbehandlades för att passa in i modelleringsproceduren och efter det skapades i enlighet med ARX-modellstrukturen modeller med en eller flera insignaler.

För att ytterligare förbättra modellernas prestanda togs icke-linjära regressorer fram med hjälp av fysikaliska relationer för luftmassflödet. Dessa användes sedan för att skapa gråboxmodeller av luftmassflödet. Avslutningsvis utvärderades

prestandan genom att det estimerade luftmassflödet från den bästa modellen jämfördes med det estimat som ett utökat kalmanfilter tillsammans med fysikaliska ekvationer genererade.

Acknowledgments

I would like to thank my supervisor at Scania CV AB and at Linköpings Universitetet, Erik Höckerdal for all helpful support and for the opportunity to make this thesis. I will also thank my examiner at Linköpings Universitetet, Erik Frisk. Thanks also to all the helpful people at Scania, in particular Erik Geijer Lundin and Mats Jennische for always helping me with my questions and giving support.

Oskar Törnqvist
Linköping, November 2009

Contents

1	Introduction	1
1.1	Background	1
1.2	Objectives	2
1.3	Outline	2
2	The Engine System	3
2.1	Engine Overview	3
2.2	The Air Mass-Flow	4
3	System Identification	7
3.1	Linear Black-Box Models	8
3.1.1	General Structure for Linear Black-Box Models	8
3.1.2	The Most Common Models	8
3.2	Model Order Selection	9
3.2.1	Loss Function	9
3.2.2	Akaike's Information Criterion	11
3.3	Model Validation	11
3.3.1	Model Fit	11
3.3.2	Residual Analysis	12
3.4	Introduction to Regressors	12
3.5	Model Nomenclature	13
3.6	System Identification Toolbox	13
4	Data Preprocessing	15
4.1	Dealing with Incorrect Data	15
4.2	Removing Trends	15
4.3	Resampling the Data	16
4.3.1	The Sampling Theorem and Alias Effect	17
4.4	Signal Energy and Periodogram	17
4.5	Estimation and Validation Data	20
5	Linear Regressor Models	21
5.1	Modeling Strategy	21
5.2	Single Input Single Output Models	22
5.2.1	Fit Calculation	22

5.3	Residual Analysis	24
5.3.1	Pole-Zero Diagram	25
5.4	Multiple Input Single Output Models	26
6	Non-linear Regressor Models	29
6.1	Physical Relations for the Air Mass-Flow	29
6.2	Custom Regressors	31
6.2.1	Turbine Speed Regressors	31
6.2.2	Pressure Regressors	32
6.2.3	Combined Regressors	34
6.3	Combination of Regressors and Signals	35
7	Results	39
7.1	The Best Model	39
7.2	Extended Kalman Filter	40
7.2.1	Physical Model	41
7.2.2	Filter Design	42
7.3	EKF Estimate	43
7.4	Root Mean Square Error	43
8	Concluding Remarks	45
8.1	Conclusions	45
8.2	Discussion of Problems	46
8.3	Future Work	46
	Bibliography	47
	A Notation	49
	B Figures	50
B.1	Measured Signals	50
B.2	Resampled Signals	53
B.3	Periodogram	55
B.4	Residuals for Linear SISO Models	56
B.5	Pole-Zero Diagram for Linear SISO Models	58
B.6	Residuals for Linear MISO Models	60
B.7	Residuals for Non-linear MISO Models	62

Chapter 1

Introduction

This master's thesis has been performed for Scania CV AB at the division of Powertrain Control System Development - Engine Performance Software (NESE). The purpose is to examine if it is possible to estimate the air mass-flow into the engine without using a sensor measuring it. In earlier work at Scania, for example in [19], physical models was developed and used to design an observer that estimates the air mass-flow.

System identification is used in this thesis to develop a black-box model in order to see how good the estimate becomes without use of the physics. One step closer to the physical models is done if some physical relations are used to find appropriate variables and regressors, resulting in grey-box models.

A concrete list of what this thesis will deal with is presented in Section 1.2.

1.1 Background

Stricter emission legislation for heavy trucks in combination with the customers demand on low fuel consumption has resulted in intensive technical development of engines and their control systems. New physical solutions have been introduced, some of these are Exhaust Gas Recirculation (EGR) and Variable Geometry Turbo (VGT) systems on diesel engines. These new solutions increase the system complexity and the demand of new control variables that can be hard, or even impossible, to measure with physical sensors. Because of that, it is desirable to develop good and reliable system models, which can be used together with the physical sensors or completely replace them. An example is the air mass-flow sensor that is important when controlling the amount of recirculated exhaust gases in the EGR system and to make sure that the air to fuel ratio is correct in the cylinders.

1.2 Objectives

The main objectives in this thesis are to:

- Find the sensor signals that correlate well with the air mass-flow, determine the bandwidth of the system in order to find a suitable sampling frequency and examine what type of model structure that gives the best estimate of the air mass-flow and use this to create a black-box model of the air mass-flow.
- Try to improve the black-box models using physical insight, so called grey-box modeling.
- Compare the performance of the resulting models with the performance of an observer based on a physical model, i.e. an extended Kalman filter (EKF).

1.3 Outline

Chapter 2 gives an overview of the engine system and introduces the air mass-flow.

Chapter 3 presents system identification, the theory for black-box model development.

Chapter 4 contains the preprocessing work, where the collected data is prepared for model development, correlation between the inputs and the bandwidth of the system are studied to find the most usable inputs respectively a suitable sample frequency.

Chapter 5 contains the development and validation of linear black-box models.

Chapter 6 contains the development and validation of non-linear grey-box models.

Chapter 7 presents and discusses the resulting model and compares its results to the results from an extended Kalman filter observer based on a physical model.

Chapter 8 contains the conclusions and discuss difficulties and future work.

Chapter 2

The Engine System

This chapter describes the engine system and its main parts. It also contains a short explanation about the sensor signals.

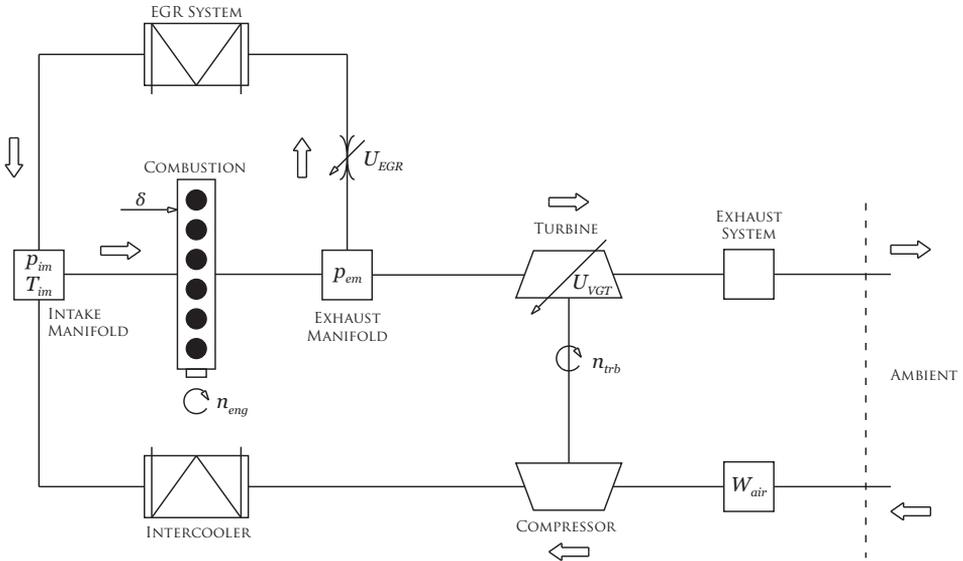


Figure 2.1: Model of the engine system.

2.1 Engine Overview

Figure 2.1 shows the main parts of the engine system and how the air and gases flows through the system.

Fresh air flows into the engine system from the ambient. The air mass-flow into the intake system is measured by the air mass-flow sensor, W_{air} , before it passes through the compressor. The compressor is driven by the turbine and increases

the density and temperature of the air. After that, the air passes the intercooler where it is cooled down. This further increase the density of the air and the result is a higher amount of air flowing into the cylinders which makes it possible to inject more fuel and get higher power.

The intake manifold connects the inlet system with the Exhaust Gas Recirculation (EGR) system and feeds the cylinders with a mixture of fresh air and exhaust gases. Here are the intake manifold pressure and temperature sensors located.

A mixture of gases and fuel is injected in the combustion chamber and is burned in the cylinders. Afterwards the exhaust gases flows into the exhaust manifold, where the exhaust manifold pressure sensor is located.

The exhaust gases flows through the Variable Geometry Turbo (VGT) that is connected to the compressor and the exhaust system. Part of the exhaust gases is recirculated through the EGR system back to the intake manifold. This is done in order to reduce NO_x formation. An increased EGR flow also results in an decreased fuel consumption [7]. The gases flowing through the EGR and the VGT are controlled by valves and the control signals to these valves (U_{EGR} and U_{VGT}) are measured.

A complete list of available sensor signals is presented in Table 2.1.

Table 2.1: Available sensor signals.

Sensor	Description	Unit
δ	Injected fuel	$[mg/stroke]$
n_{eng}	Engine speed	$[rpm]$
n_{trb}	Turbine speed	$[rpm]$
p_{im}	Intake manifold pressure	$[Pa]$
p_{em}	Exhaust manifold pressure	$[Pa]$
T_{im}	Intake manifold temperature	$[K]$
U_{EGR}	EGR control signal	$[-]$
U_{VGT}	VGT control signal	$[-]$
W_{air}	Air mass-flow before compressor	$[kg/s]$

2.2 The Air Mass-Flow

According to [5] the air mass-flow in a diesel engine is a very important quantity that has a direct impact on many control and diagnosis functions.

The air mass-flow is used for computations of λ , the air to fuel ratio during combustion,

$$\lambda = \frac{W_{air}}{W_{fuel}}. \quad (2.1)$$

It is important to keep λ above a certain level. If the ratio becomes too low, smoke is generated and the control law is forced to reduce the torque output.

The air mass-flow is also used to calculate the EGR-fraction

$$EGR_{frac} = \frac{W_{tot} - W_{air}}{W_{tot}}, \quad (2.2)$$

where W_{tot} is the total gas mass-flow into the engine and W_{air} is the part that consists of fresh air, i.e. the air mass-flow. The EGR-fraction is used to control the NO_X emissions.

The performance of the air-mass flow sensor may change over time due to sensor aging, differences in engine configurations or because of different operation conditions caused by geographical location, for example pressure, temperature and humidity of the surrounding air. In order to reduce the impact of these situations it is desirable to find a model of the air mass-flow that can be used together with or even replace the sensor.

Chapter 3

System Identification

System identification is a method to build a mathematical model of a real system when it is not possible to construct it entirely from prior knowledge and physical insight. It uses measured data, e.g. observed input and output signals, to estimate a model. There are two general ways to use measured data to identify a whole system or unknown parameters in it [10]:

1. Build a model, which describe the relation between input and output, without physical insight. This could be done either by using nonparametric methods like frequency analysis and impulse responses or by building a parametric black-box model wherein the parameters are determined to fit the data.
2. Use data to determine unknown system parameters in a physical model, called a grey-box modeling.

A commonly used approach for system identification is to use some of the non-parametric methods to determine a model of the system. Such methods can be frequency analysis, impulse responses or looking at the correlation and covariance between signals. According to [12] and [10] the frequency analysis do not give reliable results when the system works with feedback, and this is the case for the system in this thesis. It is not either possible to test the systems impulse or step response because the data was pre-collected. Taking that into consideration, parametric methods are the way to go and therefore this thesis deal with linear and nonlinear parametric methods. Mainly it deals with black-box modeling, where no physical insight is used, but it is also of interest to see if some advantage can be made from the physics and in addition some physical insight will be used to modify the regressors in regard to improve the performance of the black-box models. This is called grey-box modeling.

3.1 Linear Black-Box Models

This chapter is a summary of the theory about system identification presented in detail in [12].

Black-box models are a set of standard transfer function models that through experience are known to handle many different kinds of system dynamics. The most common among these standard models are linear. Normally these models are derived in discrete time, since the data used is sampled and therefore discrete. If a time continuous model is desired the time discrete model is transformed.

3.1.1 General Structure for Linear Black-Box Models

A general linear model in discrete time can be written as

$$y(t) = \eta(t) + w(t), \quad (3.1)$$

where $w(t)$ is a disturbance term and $\eta(t)$ is the output without disturbance. This output can be written as

$$\eta(t) = G(q, \theta)u(t). \quad (3.2)$$

$G(q, \theta)$ is a rational function of the displacement operator q ,

$$G(q, \theta) = \frac{B(q)}{F(q)} = \frac{b_1q^{-nk} + b_2q^{-nk-1} + \dots + b_{nb}q^{-nk-nb+1}}{1 + f_1q^{-1} + \dots + f_{nf}q^{-nf}}. \quad (3.3)$$

The disturbance term can be treated in the same way with

$$w(t) = H(q, \theta)e(t) \quad (3.4)$$

and

$$H(q, \theta) = \frac{C(q)}{D(q)} = \frac{1 + c_1q^{-1} + \dots + c_{nc}q^{-nc}}{1 + d_1q^{-1} + \dots + d_{nd}q^{-nd}} \quad (3.5)$$

where $e(t)$ is white noise.

Now model (3.1) can be written as

$$y(t) = G(q, \theta)u(t) + H(q, \theta)e(t) \quad (3.6)$$

where θ is the parameter vector that contains the coefficients a_i , b_i , c_i and f_i in the transfer functions. The parameters nb , nc , nd and nf are called structural parameters and represent the order of their respective polynomial. The fifth parameter, nk , represent the time delay in the system.

3.1.2 The Most Common Models

The most general case is when all polynomials are present in (3.6), i.e.

$$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t). \quad (3.7)$$

This model is known as the Box-Jenkins (BJ) model. This model can be simplified in a numerous of different ways to suit other more specific systems. If the noise signal is not modeled at all, i.e. $H(q) \equiv 1$ ($nc = nd = 0$), the model is an Output Error (OE) model,

$$y(t) = \frac{B(q)}{F(q)}u(t) + e(t). \quad (3.8)$$

If the denominators in G and H are equal, i.e.

$$F(q) = D(q) = A(q) = 1 + a_1q^{-1} + \dots + a_{na}q^{-na} \quad (3.9)$$

the model is called an ARMAX model and can be written as

$$A(q)y(t) = B(q)u(t) + C(q)e(t). \quad (3.10)$$

The last model described in this thesis is the commonly used ARX model which is a simplification of the ARMAX model with $C(q) \equiv 1$,

$$A(q)y(t) = B(q)u(t) + e(t). \quad (3.11)$$

In Figure 3.1 the different model structures are represented in block diagrams.

3.2 Model Order Selection

If the structure parameters, na, nb, nc, nd, nf and nk , are choosen, a computer can calculate the parameters in θ corresponding to the chosen model structure. The obtained model gives a prediction,

$$\hat{y}(t, \theta), \quad (3.12)$$

of $y(t)$:s value for each value of θ in combination with old input and output values. The following sections describe how to choose the structure parameters.

3.2.1 Loss Function

According to [12], the prediction $\hat{y}(t, \theta)$ can be evaluated at time t by calculating the prediction error,

$$\varepsilon(t) = \varepsilon(t, \theta) = y(t) - \hat{y}(t, \theta). \quad (3.13)$$

If there are N collected samples, the loss function can be formed as

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^N \varepsilon(t, \theta)^2. \quad (3.14)$$

This scalar is a measure of how well the model with the parameter value θ can describe the system.

It is natural to choose θ as the value which minimizes the loss function:

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta). \quad (3.15)$$

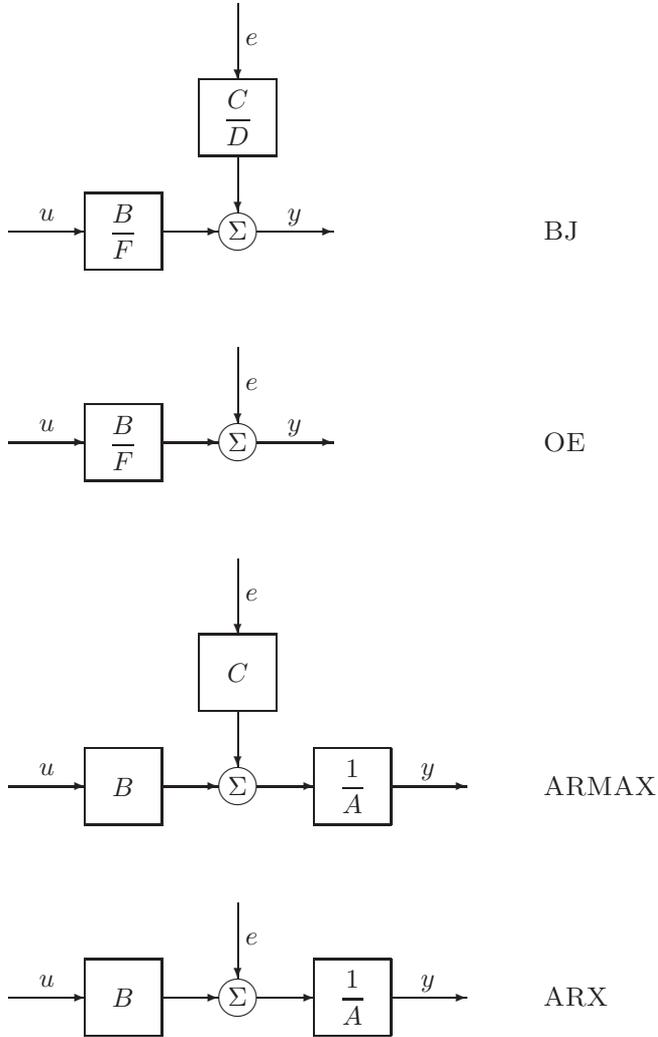


Figure 3.1: Linear Black-Box Model structures.

3.2.2 Akaike's Information Criterion

If the prediction error is calculated from the same data set used to estimate the model some problems will occur. This is because of the fact that the loss function (3.14) decreases when more parameters are added. However, if more parameters are added when the model already describes the relevant properties of the system, the only thing that happens is that the model is fitted to the specific disturbance in the data set. This is called overfit and does not serve any purpose since the model will be used when other disturbances affect the system. In fact, the model will be worse because of the overfit.

To avoid the problem with overfit mentioned above it is desirable to have two separate sets of data, one used to find the model of the system, i.e. estimation data, and one used when validate the model, i.e. validation data. Preferable are if these data is collected at different times, or runs, when different disturbances affect the system, but if that is not possible it is a possibility to split the collected data into two parts and use one part for estimation and the other for validation.

There are several methods to find the optimal number of parameters. The one that will be used here is Akaike's information criterion (AIC):

$$\hat{\theta}_N = \arg \min_{d, \theta} \left(1 + \frac{2d}{N} \right) \sum_{t=1}^N \varepsilon(t, \theta)^2, \quad (3.16)$$

where (3.14) has been modified with θ 's dimension, d , i.e. the number of estimated parameters. This will penalize models that contain many parameters.

3.3 Model Validation

When a model is built it is necessary to examine if it can be accepted for its intended use. There are several ways to do this. Some of them are presented in the following sections.

3.3.1 Model Fit

The first and most obvious validation technique is to see how well the model output correspond to the measured. This is called fit:

$$\text{fit} = 100 \left(1 - \frac{\|\hat{y} - y\|}{\|y - \bar{y}\|} \right), \quad (3.17)$$

where y is the output of the validation data, \hat{y} is the model output and \bar{y} is the mean value of y . 100 % corresponds to a perfect fit and 0 % indicates that the fit is no better than guessing the output to be a constant (i.e. $\hat{y} = \bar{y}$). A negative value is worse than 0 % fit.

3.3.2 Residual Analysis

Another way to validate the model is to look at the part of the data that the model cannot reproduce. This part is called residual and is the same as the prediction errors [12]:

$$\varepsilon(t) = \varepsilon(t, \hat{\theta}_N) = y(t) - \hat{y}(t|\hat{\theta}_N) \quad (3.18)$$

The residuals should ideally be independent of the input. If the covariance between the residuals and the past inputs (also called the cross correlation)

$$\hat{R}_{\varepsilon u}^N(\tau) = \frac{1}{N} \sum_{t=1}^N \varepsilon(t)u(t-\tau) \quad (3.19)$$

are calculated, this dependence can be investigated. If $\hat{R}_{\varepsilon u}^N$ is close enough to zero for $\tau > 0$, e.g. $\hat{R}_{\varepsilon u}^N$ is less than three standard deviations [12], it is likely that the model will be good also when applied to other input data. If they are not, it is an indication that components in the residual stem from past inputs and that the model $\hat{y}(t|\hat{\theta}_N)$ has not picked up all the system dynamics and it could be an idea to recalculate the model with another model order or a different time delay. A rule of thumb is that a slowly varying cross correlation function outside the confidence region is an indication of too few poles, while sharper peaks indicate too few zeros or an incorrect time delay. [11, p.1-17]

If calculating the auto correlation for the residuals,

$$\hat{R}_{\varepsilon}^N(\tau) = \frac{1}{N} \sum_{t=1}^N \varepsilon(t)\varepsilon(t-\tau), \quad (3.20)$$

it can be seen if the residuals themselves are correlated at different time delays. The auto correlation should be small for all $\tau \neq 0$. If not it is significant correlation between the different residuals and it is an indication that part of $\varepsilon(t)$ could have been predicted from past data. This means that the disturbance model can be improved.

3.4 Introduction to Regressors

Regressors are a transformation of past input and output signals and are used to estimate the output. Typical regressors are simply delayed input and output signals. This type of regressors will be used in Chapter 5 and are usually referred to as standard regressors. It is also possible to use more advance regressors defined by the user to fit the actual problem, usually called custom regressors. This gives the possibility to consider, for example, non-linear behavior in the system or known physical relations, for example quotients or products of two signals, and use this modified regressors together with the theory in Section 3.1 when building black-box models. In fact, it can be more efficient to use appropriate defined

custom regressors than to use standard regressors only. Custom regressors will be used in Chapter 6, when physical relations are used to construct regressors for grey-box models.

3.5 Model Nomenclature

A special nomenclature will be used to separate different models from each other. All models will be represented as in the following example.

$$M_{p_{em},arx3|4|2} \tag{3.21}$$

M means that it is a model, p_{em} represent the regressor that was used when estimating the model. One or more inputs can be used. $arx3|4|2$ means that the model has an ARX structure and that the structure parameters have the value $na = 3$, $nb = 4$ and $nk = 2$.

3.6 System Identification Toolbox

The System Identification Toolbox (SITB) is a toolbox in Matlab, that provide tools for creating mathematical models of dynamic system based on observed input and output data. It includes both linear and nonlinear standard models that can be used for creating both black-box and grey-box models. With both its flexible graphical user interface and the possibility to work at the command line it offers a great help when creating models.

This tool will be used when estimate and validate the models in this thesis.

Chapter 4

Data Preprocessing

The measurement data used in this thesis is the same that was used in [6] and was previously collected in a test cell at Scania CV AB in Södertälje. The data is from a six cylinder Scania diesel engine with EGR and VGT and was collected with a sample frequency of 100 Hz during a European Transient Cycle (ETC). The figures in Appendix B.1 shows the collected data.

4.1 Dealing with Incorrect Data

It is known that the measurements contain some incorrect data. In several of the signals the first sample differs a lot from the rest in the series, which likely depend on the startup of the measurement equipment, and therefore it has been removed from all signals. It is also known that the turbine speed (n_{trb}) is registered as zero when it drops below 20 000 rpm. These inaccuracies are not desirable when building a model of the system and therefore they need to be removed. Sometimes it can be done by prediction or interpolation, but if possible, a good idea is to choose a sequence without inaccuracies.

Figure 4.1 shows that there are two long segments of the turbine speed data that contains the measured value without any visible inaccuracies. The second one, that begins at 600 s and lasts to 1790 s, when the speed again drops below 20 000 rpm, ought to contain enough data for both estimation and validation of the models in this work and will be used. It will be left for future work to find a way to handle the inaccuracies, e.g. the loss of turbine speed below 20 000 rpm.

4.2 Removing Trends

Linear models cannot capture arbitrary differences between the input and output signal levels [10]. Therefore, a linear trend is removed from each signal. Then the estimated models focus on describing the relationship between the change in input signals and the change in output signals.

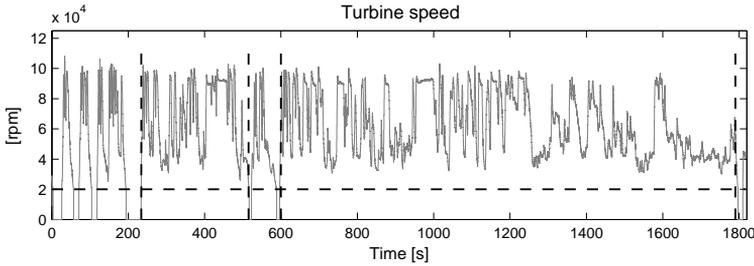


Figure 4.1: The signal n_{trb} . There are two main segments without visible inaccuracies, one between 234 s and 515 s and another between 600 s and 1790 s. The latter is used for system identification and validation.

4.3 Resampling the Data

If looking at a short sequence of the intake manifold pressure (p_{im}) shown in Figure 4.2, it can be seen that the signal contains a lot of high-frequency variations or noise. It is possible to let a good noise model handle this and estimate the high-frequency disturbance, but instead it was decided to remove the disturbance by filtering the signals with a low pass filter because it seems to simplify the modeling.

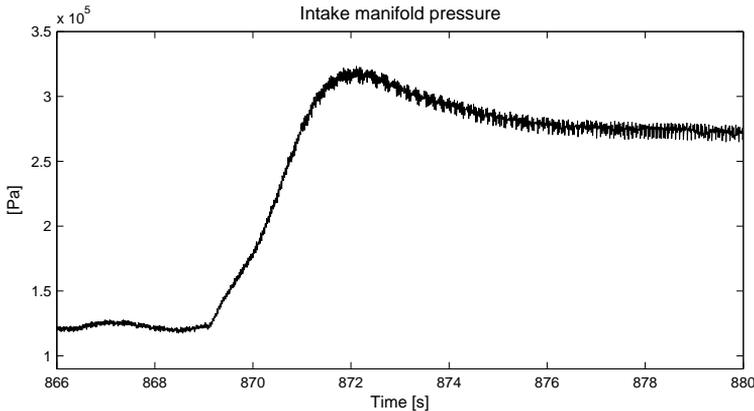


Figure 4.2: A segment of the signal p_{im} showing the high-frequency variation.

Not only low pass filtering is desirable, but rather downsampling is something to consider. The data sequences contain much more samples than is required to capture the variations in the signals. Resampling will reduce the data sequences to a length that makes the modeling process smoother but, if done correctly, still containing enough relevant information about the signal. An even bigger advantage with a lower sample frequency is when the model is implemented in a real system. If the signals are sampled faster than needed to capture the system dynamics the extra data points are redundant and the process uses resources that can be better

used of other purposes.

Low pass filtering not only remove the, above mentioned, high frequency components of the signal which is desirable. It is also necessary to avoid aliasing and get a correct resampling. Aliasing will be described in more detail in Section 4.3.1.

In the case of resampling, a rule of thumb is to use a sample frequency that is 10 times the bandwidth of the system which corresponds roughly to 4-8 samples along the rise time of a step response [12, 13]. Because the data was pre-collected it is not possible to do step responses in the data that could give an estimation of the bandwidth, and it can be difficult anyway due to the complexity of the system (it is not easy to do clean steps in most of the signals). Instead, looking at the energy of the signal can give an idea of a suitable sampling frequency. But first the alias effect will be explained.

4.3.1 The Sampling Theorem and Alias Effect

A well known term in signal processing is the Nyquist frequency [3], ω_N , that is of importance in the sampling theorem and is given by $\omega_N = \omega_s/2$, where ω_s is the sampling frequency.

According to the sampling theorem (or the Poisson's summation formula) frequencies higher than the Nyquist frequency will incorrectly be interpreted as lower ones. This is called the alias effect and is not desirable. To avoid aliasing the signals can be filtered with a low-pass filter with the cut-off frequency just at the Nyquist frequency before sampling.

4.4 Signal Energy and Periodogram

In [10] it is described that the total energy of a signal can be decomposed into energy contributions from different frequencies. The contributions from the discrete frequencies $\omega = \frac{2\pi}{NT}n$, $n = 0, 1, \dots, N-1$, could be represented in a periodogram [3]:

$$\hat{\Phi}_N(2\pi n/NT) = \frac{T}{N} |Y_N[n]|^2, \quad (4.1)$$

where N is the number of samples, T is the sampling interval and $Y_N[n]$ is the discrete Fourier transform (DFT) of the signal:

$$Y_N[n] = \sum_{k=0}^{N-1} e^{-2\pi i kn/N} y(kT). \quad (4.2)$$

In Appendix B.3, the energy distributions for the signals are represented in periodograms. They look almost the same and all of them show that the main part of the energies are concentrated at low frequencies which indicates that even if the signals are resampled to a much lower frequency, most of the energy will remain. As can be seen, the periodogram for the exhaust manifold pressure has

a small top at about 20-30 Hz. Also the periodogram for the engine speed signal shows this behavior. This can affect the frequency needed to get the remaining signal energy above a desirable level. If looking at the case when 99 % of the signal energy should remain, the cut of frequencies should be as listed in Table 4.1. As can be seen the frequency for the engine speed signal does not differ from the other ones, but the frequency for the exhaust manifold pressure does. Its value of 27.1303 Hz is very much higher than the second highest (0.9101 Hz for the VGT control signal). This means that 8 of 9 signals had to be sampled much faster than required if choosing a cut of frequency of 27 Hz that is required by p_{im} to achieve a 1 % signal energy loss, which may be a waste of resources.

Table 4.1: Cut of frequencies of low-pass filters where 99 % of the signal energy remains.

Signal	Frequency [Hz]
Air mass-flow before compressor	0.3471
Exhaust manifold pressure	27.1303
Turbine speed	0.2277
Intake manifold pressure	0.2563
Intake manifold temperature	0.0874
Engine speed	0.3269
Injected fuel	0.3555
EGR control signal	0.7067
VGT control signal	0.9101

If doing another calculation and keeping the ratio at a 95 % level the frequencies become as listed in Table 4.2. The frequencies are now much closer together and a new sample frequency can be chosen that is more suitable to all signals.

Table 4.2: Cut of frequencies of low-pass filters where 95 % of the signal energy remains.

Signal	Frequency [Hz]
Air mass-flow before compressor	0.1798
Exhaust manifold pressure	0.2992
Turbine speed	0.1328
Intake manifold pressure	0.1555
Intake manifold temperature	0.0395
Engine speed	0.1748
Injected fuel	0.1832
EGR control signal	0.3294
VGT control signal	0.3538

As mentioned above, a signal should be low-pass filtered with a cut-off frequency just at the Nyquist frequency before sampling to avoid aliasing. In this

case we want the Nyquist frequency to be higher than all the frequencies in Table 4.2 to keep 95 % of the signals energy. Because the Nyquist frequency is half the sample frequency it means that the sample frequency had to be at least 0.7076 Hz (then the Nyquist frequency are the same as the frequency for VGT control signal in Table 4.2). This frequency will be rounded to a simpler frequency of 1 Hz.

If the air mass-flow is resampled at 1 Hz with an antialias filter the resulting signal look likes in Figure 4.3.

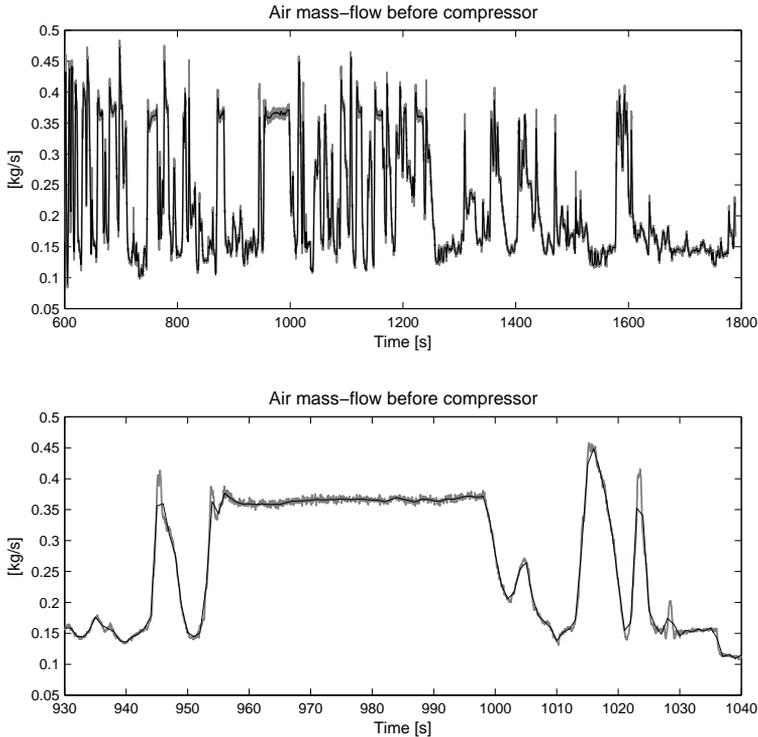


Figure 4.3: Comparison between the original 100 Hz signal (grey line) and the resampled 1 Hz signal (black line). The lower graph shows a short segment of the upper one.

With this sample frequency almost all dynamics is reproduced, only the sharpest peaks are missed. To also include those peaks the sample frequency was increased to 2 Hz and it seems to give a satisfying representation of all the signals. The peaks missed with frequency 1 Hz are now completed, the sample frequency gives a Nyquist frequency above the frequency where 95 % of the energy remains and the data is heavily reduced. Appendix B.2 contains a comparison of the original and the resampled signals and the remaining energies are listed in Table 4.3.

Table 4.3: Signal energy remaining after low-pass filtering with cut of frequency at 2 Hz.

Signal	Energy [%]
Air mass-flow before compressor	99.94
Exhaust manifold pressure	97.82
Turbine speed	99.99
Intake manifold pressure	99.87
Intake manifold temperature	99.98
Engine speed	99.88
Injected fuel	99.97
EGR control signal	99.85
VGT control signal	99.68

4.5 Estimation and Validation Data

Before the identification the sequence is divided into two parts. The reason for that is to get a fresh sequence that can be used for the model validation. The first 1500 samples in the resampled sequence will be the estimation data and the remaining 880 samples will be used for validation.

Chapter 5

Linear Regressor Models

This chapter contains the model development procedure. The models are built according to the standard linear models and criteria presented in Chapter 3. The models are validated on the basis of its fit and residual correlation analysis.

5.1 Modeling Strategy

One of the main goals is to find which signals that will be needed to get a good model of the air mass-flow. It is possible to spend a lot of time by trial and error before a good model is found and it is hard to know in advance if a new set of parameters will increase or decrease the quality of the model. Therefore, a modeling strategy is needed.

In [17] the principle “try simple things first” is used. Following this principle, some simple models are built, and if it is needed, they can be extended to more complex or advanced models. It is possible that a simpler model will give a good enough description of the system in its sector of application.

To begin with, the inputs will be separated and one model will be built for each of them. This is called a Single Input Single Output (SISO) model. This will be very simple models and hopefully it will show if some signals will be better than others to use in the model of the air mass-flow.

It is not obvious which of the model structures presented in Chapter 3 to use. The ARX structure is a bit simpler than the ARMAX and the BJ model but it still includes a noise model. Because of that, the ARX structure will be used at first. If the results indicate that the ARX structure does not fulfill the requirements, the auto- and cross-correlation of the residuals can be used to find out what the model do not capture and then the model may be adjusted to capture these missing parts or otherwise another model can be built accordingly to the other structures.

The ARX structure also has another small advantage towards the other structures, that is, in SITB there is several functions built specifically for estimating ARX models.

5.2 Single Input Single Output Models

SITB is used to develop all models. When developing the ARX models the estimation focus is set to prediction. This means that the model is determined by minimizing the prediction error, [11, p.3-31], and it is optimized for predicting the output.

To find the model order that results in the best model, 1000 models are computed with all combinations of the structure parameters na , nb and nk , in the range 1 to 10. The model with the best fit for each given number of parameters is displayed in a diagram and the model with the overall best fit calculated from the loss function and the best choice according to Akaike's Information Criterion are highlighted in dark grey and light grey respectively. Figure 5.1 shows the diagram for the models with input p_{im} as an example. The heights of the bars represent the part of the output variance which is not explained by the model. In this input, and for most of the others inputs too, the difference between the height of the bar corresponding to best fit and AIC is almost negligible.

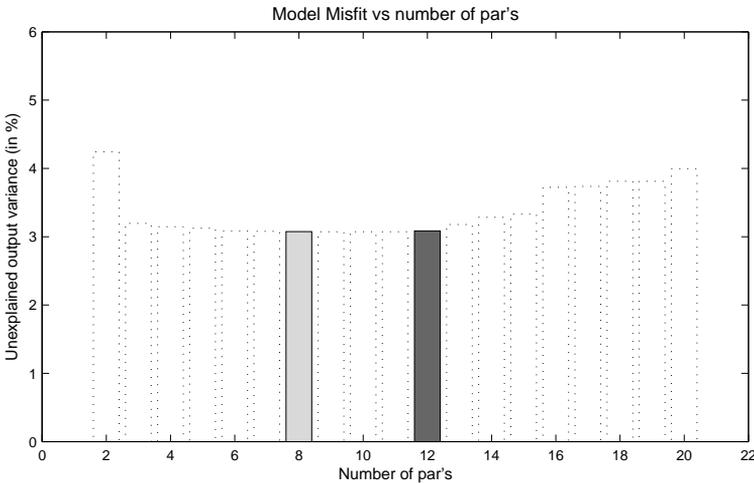


Figure 5.1: The best fit to validation data for ARX-model with input p_{im} and output W_{air} . The light grey bar is the best choice according to Akaike's Information Criterion and the dark grey bar is the choice according to best fit.

The best set of structure parameters will be chosen accordingly to Akaike's Information Criterion. The procedure described above is used for all inputs and the results become as in Table 5.1.

5.2.1 Fit Calculation

A first step of evaluating the models is to see how well they can re-create the measured output. This is done by calculating the fit between the output provided by the model and the measured output according to Equation (3.17) on page 11.

Table 5.1: Numbers of parameters in the models with one input, that are optimal according to Akaike's Information Criterion.

Model Name	Input Signal	na	nb	nk
$M_{p_{em},arx9 1 5}$	Exhaust manifold pressure	9	1	5
$M_{n_{trb},arx2 10 1}$	Turbine speed	2	10	1
$M_{p_{im},arx4 4 2}$	Intake manifold pressure	4	4	2
$M_{T_{im},arx5 3 1}$	Intake manifold temperature	5	3	1
$M_{n_{eng},arx5 5 1}$	Engine speed	5	5	1
$M_{\delta,arx7 2 1}$	Injected fuel	7	2	1
$M_{U_{EGR},arx5 6 1}$	EGR control signal	5	6	1
$M_{U_{VGT},arx7 1 6}$	VGT control signal	7	1	6

There are some options when calculating the fit. It is possible to choose the number of previous output values to include when predicting the next output value. It is possible to include all previous outputs, some of them or completely exclude them. This is referred to as k -step prediction of the output. The number k specifies which outputs that are included in the prediction of the output at time t . If $k = 1$, it means that all input values up to time t and all outputs up to time $t - 1$ are used in the prediction. This is called 1-step-ahead prediction of the output. Similarly, k -step-ahead prediction means that output values up to $t - k$ are used to estimate $\hat{y}_k(t|\theta)$ (the k -step-ahead prediction for the model with parameters θ). If $k = \infty$ it means that only the inputs are used in the prediction and this is equal to a simulation of the output.

In Table 5.2 the fit for the computed models are listed for some different values of k .

Table 5.2: Fit with different prediction horizons.

Name	1-Step	3-Step	5-Step	10-Step	Simulation
$M_{p_{em},arx9 1 5}$	82.09%	53.60%	38.42%	28.71%	28.47%
$M_{n_{trb},arx2 10 1}$	82.59%	71.46%	68.77%	67.00%	66.90%
$M_{p_{im},arx4 4 2}$	82.21%	55.28%	49.05%	48.60%	48.63%
$M_{T_{im},arx5 3 1}$	81.97%	52.95%	36.52%	22.73%	17.12%
$M_{n_{eng},arx5 5 1}$	82.07%	52.71%	34.98%	17.00%	-5.45%
$M_{\delta,arx7 2 1}$	84.62%	70.11%	64.04%	59.51%	58.28%
$M_{U_{EGR},arx5 6 1}$	82.80%	58.97%	47.31%	30.49%	7.02%
$M_{U_{VGT},arx7 1 6}$	82.04%	53.22%	37.38%	22.82%	9.77%

The question is what value k should have. Normally it should be bigger than the system delay, but in Chapter 4 it emerged that it was not trivial to find the system delay for this system and that will make it harder to correctly set the value of k . Therefore the value of k had to be chosen another way.

In Table 5.2 the fit for different values of k was presented. Not surprising,

the fit at 1-step-ahead prediction (where all previously outputs are known and included in the calculation, is the overall best fit for the model. As can be seen, this fit is about 82-84 % and that result returns in every test made. Using this value of k seems to make it harder to discern which model that is better than another because the difference between the fits are small. A bigger k -value results in a bigger difference between the resulting fits and therefore it seems easier to evaluate the models. In this thesis, $k = \infty$ was chosen, which means the simulation case where no previously outputs are used.

The result shows that the fit decreases for every model when the value of k increases, but it decreases less for some models. If looking at the fit in the case of simulation, three models are much better than the others, the one with regressor n_{trb} , p_{im} and δ . It seems to be a good idea to look deeper at models based on these inputs. If models where the simulated fit is closer to the 1-step-ahead prediction fit can be found, it means that the measured outputs do not contribute so much to the model and therefore can be omitted.

According to the paragraph above, the fit will henceforth be calculated and presented as the simulated fit.

5.3 Residual Analysis

To further validate a model, its residuals will be studied. The reason for this is to see if the model is robust or if there are things that it could not describe, as mentioned in Section 3.3.2.

The residuals for the model $M_{n_{trb}}$ are presented in Figure 5.2 and in Appendix B.4 the residuals for all models in Table 5.1 are presented.

The lower graph in these figures shows the cross-correlation between the input and the output for the models. As can be seen in Appendix B.4, for positive delays the cross-correlation is inside its confidence region (99%) for nearly all models and according to Section 3.3.2 it is desirable that this criterion will be fulfilled. The only model that does not fulfill this requirement is $M_{n_{trb}}$ which cross-correlation slowly varying outside the confidence region. According to the rule of thumb in Section 3.3.2 this is a indication of too few poles in this model. Section 5.3.1 will further deal with this. For negative time lags, the cross-correlation vary outside the confidence region for all models without $M_{n_{trb}}$. When this happens it is an indication of that the system worked with feedback when the data was collected. According to [12] this can be disregarded when validating the model.

The upper one of the graphs in Figure 5.2 and the figures in Appendix B.4 shows a models auto-correlation. As can be seen it is completely inside its confidence region for all models and it change stochastically. This means that the disturbance is correctly modeled as white noise, which means that the ARX-structures noise model is satisfying and a more advanced noise model is not needed.

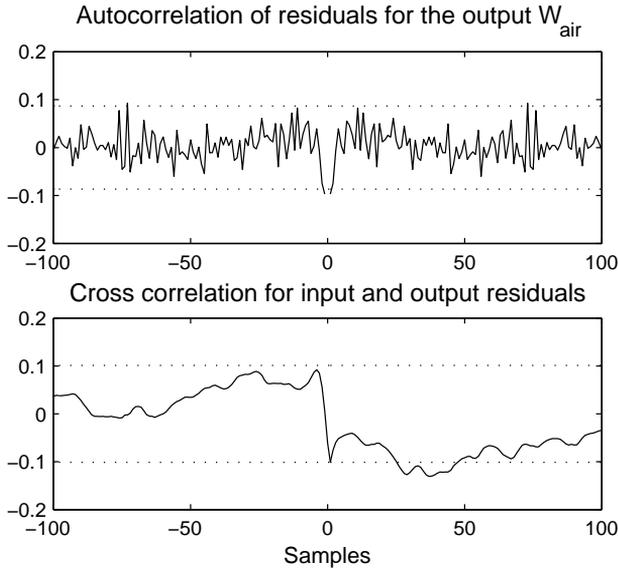


Figure 5.2: Residuals for $M_{n_{trb}}$.

5.3.1 Pole-Zero Diagram

The pole-zero diagram for the models is shown in Appendix B.5. They show that all of the models have their poles inside the unit circle with means that they all are stable. Looking at some of the diagram, it is obvious that poles and zeros could cancel each other which results in the possibility to create smaller models that gives similar results. There are several ways to reduce model order for linear systems and avoiding cancellation. The most common is model reduction by balanced truncation introduced in [14]. Unfortunately this was not known by me at this moment and therefore it was not used.

The residuals for $M_{n_{trb}}$ indicate that it could be too few poles. If the number of poles increases, the model order also increases, but because the pole-zero diagram for the model shows that it already is possible cancellations an increase of the number of poles will only result in modeling the noise [10, p.524]. No way to deal with this has been found.

The estimated models show that some signals are better than others to use as regressors when building a black-box model. The question is if the best model has been found. The residual analysis shows that the model with best fit do not has completely satisfying residuals and the simulation fit for this model is much smaller than for the 1-step-ahead prediction fit. It is possibly to use more than one input regressor when creating a black-box model and usage of more collected signal from the engine will probably be an advantage. To find out, models with multiple inputs are built in the next section.

5.4 Multiple Input Single Output Models

The next approach will be to combine the regressors used in Section 5.2 to form models with more than one input regressor. Such a model is called Multiple Input Single Output (MISO) Model. Hopefully, the use of more than one regressor will give an advantage when creating the model because information is available from different signals. The three regressors who gave the best result when it was used as a single input regressor in Section 5.2 are combined to four different models. As before, Akaike's Information Criterion is used to find the optimal set of parameters. The model name, its input regressors and the optimal numbers of parameters are presented in Table 5.3. Because of the way the tool for AIC calculations in SITB works, all input regressors to a model got the same structure parameters. Therefore, the structure parameters is not the same as for the models with one input regressor.

Table 5.3: Numbers of parameters in the models with multiple inputs, which are optimal according to Akaike's Information Criterion.

Model Name	Input Signal	na	nb	nk
$M_{n_{trb},p_{im}}$	Turbine speed	1	9	1
	Intake manifold pressure	1	9	1
$M_{n_{trb},\delta}$	Turbine speed	7	2	1
	Injected fuel	7	2	1
$M_{p_{im},\delta}$	Intake manifold pressure	2	4	1
	Injected fuel	2	4	1
$M_{n_{trb},p_{im},\delta}$	Turbine speed	1	7	1
	Intake manifold pressure	1	7	1
	Injected fuel	1	7	1

When the fit are calculated for these models the result becomes as in Table 5.4.

Table 5.4: Fit for the models with multiple inputs.

Name	Simulation Fit
$M_{n_{trb},p_{im}}$	65.68%
$M_{n_{trb},\delta}$	69.07%
$M_{p_{im},\delta}$	49.54%
$M_{n_{trb},p_{im},\delta}$	42.83%

The fit for $M_{n_{trb},p_{im}}$ and $M_{n_{trb},\delta}$ are in the same range as for the best of the models in Section 5.2.1. $M_{n_{trb},\delta}$ shows an increase of just over 2 percentage points (69.07%) whilst $M_{n_{trb},p_{im}}$, with a 65.68% fit, is slightly behind the best SISO model. There are no distinct increases in fit when adding another signal to the input of the model. Maybe the structure parameters should not be the same for all input regressors, but according to the reason above, this was not

further analyzed. Another reason could be that the regressors do not contain much additional information about the air-mass flow and then the fit does not increase.

The residuals for all MISO models, attached in Appendix B.6, looks good and are mainly inside their confidence interval of 99 %. Some of them touches the edge of the confidence interval and the residuals for $M_{n_{trb}, p_{im}}$ is slightly outside the interval, but specially the residual connected to n_{trb} looks better than it does for the SISO model $M_{n_{trb}, arx2|10|1}$.

So far, it seems that the best results will be derived with a model that contains regressors based on p_{im} , n_{trb} , δ or combinations of these. In the next chapter, models with non-linear regressors will be analyzed. Physical relations between the air mass-flow and the other signals will be used to find these non-linear regressors, i.e. the resulting models are grey-box models.

Chapter 6

Non-linear Regressor Models

Johan Wahlström presented a detailed description and physical relations between parameters in a mean value diesel engine with VGT and EGR in [20] and this model was later used in [1]. The relations for the air mass-flow will be used to form custom regressors to use together with the linear black box models to see if this can result in an improved fit for the models. The next section summarizes the relations for the air mass-flow that will be used in this thesis. Related model with another choice of regressors can be found in many other works, e.g. in [9, 1, 18, 2, 15, 16].

6.1 Physical Relations for the Air Mass-Flow

The structure of the engine system used when developing the mean value model of a diesel engine in [20] is basically the same as presented in Section 2.

The mass flow through the compressor, W_c , is the same as the air mass-flow before compressor modeled in this thesis. The denomination W_c used in [20] will be kept to separate the air mass-flow calculated from these equations from the measured one, i.e. W_{air} . The mass-flow through the compressor is modeled using two dimensionless variables. The first variable is the energy transfer coefficient

$$\Psi_c = \frac{2c_{pa}T_{amb}(\Pi_c^{1-1/\gamma_a} - 1)}{R_c^2\omega_t^2} \quad (6.1)$$

which is the quotient of the isentropic kinetic energy of the gas at the given pressure ratio Π_c and the kinetic energy of the compressor blade tip, where

$$\Pi_c = \frac{p_{im}}{p_{amb}}. \quad (6.2)$$

The second variable is the volumetric flow coefficient

$$\Phi_c = \frac{W_c/\rho_{amb}}{\pi R_c^3\omega_t} = \frac{R_a T_{amb}}{p_{amb}\pi R_c^3\omega_t} W_c \quad (6.3)$$

which is the quotient of volume flow rate of air into the compressor and the rate at which volume is displaced by the compressor blade. The relation between Ψ_c and Φ_c can be described by a part of an ellipse

$$c_{\Psi_1}(\omega_t)(\Psi_c - c_{\Psi_2})^2 + c_{\Phi_1}(\omega_t)(\Phi_c - c_{\Phi_2})^2 = 1 \quad (6.4)$$

where c_{Ψ_1} and c_{Ψ_2} varies with compressor speed ω_t and are modeled as polynomial functions.

$$c_{\Psi_1}(\omega_t) = c_{\omega\Psi_1}\omega_t^2 + c_{\omega\Psi_2}\omega_t + c_{\omega\Psi_3} \quad (6.5)$$

$$c_{\Phi_1}(\omega_t) = c_{\omega\Phi_1}\omega_t^2 + c_{\omega\Phi_2}\omega_t + c_{\omega\Phi_3} \quad (6.6)$$

The air mass flow is modeled by solving Φ_c from Equation (6.4) and solving W_c from Equation (6.3)

$$\Phi_c = \sqrt{\frac{1 - c_{\Psi_1}(\Psi_c - c_{\Psi_2})^2}{c_{\Phi_1}}} + c_{\Phi_2} \quad (6.7)$$

$$W_c = \frac{p_{amb}\pi R_c^3 \omega_t}{R_a T_{amb}} \Phi_c. \quad (6.8)$$

Table 6.1 and Table 6.2 presents a summary of the variables respective the constants used in Equation (6.1) - (6.8).

Table 6.1: Summary of the variables.

Variables	Description
p_{im}	Intake manifold pressure
ω_t	Rotational turbo speed
W_c	Compressor mass flow

Table 6.2: Summary of the constants.

Constants	Description
ρ_{amb}	Density of the ambient air
p_{amb}	Ambient pressure
T_{amb}	Ambient temperature
R_c	Compressor blade radius
R_a	Ideal gas constant for air
γ_a	Specific heat capacity ratio for air
c_{pa}	Specific heat capacity at constant pressure, for air
$c_{\Psi 2}, c_{\Phi 2}$	Parameter in the ellipse model for the compressor mass flow
$c_{\omega \Psi 1}, c_{\omega \Psi 2}, c_{\omega \Psi 3},$ $c_{\omega \Phi 1}, c_{\omega \Phi 2}, c_{\omega \Phi 3}$	Coefficients in the polynomial equation (6.5) and (6.6)

6.2 Custom Regressors

The above equations show that two signals have a close connection to the air mass-flow, the turbine speed and the intake manifold pressure. In these equations both turbine speed and intake manifold pressure occur in nonlinear expressions and these can be used as custom regressors. These two regressors both give good results in Chapter 5. The injected fuel, δ , does not appear in the equations but because it gives good results when creating models with one linear regressor and also shows a slightly improved fit when it was combined with the turbine speed to form a MISO model, it will be included further on.

These quite simple custom regressors can also be combined until they form the complete expression of the air mass-flow. At first, models with only one input will be made.

6.2.1 Turbine Speed Regressors

At first, the regressors based upon the turbine speed are examined. A closer look at the equations shows that a square, an inverse and a squared inverse of the turbine speed are locally represented in the equations. These expressions can be used as regressors.

Table 6.3 holds the turbine speed regressors examined. The linear turbine speed is used as a reference.

The models are estimated in the same way that before, the regressors are used one by one as inputs, Akaike's information criterion is used to obtain the best combination of the structure parameters and the fit is calculated. The results is presented in Table 6.4.

As can be seen, one of the models, $M_{R_{trb2}, arx1|10|1}$, perform better than the linear turbine speed regressor. The other two, with R_{trb3} and R_{trb4} as regressors, give a clearly worse result.

Table 6.3: Regressors based upon the turbine speed.

Name	Regressor
R_{trb1}	ω_t
R_{trb2}	ω_t^2
R_{trb3}	$1/\omega_t$
R_{trb4}	$1/\omega_t^2$

Table 6.4: Fit for optimal ARX models based on turbine speed regressors, due to Akaike's information criterion.

Model Name	Simulation Fit
$M_{R_{trb1}, arx2 10 1}$	66.90%
$M_{R_{trb2}, arx1 10 1}$	73.89%
$M_{R_{trb3}, arx5 5 3}$	24.81%
$M_{R_{trb4}, arx5 1 10}$	8.99%

The conclusion of this test is that a custom regressor based on the turbine speed should contain the square of the turbine speed but due to the fit test it should not be an inversed square of the turbine speed.

One remaining question is if the residuals of the model with the squared turbine speed regressor, $M_{R_{trb2}, arx1|10|1}$, are good. Figure 6.1 shows the residuals and they seem to fulfill the requirement. The cross correlation is better than for the model based only on turbine speed, i.e. $M_{n_{trb}, arx2|10|1}$, and is now completely inside the confidence interval.

6.2.2 Pressure Regressors

Whereas there were several possible turbine speed regressors, that is not the case for regressors based on the intake manifold pressure. It is just one expression that can be used and that is the non-linear pressure quotient listed in Table 6.5.

Table 6.5: Regressors based upon the intake manifold pressure.

Name	Regressor
R_{pres1}	Π_c^{1-1/γ_a}

With the same procedure as in the case with turbine speed regressors, the result becomes as in Table 6.6, where the result for the model based on just the intake manifold pressure also is listed to make it easier to compare the result.

This custom regressor gives about 10 percentage points improvement of the fit. This improvement is approximately as large as the one that the squared turbine speed regressor gives. But because the starting fit is much lower the result cannot match the results from the turbine speed regressor. However, the residuals fit inside the confidence interval, just as for the linear regressor.

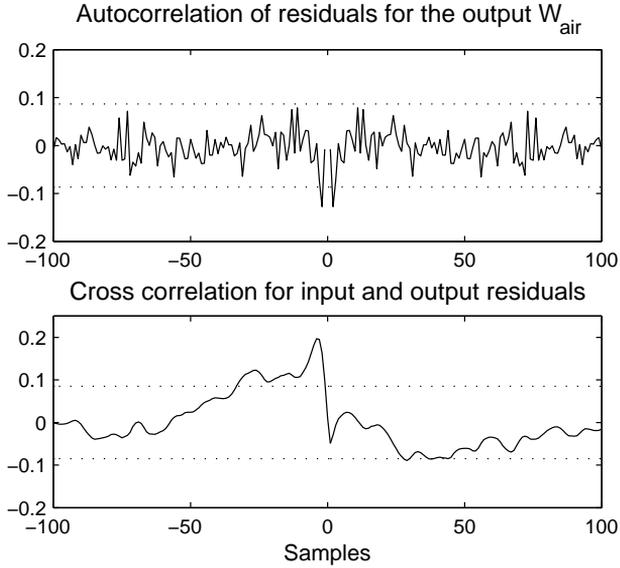


Figure 6.1: The residuals for $M_{R_{trb2,arx1}|10|1}$.

Table 6.6: Fit for optimal ARX models based on the intake manifold pressure regressors, due to Akaike’s information criterion.

Model Name	Simulation Fit
$M_{p_{im},arx4 4 2}$	48.63%
$M_{R_{pres1},arx4 6 1}$	57.23%

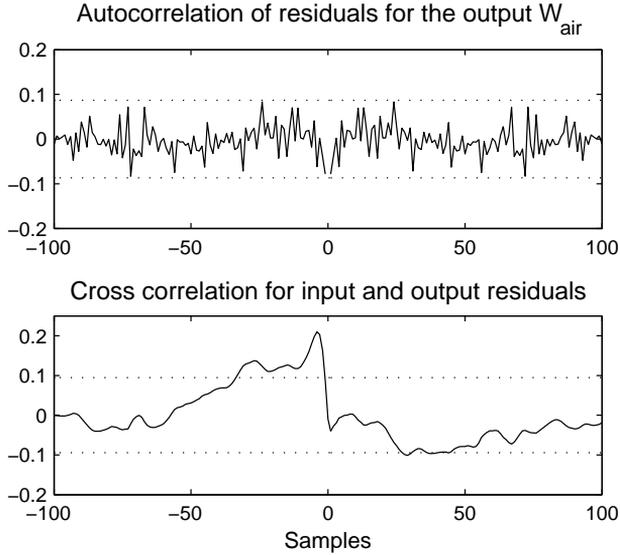


Figure 6.2: The residuals for $M_{R_{pres1,arx4|6|1}}$.

6.2.3 Combined Regressors

The question now is if the fit will improve if regressors that are even more like the complete equations will be used. This will be tested and the regressors will be equal to the three equations that form the air mass-flow, i.e. Equation (6.1), (6.7) and the complete expression (6.8). The regressors are listed in Table 6.7.

Table 6.7: Regressors based upon both turbine speed and intake manifold pressure.

Name	Regressor
R_{comb1}	Ψ_c
R_{comb2}	Φ_c
R_{comb3}	W_c

The models are estimated exactly as before and the results are presented in Table 6.8.

The first of these regressors gives a very bad result but the other two give a fit that is among the best so far. The residuals for the good models looks almost as in Figure 6.1 but some peaks are outside the confidence interval just like in the case for the linear turbine speed regressor. For all residuals, see Appendix B.7.

If consider that the regressor R_{trb2} is much simpler than R_{comb2} and R_{comb3} are, it ought to be a better choice to use the first one. This gives less calculation and contingent problems with the tuned parameters in the mean value diesel engine model can be avoided.

Table 6.8: Fit for optimal ARX models based on a combination of turbine speed and intake manifold pressure regressors, due to Akaike's information criterion.

Model Name	Simulation Fit
$M_{R_{comb1}, arx5 1 9}$	9.91%
$M_{R_{comb2}, arx1 10 1}$	73.05%
$M_{R_{comb3}, arx1 10 1}$	73.12%

6.3 Combination of Regressors and Signals

A last test will be done where the non-linear and linear, from Section 5.2, regressors that give the best results will be combined to MISO models. Thus the regressors will be kept separated to become multiple inputs to the system. The regressors to use is the squared turbine speed, the pressure quotient, the turbine speed, the intake manifold pressure and the injected fuel, where the last three are the ones that give the best result in Section 5.2.

In Table 6.9 the models and their results are presented. As before, the regressors and the model parameters can be read in the index for respective model.

Table 6.9: Fit for optimal ARX models based on a combination of standard and custom regressors, due to Akaike's information criterion.

Model Name	Simulation Fit
$M_{R_{trb2}, n_{trb}, arx1 8 1}$	72.91%
$M_{R_{trb2}, pim, arx1 2 1}$	71.71%
$M_{R_{trb2}, \delta, arx1 10 1}$	69.76%
$M_{R_{pres1}, n_{trb}, arx10 4 1}$	66.65%
$M_{R_{pres1}, pim, arx5 2 1}$	45.27%
$M_{R_{pres1}, \delta, arx3 7 1}$	61.50%
$M_{R_{trb2}, R_{pres1}, arx1 10 1}$	74.57%

The result show that this combination of regressors works well and several models give a relative good result, especially the ones that included the R_{trb2} , It seems that this regressor is the one that conduce the best results. But only one model is actually better than before, namely the one where the two 'grey' regressors, R_{trb2} and R_{pres1} , are combined, i.e. $M_{R_{trb2}, R_{pres1}, arx1|10|1}$. This results in a fit of 74.57% and both the residuals and the pole-zero diagram looks satisfying, see Figure 6.3 and Figure 6.4.

Even more advantages may be possible if this model are extended to include more than two input regressors, i.e. to include one or more of the linear regressors previously used in this section to form the two-input models in Table 6.9. If models are created upon the possible combinations of these regressors the results become as in Table 6.10.

The results for $M_{R_{trb2}, R_{pres1}, \delta, arx1|4|1}$ and $M_{R_{trb2}, R_{pres1}, pim, \delta, arx1|2|1}$ shows an

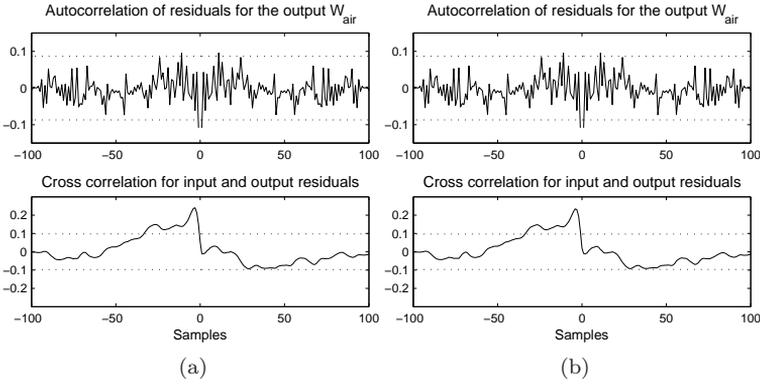


Figure 6.3: The residuals for $M_{R_{trb2}, R_{pres1}, arx1|10|1}$. Residuals for regressor R_{trb2} in (a) and for regressor R_{pres1} in (b).

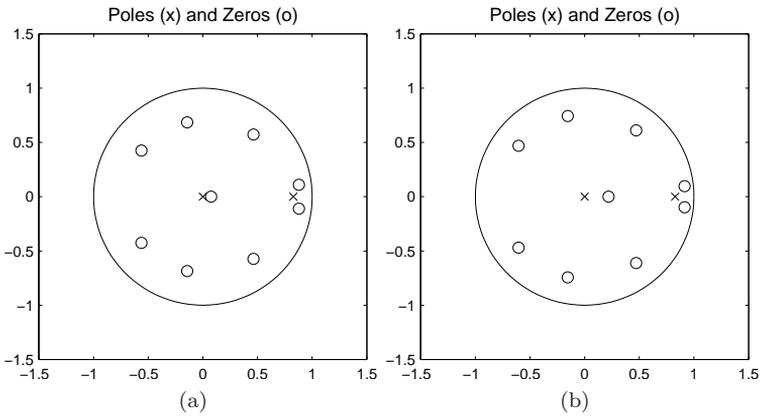


Figure 6.4: The pole-zero diagram for $M_{R_{trb2}, R_{pres1}, arx1|10|1}$. Diagram for regressor R_{trb2} in (a) and for regressor R_{pres1} in (b).

Table 6.10: Fit for optimal ARX models based on a combination of standard and custom regressors, due to Akaike’s information criterion.

Model Name	Simulation Fit
$M_{R_{trb2}, R_{pres1}, n_{trb}, arx1 5 1}$	74.47%
$M_{R_{trb2}, R_{pres1}, p_{im}, arx1 2 1}$	70.04%
$M_{R_{trb2}, R_{pres1}, \delta, arx1 4 1}$	80.70%
$M_{R_{trb2}, R_{pres1}, n_{trb}, p_{im}, arx1 5 1}$	75.46%
$M_{R_{trb2}, R_{pres1}, n_{trb}, \delta, arx4 6 1}$	Undef.
$M_{R_{trb2}, R_{pres1}, p_{im}, \delta, arx1 2 1}$	79.99%
$M_{R_{trb2}, R_{pres1}, n_{trb}, p_{im}, \delta, arx3 3 1}$	Undef.

unexpected increase in fit. These two models got a fit of about 80% and the thing they have in common is that the injected fuel, δ , has been added. Two of the models, also with δ as regressor, got an unreasonable result (an extremely big negative value) and is therefore declared undefined. If the number of structure parameters for these two models are slightly changed, this will be avoided. If doing this, the fit will become high, but not as high as for two best ones mentioned above and the model will not be optimal due to Akaike's information criterion.

The two best models are now based on three different regressors, one based on the turbine speed, one on the intake manifold pressure and one on the injected fuel. Another combination of these signals was used in Section 5.4 in the model $M_{n_{trb},p_{im},\delta}$, but this time the result was not good. Yet again the use of the non-linear regressors prove to be useful. And to point out, δ is not included in the physical relations described in Section 6.1. However, the use of this regressor also prove to be useful.

However, there are things that indicate that these models are not acceptable. The residuals for the best model, $M_{R_{trb2},R_{pres1},\delta,arx1|4|1}$, are shown in Figure 6.5. The residuals for $M_{R_{trb2},R_{pres1},p_{im},\delta,arx1|2|1}$ has the same fundamental look. Neither of the auto-correlation or the cross-correlation is satisfying. They are both way outside its own confidence interval. This is also the case for the other models in Table 6.10 that gives an improved fit. Thus, it seems to be a better choice to use $M_{R_{trb2},R_{pres1},arx1|10|1}$, which has the best fit and still has satisfying residuals and pole-zero diagram.

In summary, the previous work has resulted in a model that gives a simulation fit of nearly 75%. If looking at the resulting fit when old outputs are included when building the model, see the discussion in Section 5.2.1 and the results in Table 6.11, it can be seen that the results are close to each other even if the old outputs are used. Certainly, the fit increases when more old outputs are used, but the difference is approximately 4 percentage points between the case when $k = 3$ and no outputs are used. It is not a big difference.

Table 6.11: Fit for optimal ARX models, based on a combination of standard and custom regressors as a MISO system, due to Akaike's information criterion.

Model Name	1-Step	3-Step	5-Step	10-Step	Simulation
$M_{R_{trb2},R_{pres1},arx1 10 1}$	84.28%	78.27%	76.20%	74.83%	74.57%

In the following chapters, a summary of the results is done and the results is compared to the results from an extended Kalman filter used to estimate the air mass-flow. Then we get a measurement of how the solution created with system identification will perform.

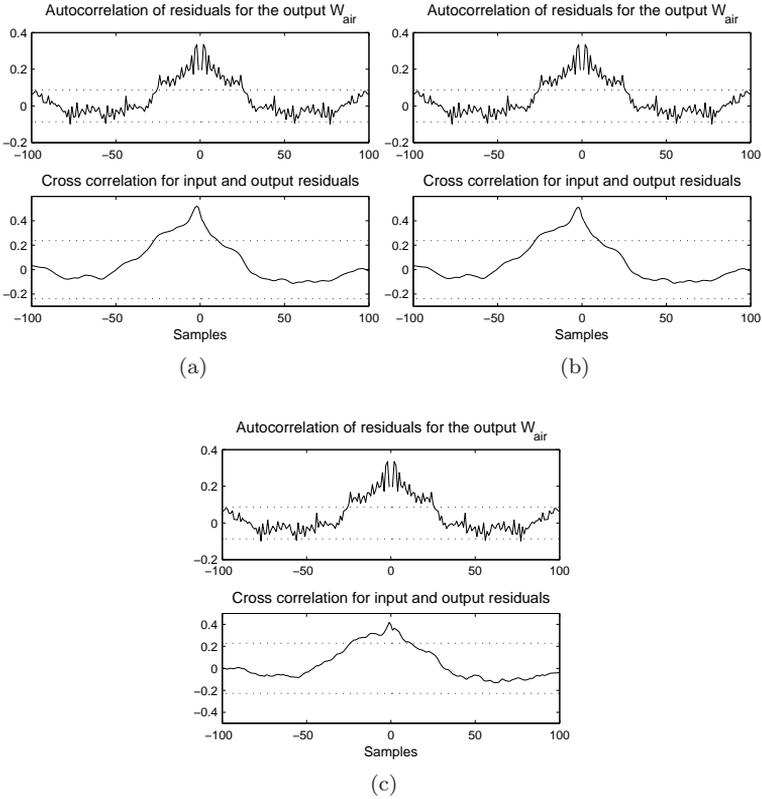


Figure 6.5: The residuals for $M_{R_{trb2}, R_{pres1}, \delta, arx1|4|1}$. Residuals for regressor R_{trb2} in (a), for regressor R_{pres1} in (b) and for regressor δ in (c).

Chapter 7

Results

In this chapter the best black-box model developed and its estimate of the air-mass flow is presented. The estimate is also compared to the estimate from the extended Kalman filter presented in Section 7.2 in order to evaluate its performance.

7.1 The Best Model

The work in this thesis to find a black-box model of the air mass-flow has result in a model based upon the ARX structure. This model, $M_{R_{trb2}, R_{pres1}, arx1|10|1}$, has the structure parameters $na = 1$, $nb = [10 \ 10]$ och $nk = [1 \ 1]$ and has two input regressors, the non-linear regressors ω_t^2 and Π_c^{1-1/γ_a} . There are two elements in nb and nk because the model has two inputs, where each element relate to one input regressor.

This model gives a fit of 74.57% and its residuals and pole-zero diagrams are satisfying. If the sequence between 600 s and 1790 s, chosen in Section 4.1, is used to simulate the model the estimated air mass-flow becomes as in Figure 7.1. It is compared with the measured air mass-flow.

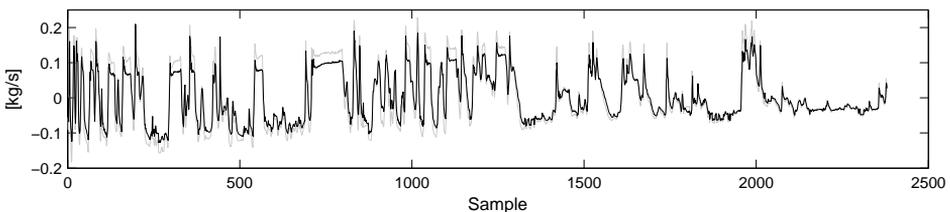


Figure 7.1: Estimated air mass-flow (black line) compared to measured air mass-flow (grey line) for the downsampled sequence between 600 s and 1790 s.

Figure 7.2 shows the part of sequence in Figure 7.1 that was used to validate

the model. Here it is a little bit easier to see the difference between the estimate and the measurement.

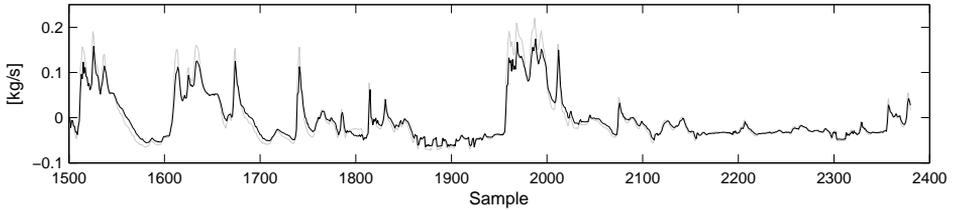


Figure 7.2: Estimated air mass-flow (black line) compared to measured air mass-flow (grey line) for the sequence used to validate the data.

If the difference between the estimated and the measured air mass-flow is calculated the results become as in Figure 7.3. As could be seen, the maximum absolute error is in the range of 0.1 kg/s which means the error is about 50 % wrong comparing to the measured value. However, some errors with this size is not a big deal and therefore it is better to look at the errors in another way, namely looking at the root mean square error. This will be done in Section 7.4, where the grey-box estimate also will be compared to the EKF estimate presented in Section 7.3.

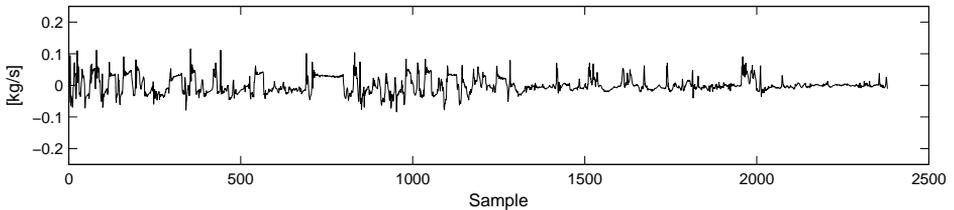


Figure 7.3: The absolute error between the estimated and the measured air mass-flow.

7.2 Extended Kalman Filter

To find out how well the best black-box model actually performs, its estimate of the air mass-flow will be compared to the estimate from an extended Kalman filter (EKF) [8].

7.2.1 Physical Model

The model used together with the extended Kalman filter is a modified version of the model presented in [20]. This slightly simplified model is used in [4].

It is a third order non-linear state-space model of an inline six cylinder Scania diesel engine with variable geometry turbo (VGT) and exhaust gas recirculation (EGR). The model states are intake manifold pressure, p_{im} , exhaust manifold pressure, p_{em} , and turbine speed, n_{trb} .

The basic structure of the model is

$$\dot{x} = \begin{pmatrix} \dot{p}_{im} \\ \dot{p}_{em} \\ \dot{n}_{trb} \end{pmatrix} = \begin{pmatrix} f_{p_{im}}(x) \\ f_{p_{em}}(x) \\ f_{n_{trb}}(x) \end{pmatrix}$$

$$y = (p_{im} \quad p_{em} \quad n_{trb} \quad h_{W_c})^T.$$

7.2.2 Filter Design

The designed EKF is based on the continuous time model of the air mass-flow in Section 7.2.1, and has feedback from all model states, that is p_{im} , p_{em} and n_{trb} . The covariance matrices for the system and measurement noise, Q and R , were used as tuning parameters. The EKF is designed as follows.

The continuous time model,

$$\begin{aligned}\dot{x} &= f(x, u) \\ y &= h(x),\end{aligned}$$

is discretized with forward Euler and a sampling time of T_s seconds,

$$x_{t+1} = x_t + T_s f(x_t, u_t) \quad (7.1a)$$

$$y_t = h(x_t). \quad (7.1b)$$

The time discrete EKF is designed on Equation (7.1b). The EKF equations for a time discrete model look as follows, starting with the internal variables

$$\begin{aligned}S_t &= H_t P_{t|t-1} H_t^T + R_t \\ K_t &= P_{t|t-1} H_t^T S_t^{-1} \\ e_t &= y_t - h(\hat{x}_{t|t-1})\end{aligned}$$

that is the innovation covariance, the Kalman gain, and the estimation error respectively. Continuing with the update equations

$$\begin{aligned}\hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t e_t \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1} H_t^T S_t^{-1} H_t P_{t|t-1}\end{aligned}$$

and the prediction equations

$$\begin{aligned}\hat{x}_{t+1|t} &= \hat{x}_{t|t} + T_s f(\hat{x}_{t|t}, u_t) \\ P_{t+1|t} &= (I - T_s A_t) P_{t|t} (I - T_s A_t)^T + Q_t\end{aligned}$$

where

$$A_t = \left. \frac{\partial f}{\partial x} \right|_{x=\hat{x}_{t|t-1}}, \quad H_t = \left. \frac{\partial h}{\partial x} \right|_{x=\hat{x}_{t|t-1}}.$$

Finally the estimator can be written

$$\hat{W}_{EKF,t} = h_W(\hat{x}_t)$$

where h_W is a nonlinear function of the states describing the air mass-flow through the compressor, see [20, 4]. The resulting estimate can be compared to the black-box estimate to evaluate its performance.

7.3 EKF Estimate

Figure 7.4 and Figure 7.5 shows the EKF estimate compared to the measured air mass-flow for the same sequences as in Figure 7.1 respectively Figure 7.2. This estimate looks like the grey-box estimate. It is possible to see some differences but it is hard to say if one is better than the other by just looking at the results. Instead, the root mean square error will be used for this.

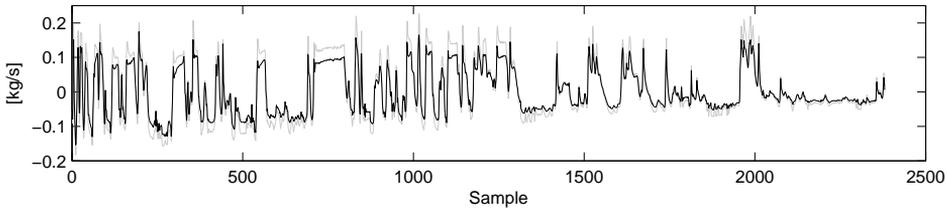


Figure 7.4: EKF estimated air mass-flow (black line) compared to measured air mass-flow (grey line) for the downsampled sequence between 600 s and 1790 s.

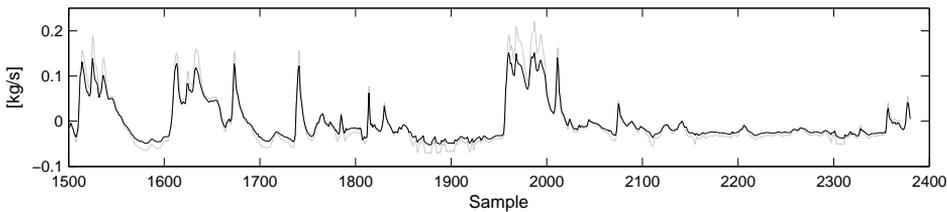


Figure 7.5: EKF estimated air mass-flow (black line) compared to measured air mass-flow (grey line) for the sequence used to validate the data.

7.4 Root Mean Square Error

The result from the grey-box model and from the extended Kalman filter can be compared if the root mean square error (RMSE) is calculated. Root mean square error is a frequently used measure of the difference between a measured and an estimated signal and gives a measure of the specificity of the whole signal, in contrast to the residuals who just gives the error in specific samples. The root mean square error is calculated according to:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad (7.2)$$

where y_i is the measured values, \hat{y}_i is the estimated values and n is the total number of samples. When the root mean square error for these two signals are

Table 7.1: The root mean square error for the grey-box estimate and the EKF estimate, for both the validation sequence and for the downsampled sequence between 600 s and 1790 s (called working sequence).

Estimate	Sequence	RMSE
Grey-Box	validate	0.0154
	working	0.0254
EKF	validate	0.0147
	working	0.0231

calculated the results become as in Table 7.1. It is calculated for both the validate sequence and for the longer sequence mentioned above.

The performance of the grey-box ARX model seems to be nearly the same as for the extended Kalman filter.

Chapter 8

Concluding Remarks

This chapter contains a summary of the conclusions made in this thesis. It also discusses problems that occur during the work and raise some questions and things to build future work on.

8.1 Conclusions

The purpose with this thesis was to use system identification theory to develop a model for the air mass-flow through the compressor. The modeling data was collected from nine different sensors on a six cylinder Scania diesel engine with EGR and VGT.

All models were built according to the ARX model structure. The reason for this was that this structure was a little bit simpler than others, but still includes a noise model. The System Identification Toolbox used also has most functions suited for ARX model development.

When first creating models with one single input regressor it turned out that the models with the intake manifold pressure, the turbine speed and delta as input regressor give clearly the best estimate of the air-mass flow through the compressor. Therefore, the continuing modeling focus on these signals. The creation of additional black-box models shows that the fit of the estimate was not improved if these regressors were combined to form models with multiple inputs.

Because of the systems non-linearity, the models had to be expanded to include some information about this. To achieve this, the non-linear regressor was used together with the linear ARX model. These non-linear regressors were based upon physical relations for the air-mass flow. It turned out that the use of these non-linear regressors improved the prediction performance. Several non-linear expressions with different complexity were tested, but it turned out that the performance of these models mostly depend on if one non-linear regressor was included in the model or not, namely the squared turbine speed.

The performance of the best model found, was compared to the estimate from an extended Kalman filter observer based on a physical model. This comparison shows that the best grey-box estimate performs just as well as the estimate from

the observer. The conclusion is that this model does not contribute with something that could not be achieved with an observer and an already existing model of the system.

8.2 Discussion of Problems

The use of this grey-box model has some restrictions. The fact is that the model is heavily dependent of the turbine speed, but as mentioned before, there are faults in the measured signal of the turbine speed (a registered zero when the turbine speed drops below 20 000 rpm) and this model do not deal with that. The effect of using an input signal that contains this error is still unknown. The work in this thesis shows that the usage of a model which does not contain the turbine speed based regressor provides a considerable lower fit and is probably no alternative. An option is to build a model that deal with the loss of turbine speed, but how to do that and the difficulties has not been dealt with.

Because of the fact that the best model in this thesis do not perform better than the physical model with the extended Kalman filter, the conclusion is that a lot more work has to be done before a grey-box model can be properly used to estimate the air mass-flow. This thesis has only dealt with a small part of the system identification research area and there is a lot more approaches, model structures and ways to choose the regressors to study.

8.3 Future Work

To directly proceed with the results in this thesis the problem with the loss of turbine speed had to be dealt with. It can possibly be done if finding a way to always measure a correct turbine speed or let the model itself deal with that.

Another important thing is to do more specific data collections that make it easier to describe the dynamics of the system. Run separate parts of the system to acquire step responses, time constants and easier find the bandwidth of the system to simplify the preprocessing of the data. If these methods to analyze the system could be more reliable than they become in this thesis, it would likely be an advantage. A lot of additional work can also be done when the models are derived. Instead of an ARX model, some of the other can be thorough analyzed and there are a lot of ways to find non-linear models, that not just use non-linear regressors, for example usage of wavelets and neural networks.

Bibliography

- [1] Lars Eriksson. Modeling and control of turbocharged SI and DI engines. *Oil & Gas Science and Technology - Rev. IFP*, 62(4):523–538, 2007.
- [2] Lars Eriksson, Lars Nielsen, Jan Brugård, Johan Bergström, Fredrik Pettersson, and Per Andersson. Modeling and simulation of a turbo charged SI engine. *Annual Reviews in Control*, 26(1):129–137, 2002.
- [3] F. Gustafsson, L. Ljung, and M. Millnert. *Signalbehandling*. Studentlitteratur, Lund, 2 edition, 2001.
- [4] Erik Höckerdal. Observer design and model augmentation for bias compensation with engine applications. Technical report, 2008. LiU-TEK-LIC-2008:48, Thesis No. 1390.
- [5] Erik Höckerdal, Lars Eriksson, and Erik Frisk. Air mass-flow measurement and estimation in diesel engines equipped with EGR and VGT. In *Electronic Engine Controls*, number 2008-01-0992 in SAE Technical paper series SP-2159, SAE World Congress, Detroit, USA, 2008.
- [6] Erik Höckerdal, Erik Frisk, and Lars Eriksson. Observer design and model augmentation for bias compensation applied to an engine. IFAC World Congress, Seoul, Korea, 2008.
- [7] Andreas Jerhammar and Erik Höckerdal. Gas flow observer for a Scania diesel engine with VGT and EGR. Master’s thesis, Linköpings Universitet, SE-581 83 Linköping, 2006.
- [8] Thomas Kailath, Ali H. Sayed, and Babak Hassibi. *Linear Estimation*. Prentice-Hall, Inc, Upper Saddle River, New Jersey 07458, 2 edition, 2000.
- [9] Oskar Leufven and Lars Eriksson. Time to surge concept and surge control for acceleration performance. IFAC World Congress, Seoul, Korea, 2008.
- [10] L. Ljung. *System Identification - Theory for the User, 2nd Edition*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 2 edition, 1999.
- [11] L. Ljung. *System Identification Toolbox - for use with MATLAB*. The Mathworks, Inc., 5 edition, 2000.

-
- [12] L. Ljung and T. Glad. *Modellbygge och Simulering*. Studentlitteratur, Lund, 2 edition, 2004.
 - [13] Lennart Ljung. System identification. Technical Report LiTH-ISY-R-2809, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, June 2007.
 - [14] B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, Feb 1981.
 - [15] Paul Moraal and Ilya Kolmanovsky. Turbocharger Modeling for Automotive Control Applications. *SAE Technical Paper 1999-01-0908*, pages 309–322, 1999.
 - [16] Martin Müller, Elbert Hendricks, and Spencer C. Sorenson. Mean Value Modelling of Turbocharged Spark Ignition Engines. *SAE SP-1330 Modeling of SI and Diesel Engines*, (SAE Technical Paper 980784):125–145, 1998.
 - [17] Jonas Sjöberg. *Non-Linear System Identification with Neural Networks*. PhD thesis, Linköping University, Linköping, 1995. Dissertation no. 381.
 - [18] Spencer C. Sorenson, Elbert Hendrick, Sigurjon Magnusson, and Allan Bertelsen. Compact and accurate turbocharger modelling for engine control. In *Electronic Engine Controls 2005 (SP-1975)*, number SAE Technical Paper 2005-01-1942, 2005.
 - [19] Fredrik Swartling. Gas flow observer for diesel engines with EGR. Master's thesis, Linköpings Universitet, SE-581 83 Linköping, 2005.
 - [20] Johan Wahlström. Control of EGR and VGT for emission control and pumping work minimization in diesel engines. Technical report, 2006. LiU-TEK-LIC-2006:52, Thesis No. 1271.

Appendix A

Notation

Table A.1: Regressors used in this thesis.

Name	Regressor(s)
p_{em}	Exhaust manifold pressure
n_{trb}	Turbine speed
p_{im}	Intake manifold pressure
T_{im}	Intake manifold temperature
n_{eng}	Engine speed
δ	Injected fuel
U_{EGR}	EGR control signal
U_{VGT}	VGT control signal
R_{trb1}	ω_t
R_{trb2}	ω_t^2
R_{trb3}	$1/\omega_t$
R_{trb4}	$1/\omega_t^2$
R_{pres1}	Π_c^{1-1/γ_a}
R_{comb1}	Ψ_c
R_{comb2}	Φ_c
R_{comb3}	W_c

Appendix B

Figures

B.1 Measured Signals

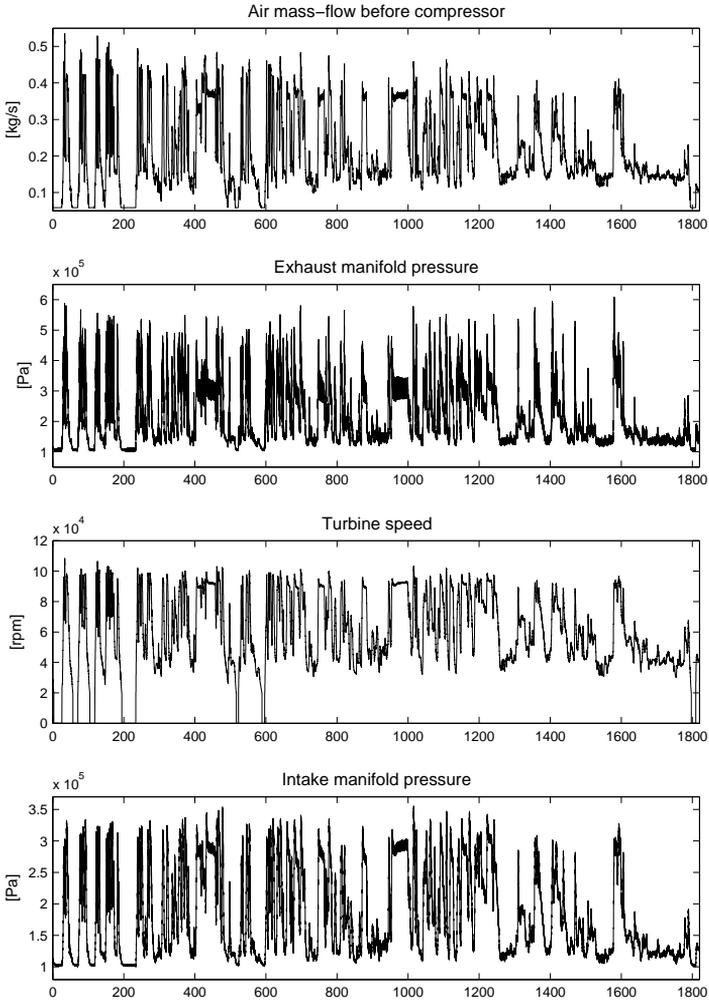


Figure B.1: The measured signals W_{air} , p_{em} , n_{trb} and p_{im} .

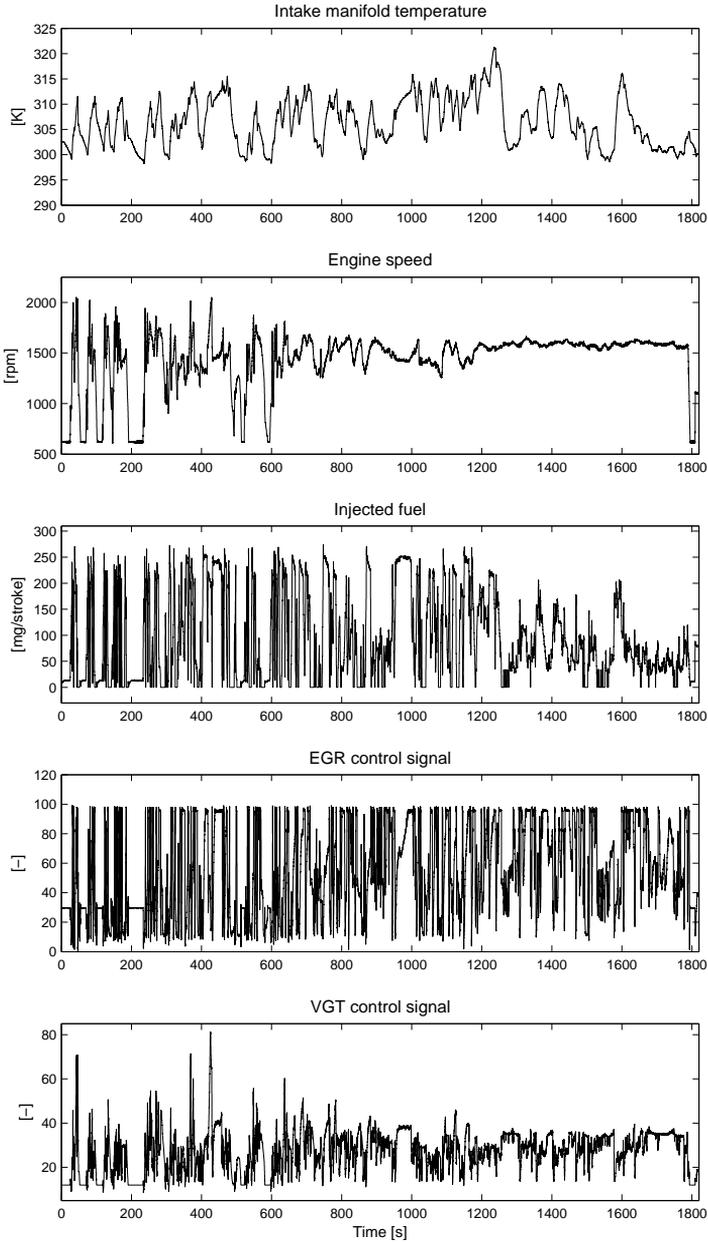


Figure B.2: The measured signals T_{im} , n_{eng} , δ , U_{EGR} and U_{VGT} .

B.2 Resampled Signals

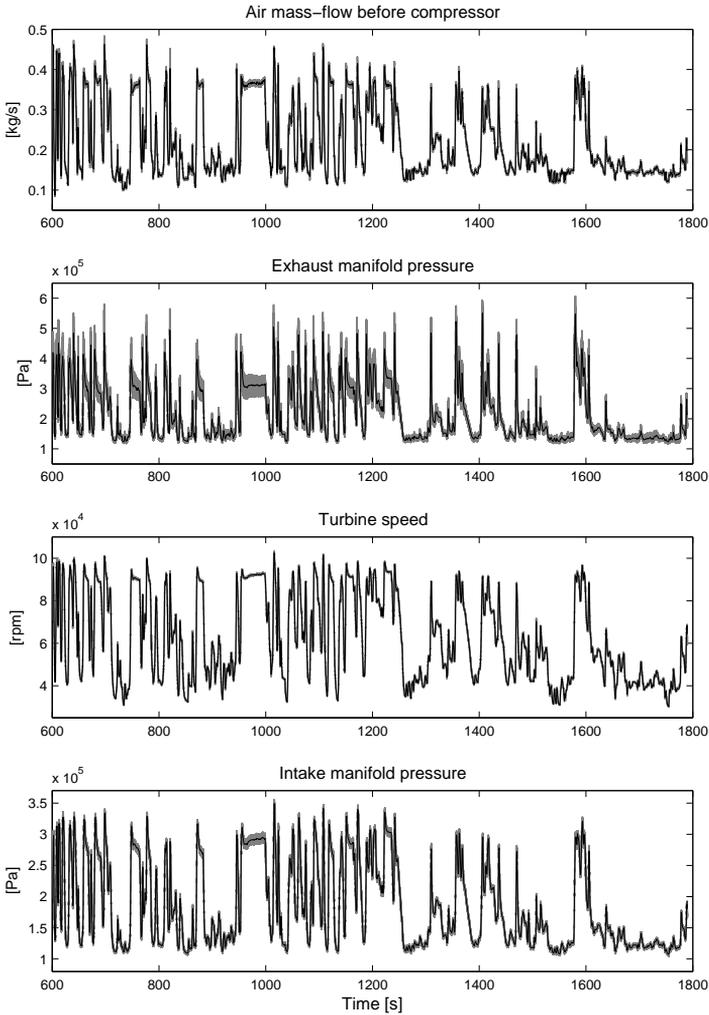


Figure B.3: Comparison between the original 100 Hz signal (grey line) and the resampled 2 Hz signal (black line) for signal W_{air} , p_{em} , n_{trb} and p_{im} .

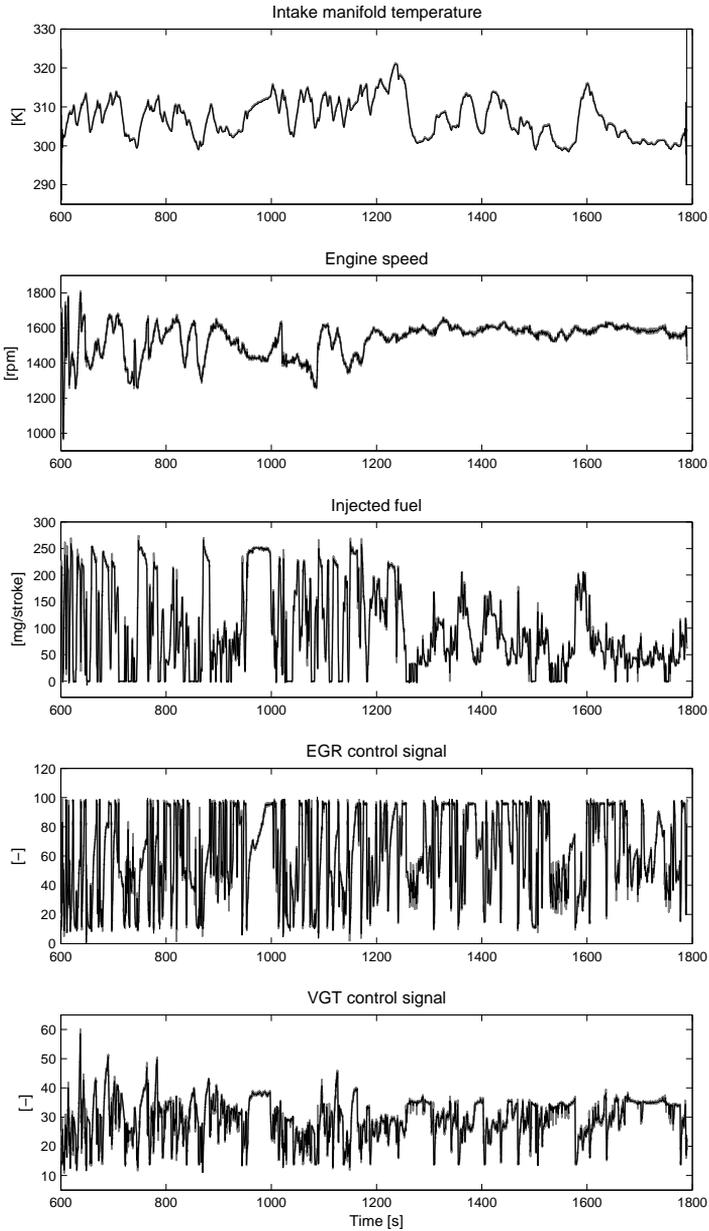


Figure B.4: Comparison between the original 100 Hz signal (grey line) and the resampled 2 Hz signal (black line) for signal T_{im} , n_{eng} , δ , U_{EGR} and U_{VGT} .

B.3 Periodogram

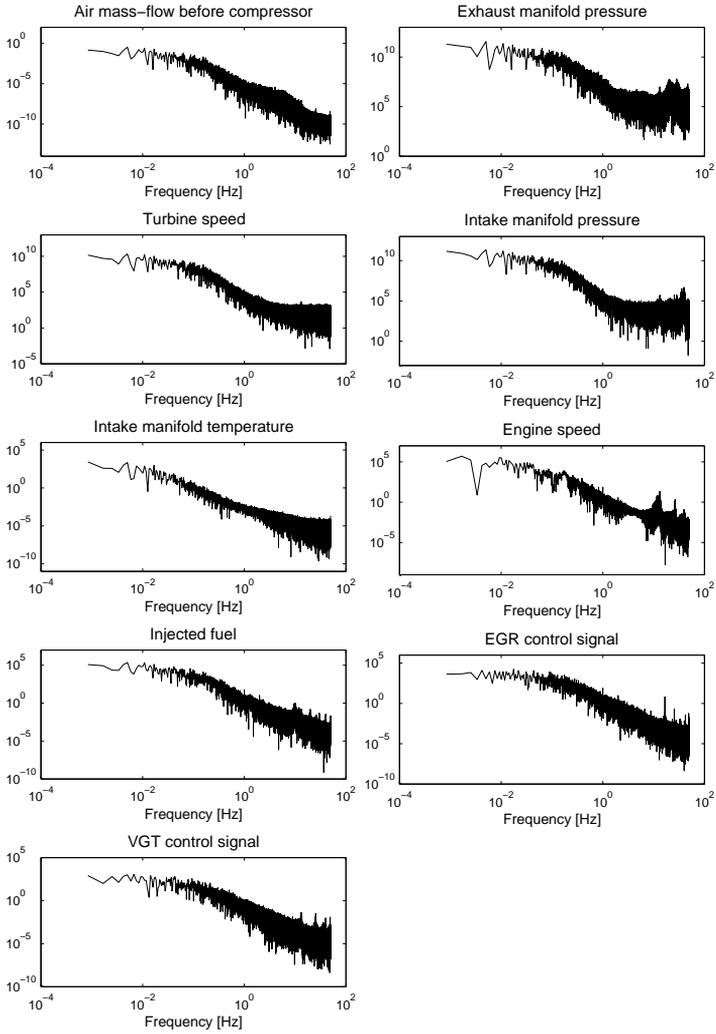


Figure B.5: Periodogram for all signals

B.4 Residuals for Linear SISO Models

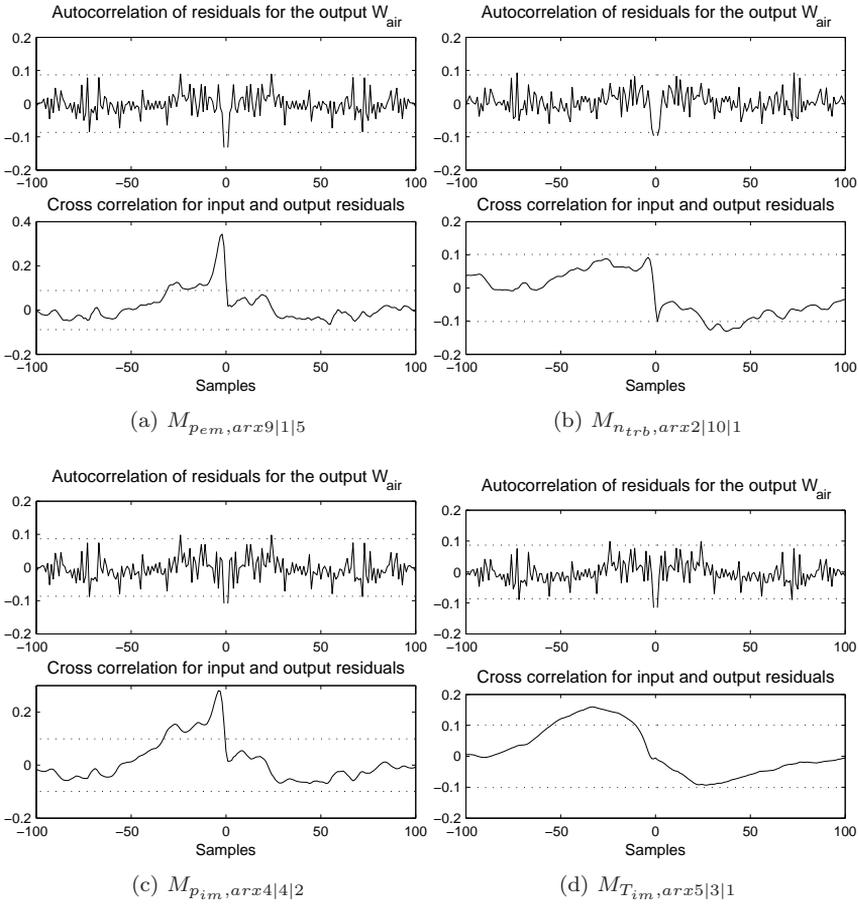


Figure B.6: Residuals for linear single input single output models.

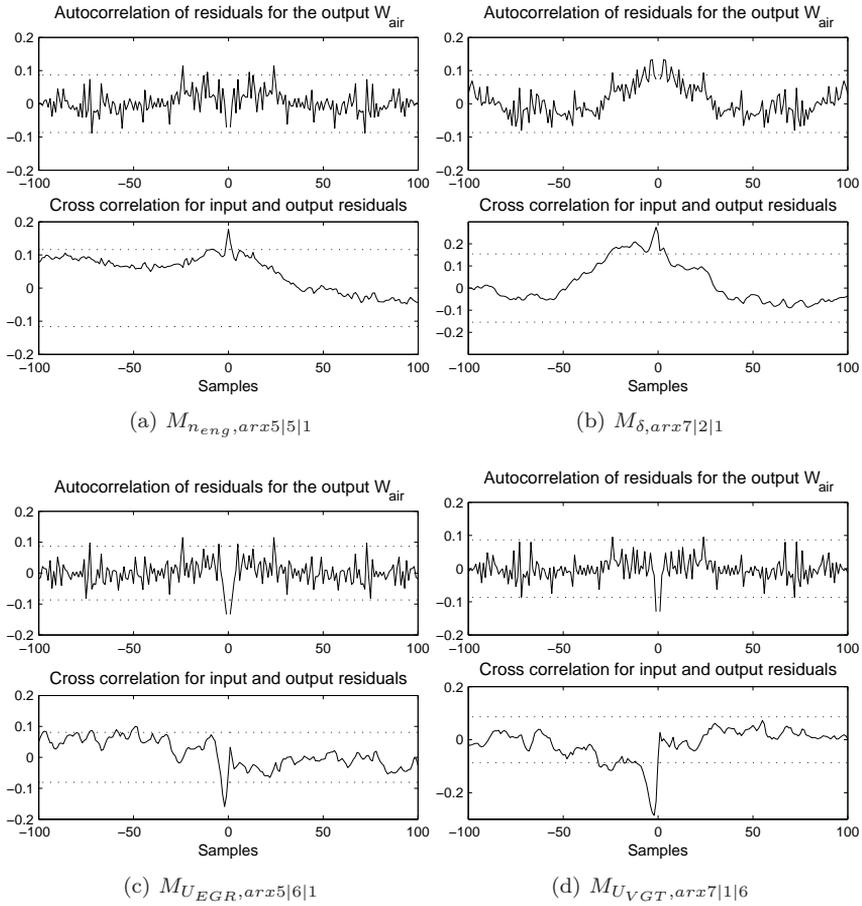


Figure B.7: Residuals for linear single input single output models.

B.5 Pole-Zero Diagram for Linear SISO Models

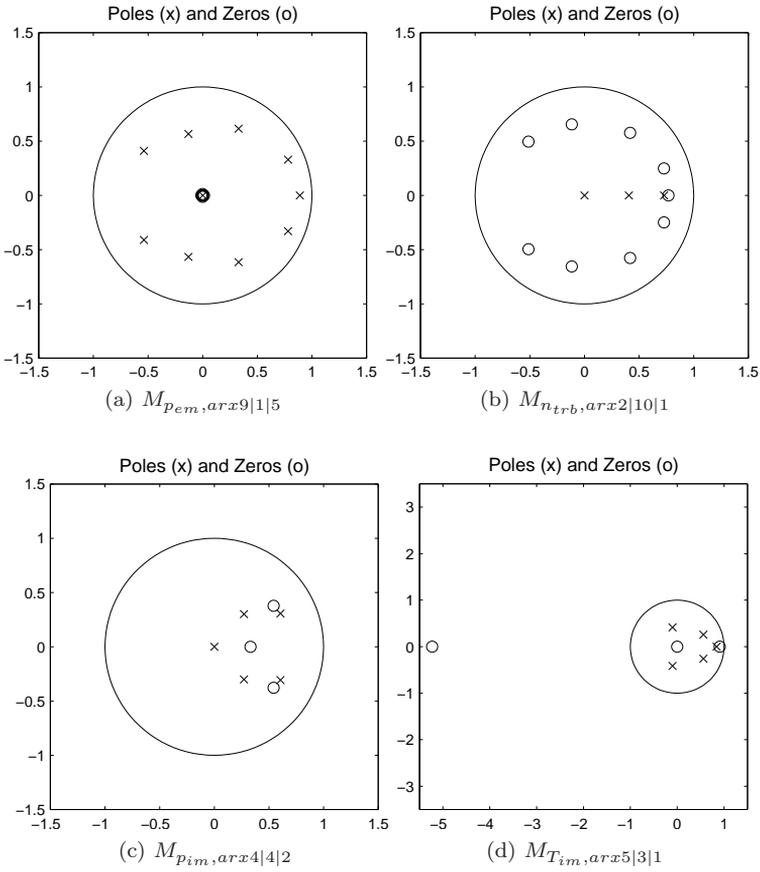


Figure B.8: Pole-zero diagram for linear single input single output models.

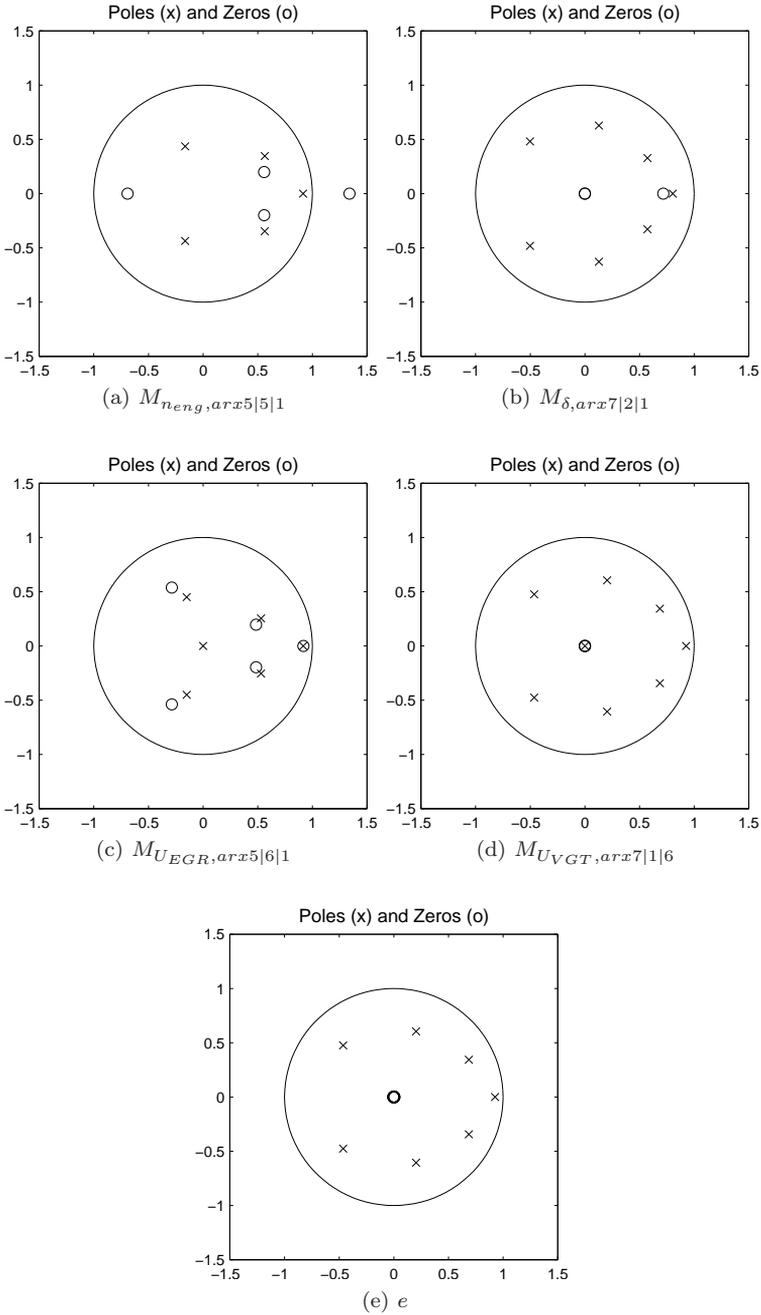


Figure B.9: Pole-zero diagram for linear single input single output models.

B.6 Residuals for Linear MISO Models

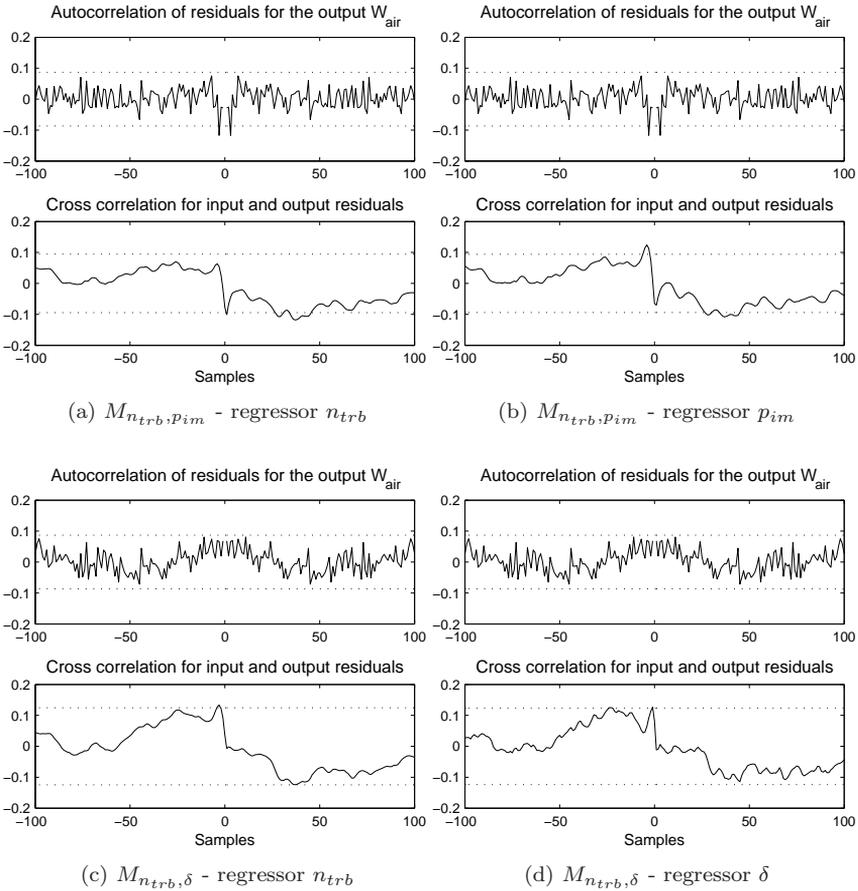


Figure B.10: Residuals for respective regressor for the MISO models in Section 5.4.

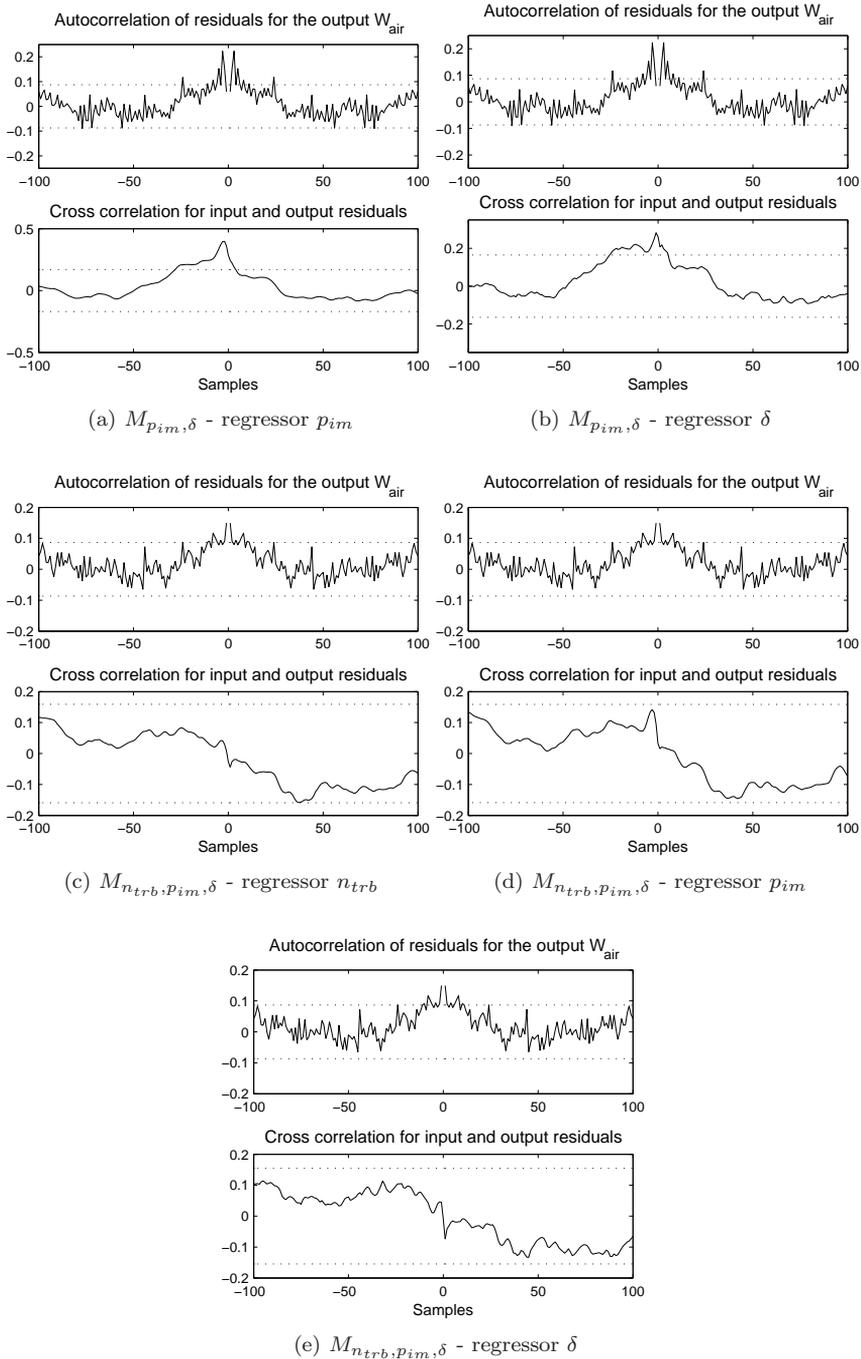


Figure B.11: Residuals for respective regressor for the MISO models in Section 5.4.

B.7 Residuals for Non-linear MISO Models

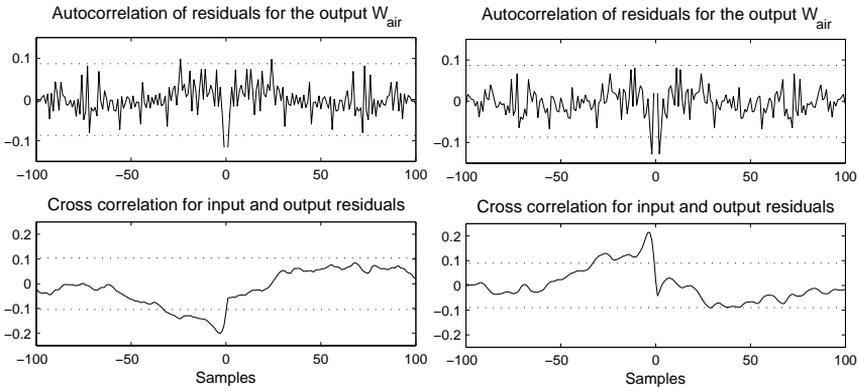
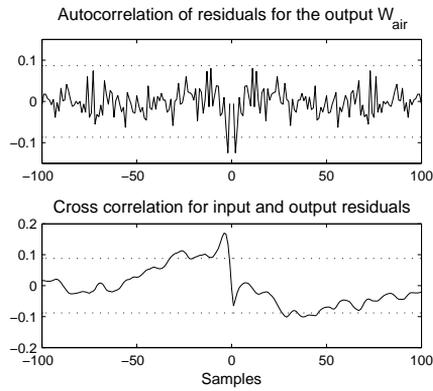
(a) $M_{R_{comb1}, arx5|1|9}$ (b) $M_{R_{comb2}, arx1|10|1}$ (c) $M_{R_{comb3}, arx1|10|1}$

Figure B.12: Residuals for the combined models in Section 6.2.3.