ISSN 0280-5316
ISRN LUTFD2/TFRT--5870--SE

Modeling of soot emission for heavy-duty diesel engines in transient operation

Mohammed Adlouni

Department of Automatic Control
Lund University
January 2011

Lund University
Department of Automatic Control
Box 118
SE-221 00 Lund Sweden

Document name
MASTER THESIS

Date of issue
January 2011

Document Number
ISRN LUTFD2/TFRT--5870--SE

Author(s)
Mohammed Adlouni

Supervisor
Christer Lundberg Scania, Södertälje
Rolf Johansson Automatic Control, Lund (Examiner)

Sponsoring organization

Title and subtitle
Modeling of soot emission for heavy-duty diesel engines in transient operation. (Modellering av transient sotflöde för tunga dieselmotorer).

Abstract
The soot emissions from diesel engines are much higher in transient operation than during steady state operation. In this thesis, a method for estimating the soot mass flow emission in transient drive cycles is presented. The proposed method uses stationary emission maps in combination with step responses. Bandpass filters are used to filter the engine torque/power to provide transient detection and quantification. The final estimation method combines the stationary emission map and the bandpass filtered signals with correction parameters determined by the step responses. Good results were obtained for the World Harmonized Transient Cycle (WHTC). Black box modeling was also conducted for the soot mass flow emission. The fitted models take two inputs: injected fuel mass per cycle and equivalence ratio. Linear models as well as nonlinear models were identified with the best results obtained for a nonlinear (Wiener) model. The best model had a fit of 95 % in accumulated soot mass emitted as a function of integrated power in the WHTC cycle. The system identification was performed using data from the first half of a WHTC cycle and validation was conducted using the whole cycle.

Keywords
Diesel engine, Soot, Soot mass flow, Transient, Stationary emission maps, Step responses, Empirical, Estimation, Black box modeling, Statistical, Nonlinear models.

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title
0280-5316

ISBN

Language
English

Number of pages
53

Recipient's notes

Security classification

Preface

The author of this thesis is at the time of writing an engineering student at the Department of Automatic Control, Lund Institute of Technology, Lund University. This Master's thesis project was performed at Scania CV AB in Södertälje, Sweden from August 2010 to January 2011. Scania is one of the leading manufacturers of heavy trucks and buses in the world. The products of Scania also include industrial and marine engines. The company is represented in more than 100 countries and employs some 34000 people. About 2400 of these work in research and development in Södertälje.The excellent work environment and helpful colleagues at Scania provided an educating experience. I would like to thank all the people who I came in contact with at Scania. Furthermore I thank the group (NMED) where I performed the thesis for the opportunity and welcoming attitude. A special thanks goes to my supervisor at Scania, Christer Lundberg, for all his valuable inputs and the helpful discussions. I also thank Professor Rolf Johansson at the Department of Automatic Control in Lund for the much appreciated feedback.

Södertälje, Sweden, January 2011.

Contents

1 Introduction **5**

2 Background **6**
2.1 Physical background . . . 6
2.1.1 Steady state emission . . . 6
2.1.2 Soot emission during transient operation . . . 7
2.2 Literature survey . . . 8

3 Data **11**
3.1 Micro Soot Sensor . . . 11
3.2 Engine data and measurement logging . . . 11

4 Empirical modeling **12**
4.1 Prerequisites . . . 12
4.2 Method . . . 13
4.2.1 Detection and quantification of transients . . . 14
4.2.2 Determining compensation . . . 16
4.2.3 Estimation formula . . . 17
4.3 Results . . . 17
4.3.1 Stationary emission maps . . . 17
4.3.2 Step response tests . . . 18
4.3.3 Parameters . . . 22
4.3.4 World Harmonized Transient Cycle . . . 24
4.3.5 Additional drive cycles . . . 27
4.3.6 Summary . . . 29
4.4 Discussion . . . 29

5 Black box modeling **31**
5.1 Data preprocessing . . . 31
5.2 Model order selection . . . 32
5.3 Choosing inputs . . . 33
5.4 Linear models . . . 35
5.4.1 ARX models . . . 35
5.4.2 ARMAX models . . . 38
5.4.3 Output Error models . . . 39
5.4.4 Coherence . . . 40
5.5 Nonlinear models . . . 41
5.5.1 Wiener models . . . 41
5.5.2 Hammerstein models . . . 44
5.5.3 Hammerstein-Wiener models . . . 45
5.6 Discussion . . . 46

6 Conclusion **48**
6.1 Future work . . . 48

7 References **49**

A Appendix **51**
A.1 Abbreviations . . . 51

1 Introduction

Heavy duty vehicles have an important role in the modern society. With increased environmental awareness and common forces striving towards minimizing pollution and emissions with global warming potential, laws concerning emission requirements are passed. These legal regulations are typically made more stringent every few years, pushing engine developers to keep up. One of the harmful emissions is soot, a carbonaceous particulate that is the result of incomplete combustion. Soot has been recognized to be environmentally harmful as well as a health hazard as it is carcinogenic. Large parts of the combustion related soot emission contribution can be assigned to diesel engines. According to [1] and references within, as much as 24 % of the worlds soot emissions come from diesel engines of which 60 % is transportation related (cars, trucks, buses, vessels etc.).

Understanding the problem is always key to making decisions about how to handle it. Models serve this purpose and many have attended to soot emission modeling over the years. It is however a difficult task and therefore simpler models are often resorted to.

The soot emission levels from a heavy duty diesel engine are quite different in transient operation as opposed to stationary operation. During engine development, emission maps are always created. These give the stationary emission level as a function of the engines operating point. A much higher soot emission level is seen in transient operation. Thus understanding the nature of the deviations and obtaining a way of estimating the soot emissions during transient operation is of great importance.

Besides stationary emission maps, performing response test on engines is common. The aim of this master thesis arose from this: Is there a way of combining the information from stationary emission maps with response tests to estimate the emissions during a transient drive cycle? Apart from this, the thesis includes black box modeling of the soot mass flow using signals available to the engine control unit (ECU) for potential online usage. The goals of this thesis can be summarized in:

1. **Literature survey**
   A short literature survey is made to get an idea of the challenges ahead and progress made by others.

2. **Development of empirical estimation method**
   An investigation of the possibility of using stationary emission maps along with step response tests to estimate the soot mass flow emission of a heavy duty diesel engine in transient operation. Proposal of the best found empirically developed method. Only engine speed and engine torque can be used as inputs to this method since it is to be used in drive cycles where these variables are the only variables known a priori.

3. **Black box modeling of soot mass flow**
   Conduction of black box modeling using inputs available to the ECU. Evaluation of different model structures and model complexities. Proposal of the best option.

2 Background

Diesel engines are appreciated for their high efficiency. From an emission standpoint, the high efficiency leads to low fuel consumption which is closely coupled with the CO_2 emissions. Unfortunately, diesel engine developers have been struggling with other emissions such as nitrogen oxides (NO_x) and soot for a long time. With increasingly stringent legal requirements on emission levels, to harvest the benefits of compression ignited diesel engines, NO_x and soot must be reduced.

2.1 Physical background

In this section a short introduction to the physical background and theory for soot formation, oxidation and net emission from a diesel engine is given, first for steady state operation and then during transient engine operation.

2.1.1 Steady state emission

The production and emission of soot is composed of many processes and stages, some of which are not completely or even poorly understood. Still most of the dominating processes are somewhat known. The first stage of soot production is the formation [6] stage. In this stage gas phase hydrocarbons produced from oxidation/pyrolysis of fuel typically would condensate, forming nuclei, the first appearance of soot particles. These soot particles that first are formed have a diameter of less than 2 nm but through surface growth, aggregation and coagulation the particles change size, mass and number. Surface growth occurs when the initial small particles incorporate gaseous compounds (various hydrocarbons such as polycyclic aromatics and polyacetylene analogues) to their particulate body, thus increasing in mass and diameter. This form of growth obviously does not change the number of particles in the combustion chamber. The formed particles can also collide and thereby coagulate. This would decrease the number of particles but leave the particulate mass unchanged. Furthermore the now fewer and larger particles can form chains and clusters through aggregation, increasing even more in size. One important aspect is the fact that soot particles have a lower hydrogen to carbon ratio than the precursors, leading to the common belief that this is achieved through dehydrogenation [6].

As soon as soot is formed it can also be oxidized [6] forming among other products carbon monoxide or carbon dioxide depending on the completeness of the reaction. The emitted soot from a diesel engine is in fact the net result of the previously described formation of soot and oxidation of the same. The oxidation rate of soot increases with temperature and oxygen presence. Since the oxidation and pyrolysis of fuel also increases with temperature (and the products from these reactions are the precursors of soot), soot is not formed below a certain temperature. The temperature dependency of soot formation is slightly different than for oxidation, the two reactions have different activation energies. This results in a temperature range where the net soot production is significant with a maximum at some temperature depending on the engine.

Given the above description, what else affects the soot emission? Obviously the amount of injected fuel increases the amount of precursors and thereby the amount of soot formed. This is assuming that no measure is taken to counteract this effect (such as decreasing the equivalence ratio). The equivalence ratio is crucial to the amount of emitted soot since it affects the oxidation rate of the formed soot. It has been shown that increasing the rail[1] pressure shifts the size distribution of soot particles towards smaller particles [18]. The in-cylinder pressure also affects the formation of soot since the condensation of precursors, the first stage of soot production, is affected by pressure. It has been shown that the pressure also affects the oxidation rate of soot. The geometry of the combustion chamber and especially the piston bowl has a great impact on the soot in the way of affecting the fuel air mixture thus affecting the presence or absence of rich and lean zones. Since the balance between the processes formation and oxidation of soot is what determines the amount of emitted soot from a diesel engine, one can also conclude that the distribution of injected fuel over the crank angle degrees (CAD) is of importance. The earlier EOI comes the longer the formed soot has to oxidize (given a certain oxygen level). Hence given this brief introduction to the physical background, the most important variables affecting soot emission are identified as:

- Injected fuel mass
- Equivalence ratio (fuel/air ratio divided by stoichiometric fuel/air ratio for the combustion)
- Engine speed
- Injection pressure
- Combustion temperature
- Cylinder manifold temperature
- In-cylinder pressure
- Injection distribution CAD
- Injector and combustion chamber geometry
- Compression ratio

2.1.2 Soot emission during transient operation

During transient operation it is of course the same factors (as mentioned in 2.1.1) that determine the soot levels. Moreover, assume that the levels of soot emission in steady state are known. Then one can instead focus on the factors that cause deviations from these levels in transient operation. Two very important factors that are affected by transient operation and that in turn cause deviations from the steady state levels are:

- Turbo inertial lag
- Temperature of cylinder and exhaust manifold

[1] Fuel rail in common rail engines.

When the power of an engine is changed, the turbo needs to assume a new speed to allow the right amount of air to be supplied to the cylinders. Since the turbo essentially is a flywheel with a moment of inertia it cannot assume the new speed momentarily. It takes some time for the turbo to accelerate or decelerate and this lag is what causes air deficiency[2] or excess. The current cylinder manifold temperature is the result of how the engine has been running during the recent time. Since the cylinder manifold is thermally connected to the combustion gases (through thermal inertia), the temperature of which affects the soot production, one realizes its importance during transient operation.

2.2 Literature survey

Developing reliable soot emission models has proved to be a challenging task. The aim of this *brief* survey is to present possible approaches to soot modeling and to outline a few common ideas in the literature. Some advantages and disadvantages of the approaches are given as well. If the reader wishes to get a more extensive introduction to general modeling principles, this is given in [4]. A comprehensive overview that is more specific to emission modeling was found in [15] and references within.

The difficulties that have challenged engineers and researchers for many years are due to the complexity of the problem. Soot production is constituted of many processes and trying to model them all and merge the results to a reliable model has proved less than trivial. Furthermore the purpose of the model is crucial (offline prediction, realtime control, realtime fault detection etc.).

The vast majority of models found in the literature are *quantitative*[3]. This is not surprising, the problem being an engineering problem and emission requirements being formulated that way. Also a distinction is present regarding the main modeling approach.

- *Deterministic models*
  Models that are based on first principles such as fundamental physical and chemical laws.
- *Empirical/stochastic models*
  Models developed based on experimental data.

Deterministic models are developed using fundamental principles from physics and chemistry. These theoretical models can describe the highly nonlinear behavior of an engine well but at a cost. The cost is computational power since the equations, often partial differential equations, describing the complex interaction between combustion zones and chemical compositions in the combustion process need to be solved numerically. Software doing just that is available but the results are at present not useful for realtime estimation, more so for offline optimization simulations with limited accuracy [14].

Empirical/stochastic models are developed from experimental data, bypassing the difficulty of first-principle modeling. This approach is a very common due to its less computationally demanding feature. Here, some distinctions are made as well:

[2]During power increase and compared to steady state turbo speed.
[3]Giving results in real numbers, not in a categorical manner.

- Pre-structured models (grey box)
- Models without pre-structures (black box)

Grey box models assume a specific model structure. For a certain engine, the included parameters are fitted using experimental data. In the publications of grey box modeling of soot emissions, two main areas are found; *phenomenological* models and *macro parameter dependent* models.

Phenomenological models use sub-models developed to describe the different processes (phenomena) during combustion. The sub-models can be empirically developed from observation or developed using basic physical and chemical relations (which could be simplified). Then a compound model describing the emission levels can be obtained by merging the sub-models. Advantages with phenomenological models is that they are globally quite good but lack in accuracy, especially if the accuracy of the included parameters is low. Examples of sub-models could be spray models, lift-off length model, mixture model, evaporation model, heat release model, ignition delay model etc. Phenomenological models (or sub-models) were developed in [8, 9, 13, 16, 17].

To give an example of how a phenomenological model performs, the content of one of the above cited articles is explained further. In [9], a phenomenological model is put forward using conceptual sub-models and typical Arrhenius equations for the soot formation and oxidation. Models were fitted and validated on two engines; a single-cylinder engine and a six-cylinder engine. The engines had different cylinder diameter, injection system, bore/stroke, number of nozzles, spray angle, compression ratio etc. For each engine, the parameters of the model were fitted for one operating point and used for several others with differing speed, load, start of injection, equivalence ratio, intake pressure, intake temperature and injection duration. A validation of the model at six operating points was performed for the single cylinder engine. The model succeeded to predict the soot values at five out of six validation points. For the six-cylinder engine, the number of validation points were 12 and at four of them the model failed to predict the soot emission. The model has a crank angle resolution, meaning that the soot formation rate and oxidation rate are predicted as a function of the crank angle degrees. It is illustrated by this article that phenomenological models have potential, they can predict the soot even when several operating conditions are changed but the accuracy can not be guaranteed.

Another category of models is that of macro parameter dependent models. These models are computationally low demanding and use macro parameters[4] such as engine speed, exhaust gas recirculation (EGR) rate, air/fuel ratio and fuel injection pressure. Having a model that uses macro parameters is desirable since they are often readily available. Macro parameter dependent models were developed in [7, 10, 11, 12].

The estimation method in the heuristic macro parameter dependent model proposed in [10] uses the macro parameters engine speed, EGR rate (external and internal), global air/fuel ratio, fuel rate etc. In addition, the method needs cycle process parameters as a function of crank angle degrees which are obtained using the engine simulation software GT-POWER. These are e.g. in-cylinder pressure,

[4]In this paragraph there is a slight abuse of notation: 'parameter' is used instead of 'signal' or 'variable'. This is due to the notation found in the corresponding literature.

temperature and burned fuel fraction. The heuristic model introduces crank angle steps (discretizes the process) and defines virtual space zones for which the heuristic sub-models (soot formation, oxidation etc.) are applied during each crank angle step. The virtual space zones are burned, unburned and burning. The model was calibrated at one operating point and validated while running with conditions varying the EGR rate and equivalence ratio.

Black box models are models that do not assume any predefined structure. This type of modeling is the fastest and does not require much knowledge about the system to be modeled. In the case of engine emission modeling usually cycle averaged data is used (time resolved).

Disclaimer
The articles cited in this section are merely a small selection and do not necessarily represent the best achievements in the field.

3 Data

The data collected for usage in this project were obtained for engines running in test beds. All the engines were warm when started and they ran according to predefined cycles in the form of engine speed and torque set points. The engines change mode depending on predetermined conditions. For purposes of this project, the engines were only allowed to operate in two modes, one normal mode and one transient mode.

3.1 Micro Soot Sensor

The soot concentration in the exhaust gases of the diesel engines were measured using the AVL 483 Micro Soot Sensor (MSS) [2, 3]. The MSS is a photo acoustic soot sensor that uses a modulated light source. The light is absorbed by the black soot particles resulting in thermal expansion. The rapid expansion generates a pressure wave which can be detected acoustically. The soot concentration is proportional to the intensity of the sound. Some relevant specifications of the MSS are given below.

Resolution	1 $\mu g/m^3$
Data rate	$<$ 5 Hz (digital)
Measuring range	5 $\mu g/m^3$ - 50 mg/m^3

3.2 Engine data and measurement logging

The logging of engine data and external measurements whether it be from the engine control unit, test cell system or from a third party measurement equipment is collected and stored by software provided by AVL. The sampling frequency or the rate at which the data is stored is 10 Hz. The channels connected to this software are vast and many, normally some restriction as to what data to include is made. The channels of importance for the work in this project are

1. Engine speed (rpm)
2. Engine torque (Nm)
3. Engine power (kW)
4. Exhaust gas flow (kg/h)
5. Soot mass concentration in exhaust gases (mg/m^3)
6. Injected fuel mass per cycle (mg)
7. Lambda (from which the equivalence ratio is computed)
8. Engine mode

4 Empirical modeling

In this section an empirical method is presented for estimating the soot mass flow emission for a heavy duty diesel engine in a transient drive cycle using a stationary emission map and step response tests. The proposed method is approximative, heuristically developed and to be used as a tool during engine development. In its simplicity the method is powerful and can give qualified indications of the emission levels in a transient drive cycle.

4.1 Prerequisites

There are some prerequisites to the method, i.e. before the method can be applied and the soot mass flow of a drive cycle be estimated, two different data sets need to be collected. As initially explained, the method uses stationary emission maps and step response tests.

A necessary part of engine development is the construction of engine emission maps. These maps are built by choosing a set of operating points (engine speed ω and engine torque τ) distributed over the operating range of an engine in a representative way. Then by running the engine at each operating point long enough for all relevant engine variables to assume stationarity (and thereby also the emissions), a measurement of the emission can be taken. By interpolation[5] a map is defined for all engine speeds ω and engine produced torques τ in the operating range of a given engine. Stationary emission maps are denoted:

$$M(\omega, \tau) \tag{1}$$

Stationary maps are of great importance as they portray the stationary behavior of an engine in a comprehensive way. Without any modeling the stationary emissions of an engine can be predicted using the constructed map. The key word here is *stationary*. If it is possible to assume quasi static behavior of the emission, i.e. that the stationary map is accurate during transient operation as well, then the map is enough for all purposes. This is seldom the case and for that reason some way of measuring an engine's transient behavior is also needed.

Soot emissions from a heavy duty diesel engine can not be treated in a quasi static way. The accumulated emissions in a transient drive cycle (30 min) are in the order of 2-10 times the quasi static emissions. Therefore to complement the stationary emission map, step responses are used in the empirical estimation method. The step responses are designed in such a way that steps are performed at six different engine speeds $\omega = [600, 800, 1000, 1200, 1500, 1900]$ rpm. For each step response test, i.e. each engine speed, three steps are made. The steps are made from no demand torque (motoring) to demanding maximum torque that the engine is calibrated to give. The duration of each step is 10 seconds. Furthermore the motoring time before the three steps is different: 60 seconds, 20 seconds and 2 seconds. The step response sequence for a given engine speed is shown in Figure 1.

[5]Not really interpolation, more a fit where $\mathcal{C}^1$-continuity is enforced.

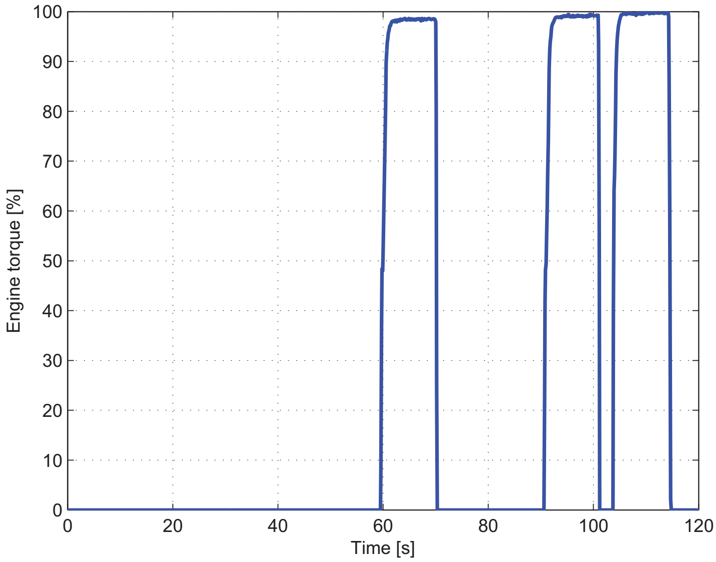


Figure 1: A typical step response test. The figure shows measured torque in percent of maximum torque that the engine is designed to produce.

4.2 Method

The idea of the method is to use the emission map $M(\omega, \tau)$ when the engine operation is stationary and based on the step response tests apply a correction or compensation whenever the operation is transient. To realize this simple idea, a number of issues need to be resolved. The first matters that come to mind are:

1. *Transient detection*
   A method for detecting transients is needed. This detection method must also be able to recognize when the operation goes from being transient to being stationary. What signal should be used for transient detection?

2. *Quantification of transient operation*
   Naturally, not all transients are the same. Both amplitude and shape of a power/torque increase can vary. Given this, a way of quantifying a transient is needed.

3. *Evaluation of step responses*
   An important question is how to extract the information from the step responses regarding the transient emission behavior. How should the compensations determined at each engine speed of the step response tests be merged for global usage at all engine speeds?

4. *Estimation formula*
   Perhaps the most important issue is how to choose the estimation formula, that is how should the stationary emission map be combined with the corrections determined from the step responses?

4.2.1 Detection and quantification of transients

A choice has to be made regarding what signal to use for detection and quantification of transients. It must be possible to derive from this signal that the engine is running in a transient manner and thus the emission level is different from what is given by the emission map. Furthermore the signal of choice must be one of the following signals:

1. Engine speed[6] ω
2. Engine produced torque τ
3. A function $f(\omega, \tau)$

The reason for this restriction in signals allowed for usage is the fact that for an arbitrary drive cycle, only the engine speed and engine torque is known a priori (in fact they define the cycle).

The third option is an interesting one since the power of an engine can be derived from engine speed and torque. If the engine speed ω is given in rad/s and the engine torque is given in Nm, the power is given by the expression:

$$P = f(\omega, \tau) = \omega\tau \tag{2}$$

From here on, the engine power P will be used in the presentation of the empirical estimation method. Keep in mind that the torque τ could just as well be used.

The next step is deciding how to use P for detection of a transient. Of course some measure of the change in power is useful. After some searching for a detection method it was decided that the best way to incorporate both detection and quantification is to use bandpass filters. They have the property of giving a zero signal as long as their input is constant and a nonzero signal when their input changes. When for instance the input to a bandpass filter is a step, the output will be a peak that settles to zero after some time, depending on the filter design. Since the filters are linear, it means that if the input step is doubled, so will the output of the filter. Therefore the filters have a transient quantification ability. For reasons explained later two filters are used. To illustrate the effect of the filters, the step responses of the filters are shown in Figure 2.

As can be seen in Figure 2, the filters[7] are of two different speeds, i.e. they operate at different frequencies. The slow filter $H_1(z)$ (z-transform of the filter) has a 10 second duration of its step response output. The fast filter $H_2(z)$ instead produces a nonzero signal for two seconds following a step. Let P_{f_1} and P_{f_2} be the filtered power signal using the designed filters, then the filtered variables are calculated according to

$$P_{f_1}(z) = H_1(z)P(z) \tag{3}$$

$$P_{f_2}(z) = H_2(z)P(z) \tag{4}$$

where

$$H_1(z) = \frac{0.0429z^2 - 0.0429}{z^2 - 1.901z + 0.9042} \tag{5}$$

[6]Theoretically the engine speed can be used but in this case the engine speed was kept constant during the step responses, hence the engine speed is not useful for transient detection.

[7]Designed using the MATLAB command `butter`.

$$H_2(z) = \frac{0.1486z^2 - 0.1486}{z^2 - 1.652z + 0.7028} \tag{6}$$

The chosen filters are second order Butterworth filters with the Bode diagram (magnitude) given in Figure 3.

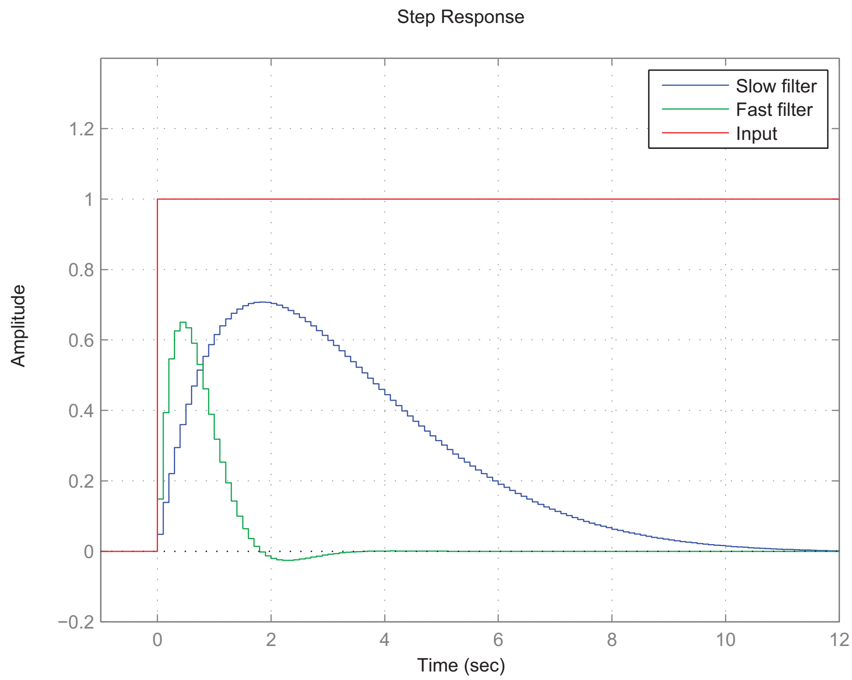


Figure 2: Step response of the filters.

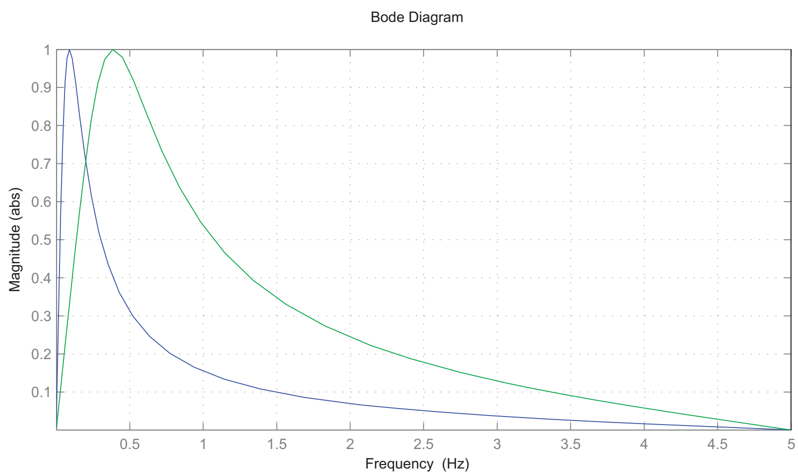


Figure 3: Bode diagram of the filters. Linear axes.

Before the filtered variables P_{f_1} and P_{f_2} can be used, they need to be modified. One can easily realize that the transient soot mass flow is not linear. With this it is meant that if the power is increased from motoring to full power, it is not the opposite of going from full power to motoring, emission wise. The filters given above will however give signals with equal amplitude but different signs in both cases. Hence some measure to distinguish or characterize a transient is also needed. The proposed way of doing so is given below. Let the modified filtered variables be denoted $P_{f_i}^+$, $i = \{1, 2\}$. Then

$$P_{f_i}^+ = \begin{cases} P_{f_i} & \text{if } P \geq 0 \text{ and } P_{f_i} \geq 0 \\ 0 & \text{else} \end{cases} \tag{7}$$

In words this translates to setting the modified filtered variable to zero when motoring or when the filtered signal is negative.

4.2.2 Determining compensation

The next step is to determine the structure/strategy of compensation when transient operation is detected using the above described method. The so far developed tools at hand are:

1. $M(\omega, \tau)$
2. $P_{f_1}^+$ and $P_{f_2}^+$

How should the the stationary contribution $M(\omega, \tau)$ be combined with the modified filtered signals $P_{f_1}^+$ and $P_{f_2}^+$? Many different structures were tested and the best one found was to simply assume that the soot mass flow S_{mf} is the sum of the stationary contribution $M(\omega, \tau)$ and a linear combination of the filtered signals.

$$S_{mf} = M(\omega, \tau) + \theta_1 P_{f_1}^+ + \theta_2 P_{f_2}^+ \tag{8}$$

For each of the step response tests, the parameters θ_1 and θ_2 need to be estimated. Luckily the chosen structure is linear and this means that an explicit solution is available using least squares (LS) estimation [4]. Solving for the parameters θ_1 and θ_2 gives (k denotes the discrete time step):

$$\begin{aligned} S_{mf}(k) - M(\omega(k), \tau(k)) &= \theta_1 P_{f_1}^+(k) + \theta_2 P_{f_2}^+(k) \\ &= \begin{pmatrix} P_{f_1}^+(k) & P_{f_2}^+(k) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \\ &= \phi_k^T \theta \end{aligned} \tag{9}$$

By letting

$$Y_N = \begin{pmatrix} S_{mf}(1) - M(\omega(1), \tau(1)) \\ S_{mf}(2) - M(\omega(2), \tau(2)) \\ \vdots \\ S_{mf}(N) - M(\omega(N), \tau(N)) \end{pmatrix} \quad \Phi_N = \begin{pmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_N^T \end{pmatrix} \tag{10}$$

where N is the number of data points, the solution to the normal equations is

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = (\Phi_N^T \Phi_N)^{-1} (\Phi_N^T Y_N) \tag{11}$$

which is the least squares estimate of θ.

4.2.3 Estimation formula

When the parameters θ_1 and θ_2 are estimated for each step response test (remember that they were run at six engine speeds), they are valid for that engine speed but of course global parameters are needed for estimation in a drive cycle. The engine speed in a drive cycle rarely stays constant. In absence of a better guess, linear variation of the parameters as a function of engine speed is assumed to give $\Theta_1(\omega)$ and $\Theta_2(\omega)$. Thereby the estimation formula for soot mass flow is

$$S_{mf} = M(\omega, \tau) + \Theta_1(\omega)P_{f_1}^{+} + \Theta_2(\omega)P_{f_2}^{+} \tag{12}$$

4.3 Results

The developed method was evaluated for four different engines. Two of the engines have a stroke volume of 13 liters. The other two engines have a stroke volume of 9 liters. The larger engines are 6 cylinder engines whilst the smaller engines are 5 cylinder engines. Furthermore for the two stroke volumes, engines with and without exhaust gas recirculation (EGR) were used. A list of the engines and their denotation is given in Table 1.

Table 1: List of engines used for evaluation of empirical method.

Reference	**Engine**	**EGR**
A	6 cyl,13 liters	Yes
B	6 cyl,13 liters	No
C	5 cyl, 9 liters	Yes
D	5 cyl, 9 liters	No

4.3.1 Stationary emission maps

For each of the used engines, a stationary emission map for the soot mass flow was constructed. $\mathcal{C}^1$-continuity was enforced for the maps, i.e. not only must the maps be continuous but also their gradient. This essentially says that the emission maps are smooth. A typical stationary soot emission map is given in Figure 4.

The measurement taken at each operating point is the soot mass concentration S_c in the exhaust gases given by the MSS. To obtain the soot mass flow, the average exhaust gas density ρ and the exhaust gas flow q are used. At each operating point (ω, τ), the stationary soot mass flow is given by

$$M(\omega, \tau) = \frac{S_c(\omega, \tau)q(\omega, \tau)}{\rho} \tag{13}$$

where q is given in kg/h, S_c is given in mg/m^3 and ρ is given in kg/m^3.

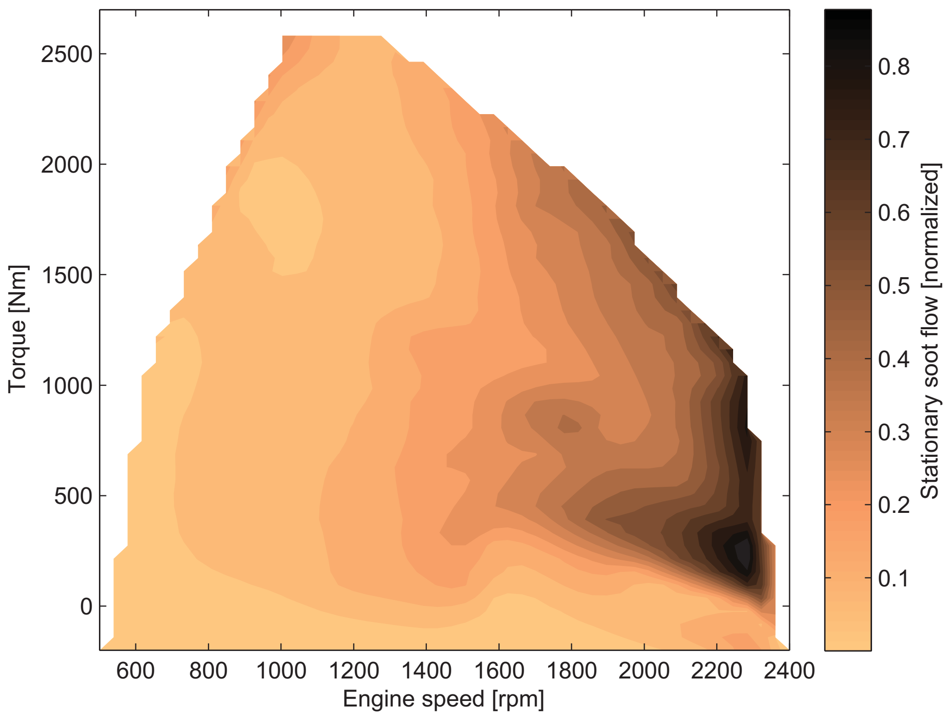


Figure 4: Example of stationary emission map of soot mass flow for a typical diesel engine.

4.3.2 Step response tests

As described previously, six step response tests were conducted for each engine. Least squares estimation of the correction parameters θ_1 and θ_2 was performed for each data set according to 4.2.2.

In Figures 5-8, the engine power, filtered power signals, measured soot mass flow, quasi static soot mass flow and estimated soot mass flow (using the proposed method) are shown for the four used engines. The results are shown for each engine running the step response test at 1900 rpm. This is not an important engine speed for overall soot emissions since it is not often used (in comparison to engine speeds in the range of 1000-1500 rpm). However 1900 rpm is shown as it well illustrates the differences between the engines.

Once again, the measurement is the soot mass concentration from the MSS and the soot mass flow S_{mf} must be computed from the concentration S_c, exhaust gas flow q and average exhaust gas density ρ. One issue that arises in the case of transient operation is that the measurement of soot mass concentration in the exhaust gases is delayed compared to the reading of the exhaust gas flow (which is calculated by measuring fuel and air flow and has no/negligible delay from the engine power reading). Furthermore the problem is complicated by the fact that this delay is not constant. It depends mainly on the engine speed[8] and the physical setup of

[8]It actually depends on the transportation time of the exhaust gases from combustion chamber

equipment.

To solve this issue, a delay needs to be specified for each step response test. This is not a cumbersome task however. Let this delay be denoted d for a given engine and speed. Then the real soot mass flow is computed as

$$S_{mf}(k) = \frac{S_c(k)q(k-d)}{\rho} \tag{14}$$

where k denotes the sample. The delays are in the order of a second.

It can be seen in the step response tests that the soot mass flow emission has a characteristic behavior. Immediately after a step, a peak arises. This peak lasts for a few seconds before the emission decreases to a level still larger than the quasi static one and then slowly decreases to the stationary level for that operating point. It is this difference in the transient behavior that prompted the use of two filters. Another noteworthy observation is that all engines but engine C consistently elicit this behavior (more or less). This observation that engine C is different shall be used later.

to measurement equipment, transportation time through the dilution tunnel and processing time of the MSS. However the varying part of all this is the exhaust gas velocity which correlates well with the engine speed.

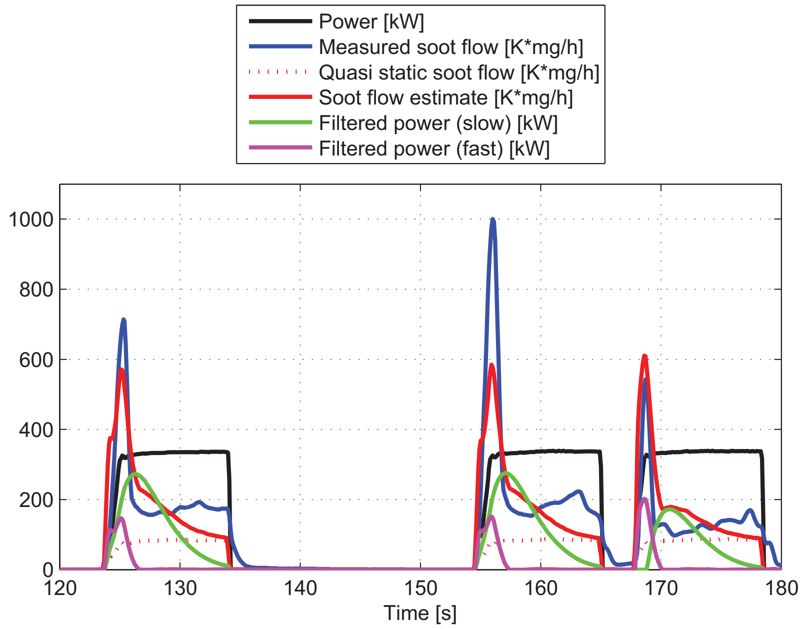


Figure 5: Step response test at 1900 rpm for engine A. K is a constant.

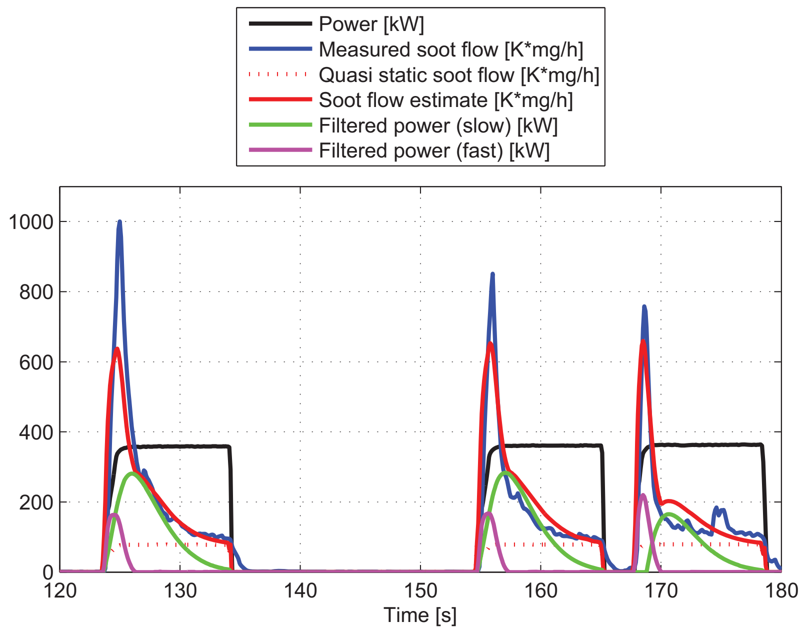


Figure 6: Step response test at 1900 rpm for engine B. K is a constant.

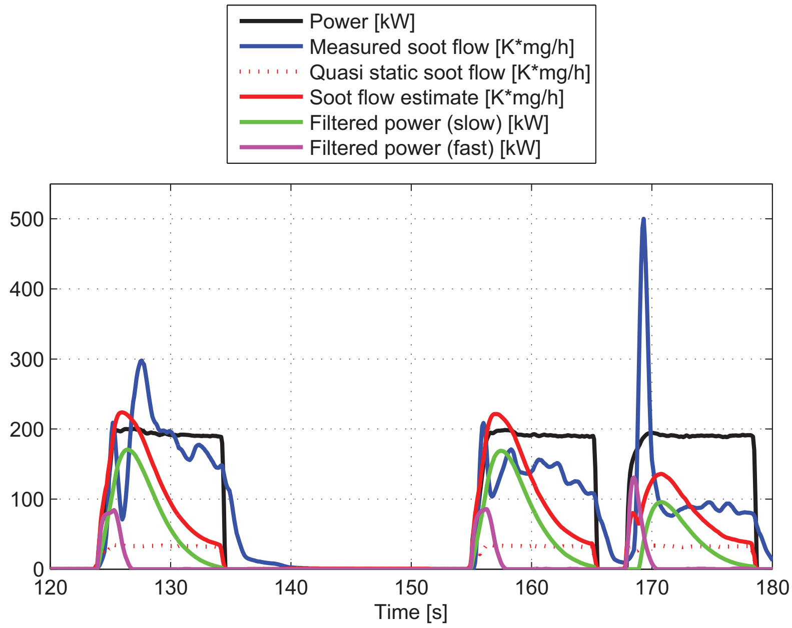


Figure 7: Step response test at 1900 rpm for engine C. K is a constant.

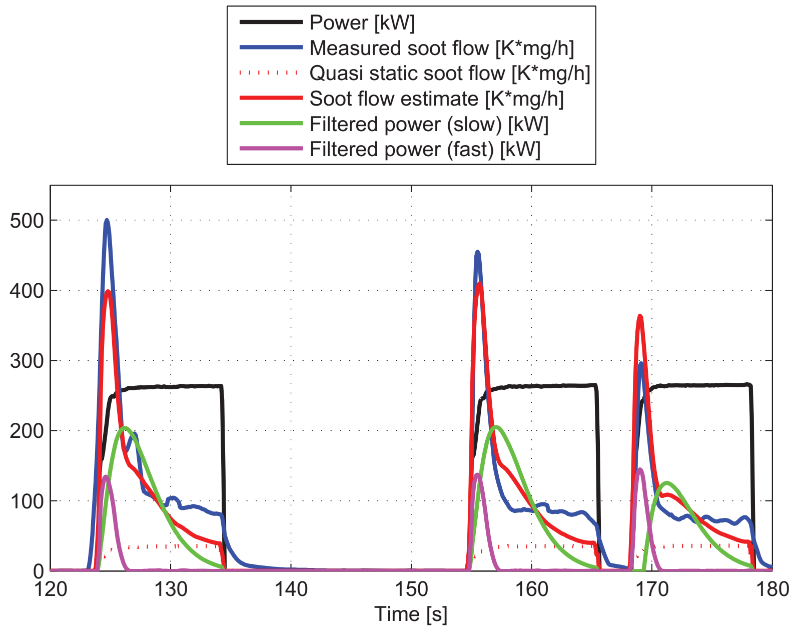


Figure 8: Step response test at 1900 rpm for engine D. K is a constant.

4.3.3 Parameters

In Figures 9-12 the parameters $\Theta_1(\omega)$ and $\Theta_2(\omega)$ are given for the engines A-D.

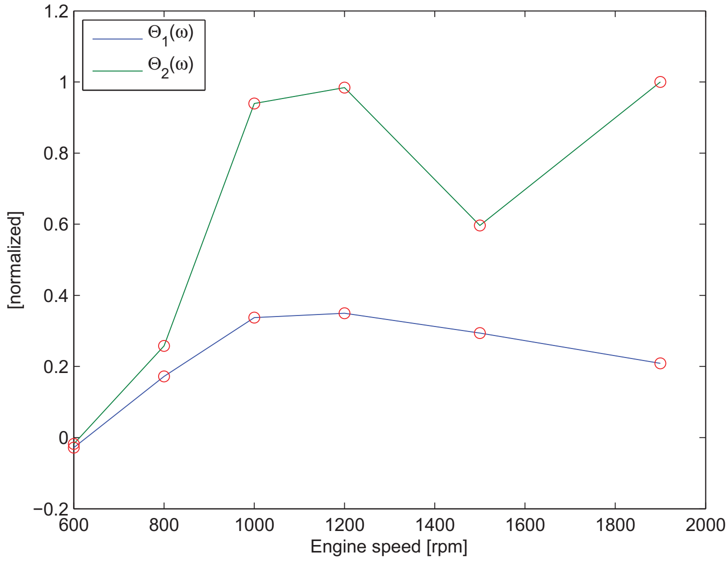


Figure 9: Parameters $\Theta_1(\omega)$ and $\Theta_2(\omega)$ for engine A. Circles indicate engine speeds where the step responses were run and the least squares estimation was done.

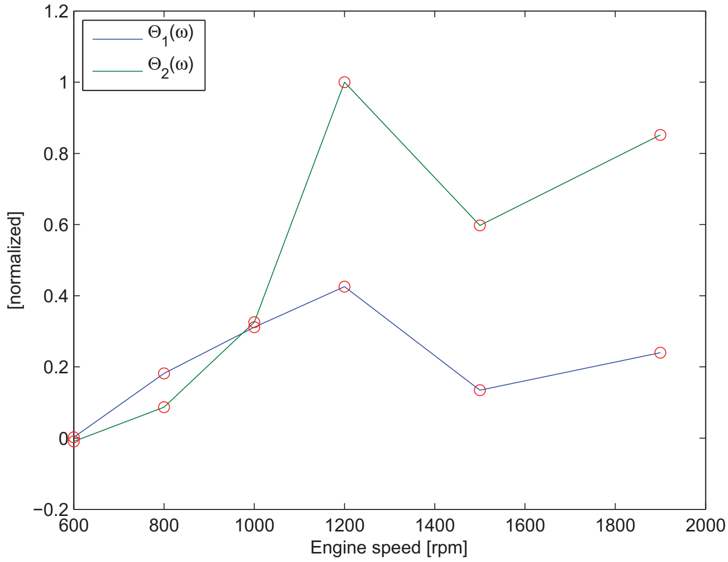


Figure 10: Parameters $\Theta_1(\omega)$ and $\Theta_2(\omega)$ for engine B. Circles indicate engine speeds where the step responses were run and the least squares estimation was done.

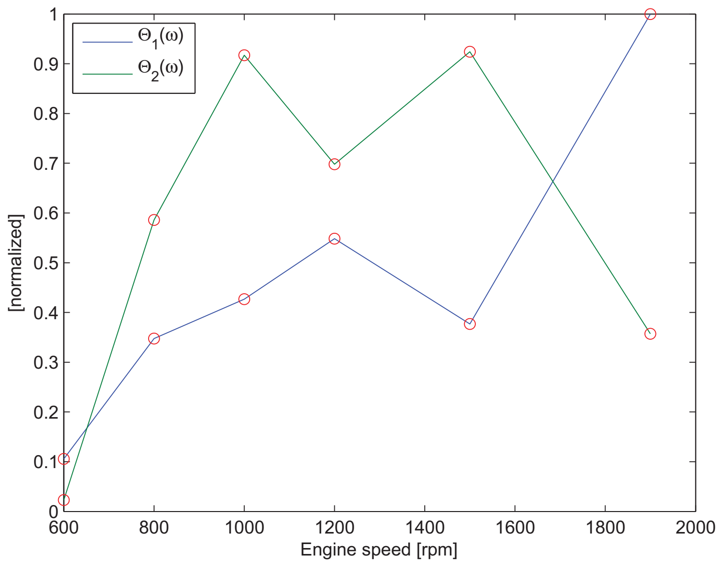


Figure 11: Parameters $\Theta_1(\omega)$ and $\Theta_2(\omega)$ for engine C. Circles indicate engine speeds where the step responses were run and the least squares estimation was done.

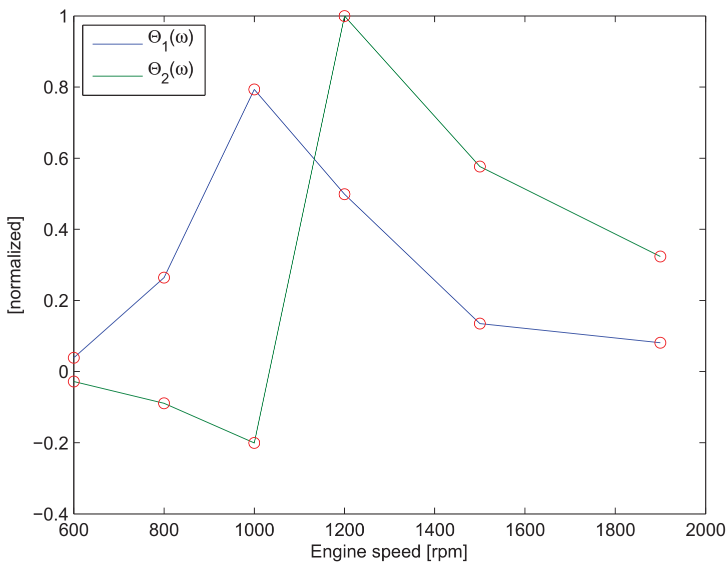


Figure 12: Parameters $\Theta_1(\omega)$ and $\Theta_2(\omega)$ for engine D. Circles indicate engine speeds where the step responses were run and the least squares estimation was done.

4.3.4 World Harmonized Transient Cycle

A commonly used transient drive cycle is the World Harmonized Drive Cycle (WHTC). This drive cycle has a duration of 30 minutes and is commonly used to evaluate the behavior of a heavy duty engine in transient operation. The drive cycle is devised is such a way as to cover several drive scenarios. It has been constructed to contain three sections corresponding to urban, rural and highway driving respectively. The drive cycle can be seen in Figure 13. It is normally scaled to fit the power level of the engine it is meant for.

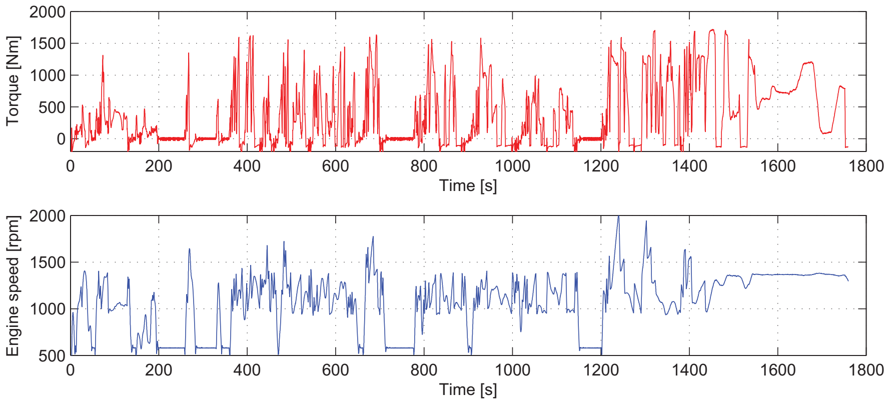


Figure 13: The World Harmonized Transient Cycle, scaled for engine D.

The developed estimation method that resulted in the formula (12) was used with the parameters given in 4.3.3 along with the respective soot emission map for each of the engines A-D to estimate the accumulated soot mass emission in the WHTC. In Figures 14-17, the results from using the empirical estimation method on the engines A-D running the WHTC is shown. Each figure contains four plots corresponding to estimation in the whole cycle, first section (urban), second section (rural) and third section (highway) respectively. Satisfactory results were obtained for engines A, B and D whilst the method applied to engine C gave unsatisfactory results. For the three engines where the method works the relative error in the complete WHTC is less than 13 %. This must be considered a good result considering that its a simulation using only engine speed, torque and power as inputs (given the parameters in 4.3.3 and emission maps in 4.3.1).

In the upper left plot in Figures 14-17, a magenta colored curve is plotted. This curve depicts the quasi static emissions using the emission map with a static compensation factor. This static compensation factor is *not* known a priori but is used to illustrate that even if the relation between dynamic and static soot emissions was known for a given engine and drive cycle, the results are not perfect since the engine behaves differently in the different sections. There is under compensation in some sections and consequently over compensation in other sections. Adversely, the proposed empirical method is consistent in this regard and the relative error is somewhat constant during the whole estimation showing that the method successfully compensates at the right time (during a power increase). In other words the shape of the estimated emission curve (red) is correct compared to the measured emission curve (blue).

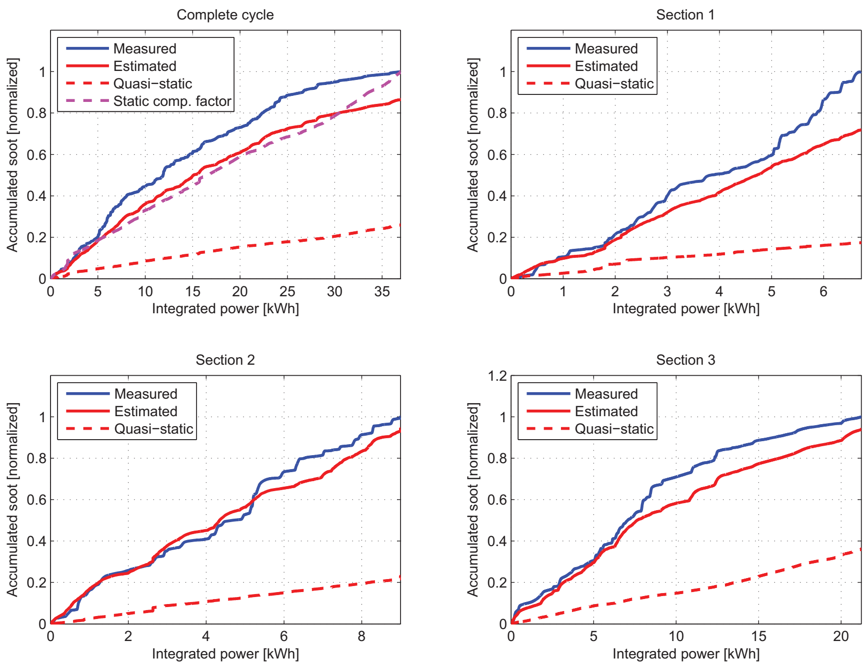


Figure 14: Evaluation of empirical method in the WHTC for engine A.

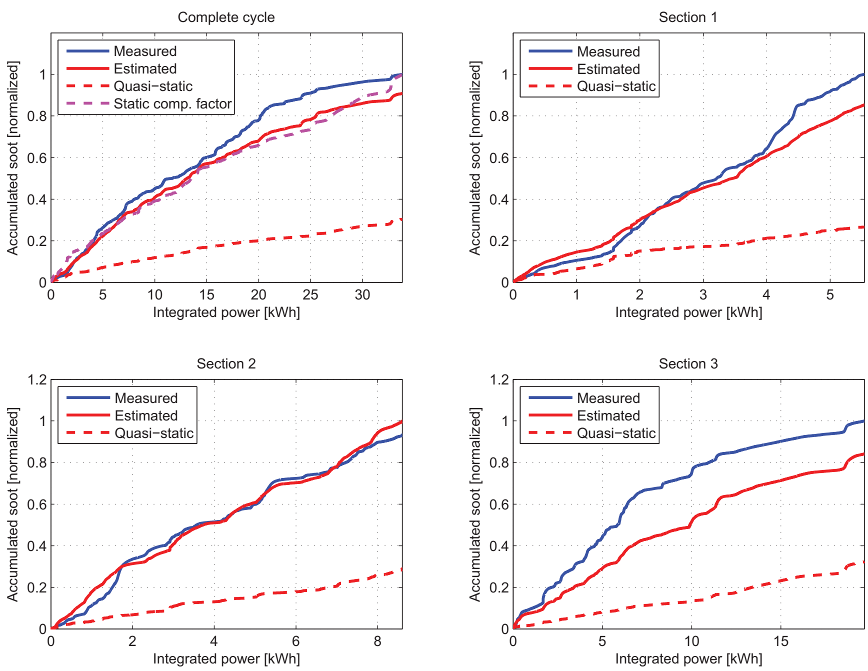


Figure 15: Evaluation of empirical method in the WHTC for engine B.

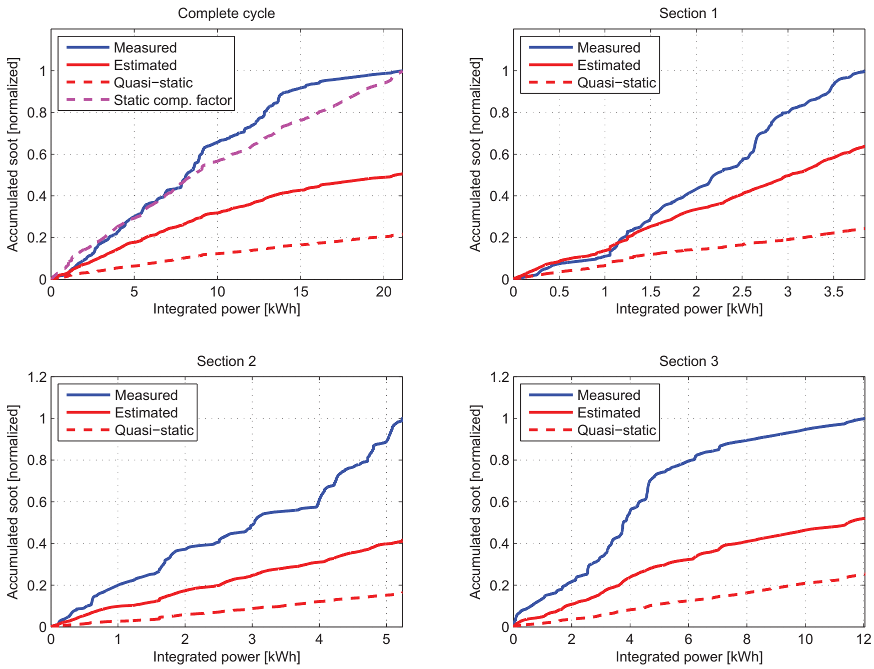


Figure 16: Evaluation of empirical method in the WHTC for engine C.

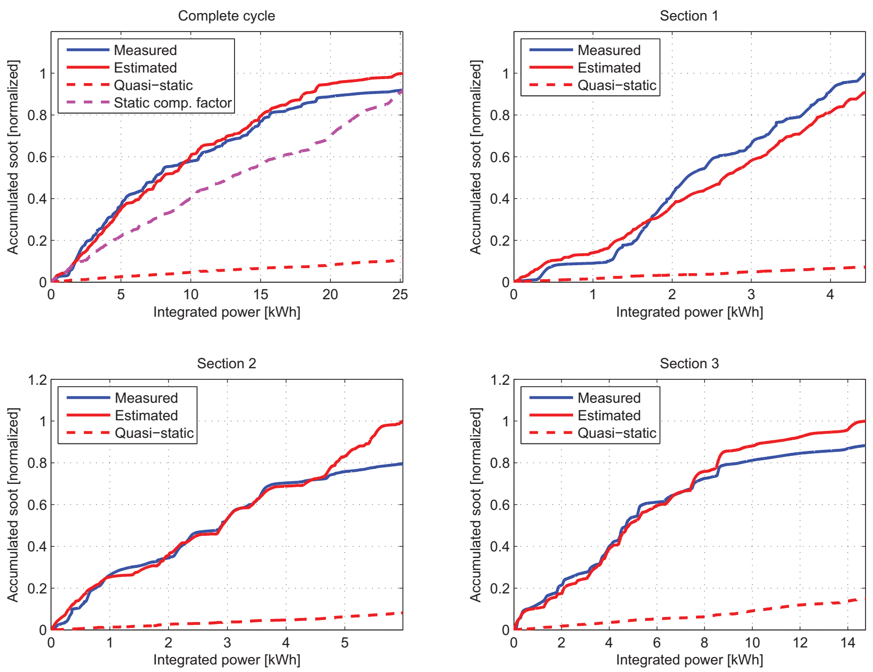


Figure 17: Evaluation of empirical method in the WHTC for engine D.

4.3.5 Additional drive cycles

To evaluate the developed method further, it was applied to a different drive cycle as well. The drive cycle is the Standardized On-Road Test (SORT) cycle which is short with many starts and stops. The cycle was looped three times to get a longer evaluation time since the duration of one iteration of the cycle is only 3 minutes. The cycle is shown in Figure 18 where it is scaled for engine C.

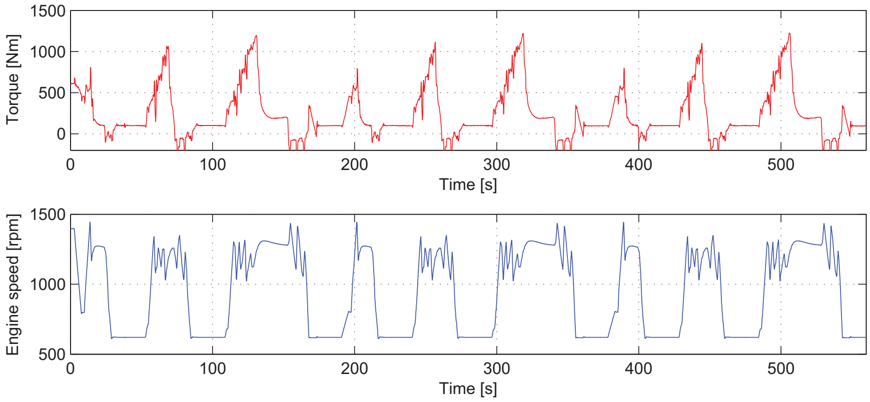


Figure 18: SORT cycle, scaled for engine C.

The evaluation of the empirical method in this drive cycle was only performed for engines B and C. Engine C is known to behave strangely from the step response tests and the evaluation in the WHTC. For that reason it is not expected that the method would work for that engine. The evaluation of the method for engine B shows over compensation of the emission due to transient operation. In the discussion section an attempt to sort out the reasons is made.

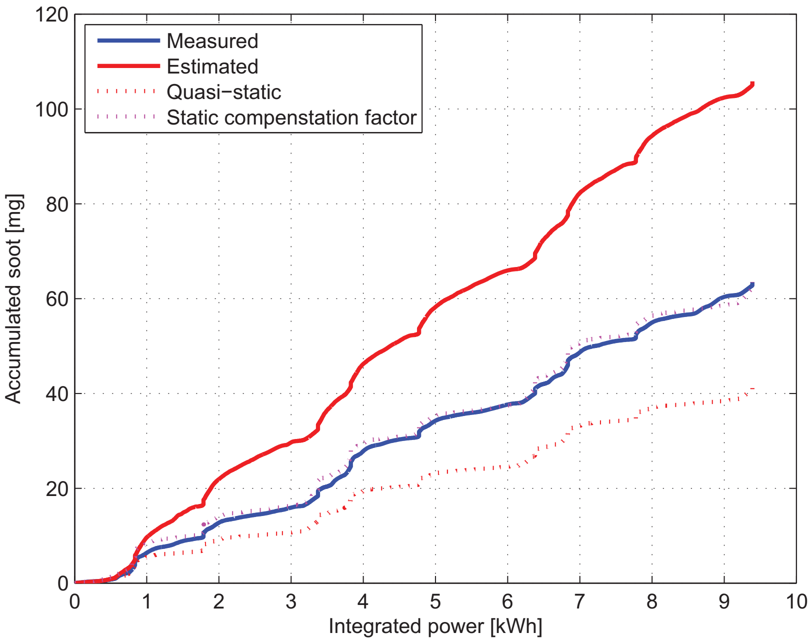


Figure 19: Evaluation of the empirical method in the SORT cycle, engine B.

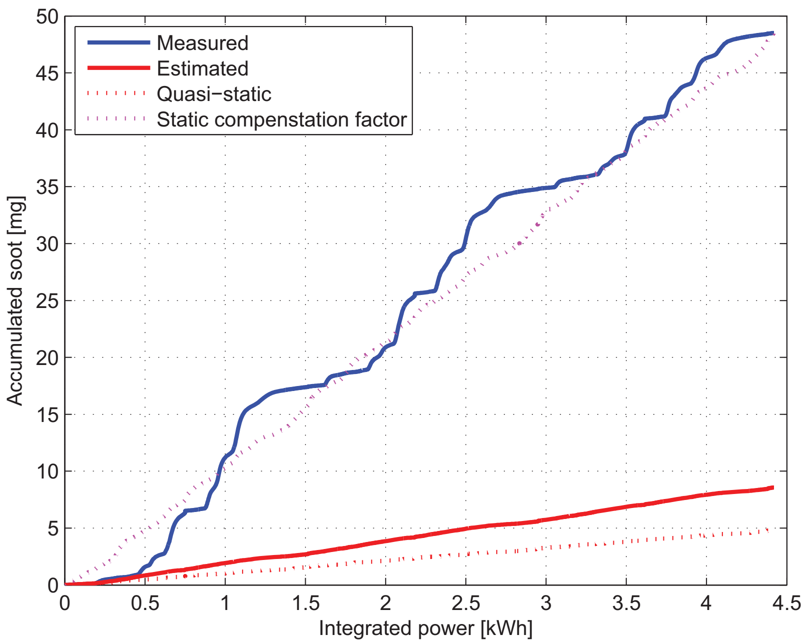


Figure 20: Evaluation of the empirical method in the SORT cycle, engine C.

4.3.6 Summary

Table 2: Relative errors at the end of drive cycles.

Engine	Cycle	Relative error	Urban[1]	Rural[1]	Highway[1]
A	WHTC	-13 %	-28 %	-6 %	-6 %
B	WHTC	-9 %	-15 %	+7 %	-16 %
C	WHTC	-49 %	-36 %	-58 %	-48 %
D	WHTC	+9 %	-9 %	+25 %	+13 %
B	SORT	+67 %	-	-	-
C	SORT	-82 %	-	-	-

4.4 Discussion

The evaluation of the empirically developed estimation method shows that the method works for three out of four engines while running the WHTC cycle. For those engines, the results are good and clear indications of the soot emission levels in the WHTC are given by the proposed method using soot emission maps and step response tests. Meanwhile the method is not perfected and it shows in the larger relative error when estimating in the SORT cycle. In this discussion section possible reasons for this and a proposed way of determining if the results for a certain engine are valid or not is given.

The parameters to the method given in 4.3.3 are the *results of the calibration* of the engine in question. The ECU controls a number of actuators using many parameters and complicated logics. Thereby together with the physical circumstances, it determines the actual emission level. This is important to know. Even negative values of the parameters are valid since two filters are used. That means that if for a certain engine speed one of the parameters gives a negative contribution together with the respective filtered variable, the other parameter with its filtered variable might very well give a positive contribution that is larger in magnitude resulting in a positive net emission estimate. Furthermore, there is nothing that restricts the ECU to controlling the actuators in such a way as to guarantee a smooth parameter curve and as seen in the results this is not the case.

In the estimation of soot mass flow emission in the WHTC, there were three different sections in the cycle corresponding to different drive scenarios or analogues of such. There seems to be a trend that the estimation method is better for rural and highway driving as opposed to urban driving. This might be part of the explanation to why the method did not give satisfying results in the SORT cycle as that cycle is very urban.

The choice of using two filterers with different speeds was made because it was found that the additional contribution of soot mass flow emission due to transient operation appear at different speeds after an increase and for different reasons. The initial peak after a power increase can be explained by the dynamics of the turbo resulting in a lag and air deficiency for a few seconds. During the following time

[1] Sections of the cycle that resemble urban, rural and highway driving.

interval the addition in soot emission is then a result of heat transfers and temperature equalization. In the midst of all this it is important to remember that the ECU is very active and its control outputs also affect the transient behavior. Examples of this is the control of EGR, start of injection (SOI), end of injection (EOI) etc. Furthermore, the choice of filter parameters is not optimal in any sense, they were heuristically found.

Another important question to discuss is the choice of inputs to the method. What justifies the use of engine speed and torque as a basis? The answer is that many functions in the ECU are mapped and controlled based on these two variables. Thereby the behavior of the engine is assumed to be a function of the two variables. This assumption is of course not always true but at least approximatively it can be used.

Engine C behaved strangely. At this point it should be stated that since the method was intended to be a tool used during engine *development*, the engines to which the method is applied will be in different stages. An engine that is still in the early stages of its development might very well behave strangely (or not in the way assumed for this method). In the step response test for one of the engines (C) it can be seen that the behavior is different compared to the engines A, B and D. It is inevitable that this can occur. However, if a rule can be advised for when to use the method and take the results as valid and when not to (in the WHTC), it would be valuable.

Rule of estimation validity: *For usage of the developed empirical estimation method, it should be assured that the step response test (e.g. at 1900 rpm) resembles*[9] *the step response tests for engine A, B, and D. If that is the case, then at least for the WHTC, the estimation results can be taken to be good (within 15 % relative error).*

A crucial assumption made in the development of this method is that the step responses are enough to excite the transient dynamics in a representative way of all transients. This is how transient compensation is determined and therefore it could be stated that perhaps these step responses (see Figure 1) are not optimal for this purpose. Also any differences in emission levels between the different steps are averaged out by the least squares method. The behavior of a given engine in a step response test seems compatible with the behavior that the engine elicits in the WHTC, but less so with the behavior in the SORT cycle.

Finally, the developed method should be viewed as a step towards better estimation of soot levels in transient drive cycles. All problems have not been solved but promising results were obtained. Furthermore the ideas from this work can be used for other emissions as well, possibly modifying the method to fit that emission type (e.g. number of filters, speed of filters etc.).

[9]The step response has a peak during a few seconds and stationarity is approximately reached after 10 seconds.

5 Black box modeling

In the previous section an estimation method was developed where an engine's specific properties were represented by stationary emission maps and step response tests were used to explain the transient deviations from steady state emission levels. The work presented in this section instead focuses on developing a model for the soot emission (soot mass flow) from a diesel engine using statistical methods. No a priori model structure is assumed which is denoted *black box* modeling. Experimental data are used throughout the modeling in making modeling choices, e.g. model order. Deriving models describing a systems behavior in relevant aspects from observed data is referred to as *system identification*.

5.1 Data preprocessing

Preprocessing soot measurements
It was discovered that the soot concentration measurements were not updated at every sample but instead a value is held for 6-9 samples. Whether this data corruption is caused by the MSS or the logging interface, it needs to be dealt with since other logged data are updated at a higher rate (10 Hz). The assumption made is that in a sequence of held data points, the first one is correct. Then the rest of the points are omitted and an interpolation between the valid measurements is made. This interpolation is performed using piecewise cubic hermite splines. The reason for choosing this interpolation as opposed to normal cubic spline interpolation is to reduce oscillations usually coupled with splines. In Figure 21, the raw and interpolated measurements are shown.

Filtering
It is often a good idea to filter the data before use in the system identification. The nature of the filtering is low pass and serves to remove unwanted measurement noise. The input data used in the system identification was low pass filtered using a Butterworth filter with a cutoff frequency of π rad/s. As for the output, there is no need for filtering due to the interpolation done.

Soot mass flow
The MSS measures the soot concentration S_c in the exhaust gases in mg/m^3 but it is the engine out soot mass flow S_{mf} that is of interest to be modeled. To relate the soot flow to the soot concentration, the exhaust gas flow is needed as well as the density of the exhaust gases. Since density inherently has a temperature dependency and the exhaust gases can elicit quite large temperature variations, some approximation is needed. The approximation done is to use the average exhaust gas density, denoted ρ (kg/m^3), at all times regardless of actual exhaust gas temperature. Let the exhaust gas flow be q kg/h, then the soot mass flow S_{mf} can be expressed as follows.

$$S_{mf} = \frac{S_c q}{\rho} \tag{15}$$

The exhaust gas flow q has negligible delay and is synced with the engine speed and engine torque. On the other hand, the soot concentration measurements given by the MSS are delayed due to transportation time of the exhaust gases from combustion

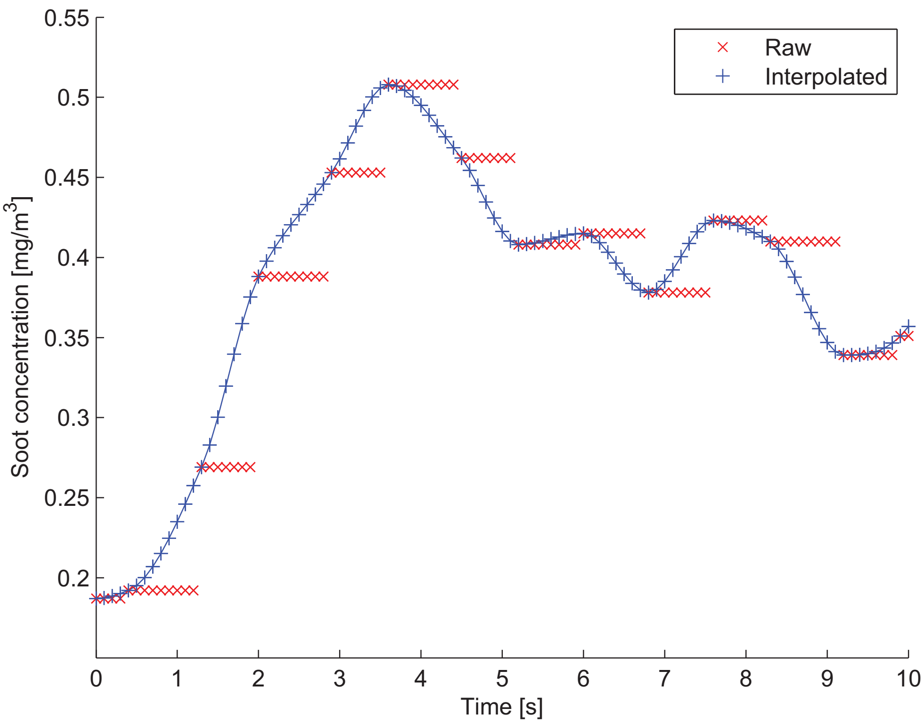


Figure 21: Reconstruction of soot concentration using piecewise cubic hermite spline interpolation. The raw data is also shown.

chamber to measuring equipment and the time it takes to produce a measurement (dilution is included). To keep the system causal and align the exhaust gas flow with the soot concentration in the exhausts, q is delayed for one second or with the 10 Hz sampling frequency; 10 samples.

$$S_{mf}(k) = \frac{S_c(k)q(k-10)}{\rho} \tag{16}$$

5.2 Model order selection

Since black box modeling by definition does not assume any model structure, the orders to use in the system identification are not known a priori. The orders of the system needs to be estimated somehow to know how dynamic the soot emissions are. Initially any inputs affecting the system are disregarded and experimental output data is used to determine how many previous soot emission values need to be considered when estimating the current soot emission level. Such models where the current value only depends on previous values is termed autoregressive (AR) models. Assuming an AR model structure, models of different orders (number of previous samples considered) are fitted. Then the errors produced by comparing the measured values and the values given by the AR model are used to create a loss function. This function is a scalar measure of the errors that result from using the model. Thereby this loss function is a way of getting an initial estimate of model orders to use by choosing the model orders that minimizes the loss function. The model that results in minimizing the loss function is denoted the *best fit*. This is however a simple validation criteria as it does not penalize the usage of high order models. Other criteria that do include penalties for using complex models are the Akaike's Information Criterion (AIC) and Rissanen's Maximum Description Length (MDL) [4, 5]. Evaluation of all three validation criteria was performed with the same results, namely the suggestion of model order 6. The results are shown in Figure 22.

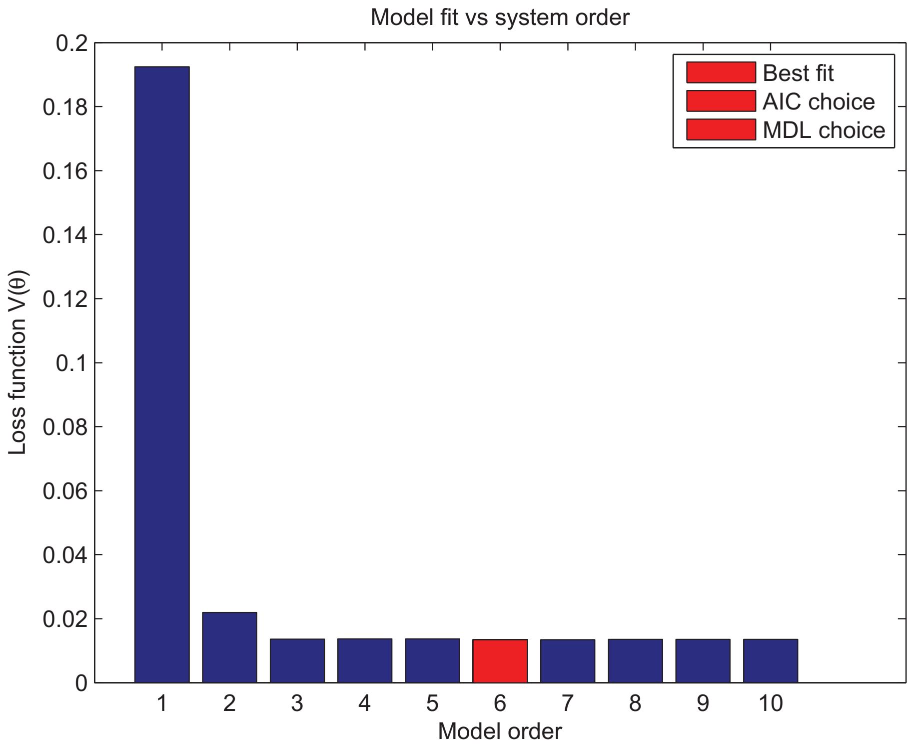


Figure 22: Loss function computed for the AR model structure. The AIC, MDL and best fit choices are highlighted.

5.3 Choosing inputs

An important task is to choose what inputs to use in the system identification. From the background section it is known that there are many inputs that affect the soot mass flow. However some limitation regarding the number of inputs must be made in order to have a feasible task ahead. Therefore a conscious choice of using two input variables was made. A further restriction is that the inputs must be available to the ECU in order for the model to be applicable for online usage. A statistical method of determining inputs that have a great impact on the output is to compute the cross correlation. This was done with several variables shown to be correlated with the soot mass flow. The two signals chosen as inputs are injected fuel mass per cycle δ and the equivalence ratio ϕ. From a physical perspective, these inputs are very plausible. In figure 23, the cross correlation function is given between δ and S_{mf} as well as between ϕ and S_{mf}. The correlation is normalized, meaning the maximum possible correlation is one.

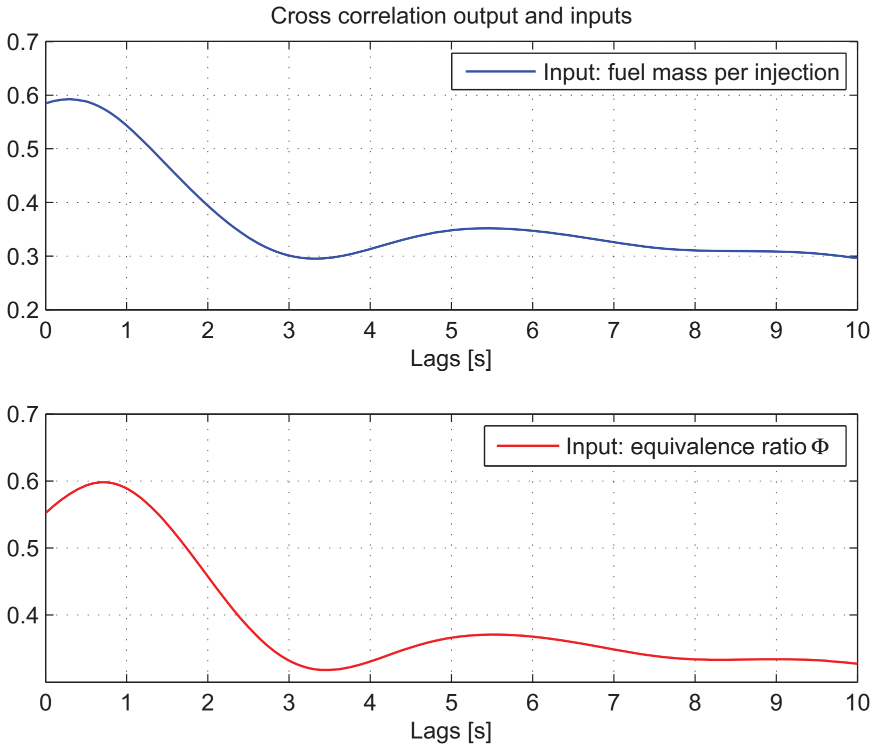
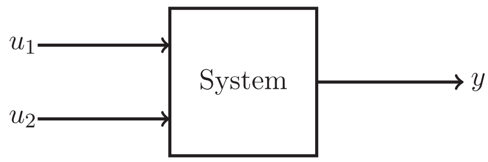


Figure 23: Cross correlation function between inputs u_1, u_2 and output y.

The system to identify is hence a two-input-single-output system.



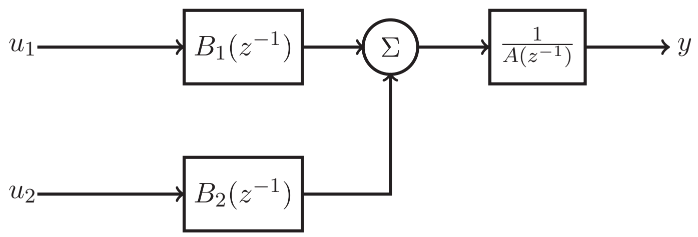
where $u_1 = \delta$, $u_2 = \phi$ and $y = S_{mf}$.

5.4 Linear models

The fact that the system that is to be modeled is nonlinear can easily be realized. For once, the soot emission can never be negative. Also it is clear from the literature that the soot emissions are nonlinear. Despite this it is a good idea to start by fitting linear models. This serves to increase the understanding of the system dynamics as well as provides a starting point for nonlinear model fitting. The validation criteria AIC, MDL and best fit suggested sixth order dynamics when an AR model structure was used. These results were used to select model order ranges in which to search for linear models using several model structures. The structures used are autoregressive models with exogenous inputs (ARX), autoregressive moving average models with exogenous inputs (ARMAX) and output error models (OE). Block diagrams showing the structures are given in the respective sections below.

5.4.1 ARX models

The first model structure to be investigated is the ARX structure. This model structure has no explicit (colored) noise model but is a good starting point for further modeling. It is basically comprised of taking the current[10] inputs and a number (determined by the orders) of previous inputs, multiplied by the coefficients of the fitted polynomials B_1 and B_2 and added together. This sum is then divided by a number (also determined by order) of previous outputs multiplied by the coefficients of A to give the current estimate. The structure is shown below in a block diagram.



Based on the results from the model order selection of an AR model structure, a range of orders for the ARX model identification is chosen. In Table 3, the selected range is given. n_a denotes the number of parameters in the A-polynomial. Analogously for the other polynomials. n_{k_i} denotes the delay for input i. This means that it is possible to not use the current input sample if there is no direct term.

Table 3: Model orders investigated with an ARX model structure.

Model structure	n_a	n_{b_1}	n_{b_2}	n_{k_1}	n_{k_2}
ARX	4-8	2-8	2-8	0-5	0-5

Models were fitted in the given range and the model *fit* computed. This was used as a validation criterion as well as the AIC, MDL and best fit criteria, the results of which are given in Table 4.

[10] Could be omitted if $n_{k_i} > 0$.

Table 4: Model orders chosen by AIC, MDL and best fit criteria.

Criterion	n_a	n_{b_1}	n_{b_2}	n_{k_1}	n_{k_2}	**Reference**
AIC	6	8	8	1	1	$\mathcal{N}_1$
MDL	6	2	8	1	1	$\mathcal{N}_2$
Best fit	6	8	8	1	1	$\mathcal{N}_1$

It was found that indeed the best performing model had the orders suggested by two of the validation criteria (AIC and best fit), namely $\mathcal{N}_1$. The polynomials are given below for the best performing ARX model. Keep in mind that the z-transform is used.

$$\begin{aligned} A(z^{-1}) =& 1 - 5.014z^{-1} + 10.73z^{-2} - 12.58z^{-3} \\ &+ 8.529z^{-4} - 3.175z^{-5} + 0.5072z^{-6} \end{aligned} \tag{17}$$

$$\begin{aligned} B_1(z^{-1}) =& 1.785z^{-1} - 6.217z^{-2} + 11.27z^{-3} - 13.87z^{-4} \\ & 13.85z^{-5} - 14.08z^{-6} + 10.65z^{-7} - 3.379z^{-8} \end{aligned} \tag{18}$$

$$\begin{aligned} B_2(z^{-1}) =& -1868z^{-1} + 7476z^{-2} - 12610z^{-3} + 11680z^{-4} \\ &- 6835z^{-5} + 3480z^{-6} + 1872z^{-7} - 556z^{-8} \end{aligned} \tag{19}$$

The output of that model (Equation 17-19) accounted for 28.1 % of the variance in the measured output. This measure of fit is called Variance Accounted For (VAF) and will be used throughout this report. The model is also judged based on its ability to simulate the accumulated soot flow for reasons explained in the discussion section. In that respect the VAF is 65.8 %. It should be noted that the validation is done on the whole WHTC drive cycle, i.e. both identification data and validation data which have been given equal parts of the original data set. For instance in Figure 24 the first half of the time resolved comparison is the part used in the identification, the second half was not used in the identification and is new to the model. As seen in Figure 25, the residuals are not uncorrelated for the ARX model with orders $\mathcal{N}_1$. The residuals are the errors, the deviations between the measured outputs and the modeled ones. Ideally, the residuals should be uncorrelated using a certain statistical confidence level. Here, a 99 % confidence level was used.

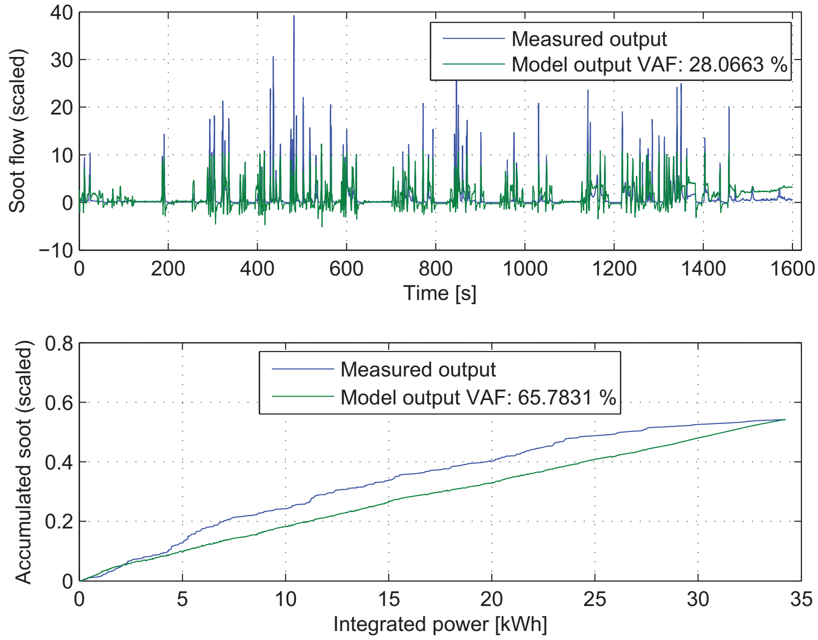


Figure 24: Comparison between measured and simulated output for the linear ARX model with orders $\mathcal{N}_1$, given in Equation 17-19. Data retrended before comparison. The results are for engine A running the WHTC cycle.

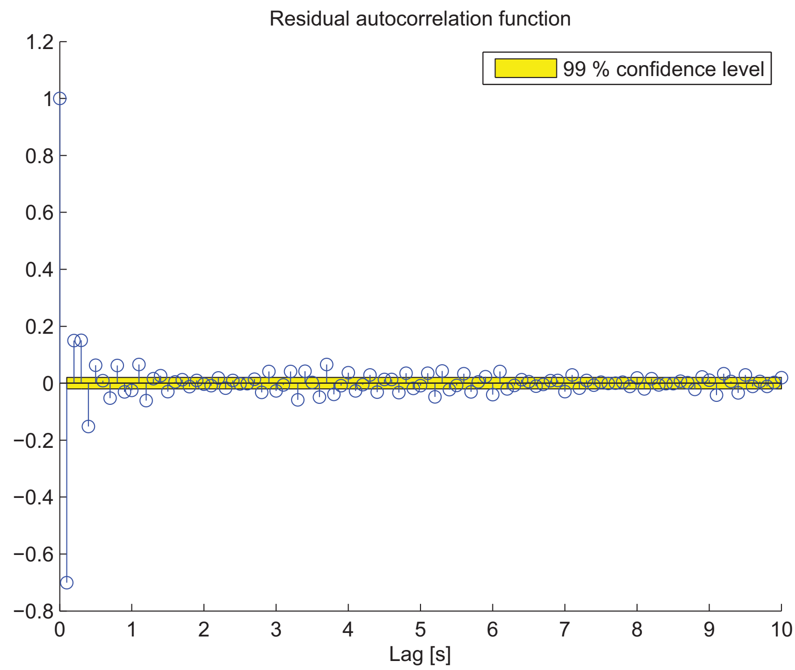
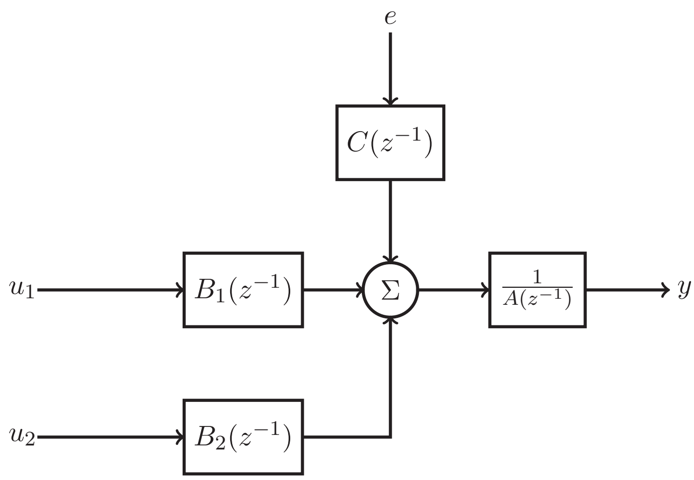


Figure 25: Residual autocorrelation function with 99 % confidence level for the model in Figure 24.

5.4.2 ARMAX models



The residuals of the best ARX model were not uncorrelated. Hence the next natural step is to add a colored noise model (the C-polynomial), yielding an ARMAX model structure. First the model orders from the best ARX model were used with the addition of C-polynomials with orders ranging n_c = 1-5. While adding a noise model served to improve the model fit somewhat, the gain in model fit was too insignificant and the reduction of correlation in the residuals likewise to motivate its usage. The performance of the model with orders $[n_a\ n_{b_1}\ n_{b_2}\ n_c\ n_{k_1}\ n_{k_2}] = [6\ 8\ 8\ 5\ 1\ 1]$ along with the residual autocorrelation is shown in Figures 26-27.

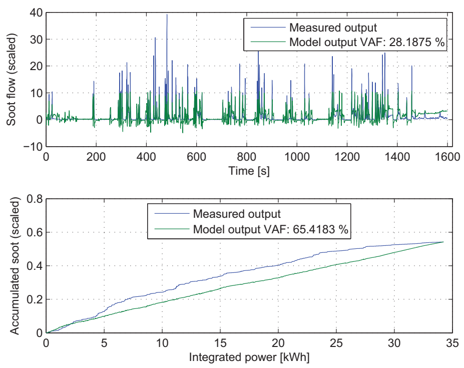


Figure 26: Model fit ARMAX model with orders $[n_a\ n_b\ n_c\ n_k] = [6\ 8\ 8\ 5\ 1\ 1]$. Data retrended before comparison. The results are for engine A running the WHTC cycle.

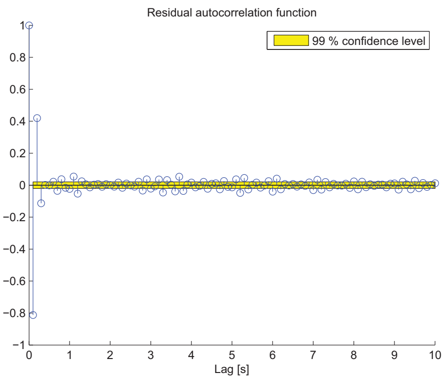


Figure 27: Residual autocorrelation function for ARMAX model with orders $[n_a\ n_b\ n_c\ n_k] = [6\ 8\ 8\ 5\ 1\ 1]$.

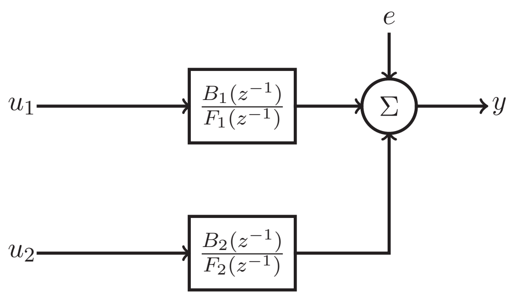
Using the orders from the best ARX model as an initial estimate for orders to use in the ARMAX model identification yielded little success. Hence a more general approach is to vary the orders of all the included polynomials in the ARMAX model (within reasonable intervals). Such a range was set up and is given in Table 5.

Table 5: Model orders investigated with an ARMAX model structure.

Model structure	n_a	n_{b_1}	n_{b_2}	n_c	n_{k_1}	n_{k_2}
ARMAX	4-8	2-8	2-8	1-5	0-5	0-5

Fitting models in the range selected in Table 5 did not result in any significant improvement over the model given in Figure 26.

5.4.3 Output Error models



The third and final linear model structure investigated was the output error (OE) model structure. The reason for choosing this structure as opposed to similar ones that include colored noise models such as Box-Jenkins models is that eventually non-linear Hammerstein-Wiener models are to be fitted where the linear part is such an OE model. For the observant this model structure resembles the ARX structure with a slight difference, this being the possibility of splitting the soot dynamics between the inputs. The A-polynomial in the ARX models (common for both channels) is split into F_1 and F_2. The range of model orders investigated is given in Table 6.

Table 6: Model orders investigated with an OE model structure.

Model structure	n_{b_1}	n_{b_2}	n_{f_1}	n_{f_2}	n_{k_1}	n_{k_2}
OE	2-8	2-8	4-8	4-8	0-5	0-5

In the given range it was found that the model orders resulting in the best fit are $[n_{b_1}\ n_{b_2}\ n_{f_1}\ n_{f_2}\ n_{k_1}\ n_{k_2}] = [5\ 7\ 4\ 6\ 2\ 1]$ (hereafter referred to as $\mathcal{N}_3$). The model output had a fit (VAF) of 30.3 % dynamically, i.e. time resolved soot mass flow as a function of time. Cumulatively, i.e. accumulated soot mass as a function of integrated power, the VAF is 88.2 %, see Figure 28.

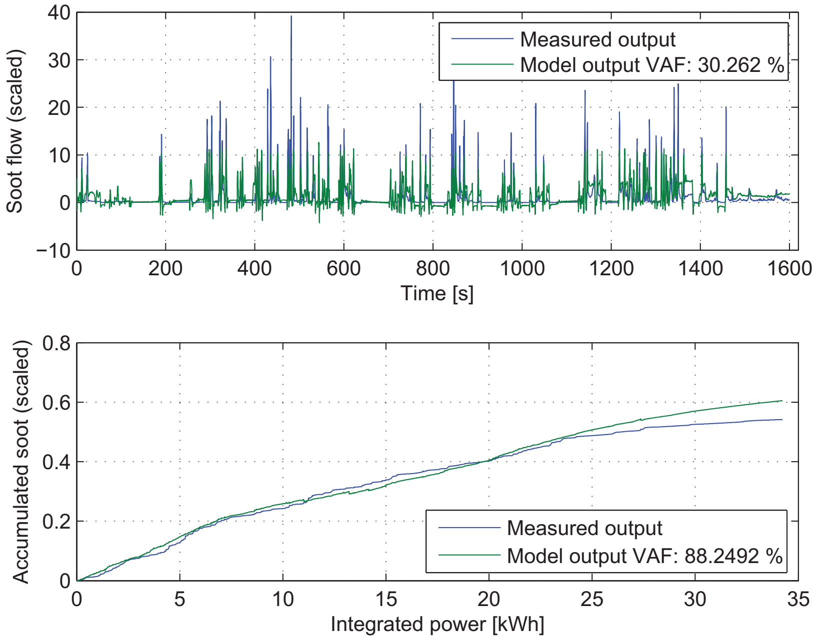


Figure 28: Comparison between measured and simulated output for the linear OE model with orders $\mathcal{N}_3$. Data retrended before comparison. The results are for engine A running the WHTC cycle.

5.4.4 Coherence

By now it is clear that linear models are inadequate to represent the soot emissions. This can be further illustrated by computing the *coherence* function. A coherence function with the value one for all frequencies of interest, i.e. where there are important dynamics, represents that the relationship between the inputs and outputs is linear, that there are no disturbances, noise or unrepresented inputs. Clearly this is not the case here, as can be seen in Figures 29-30. Therefore the identification of nonlinear models is proceeded with next.

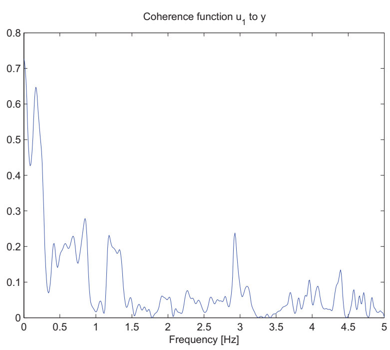


Figure 29: Windowed coherence function from u_1 to y. Chebychev window.

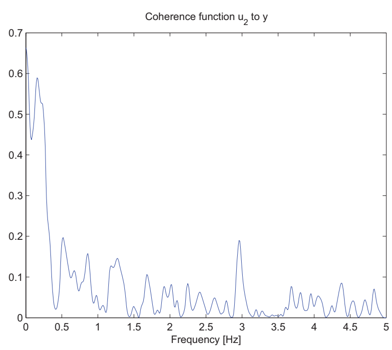


Figure 30: Windowed coherence function from u_2 to y. Chebychev window.

5.5 Nonlinear models

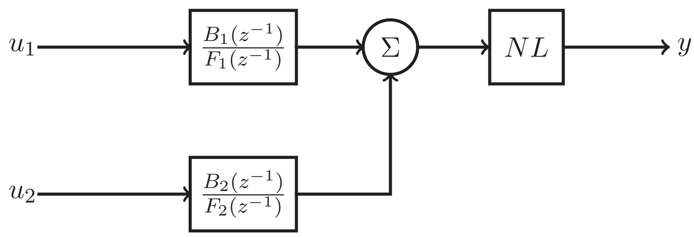
Since linear models obviously are not enough, investigating nonlinear options is the next natural step. The information obtained from the linear modeling will however be used, e.g. by providing a starting point for the orders of the nonlinear models. When it comes to nonlinear modeling, the number of possible structures and choices rapidly increases. For that reason a restriction to using three different structures is made. The structures are *Hammerstein*, *Wiener* and *Hammerstein-Wiener* models [4]. These models are comprised of a linear part, here an output error structure (see 5.4.3), complemented with static nonlinearities. Static nonlinearities are transformation functions added to the input(s) and/or output. The *static* property means that the functions are memoryless. The value of the function only depends on its current argument. Depending one where the nonlinearity is added, the model is given one of the above stated names. Wiener models only have a nonlinearity block on their output channel, whereas Hammerstein models have nonlinearity block(s) on one or both inputs. The combination of both structures, i.e. adding nonlinearities to inputs and outputs, results in Hammerstein-Wiener models.

The nonlinearities considered in this project are simple:

- *Piecewise linear functions*
  Functions that are linear in each of their sections. The number of sections can be varied.

- *One dimensional polynomials*
  Polynomials of various degrees.

5.5.1 Wiener models



The structure of a Wiener models i shown in the block diagram above. As previously stated, it is an output error model with a static nonlinearity on the output (soot mass flow). The nonlinearities are denoted *NL* in block diagrams. It was investigated whether this added block provides an improvement in the performance of the model. The comparison is to be made with the corresponding best linear models. For the nonlinear function block, piecewise linear functions with various number of units as well as polynomials of different orders were used. In Table 7, the range of orders and nonlinearities investigated are listed. For the Wiener models that were fitted with linear model orders $\mathcal{N}_1$, the best performance was achieved with a 4th order polynomial as the static output nonlinearity. The same type of nonlinearity proved to give the best model with orders $\mathcal{N}_2$ for the linear model. Since the best models for both model orders ($\mathcal{N}_1$ and $\mathcal{N}_2$) gave approximately the same fit, the one with lowest

Table 7: Wiener models fitted.

Orders	**Output Nonlinearity**
$\mathcal{N}_1$ - $\mathcal{N}_3$	Polynomial (degree 2-4)
$\mathcal{N}_1$ - $\mathcal{N}_3$	Piecewise linear function (4-10 units)

orders is of course the given choice. The model output is compared to the measured output in Figure 31. The results are quite good with a VAF of 40.2 % dynamically and 94.5 % cumulatively. It can be seen that the dynamic model output (soot mass flow as function of time) now never assumes negative values and the overall behavior of the model shows a vast improvement over the linear counterpart with the same orders. The fact that the model does not give negative values can be seen in the nonlinearity in Figure 32 where the function values are strictly positive. Wiener models having orders $\mathcal{N}_3$ gave no satisfying results. As for the best Wiener model found, it is given below (z-transform).

$$B_1(z^{-1}) = z^{-1} - 1.003z^{-2} \tag{20}$$

$$\begin{aligned} B_2(z^{-1}) = & -1136z^{-1} - 942.5z^{-2} + 6652z^{-3} - 5565z^{-4} \\ & - 525.2z^{-5} + 1311z^{-6} + 2589z^{-7} - 2304z^{-8} \end{aligned} \tag{21}$$

$$\begin{aligned} F_1(z^{-1}) = & 1 - 3.337z^{-1} + 3.376z^{-2} + 0.6922z^{-3} \\ & - 3.727z^{-4} + 2.592z^{-5} - 0.5912z^{-6} \end{aligned} \tag{22}$$

$$\begin{aligned} F_2(z^{-1}) = & 1 - 4.171z^{-1} + 7.875z^{-2} - 8.717z^{-3} \\ & + 5.989z^{-4} - 2.409z^{-5} + 0.4374z^{-6} \end{aligned} \tag{23}$$

The static output nonlinearity is as mentioned a 4th degree polynomial:

$$NL(x) = 0.000016x^4 - 0.00039x^3 + 0.0055x^2 - 0.0020x + 0.077 \tag{24}$$

where x is the input to the nonlinearity.

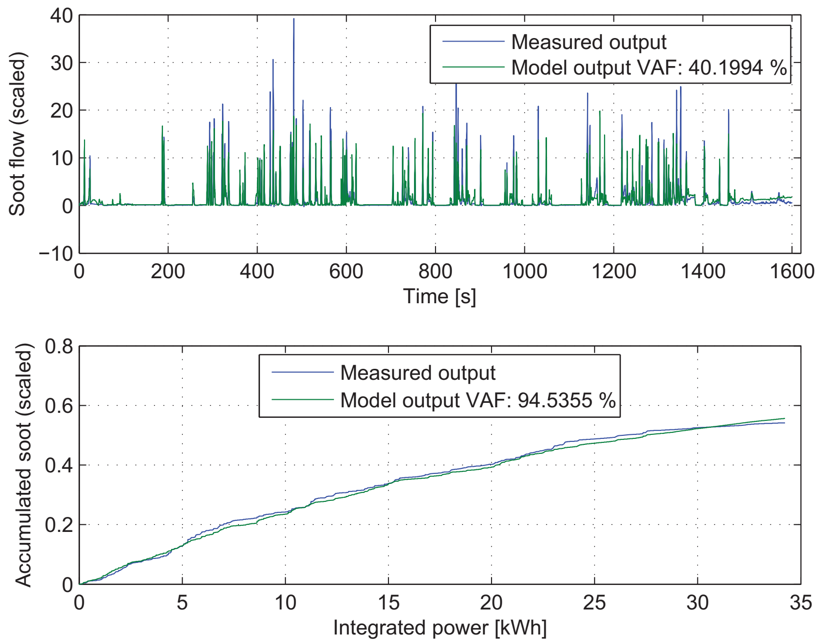


Figure 31: Wiener model with linear part using orders $\mathcal{N}_2$. Output nonlinearity is a polynomial of degree 4. The results are for engine A running the WHTC cycle.

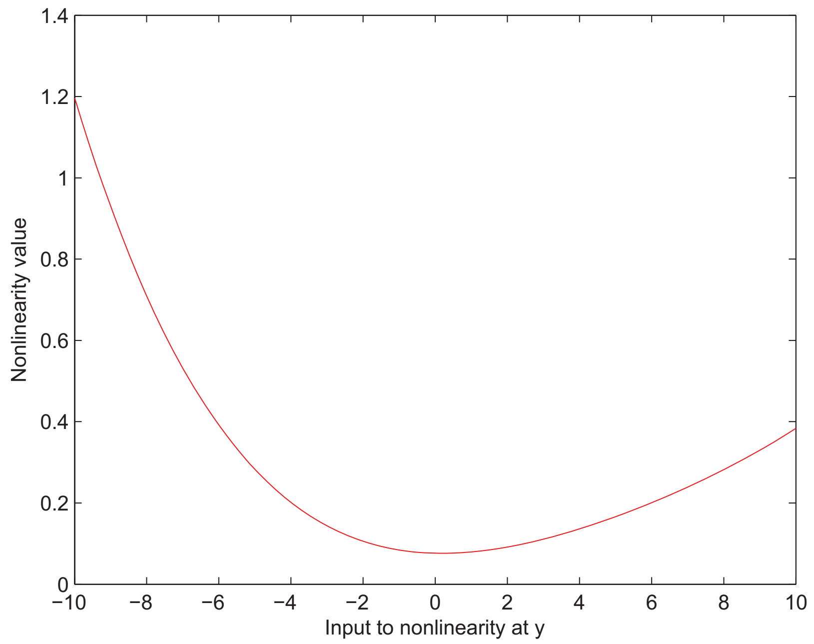
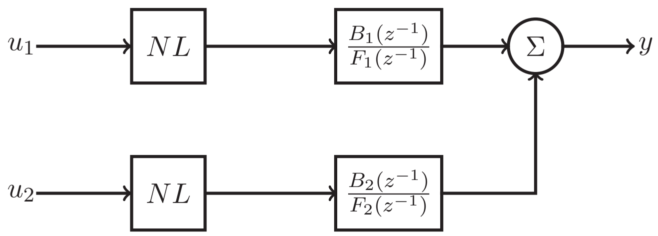


Figure 32: Output nonlinearity for model in Figure 31.

5.5.2 Hammerstein models



By placing nonlinearities on the input channels instead of on the output channel, Hammerstein models are obtained. The orders $\mathcal{N}_1$ - $\mathcal{N}_3$ for the linear part was used here as well and the input nonlinearities were polynomials and piecewise linear functions in the same range as the one used for fitting Wiener models. The results from fitting Hammerstein models were not as good as when fitting Wiener models although interesting results were obtained. It was first investigated whether a nonlinearity on either of the inputs alone could provide any good models. It was found that a nonlinearity on the first input channel alone resulted in bad models. Instead the second input channel was where the nonlinearity gave the best results. Also having nonlinearities on both input channels did not improve the results. Hence the Hammerstein modeling provided the knowledge that a piecewise linear nonlinearity on the second input channel (u_2) is what should be used for Hammerstein models. A comparison between measured and modeled output for the best Hammerstein model (order $\mathcal{N}_2$ for linear part) is given in Figure 33. In Figure 34, the input nonlinearity is shown. The piecewise linear function has 7 units. The modeled output is not strictly positive in the dynamic case, as the best Wiener model was. Also cumulatively, the shape of the output is not as good.

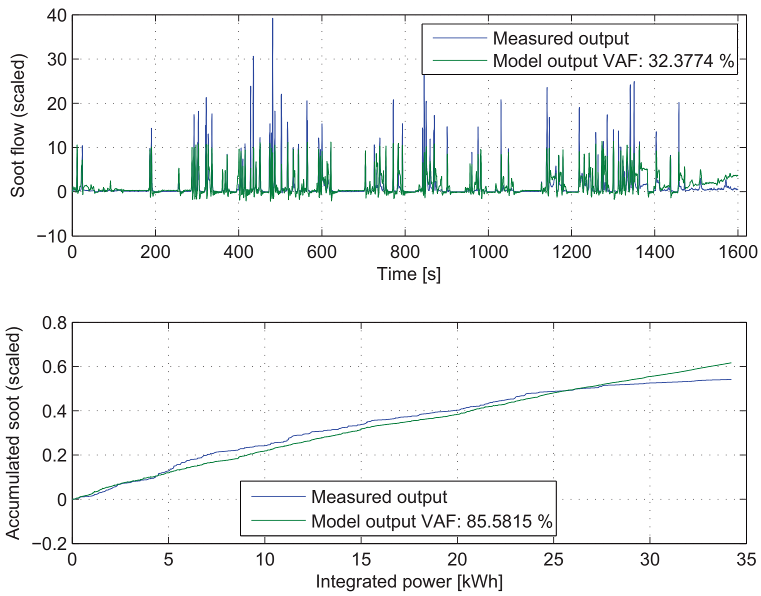


Figure 33: Comparison between measured and modeled output for a Hammerstein model with linear model having orders $\mathcal{N}_2$ and an input nonlinearity that is a piecewise linear function with 7 units on input u_2. The results are for engine A running the WHTC cycle.

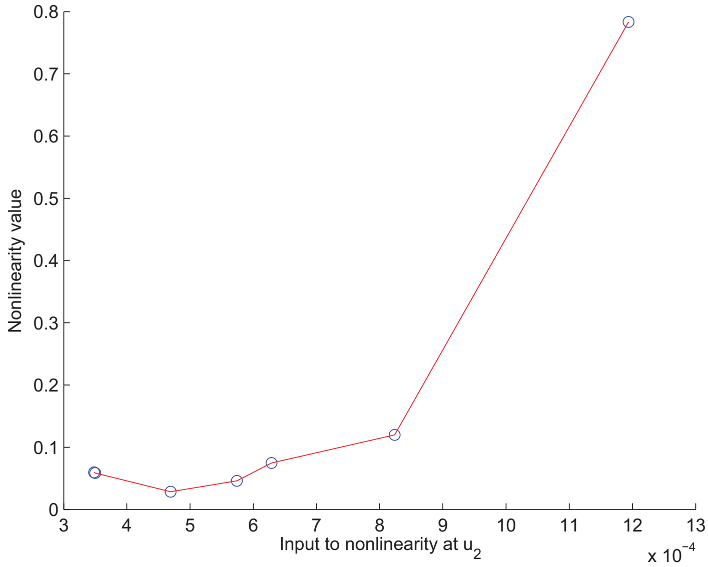
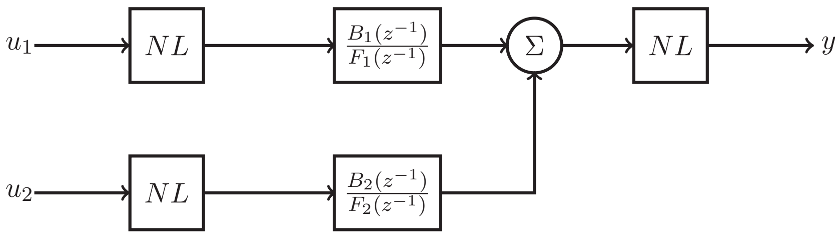


Figure 34: Input nonlinearity at second input channel for the Hammerstein model in Figure 33.

5.5.3 Hammerstein-Wiener models



While estimating Wiener models generated good results, a combination of Hammerstein and Wiener models might perform better. Such models are called Hammerstein-Wiener models and static nonlinearities can be placed on inputs as well as on the output. The static nonlinearities used in the estimation of Hammerstein-Wiener model are as before: polynomials and piecewise linear functions. The Wiener modeling suggested a polynomial as output nonlinearity and the Hammerstein modeling suggested that a piecewise linear function on the input u_2 be used as input nonlinearity. The question that arises is whether using the nonlinearity configuration from the best Wiener model and best Hammerstein model, i.e. piecewise linear nonlinearity with 7 units on u_2 and a 4th order polynomial on y, will provide the best Hammerstein-Wiener model as well. Such a model was identified and is the result is shown in Figure 35. Although an improvement is made, this improvement is not significant enough to motivate the added complexity.

Since combining the configurations from the best Wiener model with the ones from the best Hammerstein model did not provide a significant model fit gain, a systematic search for the best Hammerstein-Wiener model was conducted. This was done using orders $\mathcal{N}_1$ - $\mathcal{N}_3$ for the linear model and all three static nonlinearities were varied between piecewise linear functions with 4-10 units and polynomials with degree 2-6. No major improvement on the initial guess using the configuration from the best Wiener model and best Hammerstein model was found.

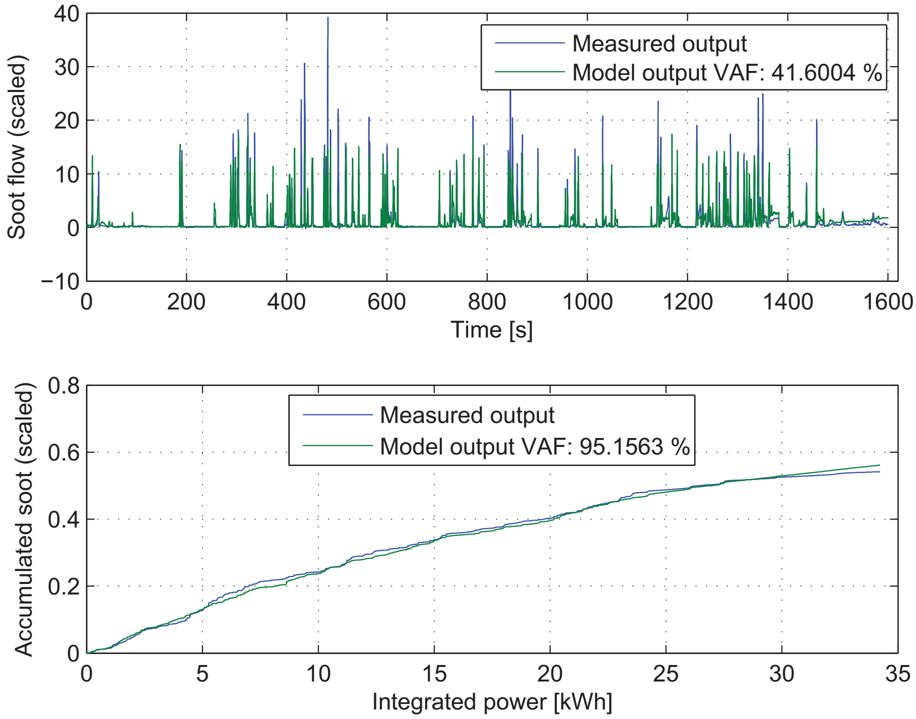


Figure 35: Hammerstein-Wiener model with linear model having orders $\mathcal{N}_2$. The input nonlinearity is placed at u_2 and is a piecewise linear function with 7 units. The output nonlinearity is a 4th order polynomial. The results are for engine A running the WHTC cycle.

5.6 Discussion

Performing black box modeling of the soot mass flow emission proved successful with some configurations and less so with others. The best model was achieved with a Hammerstein-Wiener model, *however* the best Wiener model was almost as good. For that reason it is not motivated nor recommended to use Hammerstein-Wiener models. For the lesser complexity and number of needed parameters, the Wiener model is capable of describing and representing the system of soot emission. The total number of parameters in the best Wiener model is 27. This includes the linear model and the static nonlinearity.

Some attention must be directed towards the linear models. A matter that is of importance is the detrending and retrending of data. Linear models cannot handle offsets or trends in data. Therefore before fitting a linear model any constant/linear trend is removed to yield a zero mean signal. Then linear models are fitted. When it comes to using the model, it must be known what trend was removed during identification so it can be added. For some systems the trend is constant over the entire range where the system operates. The data used in this project came from a log when engine A (see Table 1) was running a WHTC. Essentially what this means is that whenever such a model is to be used, the operating conditions must be similar to those in the WHTC (or the average soot mass flow must be known/guessed for the actual run). Furthermore, linear models failed to retain the shape of the accumulated soot mass curve. This is probably due to the fact that the retrending adds an average trend for the whole cycle, but it is known (and seen from measurements) that the average soot emissions vary in the different parts of the WHTC, as they most likely will in general operation.

Since this whole issue can be overcome by using nonlinear models and since the Wiener model *was* the better[11] model, linear models can be left aside. Fitting nonlinear models is not easy however and there are some difficulties here. First of all, whereas linear models have a unique solution, nonlinear models do not. Taking the Wiener model as an example, the procedure is the following. An initial guess for the output nonlinearity is made and the linear part is identified. The errors are computed and the nonlinearity is adjusted to decrease the errors. Then the linear part of the model is refitted. This kind of iterative numerical solution can be problematic. When the iterations have reached a minimum, there is no way of guaranteeing that the minimum is global. Some measures can be taken to try to find a lower minimum (such as trying different initial values etc.) but there are no guarantees.

The interpolation of the soot concentration measurements was basically heuristic and here some more work could have been done to develop a better method of reconstruction of the real soot concentration than the proposed spline interpolation. Another issue in the modeling is the delay between inputs and output. The problem is that the system identification performed in this project assumes constant delays, but this not the case. However, introducing varying delays is a task well beyond the scope of this project. The input data used are of course cycle averaged which might be reason of restriction to the level of model fits that are achievable.

All the models in this project were evaluated based on their ability to estimate the accumulated emitted soot mass as a function of integrated power. The reason making this important is that manufacturers of vehicles have attended to using diesel particulate filters (DPF) as part of the aftertreatment system. This is a way of reducing soot emissions and for purposes regarding DPF-functionality it is the accumulated soot mass that is of interest since filters are physical integrators (disregarding regeneration).

[11] Disregarding the Hammerstein-Wiener model due to its insignificant improvement over the Wiener model.

6 Conclusion

The initial question to be answered was whether stationary emission maps could be used with step response test to estimate the soot emissions from a heavy duty diesel engine in transient operation. Conclusively, the results are positive. The emissions can be estimated for transient operations similar to the conditions of the WHTC using the empirical method in Section 4. A working method was presented but more importantly, the idea of linear combination of stationary emission maps and bandpass filtered variables is what should be taken away from this. By modifying the method for other emissions, there is a belief that the method is useful for other purposes than soot estimation.

Black box modeling yielded that Wiener models are to recommend. They succeeded in representing the soot emissions in an adequate manner. Given that only two inputs are used, the results must be considered to be good.

6.1 Future work

The empirical estimation method showed positive results for estimation in the WHTC cycle. Future work could be to look over the method and possibly find modifications/improvements that make the method applicable in other cycles as well with acceptable accuracy. As for the black box modeling, it was conducted on data from engine A running the WHTC cycle. The continuation of that work is to study how such models succeed in other conditions, perhaps less transient. Furthermore, a larger data set is needed in the form of other engines. An advanced topic of the identification of nonlinear black box models that could be investigated is the numerical properties of the solution. During this work it was found that the scaling of signals (input signals and output signals) drastically affected the quality of the models.

7 References

[1] Institute for Governance and Sustainable Development (2008), *Reducing Black Carbon May Be the Fastest Strategy for Slowing Climate Change,* `www.indiaenvironmentportal.org.in/files/Dec-08-poznan-cc.pdf` [January 28, 2011]

[2] AVL, *Micro Soot Sensor – Continuous Measurement of Soot Concentration,* `http://www.avl.com/micro-soot-sensor` [January 28, 2011]

[3] Haisch, C; Niessner, R *Light and sound – photoacoustic spectroscopy,* Spectroscopy Europe (2002), 14(5), 10-15.

[4] Johansson, R (2010) *System Modeling and Identification*, Lund: Dept. Automatic Control, Lund University.
First edition: (1993) Prentice Hall, Englewood Cliffs, NJ.

[5] MATLAB, *System Identification Toolbox,* `http://www.mathworks.com/help/toolbox/ident/`

[6] Heywood, John B. (1988) *Internal Combustion Engine Fundamentals*, McGraw-Hill.

[7] Hirsch, M; Alberer, D; del Re, L *Grey-Box Control Oriented Emission Models* IFAC (2008).

[8] Hiroyasu, H; Kadota, T. *Models for Combustion and Formation of Nitric Oxide and Soot in Direct Injection Diesel Engines* SAE Paper 760129 (1976).

[9] Argachoy, C; Pimenta A. P. *Phenomenological Model of Particulate Matter Emission from Direct Injection Diesel Engines* Journal of the Brazilian Society of Mechanical Sciences and Engineering, July-September (2005).

[10] Zheng, J; Xin, Q *Theoretical Analysis of Diesel Engine NOx and Soot With Heuristic Macro-Parameter-Dependent Approach and Virtual Multi-Zone Real-Time Models* SAE Paper (2009).

[11] Eastwood, P. G. et al. *Estimation of deviations in NO and soot emissions between steady-state and EUDC transient operation of a common-rail diesel engine* SAE Paper (2009).

[12] Ramamurthy, R; Clark, N; Atkinson, C; Lyons, D *Models for Predicting Transient Heavy Duty Vehicle Emissions* SAE Paper (1998).

[13] Morgan, R. E. et al. *Characterisation of the Soot Formation Processes in a High Pressure Combusting Diesel Fuel Spray* SAE Paper (2003).

[14] Anderson, John D. (1995), Computational Fluid Dynamics: The Basics With Applications, Science/Engineering/Math, McGraw-Hill Science, ISBN 0070016852.

[15] Westlund, A *Measuring and Predicting Transient Diesel Engine Emissions* Lic. thesis, Royal Institute of Technology, Stockholm (2009).

[16] Wang, G et al. *A Developed Model for Emissions Prediction of a DI Diesel Engine* SAE Paper (1999).

[17] Kim, H; Sung, N *Combustion and Emission Modeling for a Direct Injection Diesel Engine* SAE Paper (2004).

[18] Bertola, A; Shubiger, R; Kasper, A; Matter, U; Forss, A. M; Mohr, M; Boulouchos, K, Lutz ,T *Characterization of Diesel Particulate Emissions in Heavy-Duty DI-Diesel Engines with Common Rail Fuel Injection Influence of Injection Parameters and Fuel Composition,* SAE Paper 2001-01-3573 (2001).

A Appendix

A.1 Abbreviations

Abbreviation	Explanation
LS	Least Squares
AR	AutoRegressive
ARX	AutoRegressive with eXogenous input
ARMAX	AutoRegressive Moving Average with eXogenous input
OE	Output Error
VAF	Variance Accounted For
MDL	Maximum Description Length
AIC	Akaike's Information Criteria
ECU	Engine Control Unit
WHTC	World Harmonized Transient Cycle
SORT	Standardized On-Road Test cycle
EGR	Exhaust Gas Recirculation
DPF	Diesel Particulate Filter
MSS	Micro Soot Sensor
CAD	Crank Angle Degrees
SOI	Start Of Injection
EOI	End Of Injection