Design of diesel engine's optimal control maps for high efficiency and emission reduction

Hoang Nguyen Khac

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 20.11.2017

Thesis supervisor:

Docent. Kai Zenger

Thesis advisor:

M.Sc. Sergey Samokhin

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Hoang Nguyen Khac

Title: Design of diesel engine's optimal control maps for high efficiency and emission reduction

Date: 20.11.2017 Language: English Number of pages: 10+77

Department of Electrical Engineering and Automation

Professorship: Control Engineering

Supervisor: Docent. Kai Zenger

Advisor: M.Sc. Sergey Samokhin

The objective of this thesis is to create static optimal control maps of diesel engines for high efficiency and emission reduction. The calibration tool to be used to create the control maps, named "Off-line parameterization tool", was designed based on the Design of Experiments method. The optimization goal is to minimize the Brake Specific Fuel Consumption (BSFC) of the engine by the engine's input parameters and under some emission constraints. The tool was designed to be able to work both fully automatically and semi-automatically. Though most of the previous researches on engine calibration have used the Design of Experiments approach, their implementations in choosing experimental design types and optimization process are different compared to this thesis. The unique aspect of this research lies on the significant properties of the Off-line parameterization tool. Firstly, this tool is flexible, it is able to work with multiple inputs and multiple outputs. Secondly, it can reduce the calibration time as the engine running time is kept as small as possible and all the data processing work is done automatically.

Keywords: diesel engine, design of experiment, static optimal control, engine calibration

Preface

First of all, I want to say a big "Thank you" to Docent Kai Zenger for giving me the oppoturnity to do this Thesis and always having faith in my abilities. Even though there were some difficult times during the last seven months, he was still having patience and kept encouraging me to finish the work.

Secondly, I want to thank Sergey Samokhin who has worked alongside me throughout this project. He was always there to help me familiarize myself to the work in the beginning and to constantly share his knowledge about the diesel engines to such a novice like me.

Moreover, my gratitude goes to Ranta Olli, Blomstedt Otto, David Bernasconi and everyone in the Internal Combustion Engine laboratory in Aalto University. Had it not been for your kindly help, I could not have finished my Thesis with valuable results.

I also want to say thanks to my friends who have been supporting and cheering me during the time. Special thanks to my girlfriend who has been there with me since day one, I could not have been able to make it without her.

Finally, my parents are the reasons why I have always been trying my best since I was a little kid. A big "Thank you" is obviously not enough to express my love and my gratefulness toward them. Mom, Dad, I love you!

Otaniemi, 20.11.2017

Hoang K. N.

Contents

List of Tables

List of Figures

Symbols and abbreviations

Symbols

Φ	Equivalent ratio
λ_{act}	Actual air-to-fuel ratio
$\lambda_{\mathrm{stoich}}$	Stoichiometric air-to-fuel ratio
Pi	Boost Pressure
P_{CR}	Common Rail pressure
SoI	Start of Injection
rpm	Revolution per minute
$\dot{m}_f$	Fuel flow rate (g/h)
P	Produced work (kW)

Operators

$\nabla f(x)$	gradient of function f(x)
$Hf(x)$	Hessian matrix of function f(x)
$\mathbf{X}^T$	transpose of matrix X
$\frac{\mathrm{d}}{\mathrm{d}t}$	derivative with respect to variable t
$\frac{\partial}{\partial t}$	partial derivative with respect to variable t
$\frac{\partial^2}{\partial t^2}$	second partial derivative with respect to variable t
$\sum_i$	sum over index i
$\mathbf{A} \otimes \mathbf{B}$	Matrix convolution of matrix **A** and **B**

Abbreviations

ECU	Engine Control Unit
VGT	Variable Geometry Turbocharger
HP	High Pressure
BSFC	Brake Specific Fuel Consumption
OP	Operating Point
dBTDC	Degree before top dead center
DoE	Design of Experiments
COST	Change Only one Separate factor at a time
CC	Central Composite design
BB	Box-Behnken design
NOx	Nitric oxide
ppm	Parts per million
SQP	Sequential Quadratic Programming
NLP	Nonlinear optimization problem
QP	Quadratic Programming
RSSE	Residual sum squared error
SST	Sum Squared Total
RSM	Response surface modeling
MSE	Mean squared error
VGT	Variable geometry turbocharger

1 Introduction

Engine calibration has been an important process since the rising of diesel engines in industry. The meaning of engine calibration (as known as engine tuning) is adjustment, modification to the internal engine's actuators or to its control unit in order to yield optimal performance and fuel economy. The need of keeping the engine running at higher efficiency and lower emissions requires the calibration work to be more sophisticated and accurate.

Besides, increasing in the number of controllable engine's parameters leads to a dramatic increase in calibration costs, hence, the new generation of engine calibration process must be capable of handling high number of parameters with reasonable costs. Using static control maps (lookup tables) in which the optimal values of the engine's actuators are contained, has been a very common control strategy in the automotive industry. Finding the accurate values for these maps is therefore really challenging for the manufacturers.

1.1 Thesis objectives

The objective of this thesis is to create static optimal control maps of diesel engines for high efficiency and emission reduction. The calibration tool to be used to create the control maps, named "Off-line parameterization tool", was designed based on the Design of Experiments method. The optimization goal is to minimize the Brake Specific Fuel Consumption (BSFC) of the engine by adjusdting the engine's input parameters and under some emission constraints. The logical structure of the calibration tool was built similarly to the engine calibration process in [1].

The chosen input parameters of the calibration tool are the following:

- Boost pressure.
- Common rail pressure.
- Start of injection.

The outcome and the emission constraints are determined respectively as:

- Optimal static control maps of the three input parameters.
- NO_x emission constraints.

The tool was designed to be able to work both fully automatically and semi-automatically depending on the engine's conditions. Using the Design of Experiments approach helps reducing the calibration process time as well as saving resources.

From a literature review, it can be concluded that several calibration processes using the Design of Experiments methods have been developed over the years. For instance, the Design of Experiments method was used in [2] to optimize the injection strategies of the diesel engines. An extensive application of the method, adaptive on-line Design of Experiments, was used in [3] with the automatic and adaptive identification of the region of interest in the high dimensional parameter space to guarantee the efficiency of the designs with highly non-linear system and irregular shaped valid regions. Moreover, the Design of Experiments was also used in the model-based engine calibration according to [4], [5] and [6].

Even though all the mentioned researches have used the Design of Experiments approach, their implementations in choosing experimental design types and optimization process are different compare to this thesis. The unique aspect of this research lies on the significant properties of the Off-line parameterization tool. Firstly, this tool is flexible, it is able to work with multiple inputs and multiple outputs. Secondly, it can also reduce the calibration time as the engine running time is kept as small as possible and all the data processing work is done automatically.

1.2 Thesis overview

This thesis is divided into five main chapters. The background information about engine working principles and the working cycle of the parameterization tool are presented in Chapter 2. In this chapter, a general introduction to diesel engine such as operating principles, the diesel combustion, the flow diagrams of air and fuel as well as some knowledge about the emissions of the diesel engine are discussed. In addition, in the last part of this chapter, the parameterization tool is described in detail discussing engine control aspects, input-output relationship and the working diagram of the tool.

The Design of Experiments method and other methods used in the tool are discussed in Chapter 3. Chapter 4 and Chapter 5 present the implementation and results of the Off-line parameterization tool on a non-road, turbocharged, common rail and direct injection 44 AWI AGCO diesel engine. Conclusions of the research and some future works are presented in Chapter 6.

2 Background

2.1 Introduction to diesel engines

2.1.1 Diesel engine operating principles

Diesel engine was invented in 1892 by Rudolph Diesel [7]. It operates using a compression ignition process. The diesel engine is different from other gasoline engines in the way of using a high compression of the air to ignite the fuel rather than using a spark plug. The operating cycle of a diesel engine is shown in the following figure 1 [8].

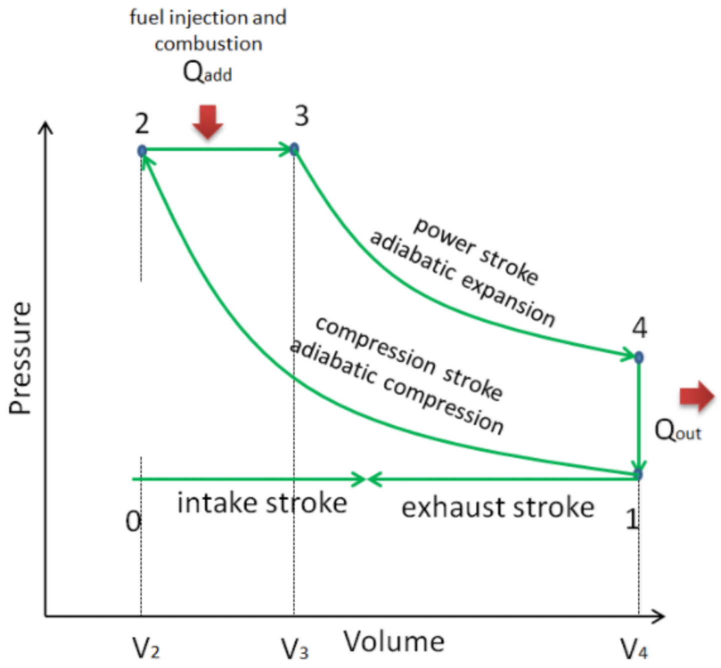


Figure 1: 4-stroke diesel engine operating cycle [8].

There are 4 strokes in an operating cycle of a diesel engine, which are: intake stroke, compression stroke, expansion stroke and exhaust stroke [9].

1. **Intakestroke**: (0 to 1) Atmospheric air after passing through the air filter gets inducted into the engine through the intake valve while the exhaust valve remains closed. This starts at top dead center, then intake valve opens when the piston moves downward and closes when the piston is at bottom dead center.
2. **Compression stroke**: (1 to 2) The stroke starts when the piston is going upward from the bottom position to compress the air-fuel mixture. As both

intake and exhaust valves are closed, the fuel mixture is trapped inside the combustion chamber and is compressed to a fraction of its volume.

3. **Expansion stroke**: (2 to 3 and 3 to 4) the stroke, also known as power stroke, begins when the air-fuel mixture is ignited by the significantly risen heat from the compression stroke. The heat is high enough to auto ignite and burn the fuel and then leads to expanding the air inside the chamber and forcing the piston to go downward. Closing both valves guarantees that all the force is exerted on the piston. The stroke ends as the piston reaches the bottom dead center.

4. **Exhaust stroke**: (4 to 1 and 1 to 0) the stroke begins near the end of the expansion stroke and the exhaust valve is opened. The piston moves upward and pushes the burnt gases from the expansion stroke out of the combustion chamber through the exhaust port. The exhaust valve closes and the intake valve opens when the piston is at the top position again. A new cycle is started with the intake stroke.

Each cylinder of a four-stroke diesel engine completes the aforementioned operation in two revolutions in which intake stroke and compression stroke happen in the first revolution and the other 2 strokes happen during the second revolution. The figure 2 shows a detailed view of the strokes[10].

2.1.2 Diesel combustion

Like most other engines, diesel engine also uses hydrocarbon based fuel. In stoichiometric conditions (perfect amount of air needed for a given amount of fuel) and under assumption that only major products of combustion are formed, fuel undergoes complete combustion [11], yielding carbon dioxide (CO_2), water (H_2O) and unreacted nitrogen (N_2). The following expression [11] explains more details about the relation. The indexes a and b in the reaction are depended on the type of fuel being used for the engine.

$$\mathrm{C_aH_b} + \left(\mathrm{a} + \frac{\mathrm{b}}{4}\right)\left(\mathrm{O_2} + 3.76\mathrm{N_2}\right) \longrightarrow \mathrm{aCO_2} + \frac{\mathrm{b}}{2}\mathrm{H_2O} + 3.76\left(\mathrm{a} + \frac{\mathrm{b}}{4}\right)\mathrm{N_2}$$

It is assumed that the simplified composition of air consists of 21 percent O_2 and 79 percent N_2 (by volume) so that for each mole of O_2 in air, there are 3.76 moles

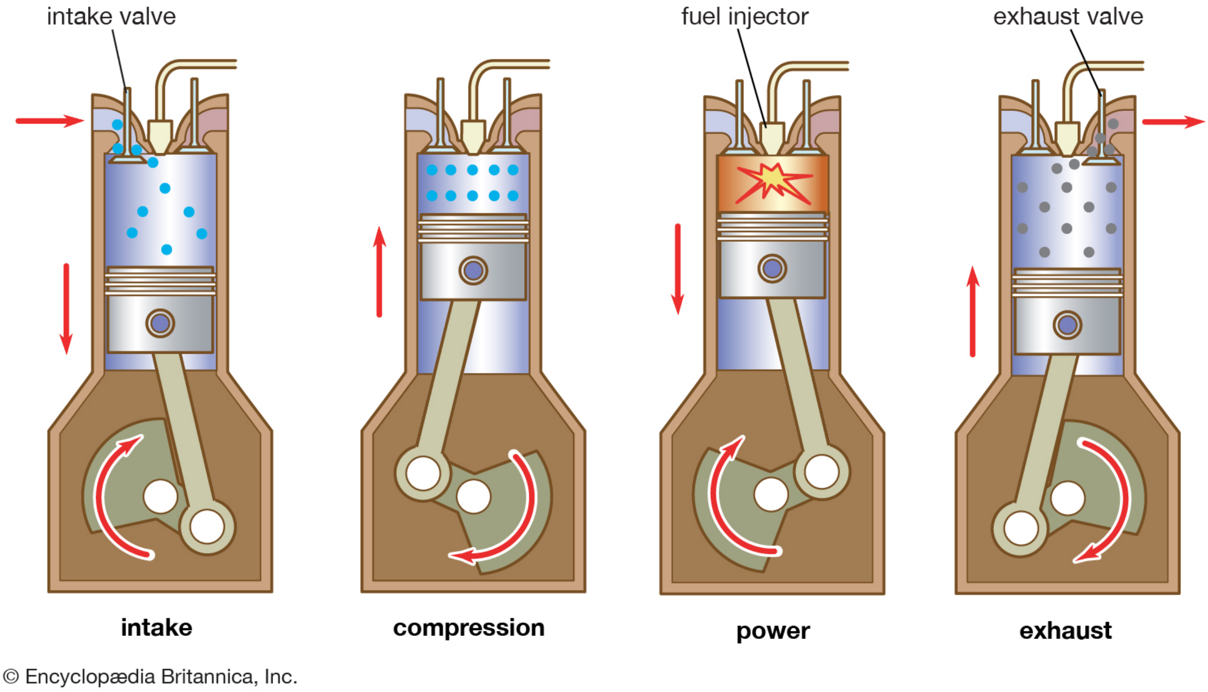


Figure 2: Four-stroke cylinder working diagram [10].

of N_2. Nevertheless, the above reaction does not happen in reality since there are hundreds of elementary reactions that make up the entire combustion process [11].

The stoichiometric quantity of oxidizer (usually referred to air) is a quantity needed to completely burn out an amount of fuel. The fuel is said to be lean if more than a stoichiometric quantity of oxidizer is supplied. On the other hand, if less than stoichiometric quantity is supplied, the fuel is said to be rich.

The air-to-fuel ratio is defined as the ratio between the air mass and the fuel mass injected into the cylinder [12]. It is denoted as $lambda(\lambda)$. There are two kinds of air-to-fuel ratio which are the actual ratio (λ_{act}) and the stoichiometric ratio (λ_{stoich}). There is a parameter called the equivalence ratio $phi(\Phi)$ which is the ratio between the actual air-to-fuel ratio (λ_{act}) and the stoichiometric ratio (λ_{stoich})

$$\Phi = \frac{\lambda_{\text{act}}}{\lambda_{\text{stoich}}} \tag{1}$$

This parameter can be used to distinguish rich mixture and lean mixture of fuel with $\Phi > 1$ representing a rich mixture while $\Phi < 1$ representing a lean mixture. The composition of the combustion using lean mixture is different than the one using rich mixture. Increasing of the fuel injected can create problems with air utilization which lead to excessive amount of soot [11].

2.1.3 Diesel engine flow diagrams

Diesel engine's working principles are based on compressed air, and the previous sections have given some knowledge about the combustion and working cycle inside the engine's cylinders. This section will discuss more the air-flow inside the engine, from the air intake port to exhaust port. Figure 3 describes how air is circulated inside the engine.

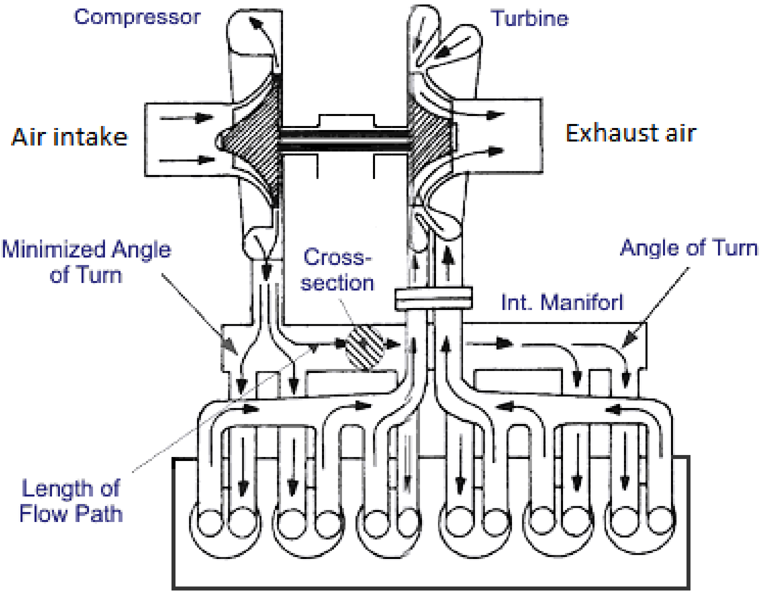


Figure 3: Air-flow diagram of a diesel engine. [13]

Air is compressed from the intake port and transfered to each cylinder. After the combustion, exhausted air goes out through the exhaust air port. The turbine in front of the exhaust port belongs to the turbocharger which is an improvement of the modern diesel engines. The turbocharger uses power of the exhausted air-flow to spin the turbine blades. This set of blades is connected to the air compressor by a rod that makes the compressor wheel spin together with the turbine, and therefore the turbocharger helps compressing air faster with less energy needed for the compressor. Control of the intake air pressure (or so called boost pressure and denoted as P_i) is then depended on control of the turbocharger.

There are several types of turbochargers to be used in engine such as:

- Single-Turbocharger

- Twin-Turbocharger
- Twin-Scroll Turbocharger
- **Variable Geometry Turbocharger**
- Variable Twin Scroll Turbocharger
- Electric Turbocharger

Each of them has both advantages and disadvantages and they are selected to serve different purposes. The one that will be considered and mentioned in this thesis is *Variable Geometry Turbocharger* (VGT). A variable geometry turbocharger is a turbocharger whose turbine is equipped with movable vanes to direct the exhaust air flow onto the blades. Figure 4 shows an example of a VGT [14]

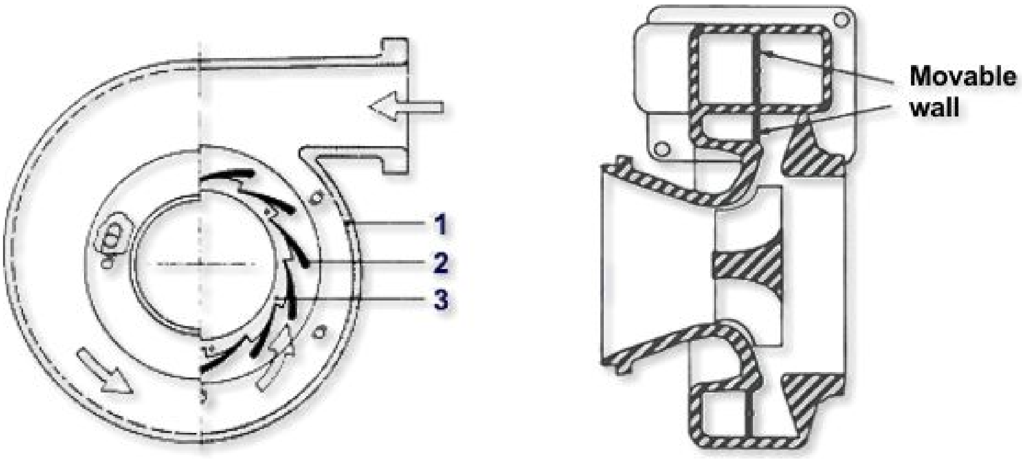


Figure 4: Example of a variable geometry turbocharger: 1. Turbine housing; 2. Variable angle vanes; 3. Adjusting ring [14].

Turbine blades are located inside the ring of vanes. Those variable vanes can rotate to open up or close down the channels between them (as described in figure 5 [14]) for the exhaust air to flow onto the turbine blades with different speeds. The moving angles are adjusted by an actuator. The reason why VGT is needed is that there are situations in which the exhaust air flow is too large, and therefore the exhaust air flows onto the turbine blades needs to be controlled.

In modern engines, fuel is injected to cylinders via common rail fuel injection system. Figure 6 shows an example of this system. It is a direct injection fuel system which includes a high pressure (over 1000 bar) common fuel rail connected to the engine's fuel injectors by separate pipes. The common rail pressure (denoted as P_{CR}) is controlled by a high pressure fuel pump. In each injector, injection timing (start of injection, denoted as SoI) and injection quantity are controlled by a programmable

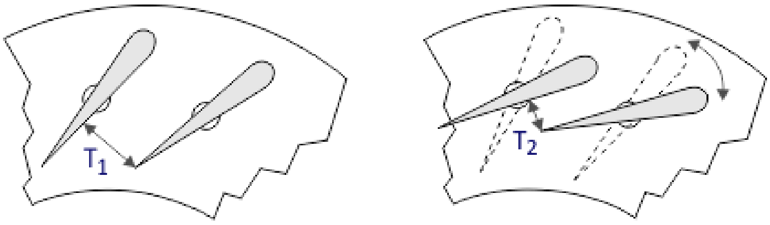


Figure 5: Rotation of variable vanes in VGT [14].

control unit. The injection timing is defined as degrees before the piston inside the cylinder reaches the top dead center.

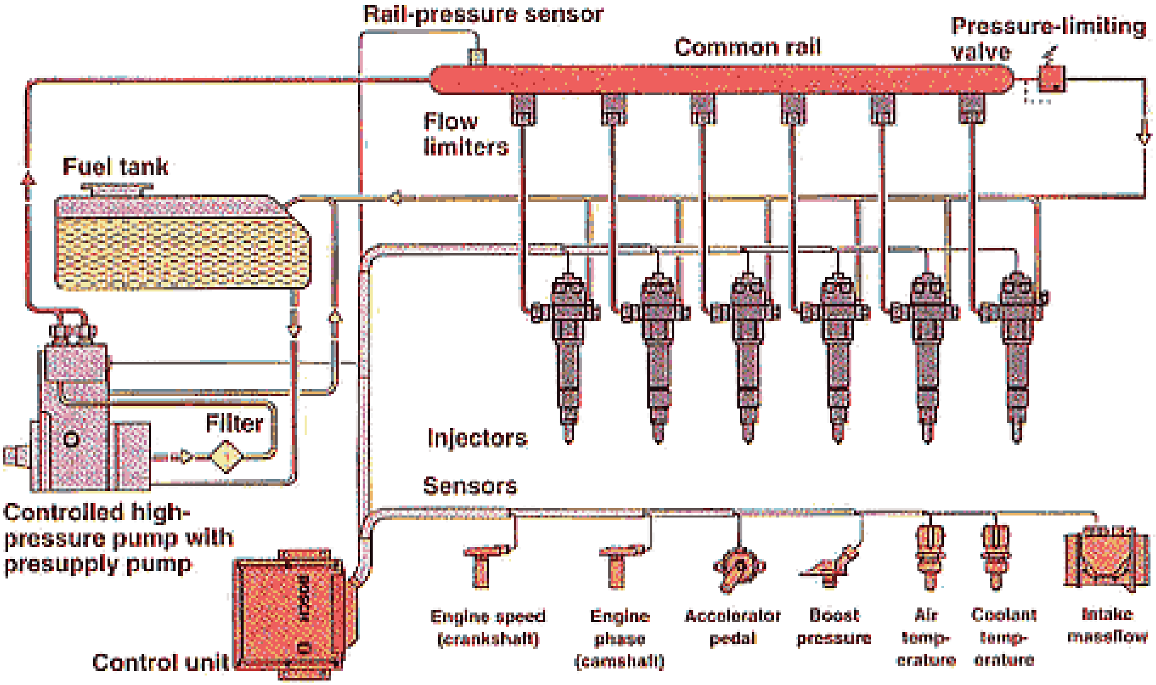


Figure 6: Common rail fuel injection system [15].

The common rail system allows multiple injections at any time so it provides flexibility to exploit and to optimize for a better engine's performance and emissions control.

2.1.4 NOx emissions from internal combustion engines

NOx refers to a mixture of nitric oxide(NO) and nitrogen dioxide (NO_2). According to [12], nitric oxide is the most dominant oxide of nitrogen formed during combustion. The amount of nitric oxide is dependent on the engine design and operating conditions, but usually in the range of 500-1000 ppm or 20 g/kg of fuel [16].

Oxidation of nitric oxide continues and leads to nitrogen dioxide which creates smog when reacting with hydrocarbon in the environment. According to [17], smog and nitrogen dioxide are dangerous as they both create some serious respiratory problems.

NOx is formed in regions in which there is enough available energy for nitrogen to be oxidized. There are several strategies that have been used to reduce NOx emissions in diesel engine. Using of ultra-lean can help reducing the NOx formation since there is less unburnt fuel, however it causes problems with sustainability of the combustion. In automotive engine, use of catalytic converters in the exhaust pipe is necessary.

2.1.5 Particulate matter emissions

In the combustion of diesel engine, incomplete oxidization of fuel can lead to forming soot and particulate emissions. Due to the size of these particles, it is easy for them to go to human's lungs by inhalation and cause serious health problems.

The objective of particulate measurement is to determine the amount of emitted particulate. Measurement devices can be smoke meters or dilution tunnels. The measurement requires a long period of sample collection and careful monitoring as the composition can be easily altered by interacting with surroundings.

2.2 Off-line parameterization tool

2.2.1 Control loops of the engine

In engine systems, there are a number of large control loops for controlling purposes. They are both feed-forward and feedback control systems. These control loops are made to fulfill main objectives such as [18]

- Target torque response, low fuel consumption and drivability.
- Avoiding damages and fatigue of the material by keeping the engine inside its allowed operating region.
- The emissions have to be controlled under limits and follow the legal regulations.

An engine's operating point is defined by its speed and torque. They represent the most important input factors of the control systems. The operating point in turn decides values for all of the engine's main variables [18] such as air mass flow, intake pressure, injection timing, etc. When the engine is fully warmed up, most of the actuator inputs remaining their value in the same operating point. In this thesis, these three following inputs are chosen for the optimization process of the engine: intake pressure (boost pressure), start of injection (injection timing) and common rail pressure (fuel pressure). Figure 7 shows an illustration of a basic feedback loop control system of the engine using the aforementioned inputs. As can be seen from figure 7, at each operating point there are three inputs, one main output and two feedback loops.

The VGT controller gives signals to control the angles of the variable vanes of the turbocharger which affect the boost pressure. Similarly, the HP pump controller sends signal to the high pressure pump to inject pressurized fuel into the cylinders. Sensors measure both pressures (Boost Pressure and Common Rail Pressure) and give feedback signals to the controllers. Unlike the other two inputs, the start of injection input is fed to the engine by feed-forward control.

The brake specific fuel consumption (BSFC) is one of the most important responses of the engine, along with other emission responses. According to [12], break specific fuel consumption represents the fuel flow rate per unit power output and it show the efficiency of using fuel to produce work of the engine.

$$BSFC = \frac{\dot{m}_f}{P}$$

With units,

$$BSFC(g/kW.h) = \frac{\dot{m}_f(g/h)}{P(kW)}$$

Set-points of all three inputs come from sets of data called static control maps. These maps are the most important outcome of this parameterization tool as well as of this thesis. The following paragraphs will give more details and meanings of control maps in the engine.

In the modern Engine Control Unit (ECU) of the engine, a map-based algorithm is usually implemented. Maps are three dimensional data tables which include steady-state optimized results of each actuator inputs at every engine's operating point (speed and load) [18]. It is noted that maps only contain optimized values of actuator inputs when the engine is fully warmed up, hence a correction map of the engine temperature is often added to the control loops to compensate the errors in cold start of the engine. Figure 8 taken from [18] is an example of using correction map along with the optimized map. It can be seen from the figure that, based on the engine speed and a relative load, a nominal spark advanced is chosen (map1),

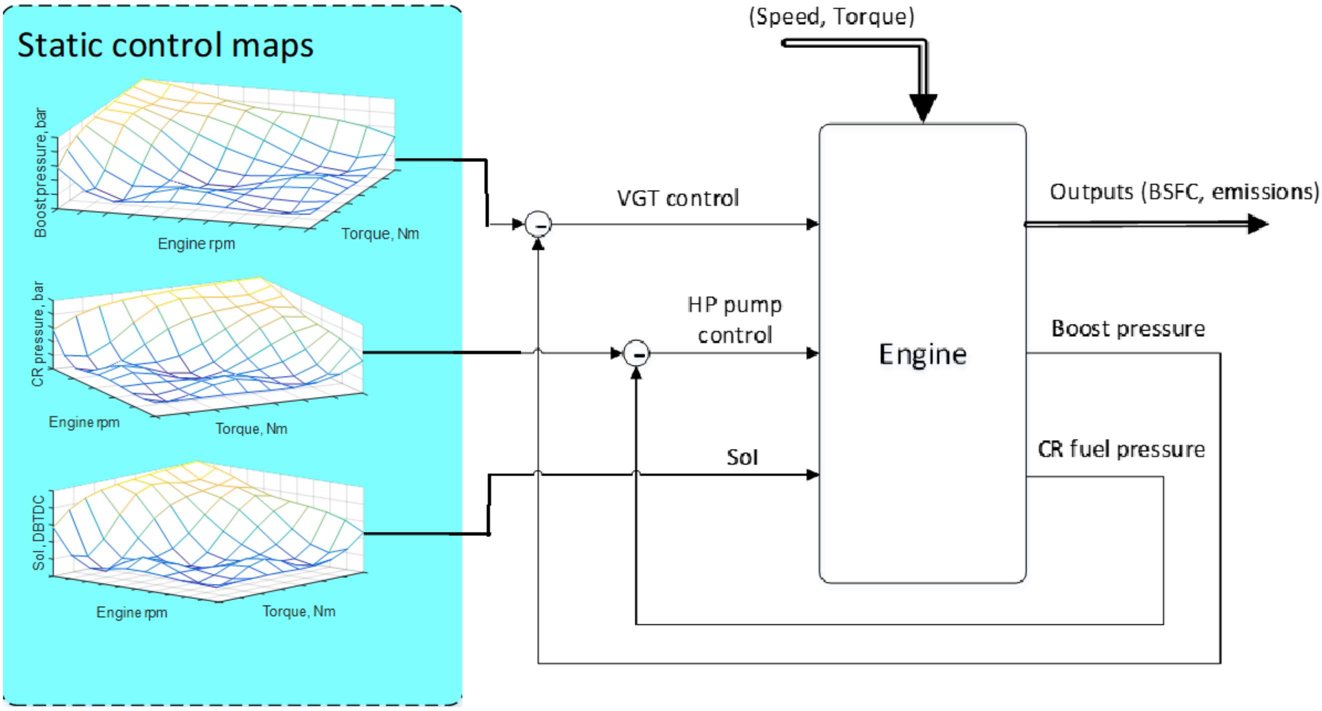


Figure 7: Example of a simple feedback loop control.

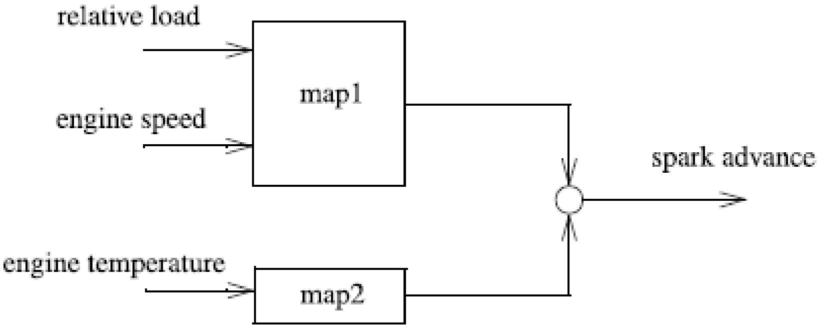


Figure 8: Example of a simple correction map structure[18].

then the correction for engine temperature is applied. In reality, there are a lot more corrections which need to be considered based on thermodynamic principles, but this is only a simple example to understand the application of correction maps.

In industry, instead of using actual torque value, the relative load is usually selected as an independent variable [18]. At each fixed engine's speed, this relative load indicates the actual percentage of air charge in the cylinder. The relative load can be used to derive the actual torque later on with the use of a full-load torque curve[18]. In some cases, the relative load can be substituted by the injection quantity, since the produced torque and the injection quantity are directly proportional. This kind of system is defined as torque-based control structure. The torque is set to a torque demand manager from which it collects and evaluates all demands for engine torque [18]. Then one signal is transferred to a block called *torque conversion manager* in which control signals will be given out to the actuators so that the torque can be "realized as best as possible".

Figure 9 shows an example of a spark advanced map in according to speed and load which is implemented in a modern ECU. The figure is extracted from Fig. 4.6 , page 197 of [18]. The map is defined by the best fuel-efficiency spark advanced with emission limits, knock avoidance, etc. included.

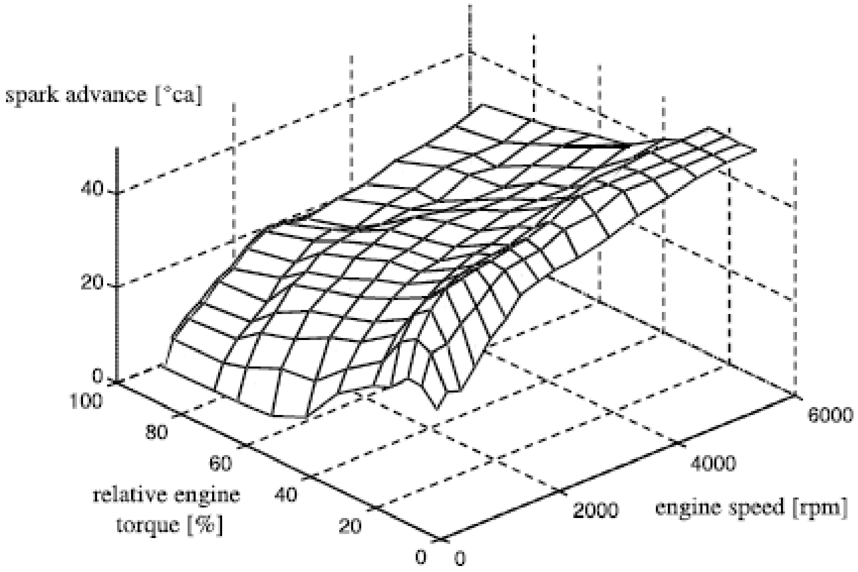


Figure 9: Example of a control map[18].

This thesis proposes a tool called *Off-line Parameterization Tool.* The goal of this tool is that it can be used in semi-automatic engine tuning to generate the optimal set-point maps. Earlier, one common method for engine calibration has been the "brute force" approach. This method requires a lot of time and work force to run many tests in all operating points to investigate the effects of separate variable on the response. What makes this tool different is the introduction of the Design of Experiment (DOE) method, due to its ability to study multiple variable effects on the output at a time rather than one effect at a time. Therefore, this parameterization tool is expected to replace the "brute force" approach in engine tuning to save time and resources. It is also required to provide better optimized engine's responses in comparison to the "brute force" guessing method. At this stage of the development, outcomes of this off-line parameterization tool are defined to be three set-point maps in according to speed and load of:

- **Boost Pressure**
- **Common Rail Pressure**
- **Start of Injection**

The maps consist of optimized values of the three mentioned inputs so that the Brake Specific Fuel Consumption (BSFC) of the engine is minimized under some emission limits. In the following sections, effects of the three inputs on the BSFC and emissions will be discussed in details.

2.2.2 Effects of intake pressure

The intake pressure (boost pressure) plays an important role in diesel engine control. It has big impact on how efficiently the engine is performing and most importantly, the intake pressure also affects the exhaust emissions of the engine.

Compressed air is used to create heat to burn the fuel and high intake pressure can increase the efficiency of fuel combustion. High efficient combustion reduces unburned components so the exhaust emissions can be improved [19]. Higher intake pressure increases the concentration of O_2 to improve the combustion, however, higher intake pressure simultaneously increases the concentration of CO_2 emission due the optimal reaction between C in the fuel and highly concentrated O_2.

Moreover, higher air intake pressure increases the NO_x emission. According to Zeldovich's mechanism [16], the NO_x formation will increase with high pressure and high temperature of the combustion. However, particle mass emission is in fact decreased significantly with high intake pressure [20].

2.2.3 Effects of common rail pressure

Common rail pressure has big effects on the engine combustion quality as the injection pressure affects the fuel spray. Rising the rail pressure to an appropriate range can help to reduce smoke and to increase the fuel economy but at the same time the NO_x emission is increased [21].

Common rail pressure can have different effects upon different working conditions of the engine. Under heavy load condition, too high rail pressure does not make clear improvement on the smoke and fuel economy but the NO_x emission is still increased.

On the other hand, too high rail pressure under light load condition makes the BSFC become to high [22]. On low and middle load, high enough rail pressure can reduce both smoke and NO_x emission.

2.2.4 Effects of start of injection/ injection timing

Start of injection decides when the fuel is injected into the cylinder and kick off the combustion. The timing is defined to be before and after the piston reaches the top dead center. Advanced start of injection happens before the top dead center and retarded start of injection happens after the top dead center. Advanced SoI can result in high in-cylinder pressure, temperature and NO_x emission while retarded SoI results in a reversed trend [23].

According to [24], the use of advanced SoI provides lower soot and higher NO_x emission compared to the the use of retarded SoI. Retarded SoI is commonly used method for effective NO_x reduction. Improving fuel atomization, filling combustion space with fuel spray to well facilitate the air and fuel mixing are fundamental principles for a low NO_x combustion [25].

2.2.5 Working procedure of the parameterization tool

The engine calibration process is classically divided into three big phases [26]:

1. Preliminary phase: choosing a set of operating points to be studied and emissions targets.
2. The optimization of engine responses on each OP under emissions targets.
3. The construction of the maps with smoothening step between optimal settings.

Based on this structure, the off-line parameterization tool's working diagram is made with some modifications as shown in the figure 10

The objective of the calibration is stated again as

$$\begin{aligned} \underset{Boost,CR,SoI}{\text{minimize}} \quad & BSFC \\ \text{subject to:} \quad & emissions \quad constraints \end{aligned} \tag{2}$$

(i) The first step is preparation for the Design of Experiments setup in which the type of experimental design and the optimization variables must be defined. The Box-Behnken design table is chosen due to its simplicity and ability to

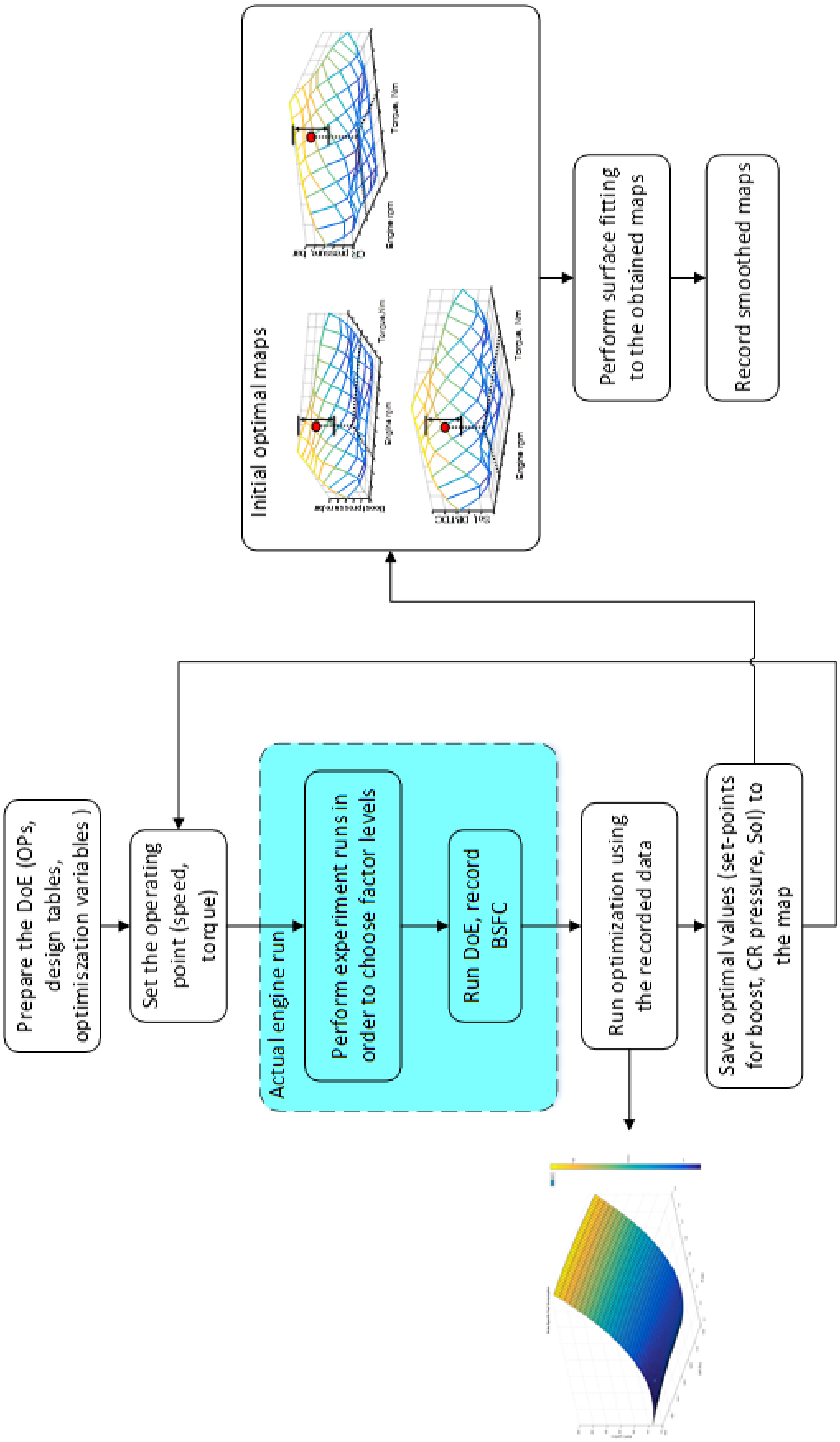

Prepare the DoE (OPs, design tables, optimiszation variables)
Set the operating point (speed, torque)
Actual engine run
Perform experiment runs in order to choose factor levels
Run DoE, record BSFC
Run optimization using the recorded data
Save optimal values (set-points for boost, CR pressure, SoI) to the map
Initial optimal maps
Boost pressure, bar
SoI, DBTDC
CR pressure, bar
Engine rpm
Torque, Nm
Perform surface fitting to the obtained maps
Record smoothed maps

Figure 10: Working diagram of the tool.

produce good experimental data. In Section 3.1, the reason why Box-Behnken design is chosen over other experimental designs will be discussed in details. The optimization variables are coded as

1. Input factors:
- Boost pressure: x_1. (unit: gauge pressure)
- Common rail pressure: x_2. (unit: bar)
- Start of injection: x_3. Unit: degree Before Top Dead Center (dBTDC).
2. Responses:
- The Break Specific Fuel Consumption (BSFC): y_1 (unit: g/kWh)
- Other possible emissions responses such as NO_x and NO responses: y_i

Though there are multiple inputs and multiple outputs, this is not the case of multiple input-multiple output modelling. Each of the output responses is modelled separately based on the inputs.

(ii) The second step is selecting the operating points to run the engine with. The points are chosen in a 300 round-per-minute interval of speed and about 8-16 load points. Interval between points may differ from the engines and should be carefully considered beforehand. In addition, domain of variations of the variables should also be predetermined. The complexity of the fitting model for engine's response depends on the size of this domain. Low order polynomials are usually sufficient to precisely model the response by using a small enough domain. Notice that choosing too small domains leads to difficulty in coherently fulfilling sub-optimal engine maps [27]. This step is also a starting point of a closed-loop process. This loop is run for each operating point, starting from the first one and finishing at the last one.

(iii) The next two steps, which are marked in blue area, require actual engine running. First task involving the engine is validation of the domains which were predetermined in the previous step. Experimenter runs the engine with predefined domains to check whether the upper and lower levels are out of engine's operating range. As all tests are predefined, the experiments can be run automatically if it is safe to let the engine run on itself. In this thesis, the engine's test bed is a 44 AGCO tractor engine in the Combustion Engine Laboratory of Aalto University. Data of inputs and response values are recorded separately for each operating point.

(iv) Modelling and optimization processes can be run right after finishing of all the experiments at each operating point. In case it is not safe to keep the engine run with the tool automatically, experiments of all operating points will be conducted without going to the modelling step. As the data has been recorded separately, modelling and optimization can be done after engine's runs. That is why the tool is called "off-line" as computation can be done without running the engine. The type of mathematical model used in modelling process depends on the complexity of the engine response and on the size of the domain. For this reason, a second order polynomial has been chosen. The model is a function of responses (BSFC and emissions responses such as NO_x and NO) with the three input variables in the form

$$\begin{aligned} y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \quad & \text{(linear terms)} \\ b_{12}x_{12} + b_{13}x_{13} + b_{23}x_{23} + \quad & \text{(interaction terms)} \\ b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \quad & \text{(quadratic terms)} \end{aligned} \tag{3}$$

with x_1, x_2, x_3 represent boost pressure, common rail pressure, start of injection respectively, Y represents each engine's response and $b_0, \ldots, b_{33}$ are regression coefficients of the model. The model has all three linear terms of the inputs plus interaction terms between each two of them and finally quadratic terms of each of them. The quadratic terms also ensure curvature in the response. Optimizing can be formulated as a classical mathematical problem of optimization under constraints. In this approach, the optimization is performed at one OP after the other, considering the responses of emissions for each OP as constraints.

(v) When optimization is done for all of the operating points, optimal values of each input factor are saved to initial optimal maps similar to the one in figure 11. They are just scatter plots of all the optimal settings and the next step

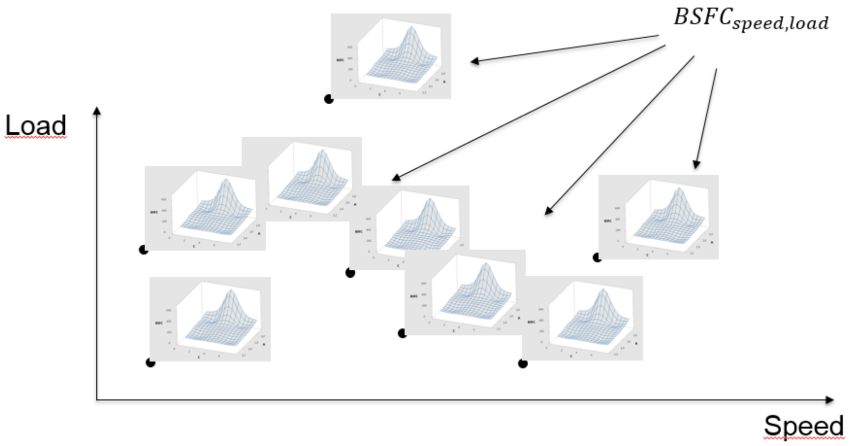


Figure 11: Initial optimal map with local optima.

is building a final map on the whole engine operating domain based on those values. Several fitting methods can be applied to build the map, such as *Robust Locally Weighted Regression* [28] method. Notice that the selected number of operating points has a big impact on the map's resolution and accuracy. The denser the map is, the more points is needed. Figure 12 shows an example of a map on the whole engine operating range.

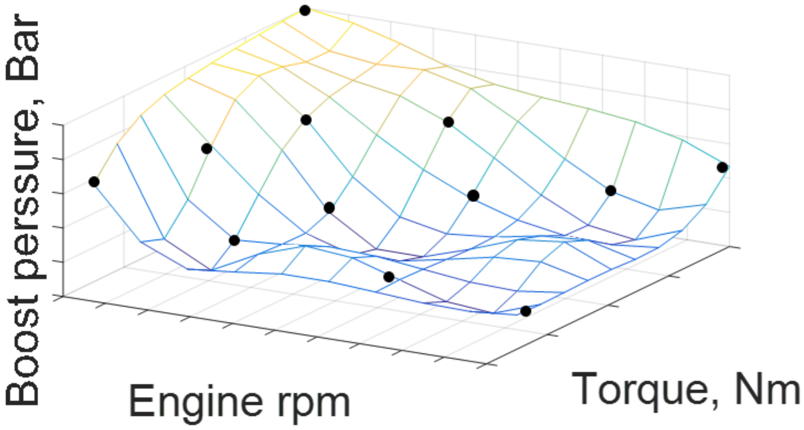


Figure 12: Example of a map on the whole operating range.

(vi) One more step must be done before the created maps can be used. The final phase of the tool is to smooth the set-point maps. Since sharp evolution of air loop parameters are not feasible during transient[1], the maps need to be smoothen to avoid rough transitions between operating points. Hence, the optimal points are often shifted away from their locations during the smoothing process. The goal is to remain as close as possible to the local optima while keeping a smooth shape of the map.

3 Research material and methods

3.1 Introduction to the Design of Experiments method

3.1.1 Preface

Experiments are used in almost every area to solve problems in an effective way. They can be found in daily life or in industry or in scientific area. Experiments are in principle comparative tests, they are used to compare between alternative options. It can be comparing the yield of a certain process to a another to prove the effect of changes made or comparing performance of an automated process with manually controlled process.

Experiments are also used in science to find significant information about a studied object. Objects can be chemical mixture or performance of engines etc. Systematic experiments can also be applied to some processes aiming to optimize their performances. In this case, available operating ranges are required in advance and experiments are designed so that when using them with some mathematical software, the optimum operation points can be achieved. Inspecting performances of an engine can be a good example as doing experiments can lead to finding suitable parameters at every operation point of the engine to improve its fuel consumption as well as to reduce emissions.

3.1.2 Meaning of Experiments

An experiment is defined as an observation that leads to characteristic information of an object. The classical purpose for such an observation is to verify a hypothesis with an investigation. The experiment setup is selected for the particular problem statement, and the experimenter tests whether the hypothesis is true or false [29].

With the concept of Design of Experiments (DoE), a set of well selected experiments is performed. The purpose of the design is to optimize a process by performing experiments orderly and systematically so that from the result of the experiments, conclusion about the significant character of the studied object is produced.

By using DoE, the number of experiments is kept as low as possible and the most informative combination of the factors is chosen [30]. Hence using DoE is an effective, economical and time-saving method. The following section gives definitions and terms used in the method.

3.1.3 Basic definitions

There are always two types of variables when experiments are performed in DoE, factors and responses. The responses can be understood as output giving information about the behavior of the studied object. Factors can be understood as inputs, which are used to manipulate the output. Factors can have two or more values and always lie in a defined range. A system or process can be manipulated by one or more different input factors. The changes put effects on the output responses and those effects can be measured. Following figure 13 shows the relationship between factors and responses to a certain system or process. Factors can be divided into 3 groups: controllable and uncontrollable factors, quantitative and qualitative factors, process and mixture factors

Inputs
Process/Product
Response

Figure 13: Factors and response relationship.

Controllable and uncontrollable factors

This is the first way to differentiate types of factors, dividing them into controllable and uncontrollable ones. Controllable factors are process inputs which can be easily monitored and investigated. These inputs can be changed during the experiments by experimenter. On the other hand, uncontrollable factors are hard to regulate since they are mostly disturbance values or external errors. They can have very high impact on the response and therefore they should always be considered during the experiments [30]. In the diesel engine case, controllable factors can be input pressure, fuel pressure and start of injection. Meanwhile uncontrollable factors can be measurement device's errors or mechanical characteristics of the engine itself.

Quantitative and qualitative factors

Another way to split the types of factors is separating them into qualitative and quantitative factors. Quantitative factors always take values from a given range with a continuous scale, meanwhile qualitative factors only have fixed values [31]. In diesel engine, the three aforementioned factors, start of injection, input pressure and fuel pressure are all quantitative factors.

Process and mixture factors

Final group of factors is process factors and mixture factors. Process factors are independent factors and they can be changed without affecting other factors in the same experiments, whereas mixture factors can be understood as amounts of ingredients in a mixture and they all add up to 100%. Since mixture factors are dependent on each other, they can not be changed independently [30].

3.1.4 Basic principles of DoE

The experiment design in the old time was usually done such that changes are made at one variable at a time; i.e. firstly the first variable is changed and its effect is measured and then the same procedure takes place for the second variable and so on. As [30] describes, the intuitive approach is to "change the value of one separate factor at a time until no further improvement is accomplished". This method is so called COST (*Change Only one Separate factor at a Time*) and was used to find the optimum. Nevertheless, it is very difficult for the experimenter since he/she does not know at which value the changing of a certain variable should be stopped due to the inability to observed any further improvement, and finding exact value of that factor is very crucial considering it in combination with other changes of other factors.

The same process can be investigated with the use of DoE and in this case a uniform set of experiments is created around a center-point. Changes are now done simultaneously and systematically according to a program decided beforehand. The following figure 14 from [30] illustrates the difference between the two approaches.

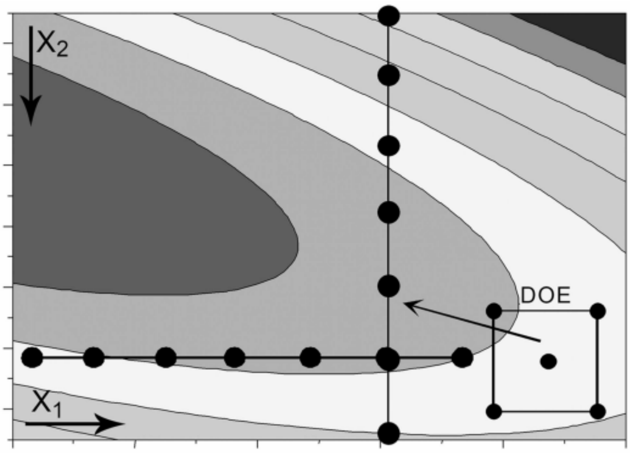


Figure 14: COST approach & DoE[30].

All factors are changed simultaneously, as shown in the low right corner. This method provides better information about the optimum of the response than result of the COST approach in which all factors are changed successively.

The basic concept of DoE is to arrange a symmetrical distribution of experiments around a center point and within a given range of input factors providing that the calculation of this center point is possible. Figure 15 shows an example of a design with three factors x_1, x_2, x_3. The three factors have their ranges and the center points are calculated. The factors are placed in a cubic pattern with an aformentioned center point in the middle.

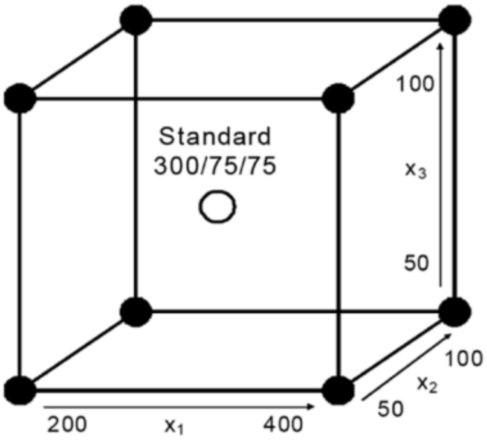


Figure 15: Symmetrical Distribution of Experiments[30].

Objectives of Experiments The purposes of designing sets of experiment are divided into two main types of designs.

1. **Screening:** A screening design is performed to characterize a process at the beginning. Its purposes are to determine the main factors and inspect the changes of responses by varying each factor. This design is meant for processes with large number of input factors and is useful for later optimization work since the experimenter only has to work with a subset of fundamental input factors.
2. **Optimization:** Screening process is followed by optimization process. It gives detailed information about effects of the chosen inputs factors to determine the best combination of factors. In other words, an optimization process is used to find the optimum point by estimating response values for all factor combinations. Response Surface Modeling (RSM) is one of the usually used methods to estimate interactions and effects between factors so that the experimenter can have an idea of the investigated response. Further details about RSM will be discussed later in Statistical Design section.

Model Concept: DoE is based on approximation of interactions and effects with the help of a mathematical model. Fundamental aspects of the studied process are represented by factors and responses. A model can not be perfectly correct but it can be helpful to transport the complexity of the process into a mathematical equation which is easy to handle [30].

The simplest approach to be used is a linear model with the n factors x affecting one response y like the following equation:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_\mathrm{n} x_\mathrm{n} + \epsilon \tag{4}$$

in which β_0, β_1,..., β_n are regression coefficients and ϵ is the model error, which is assumed to be normally distributed. In addition, the regression coefficients can be found by the help of a mathematical tool which will be described in Section 3.2. The equation (4) can be extended to N multiple responses as:

$$y_\mathrm{i} = \beta_0 + \beta_1 x_\mathrm{i1} + ... + \beta_\mathrm{n} x_\mathrm{in} + \epsilon_\mathrm{i}, i = 1, .., N \tag{5}$$

in which y_i represents ith response with the corresponding factors x_i1, x_i2,.., x_in. It can also be notated in a matrix form as:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1n} \\ 1 & x_{21} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & \dots & x_{Nn} \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_N \end{pmatrix}$$

The input factors x_i can be selected and represented in different ways depending on the type of design such as x_i being a quadratic form of another variable. The experimenter should be careful on choosing the type of design in order to acquire the best results. A further discussion about different types of design will be shown in the next section.

3.1.5 Matrix Designs

Matrix design is the way in which the experimenter systematically arranges the experiments in order. In each type of design, the order and the number of experiments are different to serve different purposes of the experimenter.

The following example illustrates a simple way of arranging experiments for a case of three input factors. The factors are A, B and C which have different ranges. Firstly, limits of the ranges are denoted as (+)/(-) or (+1)/(-1) according to maximum and minimum values. Medium values are usually denoted by the value 0.

Example.
Factor A:
(-) level is 100; (+) level is 150
Factor B:
(-) level is 5; (+) level is 10
Factor C:
(-) level is 0.1; (+) level is 0.5

In this case, the number of experiments is simply decided by the number of combination between the three factors with two levels of each factor. Hence it makes a total number of 8 (2^3) experiments needed to be run. The design matrix then can be written as:

Run number	A	B	C
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

So in the first run, factor A, B and C are kept at (-) levels and so on, their levels are changed simultaneously through 8 runs. In the following sections, matrix designs of different types will be discussed in detail.

3.1.6 Full Factorial Design

Full factorial design means that in these designs, possible combinations of all factors at every level are included. There can be more than two levels for each factor (medium level can be considered), but the number of levels can create a big influence on the number of neccessary experiments. For a simple case of 2 factors with 2 levels, there are $4(2^2)$ experiments needed for a full factorial design. If there are k factors with 2 levels of each, the number of runs is 2^k.

The aforementioned example in the previous section is a case of full factorial design with 3 factors and 2 levels of each factor. Hence there are 8 experiments needed. After all experiments are performed, a regression analysis of 2^3 factorial experiment can fit to the following model:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_{12} + b_{13}x_{13} + b_{23}x_{23} + b_{123}x_{123} \tag{6}$$

where the b_i are regression coefficients. The b_0 is referred to the model's constant; the b_1, b_2, b_3 are called main effects; the b_{12}, b_{13}, b_{23} are two-factor interactions and b_{123} is three-factor interaction. The three variables x_1, x_2, x_3 are the three corresponding factors A, B and C. Accordingly, x_{12}, x_{13}, x_{23} and x_{123} are interactions between the three factors. The interactions between factors are defined as the multiplication of their values.

In the full factorial design category, the 3^k factorial design is a special one. It is constructed by k variables with each variable having three evenly spaced levels and all combinations of levels are used [32]. This design guarantees to provide a full quadratic model due to the appearance of center levels, however, it requires so many runs. The number of runs increases exponentially with the power of k, for example with three input variables it needs 27 (3^3) run in total. For this reason, other designs with fewer runs and yet providing full quadratic models are usually used instead of 3^k factorial designs.

3.1.7 Fractional Factorial Design

Fractional factorial designs include only the most important combinations of the variables. It is very useful in cases which have many input factors in order to avoid exponential explosion. Reducing the number of runs can be beneficial for the experimenter.

3.1.8 Response Surface methods

The two-level factorial designs in the previous sections provide a powerful set of experiment design for studying complex responses; however, they are not capable of detecting curvature in the responses. Therefore they have limitations in optimization. A weakness of these designs is that they use only two levels of each variable. Adding center points can help detecting curvature but it also creates more runs and hence consumes more time and energy of the experimenter. Response surface designs are capable of resolving curvature in the response associated with each variable and guarantee a minimum of experiments needed. The power of this design is the use of quadratic terms added to their models such as:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_{12} + b_{11}x_1^2 + b_{22}x_2^2 \tag{7}$$

This equation defines the response surface, i.e. how y depends on x_1, x_2, which can be presented as a surface in a multidimensional graph.

In the response surface methods, there are two main types of designs which can be applied in practice. They are Central Composite design and Box-Behnken design.

a. Central Composite Design:

This design has a basis of the two-level factorial designs. In factorial designs, adding center cells is a good method to give them the ability to search for curvature, but it is incomplete [32]. Extra runs can be added to these designs to give them capabilities of quadratic modeling. The model which is used to fit in the case of three factors is:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_{12} + b_{13}x_{13} + b_{23}x_{23} + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \tag{8}$$

where the square terms are the source of quadratic effects.

The runs that must be added beside the two-level plus the center points fall at extreme points outside the normal limits -1/+1 of the two-level design. These extreme points are called star points. The distance from the center of the design to the star points is denoted as η. Value of η is depended on the number of points in the factorial design. Hence η is given by [32]:

$$\eta = n_{\text{cube}}^{1/4} \tag{9}$$

where n_{cube} is the number of points in a single replicate of 2^k design. Two star points are added to the original experiment for each variable, one at the -η level and the other one at +η level, while all other variables are held at their zero level [32] (medium level). This adding gives the central composite designs five levels of each variable: -η, -1, 0, 1, +η. Figure 16 shows an illustrative view of combination of 3 variables using central composite design [33].

Example. Following example illustrates the application of central composite design in a three-factor case. There are 8 points in the factorial design and hence

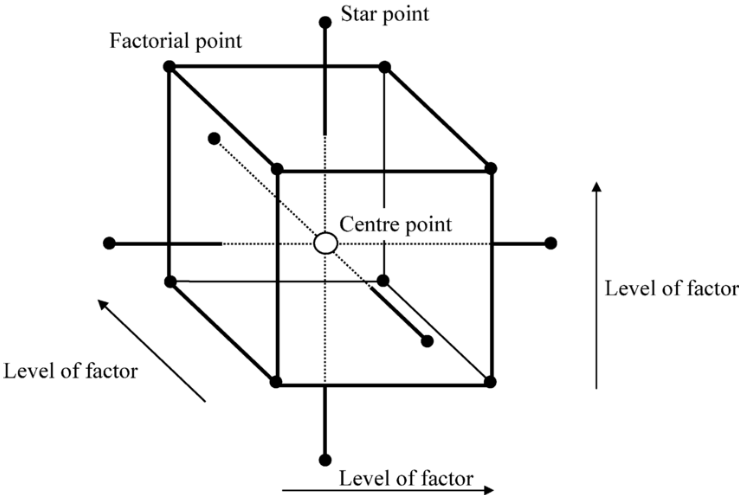


Figure 16: Central composite 3-variable design [33].

the value of η can be calculated as:

$$\eta = 8^{1/4} = 1.68 \tag{10}$$

The star points are located at -1.68 and 1.68 if the center is assumed to be zero (0). Table of runs can be described as shown in table 1.

There are 8 runs with the normal level ±1, plus 6 runs with zero level and 2 extra runs for each variable with their star values. It makes a total of 20 runs, while adding center levels to the factorial design makes a total of 27 (3^3) runs. Hence using central composite designs really reduces workload and saves time.

b. Box-Behnken design:

The Box-Behnken (BB) design is an independent quadratic design and a member of 3^k family of designs. This design are fraction of the 3^k designs with center points added to keep the balance of the design [32]. In [34], an original design was introduced for up to twelve variables. The primary idea of this design is about the location of the experimental boundaries and avoidance of extreme combinations [32]. Hence, Box-Behnken design omits all the corner points, and the star points which were included in central composite designs. By avoiding all the corner points and star points, BB design prevents the values from going beyond the low and the high limit. Figure 17 shows an illustrative view of factor combination of the BB design in the three-factor case [35]. All points are now in the middle of edges and in the center space. There are no more points at corners or star points.

Table 2 describes the matrix for BB design with three factors. In each noncenter

Table 1: Table of run for the CC 3-variable experiment.

CC (2^3)			
x_1	x_2	x_3	**No. of runs**
±1	±1	±1	8
0	0	0	6
±1.68	0	0	2
0	±1.68	0	2
0	0	±1.68	2
Total runs			20

$=$

Std	x_1	x_2	x_3
1	-	-	-
2	-	-	+
3	-	+	-
4	-	+	+
5	+	-	-
6	+	-	+
7	+	+	-
8	+	+	+
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	-1.68	0	0
16	+1.68	+	-
17	0	-1.68	+
18	0	+1.68	0
19	0	0	-1.68
20	0	0	+1.68

row (row with ±1 levels in it), there is always an experiment that resolves each two-factor interaction but another variable still remains at its zero level [32]. The center cells are added to the design to fulfill the requirements to resolve the quadratic terms. There are 12 runs with ±1 levels and only three runs with zero levels, hence the total number of experiments is significantly reduced to only fifteen. It is much less in comparison to the number of runs in Central Composite designs.

Data recorded from the experiments is fitted to the same model which was used in CC designs

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_{12} + b_{13}x_{13} + b_{23}x_{23} + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \quad (11)$$

The function includes three single terms for main effects, plus three two-factor interaction terms and three quadratic terms.

c. Comparison of the response surface designs:

Since both Central Composite and Box-Behnken designs are able to provide models

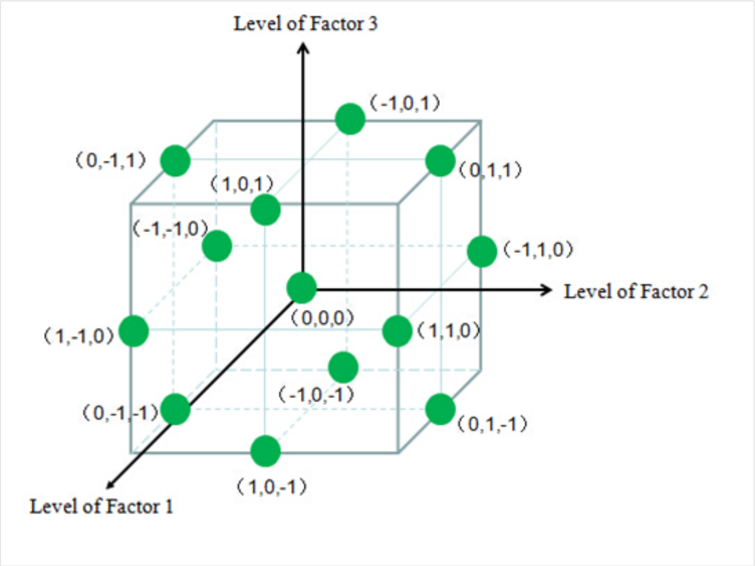


Figure 17: Box-Behnken 3-variable design[35].

Table 2: Table of run for the BB 3-variable experiment.

BB (3)			
x_1	x_2	x_3	**No. of runs**
±1	±1	0	4
±1	0	±1	4
0	±1	±1	4
0	0	0	3
Total runs			15

=

Std	x_1	x_2	x_3
1	-	-	0
2	-	+	0
3	+	-	0
4	+	+	0
5	-	0	-
6	-	0	+
7	+	0	-
8	+	0	+
9	0	-	-
10	0	-	+
11	0	+	-
12	0	+	+
13	0	0	0
14	0	0	0
15	0	0	0

with main effects, two-factor interactions and quadratic terms, other criteria must be considered to decide which design will be used [32]:

- The number of observations and number of error degrees of freedom in the design. The error degrees of freedom (df_ϵ) are defined by the difference of the total number of points in the data set and the number of coefficients in the regression model.

- The number of level required for each variable.
- The safety of te highest and lowest variable levels.

Models that have less than about eight error degrees of freedom are considered to be risky while ones those have more than about twenty degrees are considered to be wasteful [32]. Table 3 [32], is a summary of the number of runs and error degrees of freedom for both CC and BB designs.

In this table, N is number of required experiments. Total degrees of freedom df_{total} is the total number of points in the data set after running the experiments. Lastly, model degrees of freedom df_{model} is the number of coefficients in the regression model.

According to this table, of the three variable experiments, BB(3) design is very efficient with 5 error degrees of freedom compared to 10 degrees of CC(2^3) design. Although BB seems to be short on error degrees of freedom, most of the experiments have terms that can be dropped out to improve the estimation of the error. Therefore, Box-Behnken designs are mostly used in three variable experiments.

Table 3: Comparison of response surface methods

Design	**N**	df_{total}	df_{model}	df_{ϵ}
3^2	9	8	5	3
CC(2^3)	13	12	5	7
3^3	27	26	9	17
BB(3)	15	14	9	5
CC(2^3)	20	19	9	10

The next reason why BB designs are chosen over CC designs lies on the number of levels of each variable. While BB designs require only three levels from each factor, CC designs need five levels from each factor (max, min, center, star points). In some cases, getting all five needed levels for each variable is difficult or even impractical[32].

The last criterion is about the safety of variable levels. It is stated that if the levels of a variable are selected too far apart then there is possibility that one or both extreme levels will be lost. When safe limits are not known then Central Composite designs are very good selection because of the appearances of their star points. Those points can be placed in unsafe area to keep the other points in safety since the two-factorial plus the centers can still be analyzed despite the lost of those star points. In figure 16 it can be seen that if all star points are lost then the response can still be analyzed inside the cube formed by the other points. In the engine optimization problem it is assumed that the limits of the engine's parameter can be found out by running tests, and therefore the Box-Behnken designs are still safe to use.

In summary, it can be concluded that choosing Box-Behnken designs for the experimental setup can save time and resources by using less runs yet still guaranteeing to provide good quadratic models.

3.2 Data fitting method: Least squares

The next process after running the design of experiments is creating a mathematical model to fit the experimental data. The model of each operating point of the engine will then be used to predict the engine's performance at that point. There are several methods which can be used to fit the data; the one that was chosen is, the *Least Squares Regression method*.

This section is going to give a theoretical view about application of the Least Squares method as well as the statistical testing procedure for the fitted results. Fitted models must satisfy some required statistical tests to assure that the predictions on the engine's model do not go wrong. Since the parameterization tool is developed in MATLAB, implementation of the whole fitting and testing procedure is carried by the help of MATLAB fitting toolbox which uses the similar principles.

According to previous sections, the idea of the parameterization tool is that at each operating point of the engine, one set of the designed experiments will be conducted and the obtained data will be treated and fitted into a mathematical model with the form

$$Y = \beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_{12} x_1 x_2+\beta_{13} x_1 x_3+\beta_{23} x_2 x_3+\beta_{11} x_1^2+\beta_{22} x_2^2+\beta_{33} x_3^2 \quad (12)$$

The number of models to be fitted is dependent on the number of operating points chosen in advance. The more points the more accurate when it comes to creating final control maps, however, too many points require a noticeable amount of time to conduct all the experiments. The issue on selecting the number of operating points will be discussed in the result section.

3.2.1 Introduction

In general, a least squares problem is an unconstrained minimization problem
Definition 1[36] Find β^*, a local minimizer for

$$F(\beta) = \sum_{i=1}^{m} (f_i(\beta))^2, \tag{13}$$

where $f_i : R^n \mapsto R, i = 1, \ldots, m$ are given functions, and $m \geq n$.
According to [36], in a least squares fit parameters are determined to minimize the sum of squared residuals. In other words, the coefficients needed for the model are the values that make the sum of squared errors minimized.

To make it clearer, let the equation (12) be an example. This model depends on the parameters $\beta = [\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_{33}]^T$. At each operating point, the obtained data includes input values x_{1i}, x_{2i}, x_{3i} and output values y_i with $i = 1, \ldots, 15$ (Box-Behnken design has 15 experiments for each design). Hence the residual function can be written as

$$f_i(\beta) = y_i - Y(\beta, x_{1i}, x_{2i}, x_{3i}), \quad i = 1, \ldots, 15. \tag{14}$$

Now the problem becomes least squares problem which is finding β so that it minimize the function

$$F(\beta) = \sum_{i=1}^{15} (f_i(\beta))^2, \tag{15}$$

Although the least square model includes non-linear terms of the independent variables x_1, x_2, x_3, it is still linear in the parameters since the variable β appears linearly. It can be rewritten in linear form as

$$\begin{aligned} Y = \beta X, \quad X &= [1, x_1, x_2, x_3, \ldots, x_3^2]^T \\ \beta &= [\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_{33}] \end{aligned} \tag{16}$$

Therefore, this problem can be solved by basic linear calculus and the method will be discussed in the next section.

3.2.2 Basic Calculus method

Data and Matrix Notations:

The method being presented here is applied at one operating point of the engine, hence, there are 15 cases for observed data which includes output values Y and all

of the terms x_1, x_2, x_3. A set of 10 terms including intercept from equation (12) can be rewritten for easier computation as $X = [1, x_1, x_2, x_3, \ldots, x_9]^T$.
Symbols for the response and the terms of all observations using matrix notation can be written as

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{15} \end{pmatrix} \quad , \quad X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{19} \\ 1 & x_{21} & \ldots & x_{29} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \ldots & x_{n9} \end{pmatrix}$$

So Y is a 15x1 vector and X is a 15x10 matrix. In addition, β is denoted as a 10x1 vector of regression coefficients and **e** is denoted as a 15x1 vector of statistical errors.

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_9 \end{pmatrix} \quad , \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_{15} \end{pmatrix}$$

The multiple linear regression model has the form of

$$Y = X\beta + \mathbf{e}. \tag{17}$$

Variance-Covariance matrix of e:

The 15 x 1 error vector **e** is an unobservable random vector and it is assumed to have the following properties [37]

- Zero mean: $E(\mathbf{e}) = 0$.
- $\mathrm{Var}(\mathbf{e}) = \sigma^2 \mathbf{I}_n$

where Var(**e**) means covariance matrix of **e** and $\mathbf{I}_n$ is a 15 x 15 identity matrix.

Ordinary Least Squares Estimator:

The main idea of least squares estimator is to find β to minimize the residual sum of squares function in equation (15). It can be rewritten with new matrix notations as

$$RSSE(\beta) = \sum_{i=1}^{15} (y_i - x_i'\beta)^2 = (Y - X\beta)'(Y - X\beta) \tag{18}$$

in which x_i' and y_i are the i^{th} row of matrix X and i^{th} element of vector Y. The estimator can be found by using theories about finding local minimum point using derivatives of RSSE function with respect to β [37].

- First derivative of RSSE with respect to β:

$$\frac{\delta RSSE}{\delta\beta} = -2X^T(Y - X\beta) \tag{19}$$

- Second derivative with respect to β:

$$\frac{\delta^2 RSSE}{\delta\beta\delta\beta^T} = 2X^T X. \tag{20}$$

Set the first derivative to zero and solve for β, under assumption that columns of X are linearly dependent, gives

$$X^T(Y - X\beta) = 0 \tag{21}$$

Hence the normal equation is of the form

$$X^T X\beta = X^T Y. \tag{22}$$

Solve for β under the assumption that $X^T X$ exists:

$$\beta = (X^T X)^{-1} X^T Y. \tag{23}$$

Nevertheless, using equation (23) to compute for β can be inaccurate since the terms $X^T X$ and $X^T Y$ are matrices of uncorrected sums of squares and cross-products. Using uncorrected sum of squares and cross-product can possibly lead to large rounding error, and so computations can be highly inaccurate. According to [37], one alternative method can be used which is based on matrix decomposition and computations are based on corrected sum of squares and cross-products. Firstly, matrix X is redefined which excludes its first column and the column mean is subtracted from each of the remaining columns.

$$\mathcal{X} = \begin{pmatrix} (x_{11} - \bar{x_1}) & \dots & (x_{19} - \bar{x_9}) \\ (x_{21} - \bar{x_1}) & \dots & (x_{29} - \bar{x_9}) \\ \vdots & \vdots & \vdots \\ (x_{n1} - \bar{x_1}) & \dots & (x_{n9} - \bar{x_9}) \end{pmatrix}$$

Similarly, $\mathcal{Y}$ is the vector with each element being subtracted by the mean of y as

$$\mathcal{Y} = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ y_3 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}$$

Let β^* be the coefficient vector excluding the intercept β_0, then

$$\hat{\beta^*} = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\mathcal{Y} \tag{24}$$

$$\hat{\beta_0^*} = \bar{y} - \hat{\beta^*}^T\bar{x} \tag{25}$$

where $\bar{x}$ is the vector of sample means for all the terms except for the intercept.

3.2.3 Modelling Error

- If the observation Y has a variance σ^2, then

$$Var(\hat{\beta}) = (X^TX)^{-1}\sigma^2 \tag{26}$$

- The variance (mean squared error) of the regression is

$$MSE = \frac{RSSE}{n-p-1} = \frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{27}$$

  where n is number of experiments, p is number of independent variables in the regression model and $\hat{y_i}$ is the estimated value of the response.

- The standard deviation of the error

$$\hat{\sigma} = \sqrt{MSE} = \sqrt{\frac{RSSE}{n-p-1}} \tag{28}$$

- The coefficient of determination R^2 is one of important factor to show how well data is fitted to a statistical model. [32]

$$R^2 = 1 - \frac{RSSE}{SST} \tag{29}$$

in which

$$RSSE = \sum_{i=1}^{n}(\hat{y_i} - Y)^2, \quad \hat{y_i} \quad \text{is predicted value,} \quad Y \text{is measured value} \quad (30)$$

$$SST = \sum_{i=1}^{n}(\hat{y_i} - \bar{Y})^2, \quad \hat{y_i} \quad \text{is predicted value,} \quad \bar{Y} \text{is mean value of Y} \quad (31)$$

The closer of R^2 to 1, the better of the fitting result. Small R^2 indicates bad fit of data.

- In a more complex model, as more and more independent variables are carried, the R^2 values will always increase. Hence, it is necessary to compensate the complexity of the model by using a new coefficient of determination, called the *adjusted coefficient of determiantion* $R^2_{adjusted}$ [32]

$$R^2_{adjusted} = 1 - \frac{df_{total}RSSE}{df_{\epsilon}SST} \quad (32)$$

$R^2_{adjusted}$ is always smaller than R^2 and is safer to use to evaluate a complex model.

- The *t values* for the regression coefficients are calculated by dividing each coefficient by its standard deviation[32]. According to [32], "the *t value* indicates how much that the coefficient is of its standard deviations greater than zero".

$$t = \frac{\beta}{s_\beta} = \frac{\beta\sqrt{n}}{s_\epsilon} = \frac{\beta\sqrt{n}}{\sqrt{\frac{RSSE}{df_\epsilon}}} \quad (33)$$

in which df_ϵ is the number of degrees of freedom available to estimate the error after calculating the coefficients. Hence in this case, $df_\epsilon = n - 10$

3.3 Optimization method: Sequential quadratic programming

The main scope of this thesis is to optimize the performance of diesel engine based on its input factors and under emissions constraints. In the initial approach, the question of emission constraints were first neglected to simplify the problem. It then became optimizing engine's performance by the input factors. Several methods were considered such as *Gradient Descent*, *Conjugate Gradient Descent* and *Newton's method.* These methods work well with non-constrained nonlinear optimization problems, however, the presence of nonlinear constraints requires a more complex approach. A method called *Sequential Quadratic Programming* (SQP) was considered due to its ability to solve nonlinear optimization problems under some nonlinear constraints. In addition, SQP is also able to be used in non-constrained optimizing problems in which it basically becomes Newton's method (iteratively solve for the optimal point). The use of SQP method has also been mentioned in [1] for solving constrained single-objective optimization problem. The following section gives an inner view about the SQP method and how it works in real scenarios.

3.3.1 Introduction

Since its first apperance in the 1960s [38], SQP has become the most popular method for solving nonlinearly optimization problems under constraints. SQP is in fact not a single algorithm but a conceptual method [39]. With the help of solid theories and computational procedure, SQP method has been developed and used to solve a large set of important problems in practice.

The nonlinear optimization problems (NLP) in which SQP is applied usually have the form [39] of:

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{over} & x \in R^n & \\ \text{subject to} & h(x) = 0 & \quad g(x) \leq 0, \end{array} \tag{34}$$

where $f:\ R^n \rightarrow \mathrm{R}$ is the objective function, the functions *h(x)* : $R^n \rightarrow R^m$ and *g(x)* : $R^n \rightarrow R^p$ describe the equality and inequality constraints. This type of problems can be found in various application in science, engineering, industry and management. Since the great strength of SQP is its ability to solve optimization problems under nonlinear constraints, it is assumed that there is at least one nonlinear constraint function in the NLP.

The basic idea of SQP is to iteratively model the NLP at a given number of iteration x^k as a Quadratic Programming subproblem, and to use the solution from the subproblem to build a new iteration (or approximation) x^{k+1} [39]. The optimization is done when its solution converges to an optimal solution x^*. Hence, with a suitable selection of quadratic programming subproblem, this method can be considered as an extension of Newton and quasi-Newton methods with the constrained settings. Nevertheless, the constraints make the analysis and implementation of SQP more complex [39].

3.3.2 The basic SQP method

Firstly, there are some assumptions on the nonlinear problem (NLP) which need to be stated to clarify the class of the problems. Secondly, theories of nonlinear programming are used to describe the SQP algorithm. The major theory of nonlinear constrained optimization can be found in [40] and [41].

Gradient of functions:

The gradient of a function $f : R^n \to \mathrm{R}$ at $x \in R^n$ is denoted as $\nabla f(x)$, for example

$$\nabla f(x) := (\frac{\delta f(x)}{\delta x_1}, \frac{\delta f(x)}{\delta x_2}, \ldots, \frac{\delta f(x)}{\delta x_\mathrm{n}})^T. \tag{35}$$

For vector-valued functions $h : R^n \to R^m$, the symbol ∇ is also used for the Jacobian of the function h:

$$\nabla h(x) := (\nabla h_1(x), \nabla h_2(x), \ldots, \nabla h_\mathrm{m}(x)). \tag{36}$$

Hessian matrix

Hessian matrix of f at $x \in R^n$ is the matrix of second partial derivatives as given by:

$$(Hf(x))_\mathrm{ij} := \frac{\delta^2 f(x)}{\delta x_\mathrm{i} \delta x_\mathrm{j}}, \quad 1 \leq i, j \leq n. \tag{37}$$

Lagrangian function:

The key function that plays an important role in all theory of constrained optimization is the scalar-valued Lagrangian function, defined as:

$$\mathcal{L}(x, \lambda, \mu) := f(x) + \lambda^T h(x) + \mu^T g(x) \tag{38}$$

where vector $\lambda \in R^m$ and $\mu \in R^p$ are referred to as Lagrangian multipliers.

Set of active constraints:

The set of active constraints consists of the inequality constraints satisfying the equalities at given vector x with $x \in R^n$. It is denoted as:

$$\mathcal{I}_\mathrm{ac} := \{i \in \{1, \ldots, p\} \quad | \quad g_\mathrm{i}(x) = 0\} \tag{39}$$

Strict complementary slackness:

If $x^* \in R^n$ is a local minimum of the NLP, the following condition is called strict elementary slackness at x^*:

$$g_{\rm i}(x^*)\mu^*_{\rm i} = 0 \quad , \quad 1 \leq i \leq p, \tag{40}$$

$$\mu^*_i > 0 \quad , \quad i \in \mathcal{I}_{\rm ac}(x^*) \tag{41}$$

Let $q_x := |\mathcal{I}_{ac}(x)|$ and assuming $\mathcal{I}_{ac}(x) = \{i_1, \ldots, i_{q(x)}\}$, a matrix called $G(x) \in R^{n*(m+qx)}$ is denoted by:

$$G(x) := (\nabla h_1(x), \nabla h_1(x), \ldots, \nabla h_m(x), \nabla g_{i1}(x), \ldots, \nabla g_{iqx}(x)) \tag{42}$$

This matrix G(x) will be used in defining sufficient optimality conditions [39] in the next section.

First order optimality conditions:
The first order necessary conditions hold, if there exist Lagrangian multipliers $\lambda^* \in R^m$ and $\mu^* \in R^p$ such that:

$$\textbf{(A1)} \quad \nabla\mathcal{L}(x^*, \lambda^*, \mu^*) := \nabla f(x^*) + \nabla h(x^*)\lambda^* + \nabla g(x^*)\mu^* = 0 \tag{43}$$

Second order sufficient optimality conditions [40]:
In addition to (**A1**), the following conditions need to be satisfied:
(**A2**) The column of $G(x^*)$ are linearly dependent.
(**A3**) Strict elementary slackness holds at x^*
(**A4**) The Hessian of the Lagrangian respect to x is positive definite on the null space of $G(x^*)^T$ such as:

$$d^t H\mathcal{L}^* d > 0$$

for all $d \neq 0$ such that $G(x^*)^t d = 0$. The optimality conditions in second order assure that the local minimum x^* of the NLP can be confined and the Lagrange multipliers are unique.
According to [39], three standard asymptotic convergence rates respected to Euclidean 2-norm can be used to determined the convergence characteristic of the SQP methods.

Convergence rates:

Let $(x^k)_{k\in N_0}$ be a sequence of iterates converging to x^*. There are three kinds of convergence rates [42]:

- Linearly if there exists a positive constant $\xi < 1$ such that
$$\|x^{k+1} - x^*\| \leq \xi \|x^k - x^*\|$$
for all k sufficiently large.
- Superlinearly if there exists a sequence of positive constant $\xi_k \to 0$ such that
$$\|x^{k+1} - x^*\| \leq \xi_k \|x^k - x^*\|$$
for all k sufficiently large.
- Quadratically if there exists a positive constant ξ such that
$$\|x^{k+1} - x^*\| \leq \xi \|x^k - x^*\|^2$$
for all k sufficiently large.

In the next section, the construction of the quadratic programming subproblems which have to be solved in each iteration step is discussed in details.

Construction of the QP Subproblems:

Choosing a good QP subproblem is very crucial in SQP method since the step from x^k to x^{k+1} is obtained by solving the quadratic subproblem. Note that algorithms needed to solve the QP subproblems are not going to be mentioned although they are nontrivial issue. The scope of this thesis is to create an off-line tool which can optimize the performance of a diesel engine, hence computation works are carried out by using available toolboxes in MATLAB.

At a current step x^k, a reasonable choice for the constraints can be a linearization of the actual constraints about x^k and the objective function can be replaced by its local quadratic approximation. Hence the QP subproblem has the form:

$$\begin{array}{lll} \text{minimize} & \nabla f(x^k)^T d(x) + \frac{1}{2} d(x)^T B_k d(x) & \\ \text{over} & d(x) \in R^n & \\ \text{subject to} & h(x^k) + \nabla h(x^k)^T d(x) = 0 & g(x^k) + \nabla g(x^k)^T d(x) \leq 0, \\ \text{where} & d(x) := x - x^k, \quad B_k := Hf(x^k) & \end{array} \tag{44}$$

The chosen QP works well with linear constraints, however, with the presence of nonlinearity in the constraints of the original problem, the computation of the increment $d(x)$ may break down. Hence, to take nonlinearity in the constraints into account, a quadratic model of the Lagrangian function is substituted as the objective. The reason behind this substitution is that conditions **A1-A4** imply that x^* is a local minimum for this problem [39]

$$\begin{array}{ll} \text{minimize} & \mathcal{L}(x, \lambda^*, \mu^*) \\ \text{over} & x \in R^n \\ \text{subject to} & h(x) = 0 \qquad g(x) \leq 0, \end{array} \tag{45}$$

According to equation (38), the constraint functions are also included in the objective function for this equivalent problem. At a given iterate x^k, the quadratic Taylor series approximation in x for the Lagrangian function is

$$\mathcal{L}(x^k, \lambda^k, \mu^k) + \nabla\mathcal{L}(x^k, \lambda^k, \mu^k)^T d_x + \frac{1}{2} d_x^T H\mathcal{L}(x^k, \lambda^k, \mu^k) d_x. \tag{46}$$

Hence the QP subproblem can be formed as

$$\begin{array}{lll} \text{minimize} & \nabla\mathcal{L}(x^k, \lambda^k, \mu^k)^T d(x) + \frac{1}{2} d(x)^T B_k d(x) & \\ \text{subject to} & \nabla h(x^k)^T d(x) + h(x^k) = 0 & \nabla g(x^k)^T d(x) + g(x^k) \leq 0, \\ \text{where} & d(x) := x - x^k, \quad B_k := H\mathcal{L}(x^k, \lambda^k, \mu^k) & \end{array} \tag{47}$$

In the problems consisting of only equality constraints, the two equations (44) and (47) are equivalent since the derivative term of the constraint becomes constant and the two objective functions become similar as $\nabla f(x^k)^T d(x) + \frac{1}{2} d(x)^T B_k d(x)$. Whereas, in inequality-constrained cases, the two subproblems are equivalent only if a vector of slack variables $z \in R^p$ is added to the inequality constraints, changing the subproblem into equality-constrained case [39].

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0 \qquad g(x) + z = 0, \quad z \geq 0. \end{array} \tag{48}$$

Solving the QP subproblem in equation (44) gives a solution $d(x)$ which can be used to generate a new iterate x^{k+1} by adding a step α to x^k in the direction of $d(x)$. In addition, the multipliers λ and μ need to be estimated again. It can be done by using the optimal multipliers from the QP subproblem. Let the optimal multipliers of the QP subproblem be λ_{qp} and μ_{qp}. The updates of (x, λ, μ) can be defined as

[39]

$$\begin{aligned} x^{k+1} &= x^k + \alpha d(x) \\ \lambda^{k+1} &= \lambda^k + \alpha d(\lambda) \\ \mu^{k+1} &= \mu^k + \alpha d(\mu) \\ &\text{in which} \\ d(\lambda) &= \lambda_{qp} - \lambda^k \\ d(\mu) &= \mu_{qp} - \mu^k \end{aligned} \tag{49}$$

Global convergence is a term stated when convergence of the problem starts from a point which is far away from the optimal point. To guarantee global convergence, the SQP needs a measure of progress [39] called merit function ϕ from which its reduction indicates the progress toward the solution. Adjusting the step-length parameter α is introduced to guarantee that function ϕ is decreased at each iteration. A basic SQP algorithm can be stated as the following pseudo code

Basic Algorithm[39]

Given approximate starting points (x^0, λ^0, μ^0), initial Hessian B_0 and a merit function ϕ with iteration k starts from 0.

1. Form and solve the QP subproblem in eq.47 to obtain ($d(x)$, $d(\lambda)$, $d(\mu)$).

2. Choose a step-length α satisfies

$$\phi(x^k + \alpha d(x)) < \phi(x^k).$$

3. Set updates

$$\begin{aligned} x^{k+1} &= x^k + \alpha d(x) \\ \lambda^{k+1} &= \lambda^k + \alpha d(\lambda) \\ \mu^{k+1} &= \mu^k + \alpha d(\mu) \end{aligned} \tag{50}$$

4. Stop if converged.

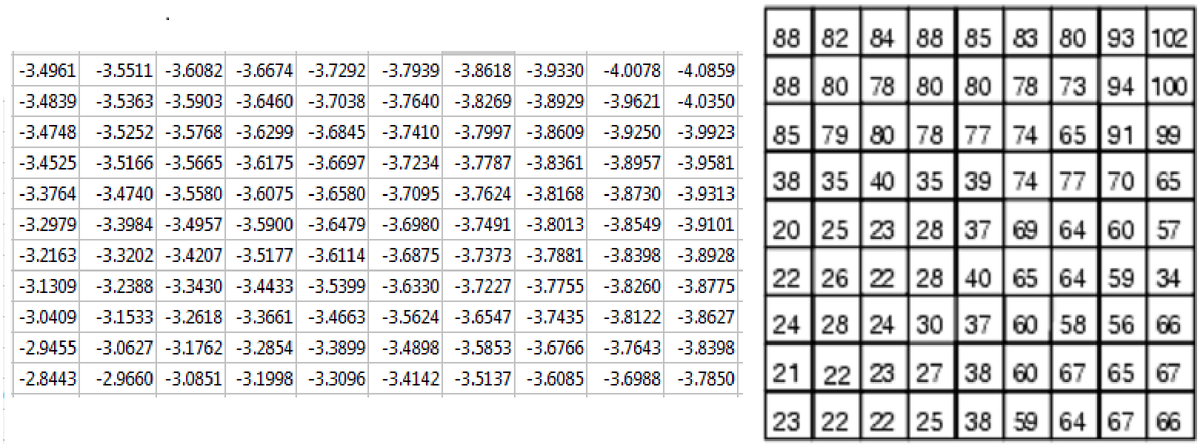
5. Compute new Hessian matrix B_{k+1}.

6. Set new iterate $k = k + 1$. Go back to step 1.

3.4 Map smoothening method

The optimization process produces a set of local optimum points from which a set-point map is created for each operating point. Nevertheless, these set-point maps consist some rough transitions between two operating points and are not good for engine performance during transient. This is why a smoothening step must be done in addition.

The initial approach of this smoothening method is considering every map as a gray-scale image due to the similarity in their structures. The 3-D set-point map corresponding to each operating point is defined as a square matrix of size m. Each element of the matrix contains the optimal value of engine's actuator that minimizes the BSFC under emission constraints. A gray-scale image is also presented as a matrix of size $m \times n$ and each element in the matrix contains a value in the range of 0 to 255. Figure 18 shows the similarity of a set-point map and a gray-scale image. On the left is the matrix presentation of a set-point map and on the right is the matrix presentation of a gray-scale image.



-3.4961	-3.5511	-3.6082	-3.6674	-3.7292	-3.7939	-3.8618	-3.9330	-4.0078	-4.0859
-3.4839	-3.5363	-3.5903	-3.6460	-3.7038	-3.7640	-3.8269	-3.8929	-3.9621	-4.0350
-3.4748	-3.5252	-3.5768	-3.6299	-3.6845	-3.7410	-3.7997	-3.8609	-3.9250	-3.9923
-3.4525	-3.5166	-3.5665	-3.6175	-3.6697	-3.7234	-3.7787	-3.8361	-3.8957	-3.9581
-3.3764	-3.4740	-3.5580	-3.6075	-3.6580	-3.7095	-3.7624	-3.8168	-3.8730	-3.9313
-3.2979	-3.3984	-3.4957	-3.5900	-3.6479	-3.6980	-3.7491	-3.8013	-3.8549	-3.9101
-3.2163	-3.3202	-3.4207	-3.5177	-3.6114	-3.6875	-3.7373	-3.7881	-3.8398	-3.8928
-3.1309	-3.2388	-3.3430	-3.4433	-3.5399	-3.6330	-3.7227	-3.7755	-3.8260	-3.8775
-3.0409	-3.1533	-3.2618	-3.3661	-3.4663	-3.5624	-3.6547	-3.7435	-3.8122	-3.8627
-2.9455	-3.0627	-3.1762	-3.2854	-3.3899	-3.4898	-3.5853	-3.6766	-3.7643	-3.8398
-2.8443	-2.9660	-3.0851	-3.1998	-3.3096	-3.4142	-3.5137	-3.6085	-3.6988	-3.7850

88	82	84	88	85	83	80	93	102
88	80	78	80	80	78	73	94	100
85	79	80	78	77	74	65	91	99
38	35	40	35	39	74	77	70	65
20	25	23	28	37	69	64	60	57
22	26	22	28	40	65	64	59	34
24	28	24	30	37	60	58	56	66
21	22	23	27	38	60	67	65	67
23	22	22	25	38	59	64	67	66

Figure 18: Similarity of set-point map and gray-scale image structures.

The rough transitions between two operating points (usually seen as peaks) can be treated an impulse noises in image since impulse noises are random variation of the brightness and that variation is similar to the rapid change between two operating points. In image processing, median and mean filtering are used as common schemes to reduce impulse noises [43] [44]. There are several researches which applied mean filtering to smooth surface as [45] or used mean filtering as a basis to develop a faster mean filter algorithm [46] [47].

The mean filtering algorithm is stated as, the value of each filtered image's pixel is the arithmetic mean computed by using the neighbor pixels of the current pixel [48]. Kernel of the mean filter is usually a 3×3 or 5×5 matrix

$$A = \frac{1}{9}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

The reason why this filter can attenuate the noise is based on the fact that averaging takes out small variations and averaging 9 values around each pixel decreases the standard deviation of the noise by $\sqrt{9} = 3$ times [49]. Larger kernel matrix hence makes larger impact in noise removal.

Matrix A is a mask which will be applied to the original image by matrix convolution operator. This operator works in an iterative procedure as a kernel matrix goes through every element(i, j) of the original matrix (image). Then for each of them, the value of the element(i, j) and values of the 8 surrounding elements are multiplied by the corresponding values of the kernel. Finally the multiplication results are added together and the element(i, j) is set to this final sum value. Figure 19 shows a simple example of a matrix convolution.

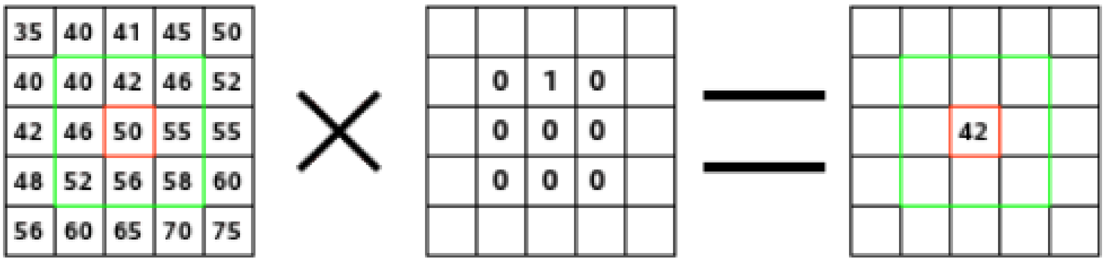


Figure 19: An example of a 2-D convolution[50].

The center element of the first matrix is being treated by the second matrix which is a 3×3 matrix. All elements inside the green box of the first matrix are then multiplied with their corresponding values in the kernel and finally the center element is set to result of the following computation

$$40 * 0 + 46 * 0 + 52 * 0 + 42 * 1 + 50 * 0 + 56 * 0 + 46 * 0 + 55 * 0 + 58 * 0 = 42$$

The fraction $\frac{1}{9}$ in the kernel matrix A represents the averaging step as the sum of all 9 multiplications is divided by 9. The simple following example shows how kernel matrix A is applied to a particular matrix using matrix convolution.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & \boxed{22} & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{pmatrix} \otimes \begin{pmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & \boxed{14} & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{pmatrix}$$

The center element (the one inside the red square) can be seen as an impulse noise since its value dramatically increases compared to its neighbor values. A kernel matrix of size 3 × 3 is applied to this matrix and the center element is being considered. All elements inside the green box of the first matrix are then multiplied with their corresponding values in the kernel and the center element is set to result of the following computation:

$$\frac{1}{9} * (7 + 8 + 9 + 12 + 22 + 14 + 17 + 18 + 19) = 14$$

The kernel matrix goes on and does the same computation for the rest of the treated matrix.

Applying mean filtering with discrete convolution has a drawback when dealing with elements at border elements of the original matrix. As shown in figure 20, when applying convolution to the top left corner element, it is assumed that values of positions which are out of the original matrix are set to 0. This is so called "zero padding" method.

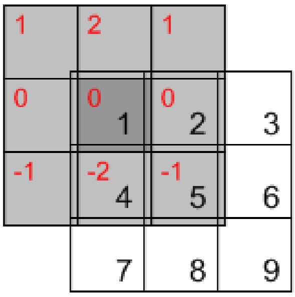


Figure 20: Example of convolution at the border element[51].

This method may lead to undesired changes in the border elements as the arithmetic mean value is affected by five neighborhood values being set to 0. These changes might not affect the filtered image much as they happen only on the edge of the image, however, with the final set-point maps, these undesired changes make a very big impact. The edges of the map will have rather big errors as shown in figure 21.

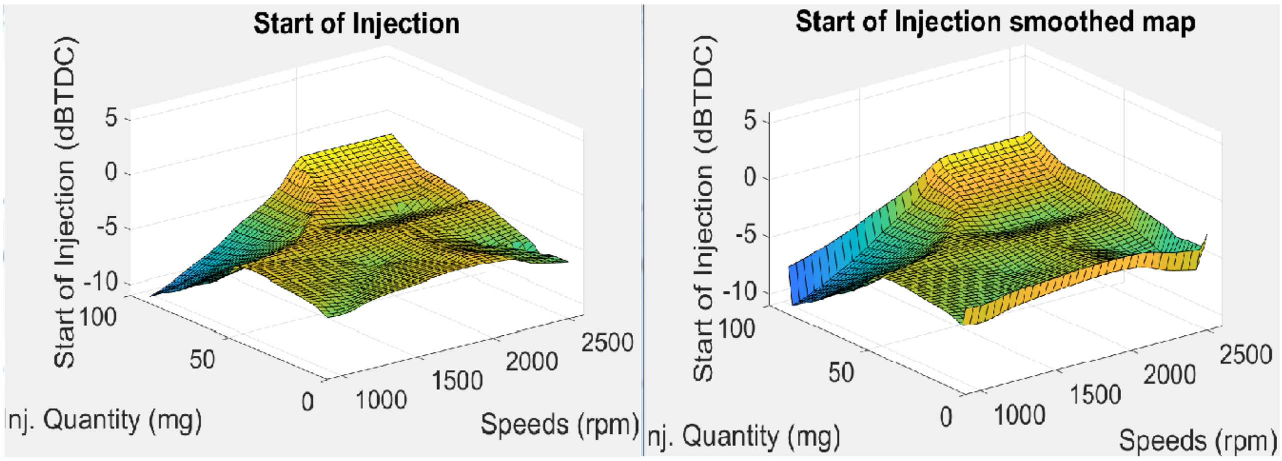


Figure 21: Example of a final map with errors at its edges.

In order to overcome this problem with the border elements, an alternative non-zero padding method is introduced to this tool. It means at each side of the unfiltered matrix, one more row/column is added and values in these rows/columns are copied from the adjacent rows/columns.

$$M = \begin{pmatrix} \boxed{\begin{matrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{matrix}} \end{pmatrix} \longrightarrow \quad M = \left(\begin{array}{c|ccc|c} 1 & 1 & 2 & 3 & 3 \\ \cline{2-4} 1 & 1 & 2 & 3 & 3 \\ 4 & 4 & 5 & 6 & 6 \\ 7 & 7 & 8 & 9 & 9 \\ \cline{2-4} 7 & 7 & 8 & 9 & 9 \end{array} \right)$$

The padded matrix on the right has one outside layer and matrix convolution will only be carried inside the red square, it means computation starts from row and column index 2 instead of 1 and finishes at the second last row/column. Padding rows and columns into the unfiltered matrix ensures that the arithmetic mean value of neighborhood elements is not affected by zero values and hence avoids big variation in the border elements. Testing result is shown in figure 22. The surface on the right is the smoothed map with non-zero padding method and it shows a decent smoothening in comparison to the original map on the left. Most importantly, the smoothed map does not consist vertical edges like the smoothed map in figure 21.

Validation of the smoothening process will be explained in the Result section. The goal of the smoothening process is to make a smooth shape of the map while

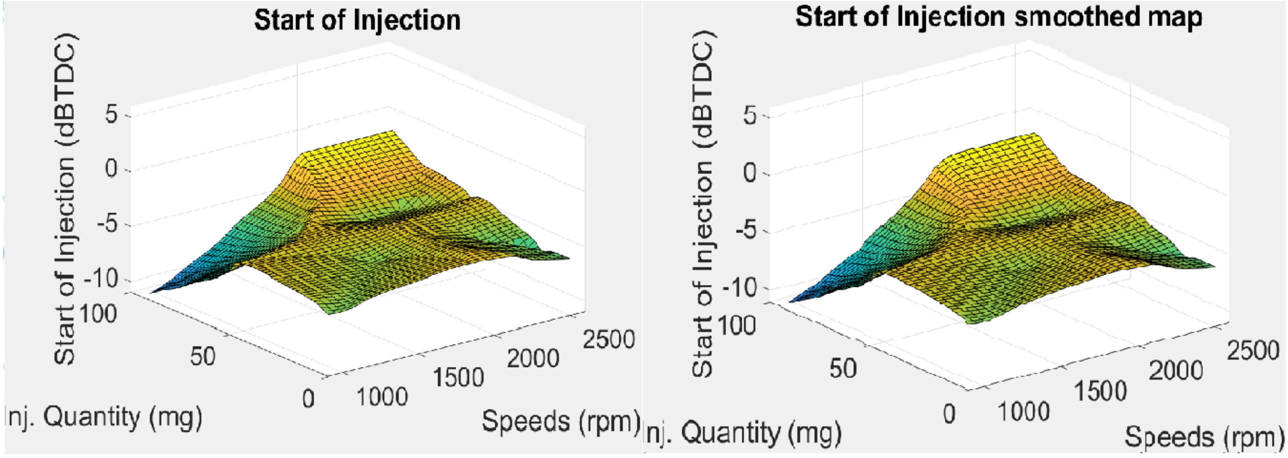


Figure 22: Testing result by using non-zero padding method.

keeping the values as close as possible to their original.

4 Experiment implementation

4.1 DoE preparations

The selected design of experiment is Box-Behnken design due to its competitive advantages mentioned in Section 3.1. This design includes more than 2 factors to capture the response curvature and its resultant model contains quadratic terms according to the following equation

$$\begin{aligned} y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + &\quad \text{(linear terms)} \\ b_{12}x_{12} + b_{13}x_{13} + b_{23}x_{23} + &\quad \text{(interaction terms)} \\ b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 &\quad \text{(quadratic terms)} \end{aligned} \tag{51}$$

According to Subsection 2.2.5, the optimization variables are coded as

1. Input factors:

- Boost pressure (P_i): x_1. (unit: gauge pressure)
- Common rail pressure (P_{CR}): x_2. (unit: bar)
- Start of injection (SoI): x_3. Unit: degree Before Top Dead Center (dBTDC).

2. Responses:

- The Break Specific Fuel Consumption (BSFC): y_1 (unit: g/kWh)
- Other possible emissions responses such as NO_x and NO responses: y_i

The resultant model can be written in physical variables as

$$\begin{aligned} BSFC = b_0 + b_1P_{\mathrm{i}} + b_2P_{\mathrm{CR}} + b_3SoI + &\quad \text{(linear terms)} \\ b_{12}P_{\mathrm{i}}P_{\mathrm{CR}} + b_{13}P_{\mathrm{i}}SoI + b_{23}P_{\mathrm{CR}}SoI + &\quad \text{(interaction terms)} \\ b_{11}P_{\mathrm{i}}^2 + b_{22}P_{\mathrm{CR}}^2 + b_{33}SoI^2 &\quad \text{(quadratic terms)} \end{aligned}$$

The design matrix at each operating point is shown in Table 4. According to the table, 15 experiments are required for each operating point and at each point, 15 values of BSFC are recorded as well as 15 values of emission measurement. The ±1 indicate minimum and maximum values of each factor while 0 indicates the central point of each range.

Table 4: Design matrix for each operating point.

No.	P_i	P_{CR}	SoI	**BSFC**	**NOx**
1	-1	-1	0		
2	-1	+1	0		
3	+1	-1	0		
4	+1	+1	0		
5	-1	0	-1		
6	-1	0	+1		
7	+1	0	-1		
8	+1	0	+1		
9	0	-1	-1		
10	0	-1	+1		
11	0	+1	-1		
12	0	+1	+1		
13	0	0	0		
14	0	0	0		
15	0	0	0		

Table 5: Specification of the engine test bed

1	Cylinder number	4
2	Bore (mm)	108
3	Stroke (mm)	120
4	Swept volume $(dm)^3$	4.4
5	Rated speed (1/min)	2200
6	Rated power (kW)	101
7	Maximum torque at rated speed (Nm)	455
8	Maximum torque at 1500 1/min (Nm)	583

4.2 Engine test bed

The engine test bed to be used in this thesis is a four-cylinder, common rail and turbocharged diesel engine in the *Internal Combustion Engine Laboratory* at Aalto University. The engine model is AGCO POWER 44 AWI and is shown in figure 23. Controller of the engine is designed entirely on LabView.

Main specifications of this engine are listed in Table 5.

Figure 24 show an example set of experiment results running on the AGCO engine at 1600 rpm and 200 Nm. As can be seen, the first three graphs show values of each factor at their maximum, minimum and center levels. The last graph is the brake specific fuel consumption index of the engine. The oscillation that happens in the boost graph is caused by the sensitivity of the VGT controller. Therefore the results of the BSFC are also affected as there are also some oscillation in the BSFC graph. In addition, running all 15 experiments of one operating point takes around one hour since the engine is very sensitive and it requires careful handling during the operation.

Due to these reasons, this engine test bed is only used for testing of operating points to find out the ranges of factors. The experiments are conducted on a

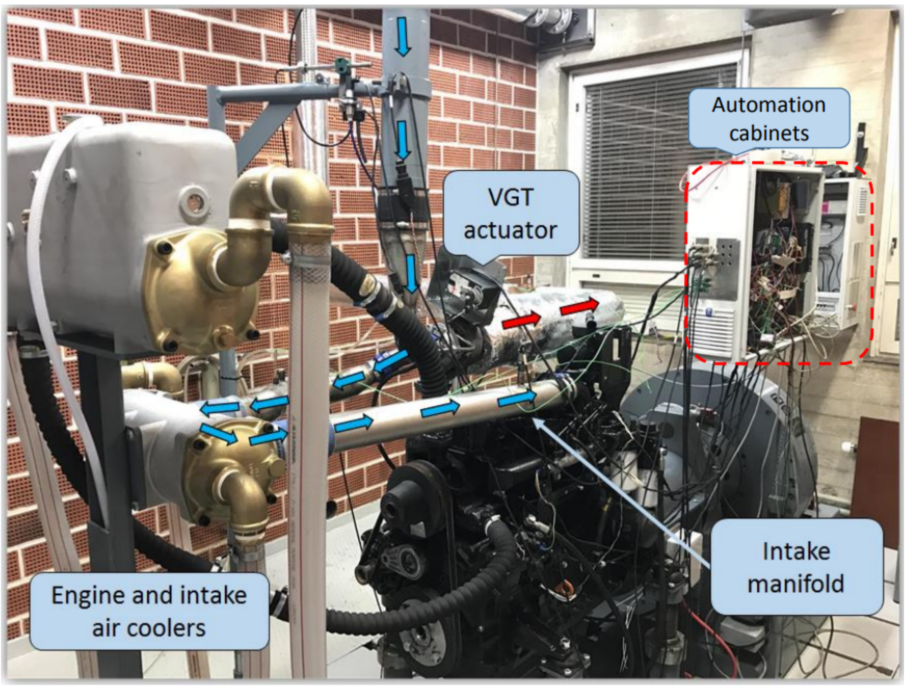


Figure 23: AGCO POWER 44 AWI.

simulation engine model which was calibrated with the AGCO engine. The model was developed in GT-SUITE software by David Bernasconi, who was a Master thesis worker in the *Internal Combustion Engine Laboratory* at Aalto University.

The simulation model will be shortly described in the next section.

4.3 Engine simulation model

4.3.1 Introduction to GT-SUITE

GT-SUITE is an industry-leading simulation tool which offers functionalities ranging from fast concept design to detailed system or sub-system analyses, optimization and investigation. The reasons why GT-SUITE is chosen over other simulation tool, such as MATLAB Simulink, are its competitive advantages and accuracy in system modeling.

- GT-SUITE is a comprehensive set of simulation tools for engine and vehicle systems with industry-standard engine simulation.
- The foundation of GT-SUITE is a versatile multi-physics platform to build up models of different systems based on many fundamental libraries such as: Mechanical library, Electric and Electromagnetic library, Thermal library,

Control library, etc.

- With a wide range of libraries, the tool is able to model boosting system, variable geometry systems and other important systems inside an engine.
- GT-SUITE has capability to run Design of Experiment tool directly on the model with built-in or customized matrix design options for users.
- In addition, it is also possible to measure the emissions (NO_x, CO and CO_2) by using GT-SUITE while it is temporarily not possible to measure in the real AGCO engine.

4.3.2 Overview of the GT-SUITE model

This simulation model was made by David Bernasconi as his Master thesis during the time he was working at the Internal Combustion Engine Laboratory. The main purpose of the work was to build a model of the real AGCO engine in the lab and implementing the controller of the engine into the model.

The model consists of two main parts: simulation execution and reports. Figure 25 shows an overview of the simulation model with fundamental blocks and components. The four cylinders are shown in the middle of the figure and are connected with piping systems. Compressor, turbine and inter-cooler are also modeled by separate blocks. All components and pipes are modeled with their real dimensions.

Figure 26 shows the result panel after the simulation is done. All the results about speed, torque, pressure, temperature, etc. are shown in this panel by both data tables and plots.

Simulations in the GT-SUITE model are run by case and each case can be configured as shown in figure 27. In the case setup option, speed, pressure, injection quantity, start of injection, etc. can be set directly.

Furthermore, DoE setup can be made directly in GT-SUITE by using an available built-in function named 'DoE setup', as shown in figure 28. Factors and their ranges can be set using ready-made designs or by customized designs.

This model has been calibrated with the real AGCO engine to mimic its performance, however, the calibration was made only at a certain range of speed and load on the real engine. Running the model inside the calibrated range or in the near area can assure that the results are quite accurate. Due to this reason, the selection of operating points for the DoE setup must be considered inside or close to the calibrated operating range:

- Speed: 1500 rpm - 2000 rpm
- Load: 150 Nm - 300 Nm

Choosing operating points which are too far away from the calibrated range can lead to errors and nonsense results. There is one drawback in this model, which is the lack of control of the common rail pressure for the engine. Modeling and calibrating the common rail pressure require in-cylinder pressure measurement, however, it is temporarily not available in the laboratory. Therefore, the rail pressure values used in the Design of Experiment method have been taken from the real AGCO engine run tests.

4.4 Operating point selection

Based on the calibrated working range of the engine model, the operating points selected for the DoE setups are inside that range and partly in the close neighborhood of that range.

Notice that in this model, the torque is controlled by the fuel injection quantity as each quantity of injected fuel gives a certain amount of torque. Therefore, several testing runs have been conducted to find out the corresponding torques. There will be 12 to 14 operating points needed, the number of necessary points affects the accuracy and resolution of the final maps. A comparison between the uses of different numbers of operating points will be conducted in section Results.

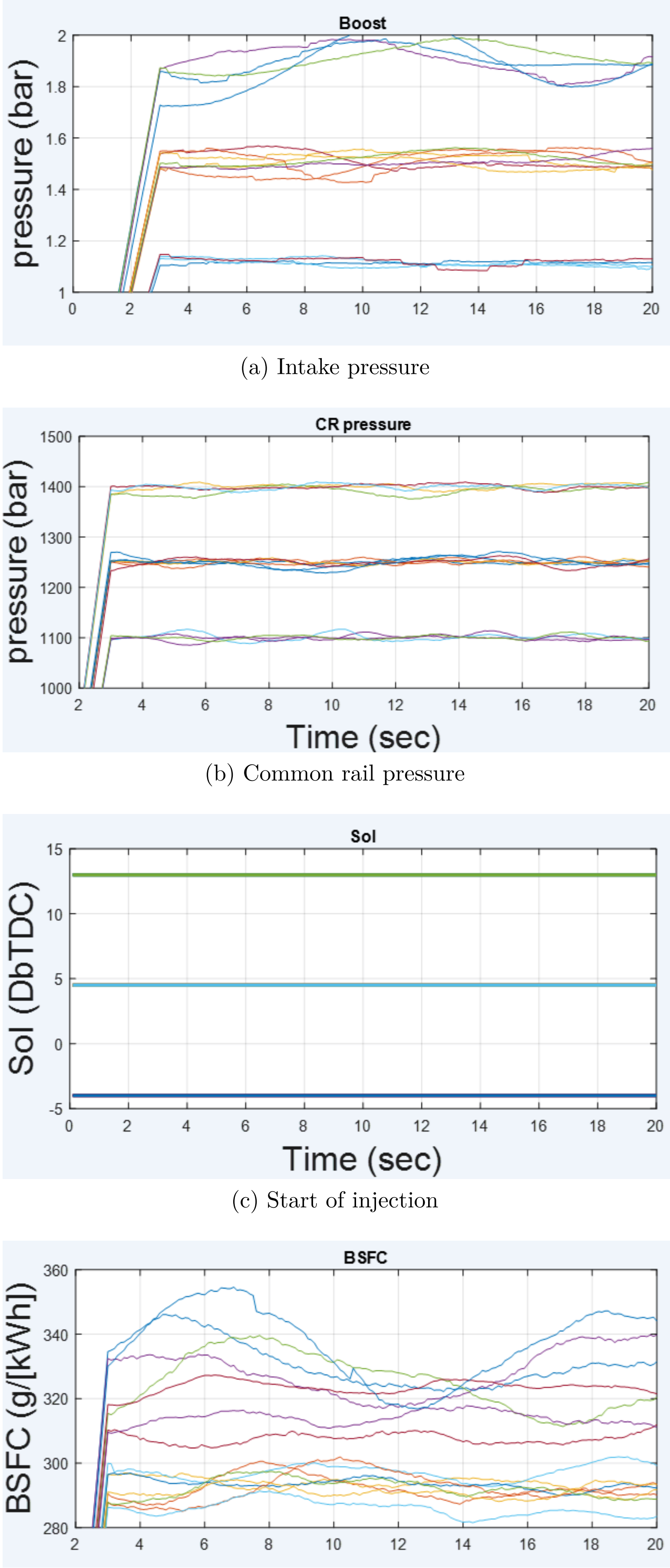
Finally, the chosen operating points are shown in the following Table 6. Points are chosen in 300 rpm interval and 15 mg interval. More points are chosen to be inside the calibrated range of the model.

Table 6: Selection of operating points

Speed (rpm) Injection quantity (mg)	1300 rpm	1600 rpm	1900 rpm	2200 rpm
15 mg ($\sim 60Nm$)				
30 mg ($\sim 100Nm$)		✓	✓	✓
45 mg ($\sim 150Nm$)	✓	✓	✓	✓
60 mg ($\sim 250Nm$)		✓	✓	✓
75 mg ($\sim 350Nm$)	✓		✓	

4.5 Simulation runs

All the points with the check mark (✓) will be run in the simulation model and results of BSFC as well as emissions are recorded for later phases. Modeling and optimization processes are executed after all operating points have been run. Results of the modeling, optimization and map constructing will be shown in details in the next section.



(a) Intake pressure

(b) Common rail pressure

(c) Start of injection

(d) Brake specific fuel consumption

Figure 24: Experiment results at operating point @(1600 rpm, 200 Nm)

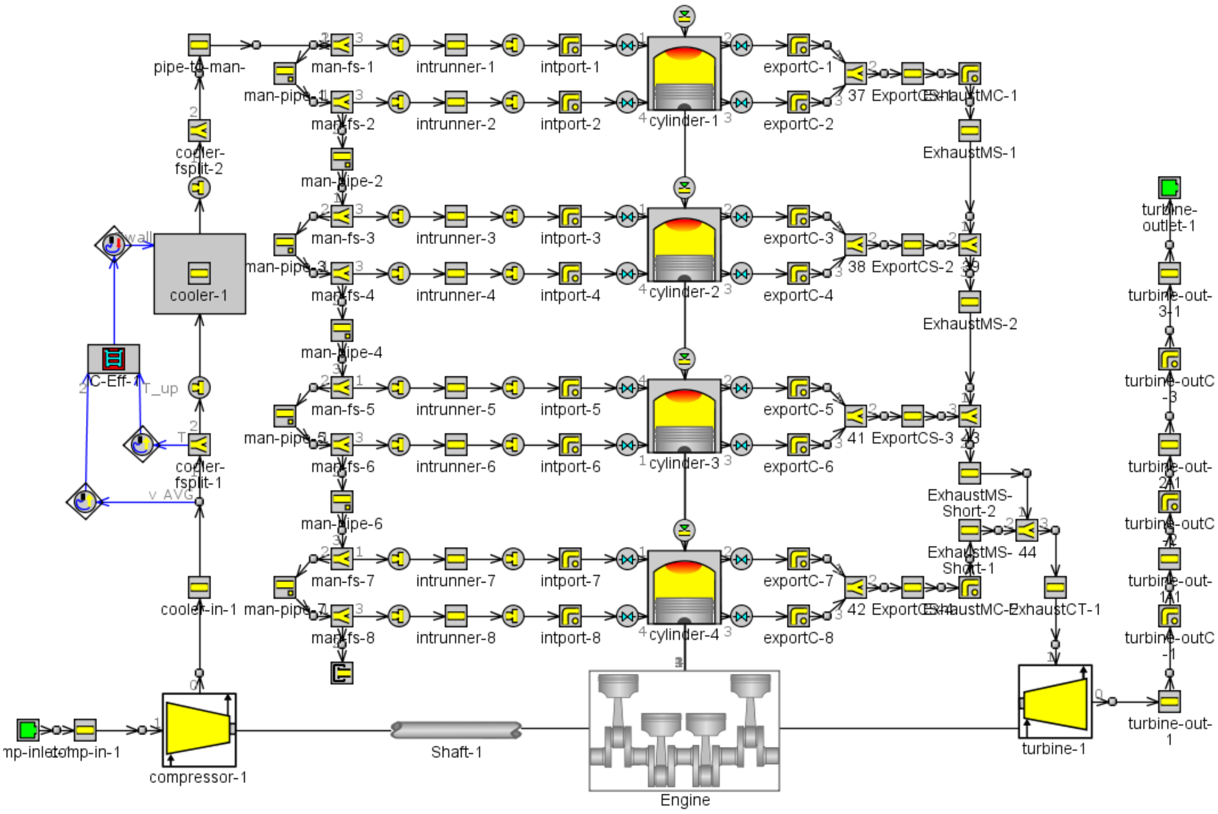
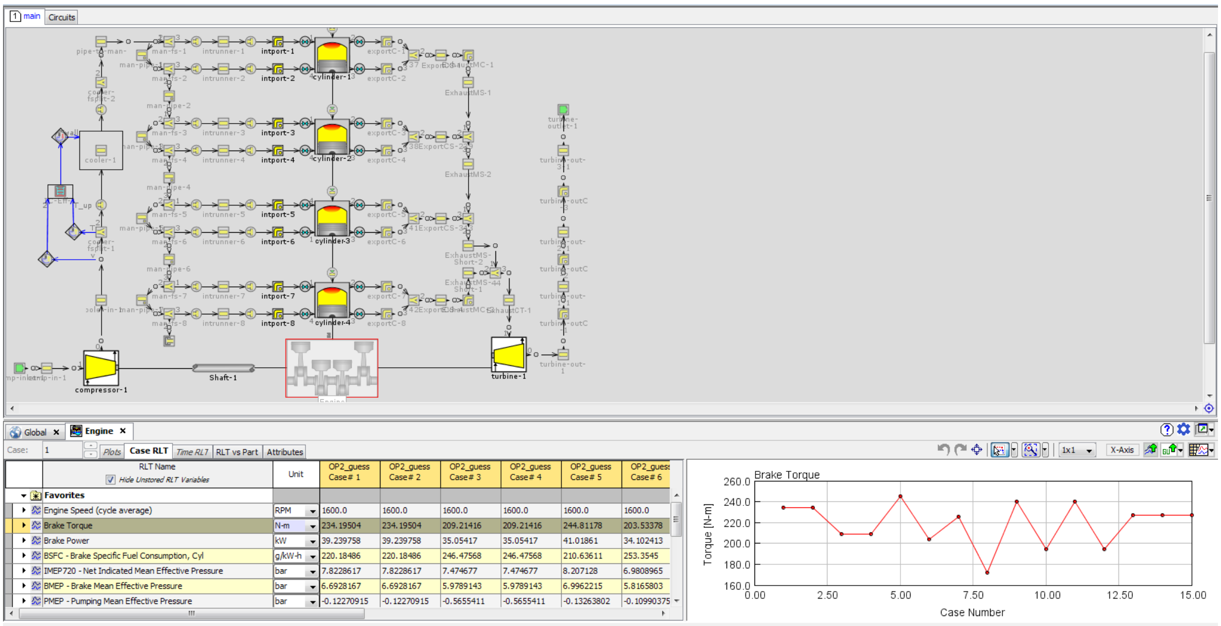
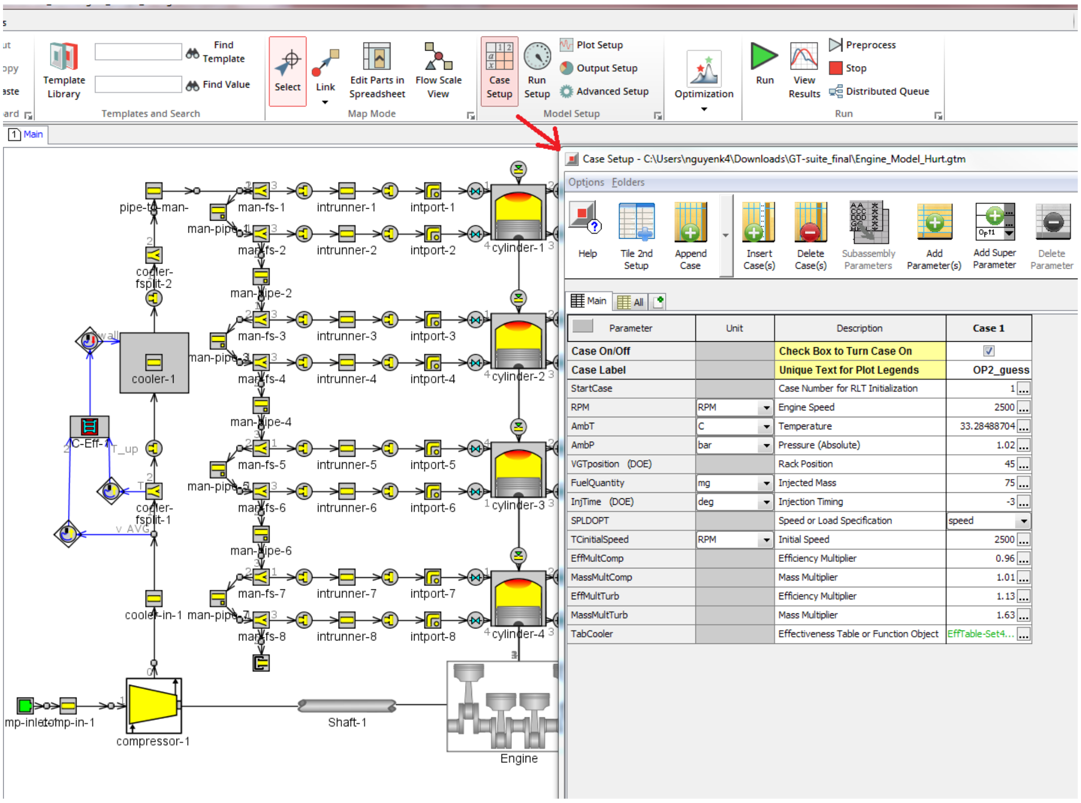


Figure 25: GT-SUITE simulation model.



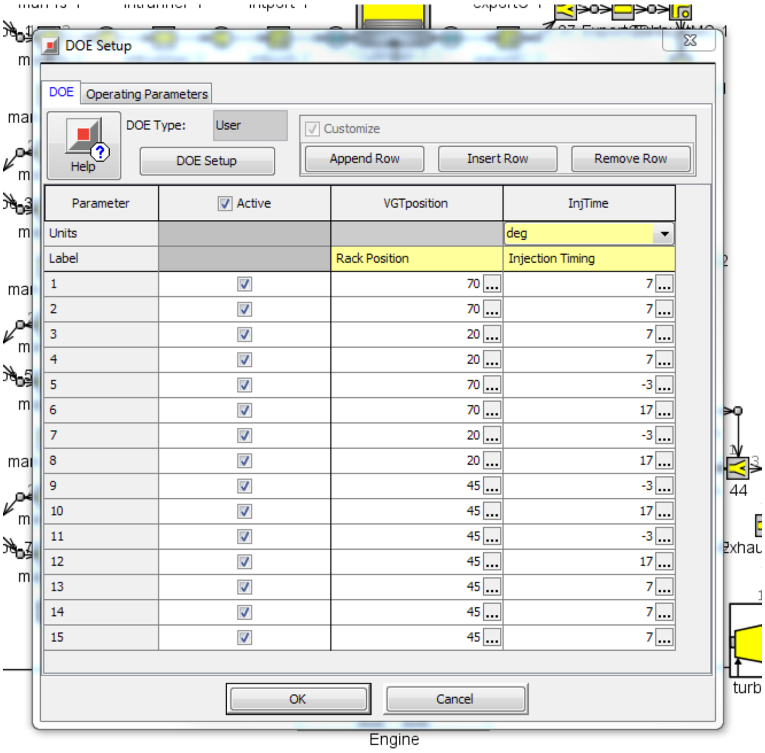
RLT Name ☑ Hide Unstored RLT Variables	Unit	OP2_guess Case# 1	OP2_guess Case# 2	OP2_guess Case# 3	OP2_guess Case# 4	OP2_guess Case# 5	OP2_guess Case# 6
Favorites							
Engine Speed (cycle average)	RPM	1600.0	1600.0	1600.0	1600.0	1600.0	1600.0
Brake Torque	N-m	234.19504	234.19504	209.21416	209.21416	244.81178	203.53378
Brake Power	kW	39.239758	39.239758	35.05417	35.05417	41.01861	34.102413
BSFC - Brake Specific Fuel Consumption, Cyl	g/kW-h	220.18486	220.18486	246.47568	246.47568	210.63611	253.3545
IMEP720 - Net Indicated Mean Effective Pressure	bar	7.8228617	7.8228617	7.474677	7.474677	8.207128	6.9808965
BMEP - Brake Mean Effective Pressure	bar	6.6928167	6.6928167	5.9789143	5.9789143	6.9962215	5.8165803
PMEP - Pumping Mean Effective Pressure	bar	-0.12270915	-0.12270915	-0.5655411	-0.5655411	-0.13263802	-0.10990375

Figure 26: GT-SUITE reports.



Parameter	Unit	Description	Case 1
Case On/Off		Check Box to Turn Case On	☑
Case Label		Unique Text for Plot Legends	OP2_guess
StartCase		Case Number for RLT Initialization	1
RPM	RPM	Engine Speed	2500
AmbT	C	Temperature	33.28488704
AmbP	bar	Pressure (Absolute)	1.02
VGTposition (DOE)		Rack Position	45
FuelQuantity	mg	Injected Mass	75
InjTime (DOE)	deg	Injection Timing	-3
SPLDOPT		Speed or Load Specification	speed
TCinitialSpeed	RPM	Initial Speed	2500
EffMultComp		Efficiency Multiplier	0.96
MassMultComp		Mass Multiplier	1.01
EffMultTurb		Efficiency Multiplier	1.13
MassMultTurb		Mass Multiplier	1.63
TabCooler		Effectiveness Table or Function Object	EffTable-Set4...

Figure 27: Speed and pressure setup.



DOE Type: User

☑ Customize

Parameter	☑ Active	VGTposition	InjTime
Units			deg
Label		Rack Position	Injection Timing
1	☑	70	7
2	☑	70	7
3	☑	20	7
4	☑	20	7
5	☑	70	-3
6	☑	70	17
7	☑	20	-3
8	☑	20	17
9	☑	45	-3
10	☑	45	17
11	☑	45	-3
12	☑	45	17
13	☑	45	7
14	☑	45	7
15	☑	45	7

Figure 28: DoE set up options.

5 Results

5.1 Modeling and Optimization processes

5.1.1 Modeling of the BSFC

Modeling process is executed after all the experiments have been done and engine responses have been fully recorded. Computational works have been done by MATLAB with Linear Model Fitting function. The recorded data is fitted to this resultant model

$$\begin{aligned} BSFC = b_0 + b_1 P_{\mathrm{i}} + b_2 P_{\mathrm{CR}} + b_3 SoI + &\quad \text{(linear terms)} \\ b_{12} P_{\mathrm{i}} P_{\mathrm{CR}} + b_{13} P_{\mathrm{i}} SoI + b_{23} P_{\mathrm{CR}} SoI + &\quad \text{(interaction terms)} \\ b_{11} P_{\mathrm{i}}^2 + b_{22} P_{\mathrm{CR}}^2 + b_{33} SoI^2 &\quad \text{(quadratic terms)} \end{aligned}$$

The outcomes of this process are a set of vectors of coefficients b_i for each operating point and corresponding model errors. Figure 29 shows relationships and effects of each factors on the BSFC response based on their coefficients. This figure shows result at (1600 rpm, 45 mg).

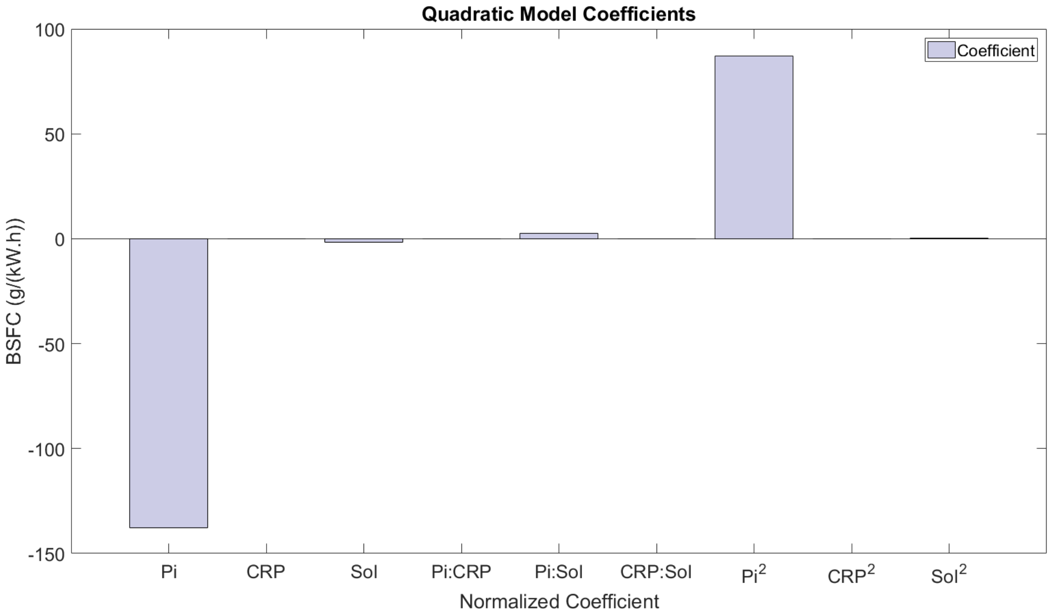


Figure 29: Coefficient effects on BSFC.

As can be seen from the graph, the common rail pressure (CRP) and start of injection (SoI) have little correlation to the BSFC response while boost pressure (Pi) has a big impact on how BSFC changes. Table 7 shows results of the modeling process applied for 12 selected operating points in Table 6.

1. 1300 rpm - 45 mg
2. 1300 rpm - 75 mg
3. 1600 rpm - 30 mg
4. 1600 rpm - 45 mg
5. 1600 rpm - 60 mg
6. 1900 rpm - 30 mg
7. 1900 rpm - 45 mg
8. 1900 rpm - 60 mg
9. 1900 rpm - 75 mg
10. 2200 rpm - 30 mg
11. 2200 rpm - 45 mg
12. 2200 rpm - 60 mg

Table 7: Modeling results of the selected operating points.

Operating Point	Mean Squared Error	R^2	$R^2_{adjusted}$
1	2.32	0.987	0.965
2	1.39	0.995	0.987
3	3.15	0.996	0.99
4	3.36	0.993	0.982
5	2.25	0.995	0.986
6	11.3	0.986	0.961
7	7.09	0.991	0.975
8	6.04	0.989	0.97
9	4.54	0.992	0.979
10	25.5	0.974	0.927
11	12.5	0.985	0.959
12	6.24	0.994	0.983

According to the result table, the model error is quite small and the goodness of fit is pretty good as all of the coefficients of determination R^2 are close to 1.

5.1.2 Modeling of the emissions

In order to model the emission response (due to limitations of the simulation model, only NO_x emission can be measured) of the engine, an investigation on how the three input factors make impacts on the amount of NO_x exerted at 1600 rpm and 45 mg was done. Result of the inspection is shown in figure 30. Horizontal axis in each graph is the running order of the executed experiments according to the design matrix in Table 4.

It can be seen that the common rail pressure and the start of injection somehow have a little correlation with the changes of NOx emission. The boost pressure on the other hand makes main impact on the NOx emission. Therefore, it is relatively safe to model the response of NOx emission based on these three input factors P_i, P_{CR} and Soi.

Furthermore, one more inspection has been done to test how NOx emission responses to different speeds and torques. Figure 31 shows results of the test and in summary, higher speeds and lower torques create the most NOx emission.

Similarly, the recorded NO_x data is also fitted to the same function that has been used to model the BSFC responses considering that the two responses are recorded in a same experiment and both have correlation with the input factors.

$$\begin{aligned} NO_x = b_0 + b_1 P_{\text{i}} + b_2 P_{\text{CR}} + b_3 SoI + &\quad \text{(linear terms)} \\ b_{12} P_{\text{i}} P_{\text{CR}} + b_{13} P_{\text{i}} SoI + b_{23} P_{\text{CR}} SoI + &\quad \text{(interaction terms)} \\ b_{11} P_{\text{i}}^2 + b_{22} P_{\text{CR}}^2 + b_{33} SoI^2 &\quad \text{(quadratic terms)} \end{aligned}$$

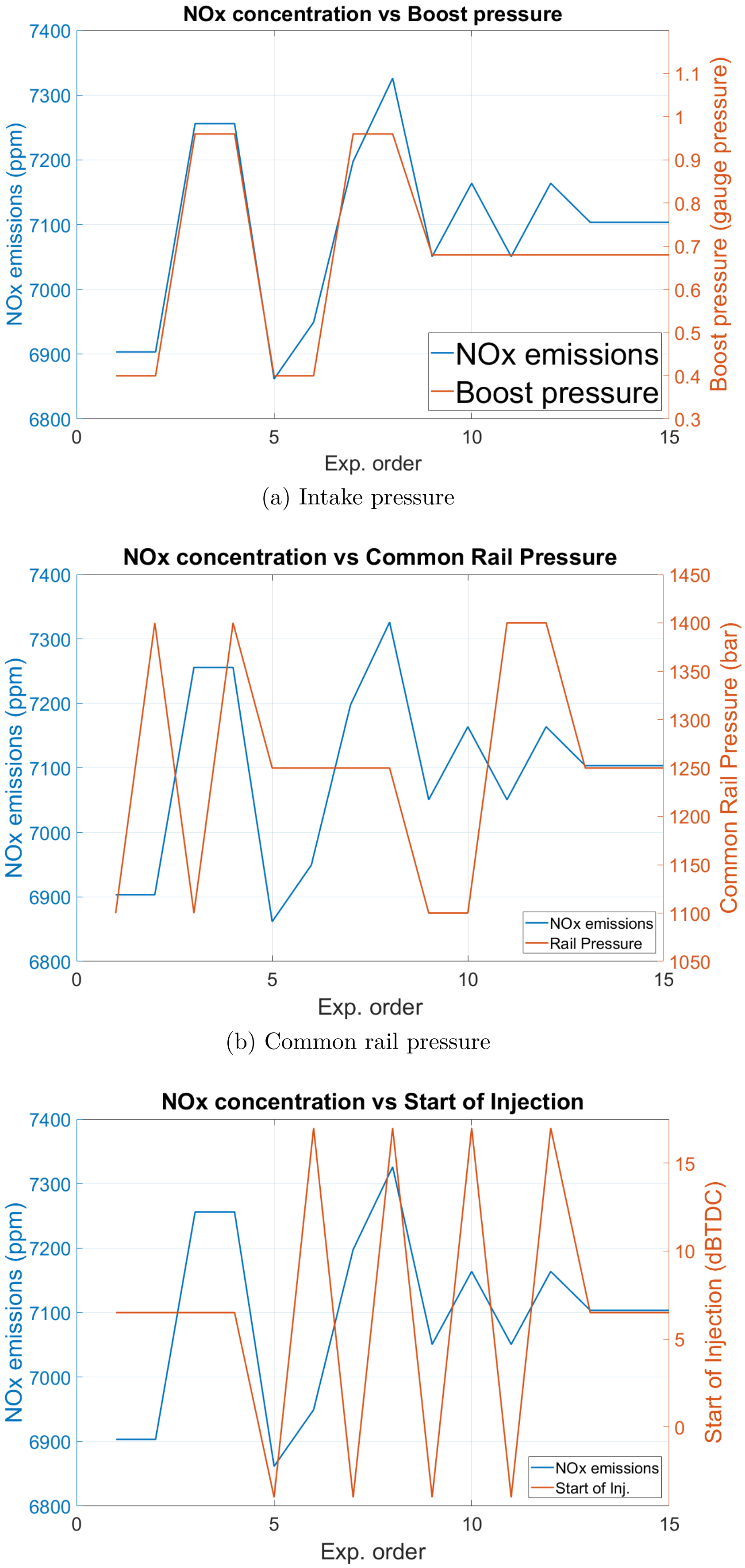
The results of this process are a set of vectors of coefficients b_i and their corresponding model errors. Emission models archived from this modeling process can be used as constraint functions in the later optimization processes.

5.1.3 Optimization processes

The optimization of BSFC is done in two different ways to compare the differences between without emission constraints and under emission constraints. The function to be optimized is the one stated in equation (51).

The first optimization problem (without emission constraints) is defined as

$$\underset{Boost,CR,SoI}{\text{minimize}} \quad BSFC$$

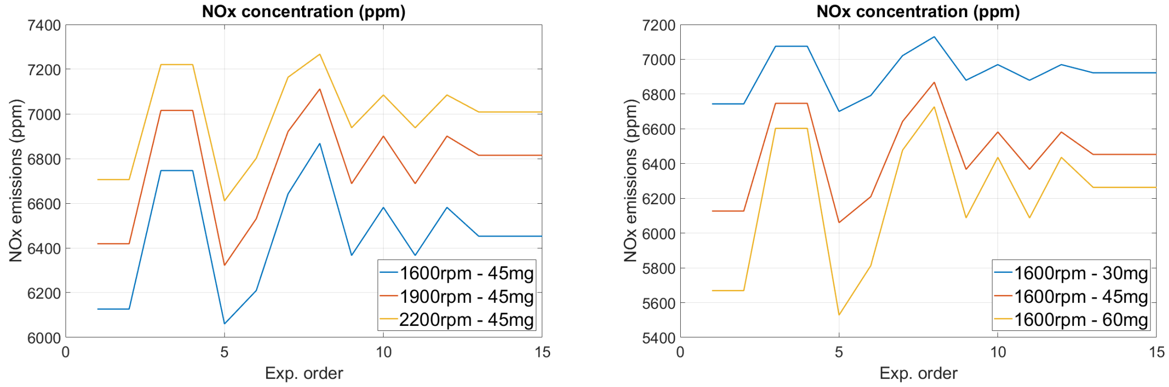


(a) Intake pressure

(b) Common rail pressure

(c) Start of injection

Figure 30: Effects of input factors on the NOx emission @(1600 rpm, 45 mg)



(a) NOx versus speeds

(b) NOx versus torques

Figure 31: Effects of different speeds and torques on the NOx emission

An example of optimized result at (1600 rpm, 45 mg) is shown in figure 32. Optimization is done with three variables, however it is not possible to create 4-D plots, the plots in the result are made by every two of the input factors with the other factor being kept at its optimal value.

The optimization of BSFC under the emission constraints is defined as

$$\begin{aligned} &\underset{Boost,CR,SoI}{\text{minimize}} && BSFC \\ &\text{subject to:} && emission \quad constraints \end{aligned}$$

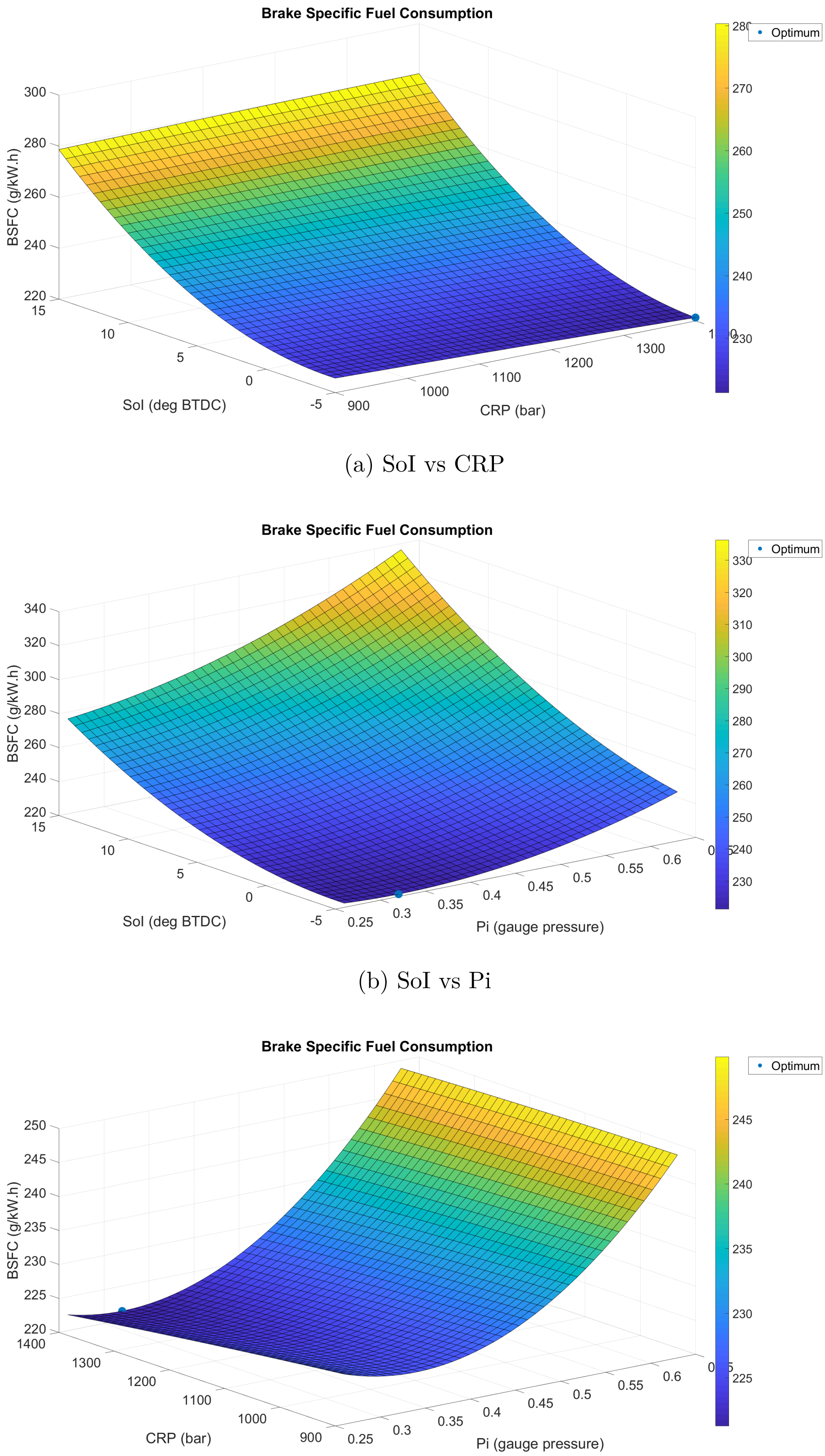
Example result at (1600 rpm, 45 mg) is shown in figure 33 and the plots are made in the similar way as in unconstrained cases.

It is shown that there are differences in optimal set-points and optimal BSFC values between the two scenarios. The following Table 8 shows detailed differences of unconstrained and constrained optimization at (1600 rpm, 45 mg).

Table 8: Differences between unconstrained and constrained optimization

	Unconstrained	**Constrained**
Boost pressure (gauge pressure)	0.5195	0.4
Common rail pressure (bar)	1400	1250
Start of injection (dBTDC)	-3.6643	-4
BSFC (g/kW.h)	207.4699	213.2500

The constrained optimal values of BSFC tend to be a bit larger than the unconstrained ones because higher boost pressure and later injection timing tend to decrease the amount of NOx but at the same time increase the BSFC of the engine.

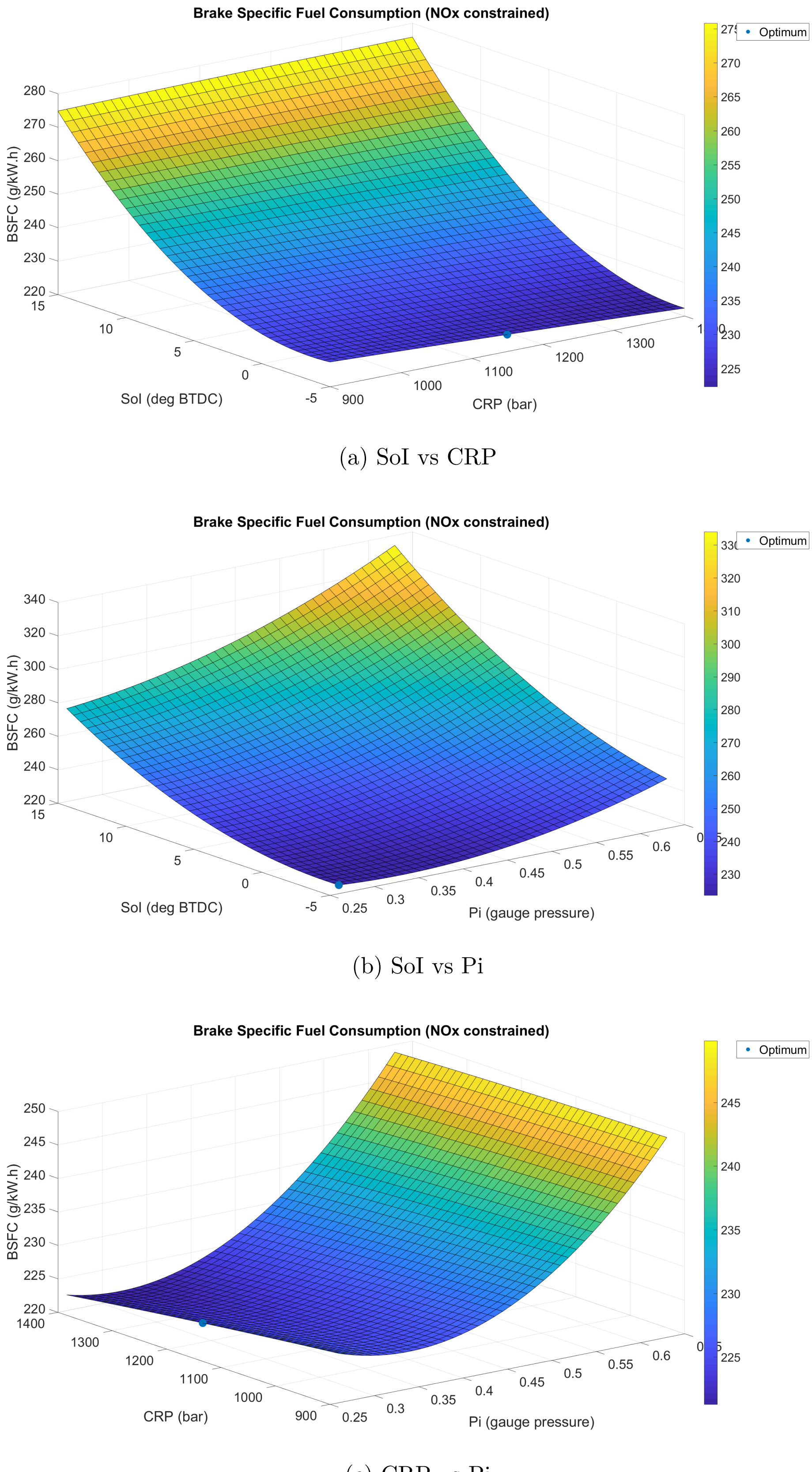


(a) SoI vs CRP

(b) SoI vs Pi

(c) CRP vs Pi

Figure 32: Optimization without emission constraints @(1600 rpm, 45 mg)



(a) SoI vs CRP

(b) SoI vs Pi

(c) CRP vs Pi

Figure 33: Optimization under emission constraints @(1600 rpm, 45 mg)

5.2 Map construction

The optimization process creates 3 sets of optimal values for all three factors P_i, P_{CR} and Soi at the selected operating points. Applying 'Robust Locally Weighted Regression' (LOESS) [28] method to each set of optimal values produces an initial map for every factor.

The name LOESS is derived from the 'locally weighted scatter plot smoothing'. The main idea of LOESS is to use the values of neighbor data points within a defined span to estimate the value at a new point. In other words, LOESS creates a continuous curve that represent the relationship between the inputs and the response. At a given point, LOESS fits a second degree polynomial using weighted linear least squares regression to the data points within the span. The weights W_i for each data point is defined by

$$W_i = \left(1 - \left|\frac{x - p_i}{d(x)}\right|^3\right)^3$$

where p_i are the neighbor points of $\mathbf{x}$ within the span and $d(x)$ is the distance from $\mathbf{x}$ to the furthest point in the span. The LOESS regression was implemented by using MATLAB toolbox. Figure 34 shows the results of LOESS-regression on the three constrained optimal sets of values of P_i, P_{CR} and Soi

In practice, static control maps of diesel engines can have many different resolutions (intervals between chosen operating points) based on the properties of the engine (high speed or low speed). In this AGCO engine, the control maps with respect to speed and injection quantity will have resolution of 125-rpm and 3.25-mg intervals. Dividing the maps into smaller intervals requires more operating points to be run for better accuracy of the created maps. Figure 35 shows set-point maps created in both unconstrained and constrained optimization.

There are definitely some differences between the constrained and unconstrained maps showing in these graphs. Figure 36 shows a clearer differences between maps when emission constraints are presented.

The colored surfaces are unconstrained and the white ones are made under emission constraints. Emission constraints tend to lower the values of the optimal maps and avoid sudden high peaks. Though the emission constraints can reduce the high peaks for some extend, these constrained maps still needs smoothening to achieve a smoother transition between operating points. Results of the smoothening process, including validation test results, will be discussed in the next section.

5.3 Map smoothening

The initial set-point maps will be treated by modified mean filtering method to smoothen all the rough transitions and sudden peaks. Figure 37 shows the result of smoothening on the Pi map. The peaks appear at medium speed-low torque and high speed-low torque positions have been smoothed to some extend.

Similarly, figure 38 and figure 39 show the same result for the common rail pressure and start of injection maps.

It seems that the smoothening process visually does not change the values of the original maps so much, however, some validation tests are still needed to be carried out at locations where the changes are significant. The tests help to assure that the smoothening process does not shift the optimal points too far away from their original locations and lead to dramatical changes in the optimized BSFC values. Nevertheless, due to the drawback of the simulation model which was mentioned in Subsection 4.3.2, it is not possible to run the model with different values of common rail pressure. Therefore, the validation tests can only be taken with the boost pressure and the start of injection maps.

Table 9 shows the BSFC results of validation tests at the following operating points, where changes are the biggest:

- Boost pressure map:

(OP1) 1750 rpm - 15 mg, (OP2) 1750 rpm - 18.25 mg

(OP3) 1750 rpm - 21.5 mg, (OP4) 2000 rpm - 18.25 mg

- Start of injection map:

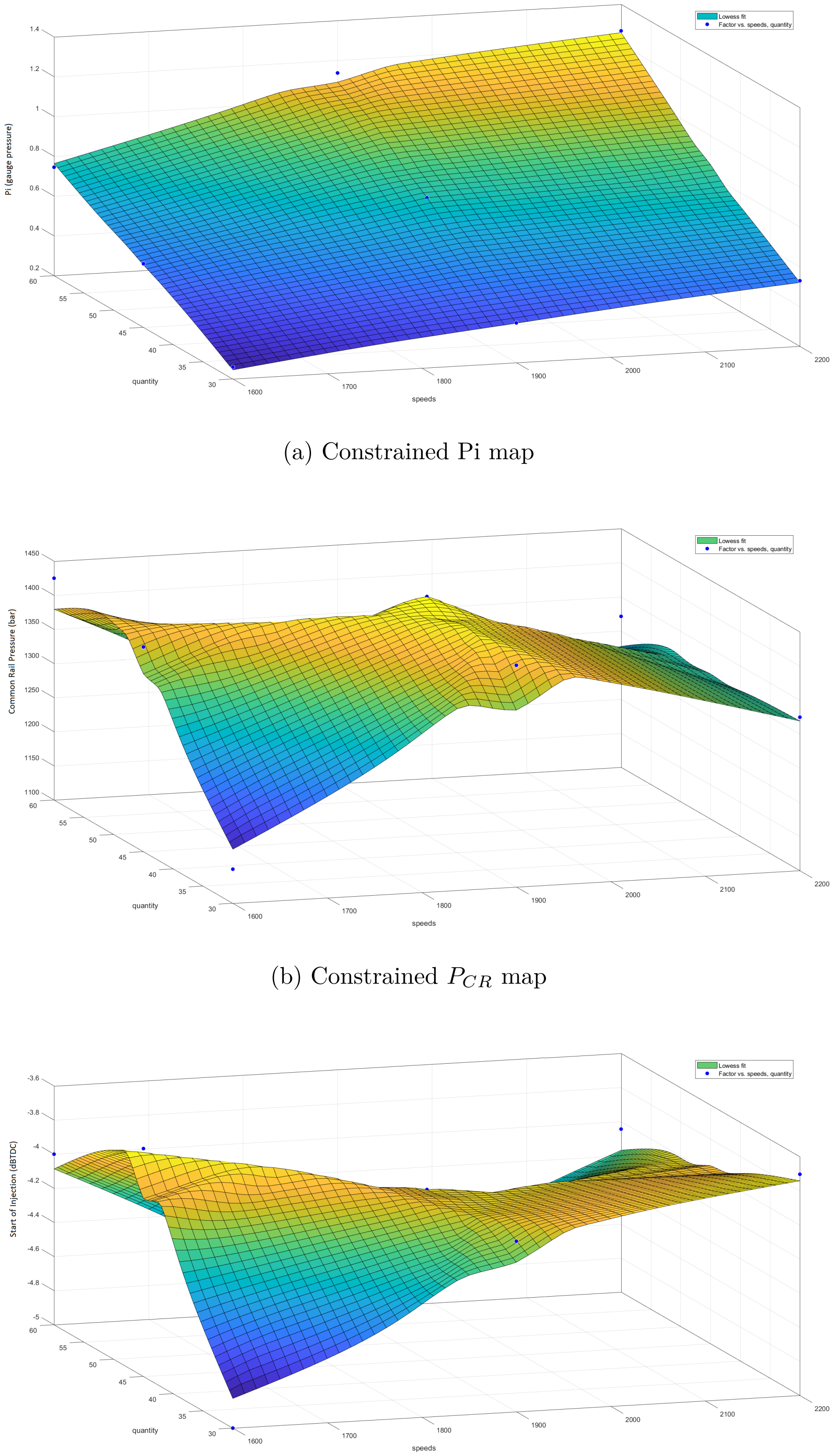
(OP5) 1750 rpm - 15 mg, (OP6) 2000 rpm - 15 mg

(OP7) 1625 rpm - 70.25 mg, (OP8) 1500 rpm - 63.75 mg

Table 9: The BSFC (g/kW.h) results of validation tests for smoothed maps.

	OP1	OP2	OP3	OP4	OP5	OP6	OP7	OP8
Initial map	327	280	256	285	326	345	206	206
Smoothed map	333	286.5	262	303	325	344	206.4	206.35

The BSFC values do not change much after the maps have been smoothed. Hence, it is safe to use this smoothening method to make the initial set-point maps smoother and not to shift the local optimum too far.

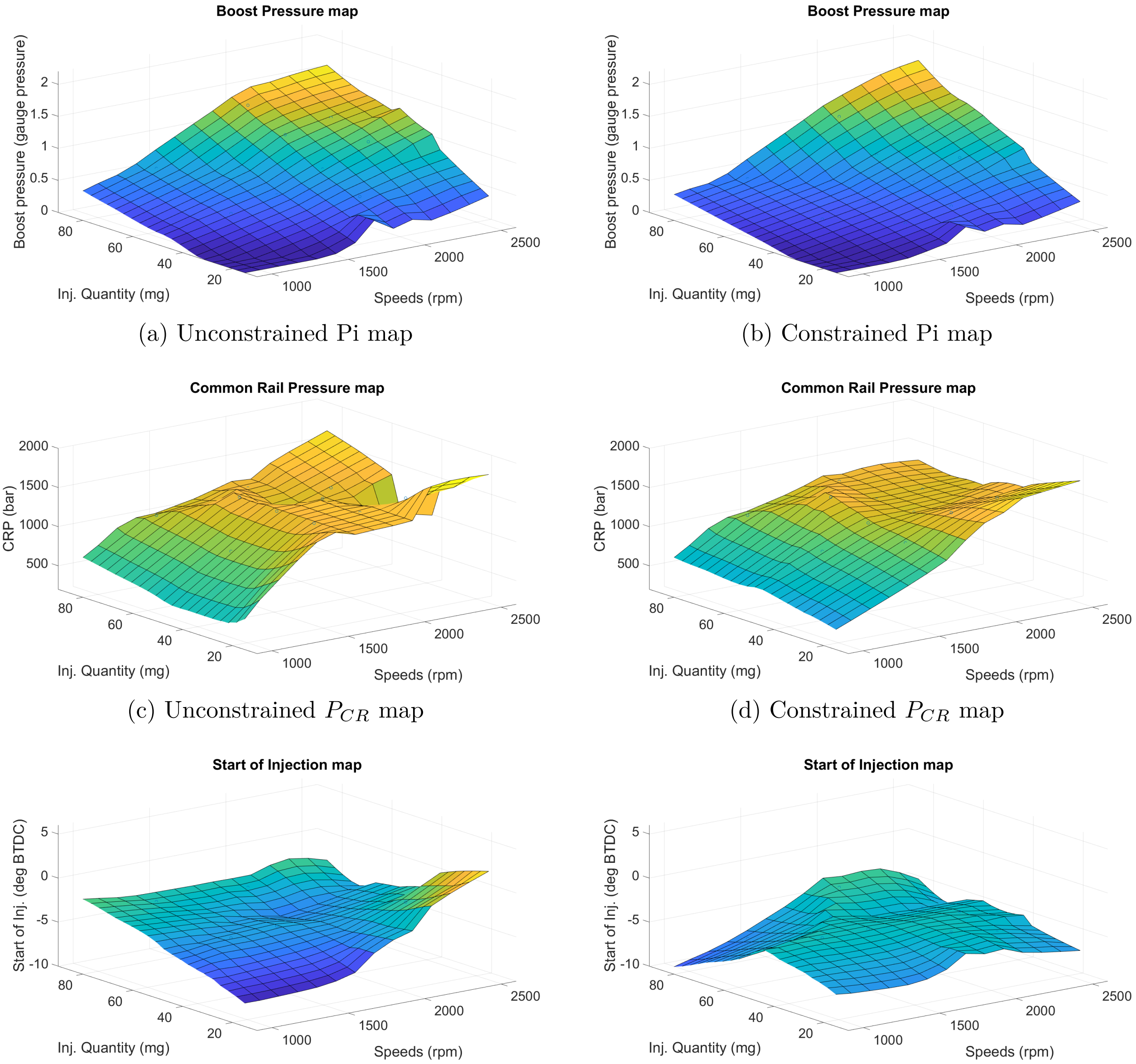


(a) Constrained Pi map

(b) Constrained P_{CR} map

(c) Constrained *Soi* map

Figure 34: LOESS fitting results



(a) Unconstrained Pi map

(b) Constrained Pi map

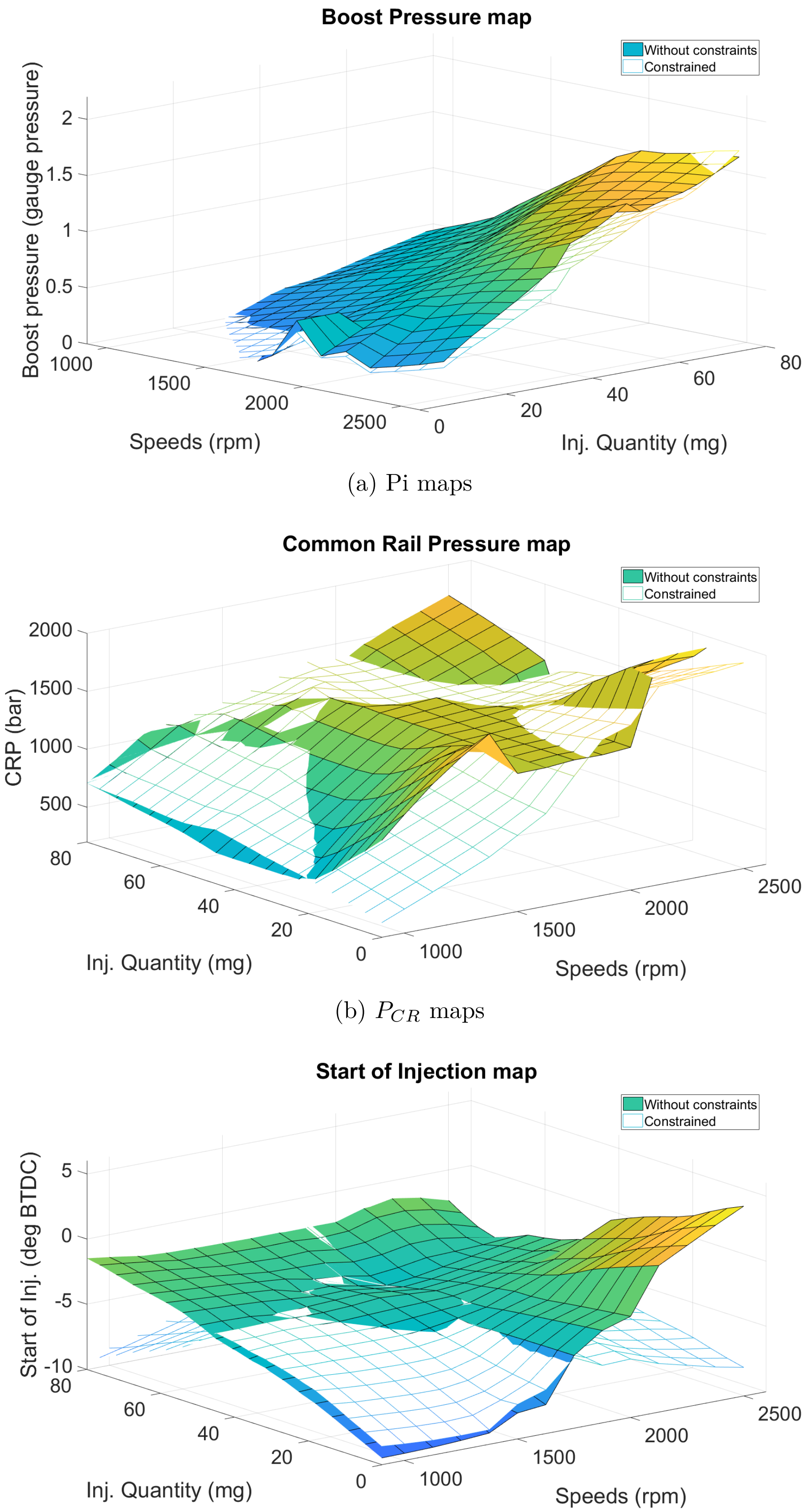
(c) Unconstrained P_{CR} map

(d) Constrained P_{CR} map

(e) Unconstrained SoI map

(f) Constrained SoI map

Figure 35: Initial set-point maps

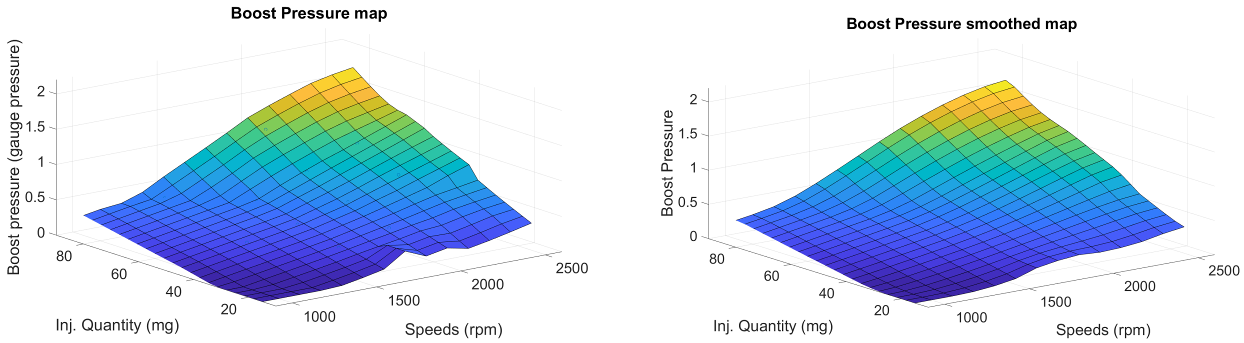


(a) Pi maps

(b) P_{CR} maps

(c) Soi maps

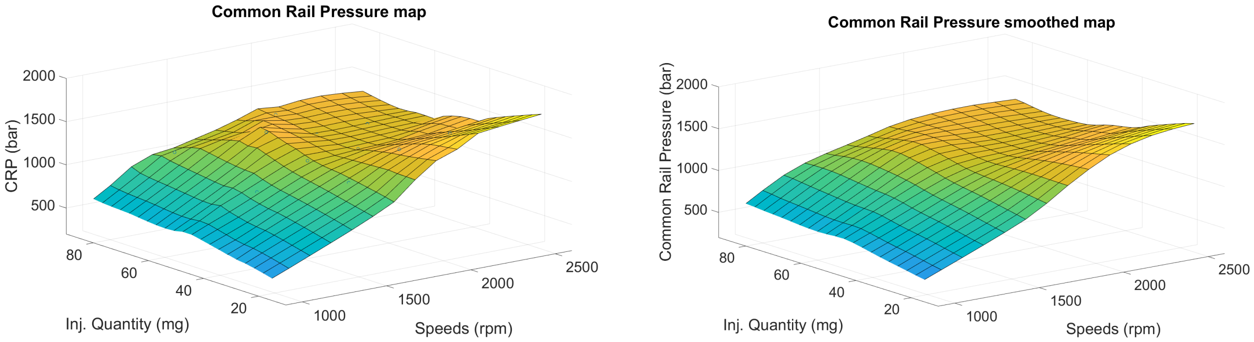
Figure 36: Differences between constrained and unconstrained maps



(a) Initial Pi map

(b) Smoothed Pi map

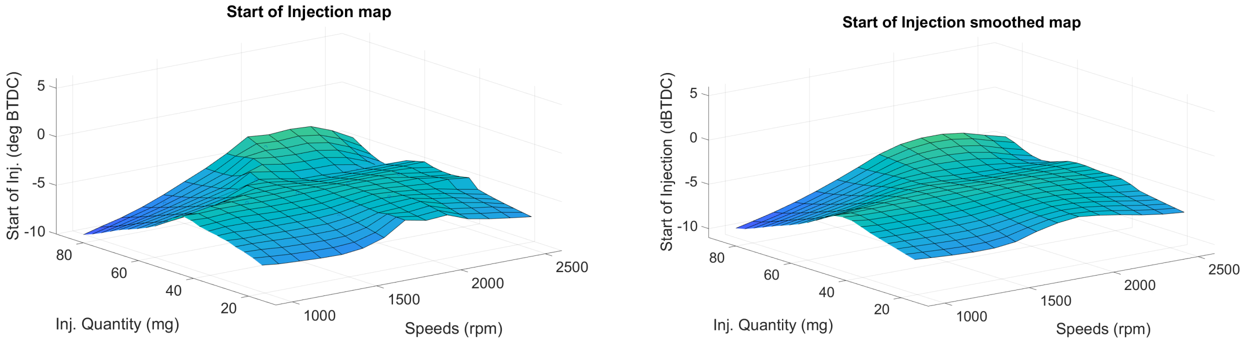
Figure 37: Differences between initial and smoothed Pi maps.



(a) Initial P_{CR} map

(b) Smoothed P_{CR} map

Figure 38: Differences between initial and smoothed P_{CR} maps.



(a) Initial Soi map

(b) Smoothed Soi map

Figure 39: Differences between initial and smoothed Soi maps.

6 Conclusions and Discussion

6.1 Conclusions

This thesis presents a study of designing static optimal control maps for high efficiency and emission reduction on diesel engine. The study leads to a finding of an engine calibration method which reduces necessary time and resources. The method has also been implemented on a non-road 44 AWI AGCO engine and on a GT-Power simulation model of that engine. Several conclusions were made from this work:

- An off-line parameterization tool which can be used for semi and fully automatic engine tuning was proposed and developed.
- The Design of Experiments method, which is the core component of the off-line tool, provides an organized and economical way of engine calibration. By using this method a considerable amount of time and resources can be saved.
- The engine response is better optimized in comparison to the traditional "brute force" method. Moreover, the response is even optimized under emission constraints to guarantee environmental protection.
- The off-line parameterization tool outputs the optimal control maps with smoothing transitions between the operating points. This smoothening work assures a smooth run for the engine in speed and load changing conditions.

6.2 Discussion

Though this study has given some promising results, there are still several aspects which need further improvements in the future for a better engine efficiency and lower emissions.

- Improve the Off-line parameterization tool with more important engine parameters. For instance, more injection strategy optimization (pre- and post-injection) should be considered instead of only the main injection in the current version of the tool.
- More constraints should be studied and included into the optimization of the engine responses to assure the engine's safety. The peak in-cylinder pressure can be an example of the addition constraints.
- The most important thing that needs improving in the future is the selection of the ranges of treated parameters at each operating point. In this thesis, the

selection is made mainly by running tests in the engine to find the limits of each parameter. However, this method does not guarantee that the combinations of the selected ranges are totally safe for the engine to run. Therefore, more research needs to be done on this issue to make sure that all the chosen ranges are safe for the engine.

- Last but not least, more full-scale tests of the parameterization tool must be carried out on other diesel engines to improve the working of the tool and to inspect for some potential drawbacks and errors.

References

[1] H. Langouët, L. Métivier, D. Sinoquet, and Q.-H. Tran, "Engine calibration: multi-objective constrained optimization of engine maps," *Optimization and Engineering*, vol. 12, no. 3, pp. 407–424, 2011.

[2] F. Mallamo, M. Badami, and F. Millo, "Application of the design of experiments and objective functions for the optimization of multiple injection strategies for low emissions in cr diesel engines," SAE Technical Paper, Tech. Rep., 2004.

[3] H. Stuhler, T. Kruse, A. Stuber, K. Gschweitl, W. Piock, H. Pfluegl, and P. Lick, "Automated model-based gdi engine calibration adaptive online doe approach," SAE Technical Paper, Tech. Rep., 2002.

[4] M. Deflorian, F. Klöpper, and J. Rückert, "Online dynamic black box modelling and adaptive experiment design in combustion engine calibration," *IFAC Proceedings Volumes*, vol. 43, no. 7, pp. 703–708, 2010.

[5] C. Atkinson, M. Allain, and H. Zhang, "Using model-based rapid transient calibration to reduce fuel consumption and emissions in diesel engines," SAE Technical Paper, Tech. Rep., 2008.

[6] M. Guerrier and P. Cawsey, "The development of model based methodologies for gasoline ic engine calibration," SAE Technical Paper, Tech. Rep., 2004.

[7] K. G. Johnson, K. Mollenhauer, and H. Tschöke, *Handbook of diesel engines.* Springer Science & Business Media, 2010.

[8] "Diesel cycle-diesel engine," http://www.nuclear-power.net/nuclear-engineering/thermodynamics/thermodynamic-cycles/diesel-cycle-diesel-engine/, accessed: 2017-05-25.

[9] V. Ganesan, *Internal combustion engines.* McGraw Hill Education (India) Pvt Ltd, 2012.

[10] "Four-stroke diesel engine," https://www.britannica.com/media/full/290504/19423, accessed: 2017-05-25.

[11] S. R. Turns, "An introduction to combustion, 2000," *MacGraw Hill, Boston, Massachusetts, US*, 2000.

[12] J. B. Heywood *et al.*, *Internal combustion engine fundamentals.* Mcgraw-hill New York, 1988, vol. 930.

[13] H. Jääskeläinen and M. K. Khair, "Air-flow diagram of a diesel engine," https://www.dieselnet.com/tech/diesel_air.php, accessed: 2017-10-03.

[14] H. Jääskeläinen, "Variable geometry turbochargers," https://www.dieselnet.com/tech/air_turbo_vgt.php, accessed: 2017-10-03.

[15] H. Jääskeläinen and M. K. Khair, "Common rail fuel injection," https://www.dieselnet.com/tech/diesel_fi_common-rail.php, accessed: 2017-10-03.

[16] R. Stone, *Introduction to internal combustion engines.* Palgrave Macmillan, 2012.

[17] F. Dryer and R. Sawyer, *Physical and Chemical Aspects of Combustion: A Tribute to Irvin Glassman.* CRC Press, 1997, vol. 4.

[18] L. Guzzella and C. Onder, *Introduction to modeling and control of internal combustion engine systems.* Springer Science & Business Media, 2009.

[19] N. R. Abdullah, N. S. Shahruddin, R. Mamat, A. Ihsan Mamat, and A. Zulkifli, "Effects of air intake pressure on the engine performance, fuel economy and exhaust emissions of a small gasoline engine," *Journal of Mechanical Engineering and Sciences*, vol. 6, pp. 949–58, 2014.

[20] J. M. Desantes, J. Benajes, J. M. García-Oliver, and C. P. Kolodziej, "Effects of intake pressure on particle size and number emissions from premixed diesel low-temperature combustion," *International Journal of Engine Research*, vol. 15, no. 2, pp. 222–235, 2014.

[21] Z. Xu, X. Li, C. Guan, and Z. Huang, "Effects of injection pressure on diesel engine particle physico-chemical properties," *Aerosol Science and Technology*, vol. 48, no. 2, pp. 128–138, 2014.

[22] T. Xu-Guang, S. Hai-Lang, Q. Tao, F. Zhi-Qiang, and Y. Wen-Hui, "The impact of common rail system's control parameters on the performance of high-power diesel," *Energy Procedia*, vol. 16, pp. 2067–2072, 2012.

[23] B. Jayashankara and V. Ganesan, "Effect of fuel injection timing and intake pressure on the performance of a di diesel engine–a parametric study using cfd," *Energy Conversion and Management*, vol. 51, no. 10, pp. 1835–1848, 2010.

[24] N. Raeie, S. Emami, and O. K. Sadaghiyani, "Effects of injection timing, before and after top dead center on the propulsion and power in a diesel engine," *Propulsion and Power Research*, vol. 3, no. 2, pp. 59–67, 2014.

[25] J. Babicz, *Wärtsilä encyclopedia of ship technology.* Baobab Naval Consultancy, 2008.

[26] K. Ropke, "Doe-design of experiments method and applications in engine development," *SV Corporate Media*, vol. 889503, 2005.

[27] M. Castagné, Y. Bentolila, F. Chaudoye, A. Hallé, F. Nicolas, and D. Sinoquet, "Comparison of engine calibration methods based on design of experiments (doe)," *Oil & Gas Science and Technology-Revue de l'IFP*, vol. 63, no. 4, pp. 563–582, 2008.

[28] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American statistical association*, vol. 74, no. 368, pp. 829–836, 1979.

[29] W. M. Trochim, "Research methods knowledge base," http://www.socialresearchmethods.net/kb/design.php, accessed: 20017-05-15.

[30] R. Carlson, *Design of Experiments, Principles and Applications, L. Eriksson, E. Johansson, N. Kettaneh-Wold, C. Wikström and S. Wold, Umetrics AB, Umeå Learnways AB, Stockholm, 2000, ISBN 91-973730-0-1, xii+ 329 pp.* Wiley Online Library, 2001, vol. 15, no. 5.

[31] C.-F. J. Wu and M. Hamada, *Experiments: planning, analysis, and parameter design optimization.* John Wiley, 2000.

[32] P. G. Mathews, *Design of Experiments with MINITAB.* ASQ Quality Press, 2005.

[33] R. H. Myers and D. C. Montgomery, *Response surface methodology: process improvement with steepest ascent, the analysis of response surfaces, experimental designs for fitting response surfaces.* New York: John Wiley and Sons, Inc, 1995.

[34] G. Box and D. Behnken, "Simplex-sum designs: a class of second order rotatable designs derivable from those of first order," *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 838–864, 1960.

[35] "Three-level box-behnken design for three factors," http://www.guosongtaohome.com/?page_id=96, accessed: 2017-07-15.

[36] K. Madsen, H. B. Nielsen, and O. Tingleff, "Methods for non-linear least squares problems," Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, p. 56, 1999. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?660

[37] S. Weisberg, *Applied linear regression.* John Wiley & Sons, 2005, vol. 528.

[38] E. Philip and W. Elizabeth, "Sequential quadratic programming methods," *UCSD Department of Mathematics Technical Report NA-10-03 August*, 2010.

[39] P. T. Boggs and J. W. Tolle, "Sequential quadratic programming," *Acta numerica*, vol. 4, pp. 1–51, 1995.

[40] D. G. Luenberger, Y. Ye *et al.*, *Linear and nonlinear programming.* Springer, 1984, vol. 2.

[41] S. G. Nash and A. Sofer, *Linear and nonlinear programming.* McGraw-Hill Inc., 1996.

[42] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables.* SIAM, 2000.

[43] G. R. Arce and J. L. Paredes, "Image enhancement and analysis with weighted medians," *Nonlinear Image Processing*, pp. 27–67, 2000.

[44] N. Nikolaidis and I. Pitas, "3-d image processing algorithms john wiley & sons," *Inc. New York, NY Google Scholar*, 2001.

[45] H. Yagou, Y. Ohtake, and A. Belyaev, "Mesh smoothing via mean and median filtering applied to face normals," in *Geometric Modeling and Processing, 2002. Proceedings.* IEEE, 2002, pp. 124–131.

[46] J.-J. Pan, Y.-Y. Tang, and B.-C. Pan, "The algorithm of fast mean filtering," in *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on*, vol. 1. IEEE, 2007, pp. 244–248.

[47] S. Rakshit, A. Ghosh, and B. U. Shankar, "Fast mean filtering technique (fmft)," *Pattern Recognition*, vol. 40, no. 3, pp. 890–897, 2007.

[48] R. C.Gonzalez and R. E. Woods, *Digital Image Processing (2^{nd} edition).* Prentice Hall, 1992.

[49] E. Trucco and A. Verri, *Introductory techniques for 3-D computer vision.* Prentice Hall Englewood Cliffs, 1998, vol. 201.

[50] "Convolution matrix," https://docs.gimp.org/en/plug-in-convmatrix.html, accessed: 2017-10-03.

[51] "Example of 2-d convolution," http://www.songho.ca/dsp/convolution/convolution2d_example.html, accessed: 2017-10-03.